

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2017235333 B2**

(54) Title
CRISPR/CAS-related methods and compositions for treating beta hemoglobinopathies

(51) International Patent Classification(s)
C12N 15/113 (2010.01) **C12N 15/09** (2006.01)
A61K 31/713 (2006.01) **C12N 15/63** (2006.01)
A61P 7/06 (2006.01)

(21) Application No: **2017235333** (22) Date of Filing: **2017.03.14**

(87) WIPO No: **WO17/160890**

(30) Priority Data

(31) Number	(32) Date	(33) Country
62/456,615	2017.02.08	US
62/308,190	2016.03.14	US

(43) Publication Date: **2017.09.21**

(44) Accepted Journal Date: **2023.08.24**

(71) Applicant(s)
Editas Medicine, Inc.

(72) Inventor(s)
Gori, Jennifer Leah;Barrera, Luis A.

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

(56) Related Art
WO 2013/126794 A1
WO 2014/186585 A2
WO 2014/197748 A2
WO 2016/135558 A2



(43) International Publication Date
21 September 2017 (21.09.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/160890 A1

(51) International Patent Classification:

C12N 15/113 (2010.01) *C12N 15/63* (2006.01)
A61K 31/713 (2006.01) *A61P 7/06* (2006.01)
C12N 15/09 (2006.01)

(21) International Application Number:

PCT/US2017/022377

(22) International Filing Date:

14 March 2017 (14.03.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/308,190 14 March 2016 (14.03.2016) US
62/456,615 8 February 2017 (08.02.2017) US

(71) Applicant: **EDITAS MEDICINE, INC.** [US/US]; 11
Hurley Street, Cambridge, Massachusetts 02141 (US).

(72) Inventors: **GORI, Jennifer Leah**; 20 Spalding Street,
Unit 1, Jamaica Plain, Massachusetts 02130 (US). **BAR-**
RERA, Luis A.; 11 Morton Street #1, Somerville, Mas-
sachusetts 02145 (US).

(74) Agent: **PROCHNOW, Courtney**; Perkins Coie LLP, P.O.
Box 1247, Seattle, Washington 98111-1247 (US).

(81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM,

DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN,
KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA,
MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG,
NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS,
RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY,
TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN,
ZA, ZM, ZW.

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

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a
patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii))*

Published:

- *with international search report (Art. 21(3))*
- *with sequence listing part of description (Rule 5.2(a))*

(54) Title: CRISPR/CAS-RELATED METHODS AND COMPOSITIONS FOR TREATING BETA HEMOGLOBINOPATHIES

(57) Abstract: CRISPR/CAS-related compositions and methods for treatment of beta hemoglobinopathies are disclosed.



WO 2017/160890 A1

CRISPR/CAS-RELATED METHODS AND COMPOSITIONS FOR TREATING BETA HEMOGLOBINOPATHIES

REFERENCE TO RELATED APPLICATIONS

The present application claims the benefit of United States Provisional Application No. 62/308,190, filed March 14, 2016, and United States Provisional Application No. 62/456,615, filed February 8, 2017, the contents of each of which are hereby incorporated by reference in their entirety.

SEQUENCE LISTING

This application contains a Sequence Listing, which was submitted in ASCII format via EFS-Web, and is hereby incorporated by reference in its entirety. The ASCII copy, created on March 14, 2017, is named 8009WO00_SequenceListing.txt and is 335 KB in size.

FIELD OF THE INVENTION

The invention relates to CRISPR/Cas-related methods and components for editing a target nucleic acid sequence, or modulating expression of a target nucleic acid sequence, and applications thereof in connection with β -hemoglobinopathies including sickle cell disease and β -thalassemia.

BACKGROUND

Hemoglobin (Hb) carries oxygen from the lungs to tissues in erythrocytes or red blood cells (RBCs). During prenatal development and until shortly after birth, hemoglobin is present in the form of fetal hemoglobin (HbF), a tetrameric protein composed of two alpha (α)-globin chains and two gamma (γ)-globin chains. HbF is largely replaced by adult hemoglobin (HbA), a tetrameric protein in which the γ -globin chains of HbF are replaced with beta (β)-globin chains, through a process known as globin switching. HbF is more efficient than HbA at carrying oxygen. The average adult makes less than 1% HbF out of total hemoglobin (Thein 2009). The α -hemoglobin gene is located on chromosome 16, while the β -hemoglobin gene (HBB), A gamma (γ^A)-globin chain (HBG1, also known as gamma globin A), and G gamma (γ^G)-globin chain (HBG2, also known as gamma globin G) are located on chromosome 11 within the globin gene cluster (i.e., globin locus).

Mutations in HBB can cause hemoglobin disorders (i.e., hemoglobinopathies) including sickle cell disease (SCD) and beta-thalassemia (β -Thal). Approximately 93,000

people in the United States are diagnosed with a hemoglobinopathy. Worldwide, 300,000 children are born with hemoglobinopathies every year (Angastiniotis 1998). Because these conditions are associated with HBB mutations, their symptoms typically do not manifest until after globin switching from HbF to HbA.

SCD is the most common inherited hematologic disease in the United States, affecting approximately 80,000 people (Brousseau 2010). SCD is most common in people of African ancestry, for whom the prevalence of SCD is 1 in 500. In Africa, the prevalence of SCD is 15 million (Aliyu 2008). SCD is also more common in people of Indian, Saudi Arabian and Mediterranean descent. In those of Hispanic-American descent, the prevalence of sickle cell disease is 1 in 1,000 (Lewis 2014).

SCD is caused by a single homozygous mutation in the HBB gene, c.17A>T (HbS mutation). The sickle mutation is a point mutation (GAG → GTG) on HBB that results in substitution of valine for glutamic acid at amino acid position 6 in exon 1. The valine at position 6 of the β -hemoglobin chain is hydrophobic and causes a change in conformation of the β -globin protein when it is not bound to oxygen. This change of conformation causes HbS proteins to polymerize in the absence of oxygen, leading to deformation (i.e., sickling) of RBCs. SCD is inherited in an autosomal recessive manner, so that only patients with two HbS alleles have the disease. Heterozygous subjects have sickle cell trait, and may suffer from anemia and/or painful crises if they are severely dehydrated or oxygen deprived.

Sickle shaped RBCs cause multiple symptoms, including anemia, sickle cell crises, vaso-occlusive crises, aplastic crises, and acute chest syndrome. Sickle shaped RBCs are less elastic than wild-type RBCs and therefore cannot pass as easily through capillary beds and cause occlusion and ischemia (i.e., vaso-occlusion). Vaso-occlusive crisis occurs when sickle cells obstruct blood flow in the capillary bed of an organ leading to pain, ischemia, and necrosis. These episodes typically last 5-7 days. The spleen plays a role in clearing dysfunctional RBCs, and is therefore typically enlarged during early childhood and subject to frequent vaso-occlusive crises. By the end of childhood the spleen in SCD patients is often infarcted, which leads to autosplenectomy. Hemolysis is a constant feature of SCD and causes anemia. Sickle cells survive for 10-20 days in circulation, while healthy RBCs survive for 90-120 days. SCD subjects are transfused as necessary to maintain adequate hemoglobin levels. Frequent transfusions place subjects at risk for infection with HIV, hepatitis B, and hepatitis C. Subjects may also suffer from acute chest crisis and infarcts of extremities, end organs, and the central nervous system.

Subjects with SCD have decreased life expectancies. The prognosis for patients with SCD is steadily improving with careful, life-long management of crises and anemia. As of 2001, the average life expectancy of subjects with sickle cell disease was the mid-to-late 50's. Current treatments for SCD involve hydration and pain management during crises, and transfusions as needed to correct anemia.

Thalassemias (e.g., β -Thal, δ -Thal, and β/δ -Thal) cause chronic anemia. β -Thal is estimated to affect approximately 1 in 100,000 people worldwide. Its prevalence is higher in certain populations, including those of European descent, where its prevalence is approximately 1 in 10,000. β -Thal major, the more severe form of the disease, is life-threatening unless treated with lifelong blood transfusions and chelation therapy. In the United States, there are approximately 3,000 subjects with β -Thal major. β -Thal intermedia does not require blood transfusions, but it may cause growth delay and significant systemic abnormalities, and it frequently requires lifelong chelation therapy. Although HbA makes up the majority of hemoglobin in adult RBCs, approximately 3% of adult hemoglobin is in the form of HbA₂, an HbA variant in which the two γ -globin chains are replaced with two delta (Δ)-globin chains. δ -Thal is associated with mutations in the Δ hemoglobin gene (HBD) that cause a loss of HBD expression. Co-inheritance of the HBD mutation can mask a diagnosis of β -Thal (i.e., β/δ -Thal) by decreasing the level of HbA₂ to the normal range (Bouva 2006). β/δ -Thal is usually caused by deletion of the HBB and HBD sequences in both alleles. In homozygous ($\delta^0/\delta^0 \beta^0/\beta^0$) patients, HbG is expressed, leading to production of HbF alone.

Like SCD, β -Thal is caused by mutations in the HBB gene. The most common HBB mutations leading to β -Thal are: c.-136C>G, c.92+1G>A, c.92+6T>C, c.93-21G>A, c.118C>T, c.316-106C>G, c.25_26delAA, c.27_28insG, c.92+5G>C, c.118C>T, c.135delC, c.315+1G>A, c.-78A>G, c.52A>T, c.59A>G, c.92+5G>C, c.124_127delTTCT, c.316-197C>T, c.-78A>G, c.52A>T, c.124_127delTTCT, c.316-197C>T, c.-138C>T, c.-79A>G, c.92+5G>C, c.75T>A, c.316-2A>G, and c.316-2A>C. These and other mutations associated with β -Thal cause mutated or absent β -globin chains, which causes a disruption of the normal Hb α -hemoglobin to β -hemoglobin ratio. Excess α -globin chains precipitate in erythroid precursors in the bone marrow.

In β -Thal major, both alleles of HBB contain nonsense, frameshift, or splicing mutations that leads to complete absence of β -globin production (denoted β^0/β^0). β -Thal major results in severe reduction in β -globin chains, leading to significant precipitation of α -globin chains in erythroid cells and more severe anemia.

β -Thal intermedia results from mutations in the 5' or 3' untranslated region of HBB, mutations in the promoter region or polyadenylation signal of HBB, or splicing mutations within the HBB gene. Patient genotypes are denoted β^0/β^+ or β^+/ β^+ . β^0 represents absent expression of a β -globin chain; β^+ represents a dysfunctional but present β -globin chain. Phenotypic expression varies among patients. Since there is some production of β -globin, β -Thal intermedia results in less precipitation of α -globin chains in the erythroid precursors and less severe anemia than β -Thal major. However, there are more significant consequences of erythroid lineage expansion secondary to chronic anemia.

Subjects with β -Thal major present between the ages of 6 months and 2 years, and suffer from failure to thrive, fevers, hepatosplenomegaly, and diarrhea. Adequate treatment includes regular transfusions. Therapy for β -Thal major also includes splenectomy and treatment with hydroxyurea. If patients are regularly transfused, they will develop normally until the beginning of the second decade. At that time, they require chelation therapy (in addition to continued transfusions) to prevent complications of iron overload. Iron overload may manifest as growth delay or delay of sexual maturation. In adulthood, inadequate chelation therapy may lead to cardiomyopathy, cardiac arrhythmias, hepatic fibrosis and/or cirrhosis, diabetes, thyroid and parathyroid abnormalities, thrombosis, and osteoporosis. Frequent transfusions also put subjects at risk for infection with HIV, hepatitis B and hepatitis C.

β -Thal intermedia subjects generally present between the ages of 2-6 years. They do not generally require blood transfusions. However, bone abnormalities occur due to chronic hypertrophy of the erythroid lineage to compensate for chronic anemia. Subjects may have fractures of the long bones due to osteoporosis. Extramedullary erythropoiesis is common and leads to enlargement of the spleen, liver, and lymph nodes. It may also cause spinal cord compression and neurologic problems. Subjects also suffer from lower extremity ulcers and are at increased risk for thrombotic events, including stroke, pulmonary embolism, and deep vein thrombosis. Treatment of β -Thal intermedia includes splenectomy, folic acid supplementation, hydroxyurea therapy, and radiotherapy for extramedullary masses. Chelation therapy is used in subjects who develop iron overload.

Life expectancy is often diminished in β -Thal patients. Subjects with β -Thal major who do not receive transfusion therapy generally die in their second or third decade. Subjects with β -Thal major who receive regular transfusions and adequate chelation therapy can live into their fifth decade and beyond. Cardiac failure secondary to iron toxicity is the leading cause of death in β -Thal major subjects due to iron toxicity.

A variety of new treatments are currently in development for SCD and β -Thal. Delivery of a corrected HBB gene via gene therapy is currently being investigated in clinical trials. However, the long-term efficacy and safety of this approach is unknown. Transplantation with hematopoietic stem cells from an HLA-matched allogeneic stem cell donor has been demonstrated to cure SCD and β -Thal, but this procedure involves risks including those associated with ablation therapy to prepare the subject for transplant and risk of graft vs. host disease after transplantation. In addition, matched allogeneic donors often cannot be identified. Thus, there is a need for improved methods of managing these and other hemoglobinopathies.

SUMMARY OF THE INVENTION

According to one aspect, the present invention provides a method of increasing the level of fetal hemoglobin in a human cell by genome editing, the method comprising the step of introducing into the human cell a ribonucleoprotein (RNP) complex comprising: (a) an RNA-guided nuclease; and (b) a guide RNA (gRNA) comprising a targeting domain comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:310, SEQ ID NO:333, SEQ ID NO:338, SEQ ID NO:339, SEQ ID NO:340, SEQ ID NO:370, SEQ ID NO:371, SEQ ID NO:372, and SEQ ID NO:379.

According to a second aspect, the present invention provides a cell modified using a ribonucleoprotein (RNP) complex comprising: (a) an RNA-guided nuclease protein; and (b) a guide RNA (gRNA) comprising a targeting domain comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:310, SEQ ID NO:333, SEQ ID NO:338, SEQ ID NO:339, SEQ ID NO:340, SEQ ID NO:370, SEQ ID NO:371, SEQ ID NO:372, and SEQ ID NO:379.

According to a third aspect, the present invention provides a genome editing system comprising a ribonucleoprotein (RNP) complex comprising: (a) an RNA-guided nuclease; and (b) a guide RNA (gRNA) comprising a targeting domain comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:310, SEQ ID NO:333, SEQ ID NO:338, SEQ ID NO:339, SEQ ID NO:340, SEQ ID NO:370, SEQ ID NO:371, SEQ ID NO:372, and SEQ ID NO:379.

Provided herein in certain embodiments are methods for increasing expression (i.e., transcriptional activity) of one or more γ -globin genes (e.g., HBG1, HBG2, or HBG1 and HBG2) in a subject or cell using a genome editing system (e.g., CRISPR/Cas-mediated genome

editing system). In certain embodiments, these methods may utilize any repair mechanism to alter (e.g., delete, disrupt, or modify) all or a portion of one or more γ -globin gene regulatory elements. In certain embodiments, these methods may utilize a DNA repair mechanism, e.g., NHEJ or HDR to delete or disrupt one or more γ -globin gene regulatory elements (e.g., silencer, enhancer, promoter, or insulator). In certain embodiments, these methods utilize a DNA repair mechanism, e.g., HDR, to alter, including mutate, insert, delete or disrupt, the sequence of one or more nucleotides in γ -globin gene regulatory element (e.g., silencer, enhancer, promoter, or insulator). In certain embodiments, these methods utilize a combination of one or more DNA repair mechanisms, e.g., NHEJ and HDR. In certain embodiments, these methods result in a mutation or variation in an γ -globin regulatory element that is associated with a naturally occurring HPFH variant, including, for example, HBG1 13 bp del c.-114 to -102; 4 bp del c.-225 to -222; c.-114 C>T; c.-117 G>A; c.-158 C>T; c.-167 C>T; c.-170 G>A; c.-175 T>G; c.-175 T>C; c.-195 C>G; c.-196 C>T; c.-198 T>C; c.-201 C>T; c.-251 T>C; or c.-499 T>A; or HBG2 13 bp del c.-114 to -102; c.-109 G>T; c.-114 C>A; c.-114 C>T; c.-157 C>T; c.-158 C>T; c.-167 C>T; c.-167 C>A; c.-175 T>C; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-255 C>G; c.-309 A>G; c.-369 C>G; or c.-567 T>G.

Provided herein in certain embodiments are methods for treating a β -hemoglobinopathy in a subject in need thereof using CRISPR/Cas-mediated genome editing to increase expression (i.e., transcriptional activity) of one or more γ -globin genes (e.g.,

HBG1, HBG2, or HBG1 and HBG2). In certain embodiments, these methods utilize a DNA repair mechanism, e.g., NHEJ or HDR to delete or disrupt one or more γ -globin gene regulatory elements (e.g., silencer, enhancer, promoter, or insulator). In certain embodiments, these methods utilize a DNA repair mechanism, e.g., HDR to alter, including mutate, insert, delete or disrupt, the sequence of one or more nucleotides in γ -globin gene regulatory element (e.g., silencer, enhancer, promoter, or insulator). In certain embodiments, these methods utilize a combination of one or more DNA repair mechanisms, e.g., NHEJ and HDR. In certain embodiments, these methods result in a mutation or variation in an γ -globin regulatory element that is associated with a naturally occurring HPFH variant, including for example HBG1 13 bp del c.-114 to -102; 4 bp del c.-225 to -222; c.-114 C>T; c.-117 G>A; c.-158 C>T; c.-167 C>T; c.-170 G>A; c.-175 T>G; c.-175 T>C; c.-195 C>G; c.-196 C>T; c.-198 T>C; c.-201 C>T; c.-251 T>C; or c.-499 T>A; or HBG2 13 bp del c.-114 to -102; c.-109 G>T; c.-114 C>A; c.-114 C>T; c.-157 C>T; c.-158 C>T; c.-167 C>T; c.-167 C>A; c.-175 T>C; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-255 C>G; c.-309 A>G; c.-369 C>G; or c.-567 T>G. In certain embodiments, the β -hemoglobinopathy is SCD or β -Thal.

Provided herein in certain embodiments are gRNAs for use in CRISPR/Cas-mediated methods of increasing expression (i.e., transcriptional activity) of one or more γ -globin genes (e.g., HBG1, HBG2, or HBG1 and HBG2). In certain embodiments, these gRNAs comprise a targeting domain comprising a nucleotide sequence set forth in SEQ ID NOs:251-901. In certain embodiments, these gRNAs further comprise one or more of a first complementarity domain, second complementarity domain, linking domain, 5' extension domain, proximal domain, or tail domain. In certain embodiments, the gRNA is modular. In other embodiments, the gRNA is unimolecular (or chimeric).

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A-II 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:39 and 40, respectively, in order of appearance);

Fig. 1B depicts a unimolecular gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO:41);

Fig. 1C depicts a unimolecular gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO:42);

Fig. 1D depicts a unimolecular gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO:43);

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

Fig. 1F depicts a modular gRNA molecule derived in part from *Streptococcus thermophilus* (*S. thermophilus*) as a duplexed structure (SEQ ID NOs:45 and 46, respectively, in order of appearance);

Fig. 1G depicts an alignment of modular gRNA molecules of *S. pyogenes* and *S. thermophilus* (SEQ ID NOs:39, 45, 47, and 46, respectively, in order of appearance).

Figs. 1H-1I depicts additional exemplary structures of unimolecular gRNA molecules.

Fig. 1H shows an exemplary structure of a unimolecular gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO:42).

Fig. 1I shows an exemplary structure of a unimolecular gRNA molecule derived in part from *S. aureus* as a duplexed structure (SEQ ID NO:38).

Figs. 2A-2G depict an alignment of Cas9 sequences (Chylinski 2013). 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:4); and Li: *L. innocua* (SEQ ID NO:5). "Motif" (SEQ ID NO:14) is a consensus sequence based on the four sequences. Residues conserved in all four sequences are indicated by single letter amino acid abbreviation; "*" indicates any amino acid found in the corresponding position of any of the four sequences; and "-" indicates absent.

Figs. 3A-3B show an alignment of the N-terminal RuvC-like domain from the Cas9 molecules disclosed in Chylinski 2013 (SEQ ID NOs:52-95, 120-123). 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 2013 with sequence outliers removed (SEQ ID NOs:52-123). 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 2013 (SEQ ID NOs:124-198). 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 2013 with sequence outliers removed (SEQ ID NOs:124-141, 148, 149, 151-153, 162, 163, 166-174, 177-187, 194-198). The last line of **Fig. 6B** identifies 3 highly conserved residues.

Fig. 7 illustrates gRNA domain nomenclature using an exemplary gRNA sequence (SEQ ID NO:42).

Figs. 8A and 8B provide schematic representations of the domain organization of *S. pyogenes* Cas9. **Fig. 8A** 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. 8B** shows the percent homology of each domain across 83 Cas9 orthologs.

Figs. 9A-9C provide schematics of the HBG1 and HBG2 gene(s) in the context of the globin locus. The coding sequences (CDS), mRNA regions, and genes are indicated. (A) Regions that were targeted for gRNA design (dashed lines and brackets indicating the genetic regions proximal to the HBG1 and HBG2 genes) are shown. (B) Core promoter elements are indicated. (C) Motifs in the gene regulatory regions to which transcriptional activators and transcriptional repressors may bind to regulate gene expression are indicated. Note the overlap between the motifs and the genomic region targeted for gRNA design. Examples of deletions in the HBG1 and HBG2 gene regulatory regions that cause HPFH are indicated, as well as the % HbF associated with each.

Figs. 10A-F shows data from gRNA screening for incorporation of the 13 bp del c.-114 to -102 HPFH mutation in human K562 erythroleukemia cells. (A) Gene editing as determined by T7E1 endonuclease assay analysis of HBG1 and HBG2 locus-specific PCR products amplified from genomic DNA extracted from K562 cells after electroporation with DNA encoding *S. pyogenes*-specific gRNAs and plasmid DNA encoding *S. pyogenes* Cas9. (B) Gene editing as determined by DNA sequence analysis of PCR products amplified from the HBG1 locus in genomic DNA extracted from K562 cells after electroporation with DNA encoding the indicated gRNA and Cas9 plasmid. (C) Gene editing as determined by DNA sequence analysis of PCR products amplified from the HBG2 locus in genomic DNA extracted from K562 cells after electroporation with DNA encoding the indicated gRNA and Cas9 plasmid. For (B) and (C), the types of editing events (insertions, deletions) and subtypes of deletions (13 nt target partially [12 nt HPFH] or fully [13-26 nt HPFH] deleted, other sequences deleted [other deletions]) are indicated by the differently shaded/patterned bars. (D)-(F) Examples of HBG1 gene regulatory region deletions.

Figs. 11A-C depict results of gene editing in human cord blood (CB) and human adult

CD34⁺ cells after electroporation with RNPs complexed to *in vitro* transcribed *S. pyogenes* gRNAs that target a specific 13 nt sequence for deletion (*HBG* gRNAs Sp35 (comprising SEQ ID NO:339) and Sp37 (comprising SEQ ID NO:333)). **Fig. 11A** depicts the percentage of indels detected by T7E1 analysis of *HBG1* and *HBG2* specific PCR products amplified from gDNA extracted from CB CD34⁺ cells treated with the indicated RNPs or donor matched untreated control cells (n=3 CB CD34⁺ cells, 3 separate experiments). Data shown represent the mean and error bars correspond to standard deviation across the three separate donors/experiments. **Fig. 11B** depicts the percentage of indels detected by T7E1 analysis of *HBG2* specific PCR product amplified from gDNA extracted from CB CD34⁺ cells or adult CD34⁺ cells treated with the indicated RNPs or donor matched untreated control cells (n=3 CB CD34⁺ cells, n=3 mPB CD34⁺ cells, three separate experiments). Data shown represent the mean and error bars correspond to standard deviation across the three separate donors/experiments. **Fig. 11C** (top panel) depicts edits as detected by T7E1 analysis of *HBG2* PCR products amplified from gDNA extracted from human CB CD34⁺ cells electroporated with HBG Sp35 RNP or HBG Sp37 RNP +/- ssODN1 (SEQ ID NO:906) or PhTx ssODN1 (SEQ ID NO:909). **Fig. 11C** (lower left panel) shows the level of gene editing as determined by Sanger DNA sequence analysis of gDNA from cells edited with HBG Sp37 RNP and ssODN1 and PhTx ssODN1. **Fig. 11C** (lower right panel) shows the specific types of deletions detected within total deletions from the data presented in the lower left panel.

Figs. 12A-C depict gene editing of *HBG1* and *HBG2* in K562 erythroleukemia cells. **Fig. 12A** depicts NHEJ (indels) detected by T7E1 analysis of *HBG1* and *HBG2* PCR products amplified from gDNA extracted from K562 cells three days after nucleofection with RNPs complexed to the indicated gRNAs. **Fig. 12B** depicts Sanger DNA sequence analysis of PCR products amplified from the *HBG1* locus for cells nucleofected with Cas9 protein complexed to gRNAs targeting the 13 nt HPFH sequence (Sp35 (comprising SEQ ID NO:339), Sp36 (comprising SEQ ID NO:338), Sp37 (comprising SEQ ID NO:333)). **Fig. 12C** depicts Sanger DNA sequence analysis of PCR products amplified from the *HBG2* locus for cells nucleofected with Cas9 protein complexed to gRNAs targeting the 13 bp HPFH sequence (Sp35, Sp36, Sp37). For **Fig. 12B** and **Fig. 12C** the deletions were subdivided into deletions that contained the 13 bp targeted deletion (HPFH deletion, 18-26 nt deletion, >26 nt deletion) and deletions that did not contain the 13 bp deletion (<12 nt deletion, other deletion, insertion).

Fig. 13 depicts gene editing of *HBG* in adult human mobilized peripheral blood (mPB) CD34⁺ cells and induction of fetal hemoglobin in erythroid progeny of RNP treated cells after electroporation of mPB CD34⁺ cells with HBG Sp37 RNP +/- ssODN encoding the 13 bp deletion. **Fig. 13A** depicts the percentage of edits detected by T7E1 analysis of *HBG2* PCR product amplified from gDNA extracted from mPB CD34⁺ cells treated with the RNP or donor matched untreated control cells. **Fig. 13B** depicts the fold change in *HBG* mRNA expression in day seven erythroblasts that were differentiated from RNP treated and untreated donor matched control mPB CD34⁺ cells. mRNA levels are normalized to GAPDH and calibrated to the levels detected in untreated controls on the corresponding days of differentiation.

Fig. 14 depicts the *ex vivo* differentiation potential of RNP treated and untreated mPB CD34⁺ cells from the same donor. **Fig. 14A** shows hematopoietic myeloid/erythroid colony forming cell (CFC) potential, where the number and subtype of colonies are indicated (GEMM: granulocyte-erythroid-monocyte-macrophage colony, E: erythroid colony, GM: granulocyte-macrophage colony, M: macrophage colony, G: granulocyte colony). **Fig. 14B** depicts the percentage of Glycophorin A expressed over the time course of erythroid differentiation as determined by flow cytometry analysis at the indicated time points and for the indicated samples.

DETAILED DESCRIPTION

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 frame shift 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.

“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.

“Alt-HDR,” “alternative homology-directed repair,” or “alternative HDR” as used herein refers to the process of repairing DNA damage using a homologous nucleic acid (e.g., an endogenous homologous sequence, e.g., a sister chromatid, or an exogenous nucleic acid, e.g., a template nucleic acid). Alt-HDR is distinct from canonical HDR in that the process utilizes different pathways from canonical HDR, and can be inhibited by the canonical HDR mediators, RAD51 and BRCA2. Also, alt-HDR uses a single-stranded or nicked homologous nucleic acid for repair of the break.

“Canonical HDR” or “canonical homology-directed repair” as used herein refers to the process of repairing DNA damage using a homologous nucleic acid (e.g., an endogenous homologous sequence, e.g., a sister chromatid, or an exogenous nucleic acid, e.g., a template nucleic acid). Canonical HDR typically acts when there has been significant resection at the double strand break, forming at least one single stranded portion of DNA. In a normal cell, HDR typically involves a series of steps such as recognition of the break, stabilization of the break, resection, stabilization of single stranded DNA, formation of a DNA crossover intermediate, resolution of the crossover intermediate, and ligation. The process requires RAD51 and BRCA2, and the homologous nucleic acid is typically double-stranded.

Unless indicated otherwise, the term “HDR” as used herein encompasses both canonical HDR and alt-HDR.

“Non-homologous end joining” or “NHEJ” as used herein refers to ligation mediated repair and/or non-template mediated repair including canonical NHEJ (cNHEJ), alternative NHEJ (altNHEJ), microhomology-mediated end joining (MMEJ), single-strand annealing (SSA), and synthesis-dependent microhomology-mediated end joining (SD-MMEJ).

A “reference molecule” as used herein refers to a molecule to which a modified or candidate molecule is compared. For example, a reference Cas9 molecule refers to a Cas9 molecule to which a modified or candidate Cas9 molecule is compared. Likewise, a reference gRNA refers to a gRNA molecule to which a modified or candidate gRNA molecule is compared. The modified or candidate molecule may be compared to the reference molecule on the basis of sequence (e.g., the modified or candidate molecule may have X% sequence identity or homology with the reference molecule) or activity (e.g., the modified or candidate molecule may have X% of the activity of the reference molecule). For

example, where the reference molecule is a Cas9 molecule, a modified or candidate molecule may be characterized as having no more than 10% of the nuclease activity of the reference Cas9 molecule. Examples of reference Cas9 molecules include naturally occurring unmodified Cas9 molecules, e.g., a naturally occurring Cas9 molecule from *S. pyogenes*, *S. aureus*, *S. thermophilus*, or *N. meningitidis*. In certain embodiments, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology with the modified or candidate Cas9 molecule to which it is being compared. In certain embodiments, the reference Cas9 molecule is a parental molecule having a naturally occurring or known sequence on which a mutation has been made to arrive at the modified or candidate Cas9 molecule.

The term “genome editing system” refers to any system having RNA-guided DNA editing activity. Genome editing systems of the present disclosure include at least two components adapted from naturally occurring CRISPR systems: a guide RNA (gRNA) and an RNA-guided nuclease. These two components form a complex that is capable of associating with a specific nucleic acid sequence and editing the DNA in or around that nucleic acid sequence, for instance by making one or more of a single-strand break (an SSB or nick), a double-strand break (a DSB) and/or a point mutation.

“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.

“Subject” as used herein may mean a human, mouse, or non-human primate.

“Treat,” “treating,” and “treatment” as used herein mean the treatment of a disease in a subject, e.g., in a human, including (a) inhibiting the disease, i.e., arresting or preventing its development or progression; (b) relieving the disease, i.e., causing regression of the disease state; (c) relieving one or more symptoms of the disease; and (d) curing the disease. For example, “treating” SCD or β -Thal may refer to, among other possibilities, preventing development or progression of SCD or β -Thal, relieving one or more symptoms of SCD or β -Thal (e.g., anemia, sickle cell crises, vaso-occlusive crises), or curing SCD or β -Thal.

“Prevent,” “preventing,” and “prevention” as used herein means the prevention of a disease in a subject, e.g., in a human, including (a) avoiding or precluding the disease; (b) affecting the predisposition toward the disease; and (c) preventing or delaying the onset of at least one symptom of 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.

“Regulatory region” as used herein refers to a DNA sequence comprising one or more regulatory elements (e.g., silencer, enhancer, promoter, or insulator) controlling or regulating expression of a gene. For example, a γ -globin gene regulatory region comprises one or more regulatory elements controlling or regulating expression of a γ -globin gene. In certain embodiments, a regulatory region is adjacent to the gene being controlled or regulated. For example, a γ -globin gene regulatory region may be adjacent to or associated with the γ -globin gene. In other embodiments, the regulatory region may be adjacent to or associated with another gene, the expression of which can lead to up- or down-regulation of the gene being controlled or regulated. For example, a γ -globin gene regulatory region may be adjacent to a gene expressing a repressor of γ -globin gene expression. For HBG1, the regulatory region comprises at least nucleotides 1-2990 in SEQ ID NO:902. For HBG2, the regulatory region comprises at least nucleotides 1-2914 in SEQ ID NO:903.

“HBG target position” as used herein refers to a position in an HBG1 or HBG2 regulatory region (“HBG1 target position” and “HBG2 target position,” respectively) containing a target site (e.g., target sequence to be deleted or mutated) which, when altered (e.g., disrupted or deleted by introduction of a DNA repair mechanism-mediated (e.g., an NHEJ- or HDR-mediated) insertion or deletion, modified by a DNA repair mechanism-mediated (e.g., HDR-mediated) sequence alteration)) results in increased expression (e.g., de-repression) of HBG1 or HBG2 gene product (i.e., γ -globin). In certain embodiments, the HBG target position is in an HBG1 or HBG2 regulatory element (e.g., silencer, enhancer, promoter, or insulator) in a regulatory region adjacent to HBG1 or HBG2. In certain of these embodiments, alteration of the HBG target position results in decreased repressor binding, i.e., de-repression, leading to increased expression of HBG1 or HBG2. In other embodiments, the HBG target position is in a regulatory element of a gene other than HBG1 or HBG2 that encodes a gene product involved in controlling HBG1 or HBG2 gene expression (e.g., a repressor of HBG1 or HBG2 gene expression). In certain embodiments, the HBG target position is that region of an HBG1 or HBG2 regulatory region with the greatest density of binding motifs involved in the regulation of HBG1 or HBG2 expression. In certain embodiments, the methods provided herein target multiple HBG target positions simultaneously or sequentially.

“Target sequence” as used herein refers to a nucleic acid sequence comprising an HBG target position.

A “Cas9 molecule” or “Cas9 polypeptide” as used herein refers to a molecule or polypeptide, respectively, that can interact with a gRNA molecule and, in concert with the

gRNA molecule, localize to a site comprising a target domain and, in certain embodiments, a PAM sequence. Cas9 molecules and Cas9 polypeptides include both naturally occurring Cas9 molecules and Cas9 polypeptides and 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.

Overview

Provided herein are methods for increasing expression (i.e., transcriptional activity) of one or more γ -globin genes (e.g., HBG1, HBG2, or HBG1 and HBG2) using a genome editing system (e.g., CRISPR/Cas-mediated genome editing). These methods utilize a genome editing system (e.g., CRISPR/Cas-mediated genome editing) to alter (e.g., delete, disrupt, or modify) one or more γ -globin gene regulatory regions to increase (e.g., de-repress, enhance) γ -globin gene expression. In certain of these embodiments, the methods alter one or more regulatory elements (e.g., silencer, enhancer, promoter, or insulator) associated with the γ -globin gene being targeted. In other embodiments, the methods alter one or more regulatory elements in genes other than the γ -globin gene being targeted (e.g., genes encoding γ -globin gene repressors). In certain embodiments, a genome editing system (e.g., CRISPR/Cas-mediated genome editing) is used to alter a regulatory element (e.g., silencer, enhancer, promoter, or insulator) of HBG1, HBG2, or both HBG1 and HBG2. In certain embodiments, a genome editing system (e.g., CRISPR/Cas-mediated genome editing) results in a mutation or variation in an γ -globin regulatory element that is associated with a naturally occurring HPFH variant, including, for example, HBG1 13 bp del c.-114 to -102; 4 bp del c.-225 to -222; c.-114 C>T; c.-117 G>A; c.-158 C>T; c.-167 C>T; c.-170 G>A; c.-175 T>G; c.-175 T>C; c.-195 C>G; c.-196 C>T; c.-198 T>C; c.-201 C>T; c.-251 T>C; or c.-499 T>A; or HBG2 13 bp del c.-114 to -102; c.-109 G>T; c.-114 C>A; c.-114 C>T; c.-157 C>T; c.-158 C>T; c.-167 C>T; c.-167 C>A; c.-175 T>C; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-255 C>G; c.-309 A>G; c.-369 C>G; or c.-567 T>G.

In some embodiments, the methods using a genome editing system (e.g., CRISPR/Cas-mediated genome editing) described herein may utilize any repair mechanism to alter (e.g., delete, disrupt, or modify) all or a portion of one or more γ -globin gene regulatory elements. In certain embodiments, the methods utilize DNA repair mechanism-mediated (e.g., NHEJ or HDR-mediated) insertions or deletions to disrupt all or a portion of one or more γ -globin gene regulatory elements. For example, the methods may utilize a DNA repair mechanism (e.g., NHEJ or HDR) to delete all or a portion of a γ -globin gene

negative regulatory element (e.g., silencer), resulting in inactivation of the negative regulatory element (e.g., loss of binding between a silencer and repressor) and increased expression of the γ -globin gene. In other embodiments, the methods utilize DNA repair mechanism-mediated (e.g., NHEJ or HDR-mediated) insertions or deletions to disrupt all or a portion of one or more regulatory elements associated with a gene encoding a γ -globin gene repressor. For example, the methods may utilize a DNA repair mechanism (e.g., NHEJ or HDR) to delete all or a portion of a positive regulatory element (e.g., promoter) of a γ -globin repressor gene, resulting in decreased expression of the repressor, decreased binding of the repressor to a γ -globin gene silencer, and increased expression of the γ -globin gene. In other embodiments, the methods utilize a DNA repair mechanism (e.g., HDR) to modify the sequence of one or more γ -globin gene regulatory elements (e.g., inserting a mutation in an HBG1 and/or HBG2 regulatory element corresponding to a naturally occurring HPFH mutation or deleting all or a portion of an HBG1 and/or HBG2 regulatory element). In some embodiments, the methods may use a combination of one or more DNA repair mechanisms (e.g., NHEJ and HDR). In certain embodiments, the methods create persistence of HbF in a subject. Also provided herein are compositions (e.g., gRNAs, Cas9 polypeptides and molecules, template nucleic acids, vectors) and kits for use in these methods.

The transition from expression of γ -globin genes (i.e., HBG1, HBG2) to expression of HBB (i.e., globin switching) is associated with the onset of symptoms of β -hemoglobinopathies including SCD and β -Thal. Therefore, in certain embodiments, methods, compositions, and kits are provided herein for treating or preventing β -hemoglobinopathies including SCD and β -Thal using CRISPR/Cas-mediated genome editing to increase expression of one or more γ -globin genes (e.g., HBG1, HBG2, or HBG1 and HBG2). In certain of these embodiments, the methods alter one or more regulatory elements (e.g., silencer, enhancer, promoter, or insulator) associated with the γ -globin gene being targeted. In other embodiments, the methods alter one or more regulatory elements in genes other than the γ -globin gene being targeted (e.g., genes encoding γ -globin gene repressors). In certain embodiments, CRISPR/Cas-mediated genome editing is used to alter a regulatory element (e.g., silencer, enhancer, promoter, or insulator) of HBG1, HBG2, or both HBG1 and HBG2. In some embodiments, the methods utilize DNA repair mechanism-mediated (e.g., NHEJ or HDR-mediated) insertions or deletions to disrupt all or a portion of one or more γ -globin gene regulatory elements. For example, the methods may utilize a DNA repair mechanism (e.g., NHEJ or HDR) to delete all or a portion of a γ -globin gene negative regulatory element (e.g., silencer), resulting in inactivation of the negative regulatory element

(e.g., loss of binding between a silencer and repressor) and increased expression of the γ -globin gene. In other embodiments, the methods utilize DNA repair mechanism-mediated (e.g., NHEJ or HDR-mediated) insertions or deletions to disrupt all or a portion of one or more regulatory elements associated with a gene encoding a γ -globin gene repressor. For example, the methods may utilize a DNA repair mechanism (e.g., NHEJ or HDR) to delete all or a portion of a positive regulatory element (e.g., promoter) of a γ -globin repressor gene, resulting in decreased expression of the repressor, decreased binding of the repressor to a γ -globin gene silencer, and increased expression of the γ -globin gene. In other embodiments, the methods utilize a DNA repair mechanism (e.g., HDR) to modify the sequence of one or more γ -globin gene regulatory elements (e.g., inserting a mutation in an HBG1 and/or HBG2 regulatory element corresponding to a naturally occurring HPFH mutation or deleting all or a portion of an HBG1 and/or HBG2 regulatory element). In some embodiments, the methods may use a combination of one or more DNA repair mechanisms (e.g., NHEJ and HDR). In certain embodiments, the methods create persistence of HbF in a subject.

In certain embodiments, increased expression of one or more γ -globin genes (e.g., HBG1, HBG2) using the methods provided herein results in preferential formation of HbF over HbA and/or increased HbF levels as a percentage of total hemoglobin. Accordingly, further provided herein are methods of using CRISPR/Cas-mediated genome editing to increase total HbF levels, increase HbF levels as a percentage of total hemoglobin levels, or increase the ratio of HbF to HbA in a subject by increasing the expression of one or more γ -globin genes (e.g., HBG1, HBG2, or HBG1 and HBG2). Similarly, in certain embodiments increased expression of one or more γ -globin genes results in preferential formation of HbF versus HbS and/or decreased percentage of HbS as a percentage of total hemoglobin. Accordingly, further provided herein are methods of using CRISPR/Cas-mediated genome editing to decrease total HbS levels, decrease HbS levels as a percentage of total hemoglobin levels, or increase the ratio of HbF to HbS in a subject by increasing the expression of one or more γ -globin genes (e.g., HBG1, HBG2, or HBG1 and HBG2).

Provided herein in certain embodiments are gRNAs for use in the methods disclosed herein. In certain embodiments, these gRNAs comprise a targeting domain complementary or partially complementary to a target domain in or adjacent to an HBG target position. In certain embodiments, the targeting domain comprises, consists of, or consists essentially of a nucleotide sequence set forth in one of SEQ ID NOs:251-901.

Genomic studies have led to the identification of several genes that regulate globin switching, including BCL11A, Kruppel-like factor 1 (KLF1), MYB, and genes within the β

globin locus. Mutations in certain of these genes may result in inhibited or incomplete globin switching, also known as hereditary persistence of fetal hemoglobin (HPFH). HPFH mutations may be deletional or non-deletional (e.g., point mutations). Subjects with HPFH exhibit lifelong expression of HbF, i.e., they do not undergo or undergo only partial globin switching, with no symptoms of anemia. Heterozygous subjects exhibit 20-40% pancellular HbF, and co-inheritance results in alleviation of β -hemoglobinopathies (Thein 2009; Akinbami 2016). Compound heterozygotes for hemoglobinopathies and HPFH, e.g., subjects who are compound heterozygotes for SCD and HPFH, β -Thal and HPFH, sickle cell trait and HPFH, or delta- β -Thal and HPFH, have milder disease and symptoms relative to subjects without HPFH mutations. Patients homozygous for HbS who also co-inherit an HPFH mutation, e.g., a mutation that induces expression of HbF through de-repression of HBG1 or HBG2, do not develop SCD symptoms or β -Thal symptoms (Steinberg et al., Disorders of Hemoglobin, Cambridge Univ. Press, 2009, p. 570). HPFH is clinically benign (Chassanidis 2009).

While the occurrence of HPFH is rare in the global population, it is more common in populations with greater prevalence of hemoglobinopathies, including those of Southern European, South American, and African descent. In these populations, the prevalence of HPFH can reach 1-2 in 1,000 individuals (Costa 2002; Ahern 1973). Theoretically, HPFH mutations persist in these populations because they ameliorate disease in subjects with hemoglobinopathies.

The most common naturally occurring HPFH mutations are deletions within the β globin locus. Common examples of deletional HPFH mutations include French HPFH (23 kb deletion), Caucasian HPFH (19 kb deletion), HPFH-1 (84 kb deletion), HPFH-2 (84 kb deletion), and HPFH-3 (50 kb deletion). In subjects with these mutations, β -globin synthesis is reduced, and γ -globin synthesis is secondarily increased.

Other HPFH mutations are located in γ -globin gene regulatory regions. One such mutation is a 13 nucleotide deletion (13 base pair (bp) del c.-114 to -102; CAATAGCCTTGAC del, sequence based on reverse complement of HBG1/HBG2) located upstream of both the HBG1 and HBG2 genes. This deletion disrupts a silencer element that normally prevents HBG1/HBG2 expression, and adult subjects heterozygous for this deletion exhibit approximately 30% HbF. Another HPFH mutation is a 4 nucleotide deletion (4 base pair (bp) del c.-225 to -222 (AGCA del)). Other HPFH mutations found in both HBG1 and HBG2 regulatory elements include, for example, non-deletional point mutations (non-del HPFH) such as c.-114 C>T; c.-158 C>T; c.-167 C>T; and c.-175 T>C.

Non-del HPFH mutations associated with HBG1 regulatory elements include, for example, c.-117 G>A; c.-170 G>A; c.-175 T>G; c.-195 C>G; c.-196 C>T; c.-198 T>C; c.-201 C>T; c.-251 T>C; and c.-499 T>A.

Non-del HPFH mutations associated with HBG2 regulatory elements include, for example, c.-109 G>T; c.-114 C>A; c.-157 C>T; c.-167 C>A; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-255 C>G; and c.-567 T>G.

Additional polymorphisms in HBG1 and HBG2 promoter regions have been identified in a cohort of Brazilian SCD patients that corrected HbF levels >5% (Barbosa 2010). These include c.-309 A>G and c.-369 C>G in the HBG2 promoter.

HBG1 and HBG2 promoter elements that may be altered to recreate HPFH mutations include, for example, erythroid Kruppel-like factor (EKLF-2) and fetal Kruppel-like factor (FKLF) transcription factor binding motifs (CTCCACCCA), CP1/Coup TFII binding motifs (CCAATAGC), GATA1 binding motifs (CTATCT, ATATCT), or stage selector element (SSE) binding motifs. HBG1 and HBG2 enhancer elements that may be altered to recreate HPFH mutations include, for example, SOX binding motifs, e.g., SOX14, SOX2, or SOX1 (CCAATAGCCTTGA).

In certain embodiments of the methods provided herein, CRISPR/Cas-mediated alteration is used to alter one regulatory element or motif in a γ -globin gene regulatory region, e.g., a silencer sequence in an HBG1 or HBG2 regulatory region, or a promoter or enhancer sequence associated with a gene encoding an HBG1 or HBG2 repressor. In other embodiments, CRISPR/Cas-mediated alteration is used to alter two or more (e.g., three, four, or five or more) regulatory elements or motifs in a γ -globin gene regulatory region, e.g., an HBG1 or HBG2 silencer sequence and an HBG1 or HBG2 enhancer sequence; an HBG1 or HBG2 silencer sequence and a promoter or enhancer sequence associated with a gene encoding an HBG1 or HBG2 repressor; or an HBG1 or HBG2 silencer sequence and a promoter or enhancer sequence associated with a gene encoding an HBG1 or HBG2 repressor. The introduction of multiple variants into the regulatory region of a single gene or the introduction of one variant into the regulatory regions of two or more genes is referred to herein as “multiplexing.” Thus, multiplexing constitutes either (a) the modification of more than one location in one gene regulatory region in the same cell or cells or (b) the modification of one location in more than one gene regulatory region.

In certain embodiments of the methods provided herein, CRISPR/Cas-mediated alteration of one or more γ -globin gene regulatory elements produces a phenotype the same as or similar to a phenotype associated with a naturally occurring HPFH mutation. In certain

embodiments, CRISPR/Cas-mediated alteration results in a γ -globin gene regulatory element comprising an alteration corresponding to a naturally occurring HPFH mutation. In other embodiments, alterations of one or more γ -globin gene regulatory elements results in an alteration that is not observed in a naturally occurring HPFH mutation (i.e., a non-naturally occurring variant).

In certain embodiments of the methods provided herein, CRISPR/Cas-mediated alteration of one or more γ -globin gene regulatory elements produces a mutation or variation in an γ -globin regulatory element that is associated with a naturally occurring HPFH variant, including, for example, HBG1 13 bp del c.-114 to -102; 4 bp del c.-225 to -222; c.-114 C>T; c.-117 G>A; c.-158 C>T; c.-167 C>T; c.-170 G>A; c.-175 T>G; c.-175 T>C; c.-195 C>G; c.-196 C>T; c.-198 T>C; c.-201 C>T; c.-251 T>C; or c.-499 T>A; or HBG2 13 bp del c.-114 to -102; c.-109 G>T; c.-114 C>A; c.-114 C>T; c.-157 C>T; c.-158 C>T; c.-167 C>T; c.-167 C>A; c.-175 T>C; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-255 C>G; c.-309 A>G; c.-369 C>G; or c.-567 T>G.

In certain embodiments, the methods provided herein comprise altering one or more transcription factor binding motifs (e.g., gene regulatory motif) in a γ -globin gene regulatory element. These transcription factor binding motifs include, for example, binding motifs that are occupied by transcription factors (TFs), TF complexes, and transcriptional repressors within the promoter regions of HBG1 and/or HBG2. In certain embodiments of the methods provided herein, introduction of a CRISPR/Cas-mediated alteration in one or more γ -globin gene regulatory elements alters binding of a transcription factor, e.g., a repressor, at one, two, three, or more than three motifs. In certain embodiments, introduction of a CRISPR/Cas-mediated alteration in one or more γ -globin gene regulatory elements results in increased RNA polymerase II initiation of transcription proximal to or at a γ -globin gene promoter region, e.g., by increasing transcription factor binding to an enhancer region, e.g., by decreased repressor binding at a silencer region.

In certain embodiments, the methods provided herein utilize a DNA repair mechanism-mediated (e.g., NHEJ- or HDR-mediated) deletion to delete all or a portion of nucleotides -114 to -102 in one or both alleles of HBG1, HBG2, or both HBG1 and HBG2, resulting in an HPFH phenotype the same as or similar to that associated with the naturally occurring 13 bp del c.-114 to -102 mutation. In other embodiments, a DNA repair mechanism-mediated (e.g., NHEJ- or HDR-mediated) deletion is utilized to delete all or a portion of nucleotides -225 to -222 of one or both alleles of HBG1, resulting in an HPFH phenotype the same as or similar to that associated with the naturally occurring HBG1 4 bp

del -225 to -222 mutation. In other embodiments, a DNA repair mechanism-mediated (e.g., NHEJ- or HDR-mediated) deletion is utilized to delete all or a portion of nucleotides -225 to -222 of one or both alleles of HBG2.

In certain embodiments, the methods provided herein utilize a DNA repair mechanism-mediated (e.g., NHEJ- or HDR-mediated) deletion to delete all or a portion of nucleotides -114 to -102 in one or both alleles of HBG1 and one or both alleles of HBG2.

In certain embodiments, the methods provided herein utilize a DNA repair mechanism-mediated (e.g., NHEJ or HDR-mediated) deletion to delete all or a portion of nucleotides -225 to -222 in one or both alleles of HBG1 and all or a portion of nucleotides -114 to -102 in one or both HBG2 alleles. In other embodiments, a DNA repair mechanism (e.g., NHEJ- or HDR-mediated deletion) is utilized to delete all or a portion of nucleotides -225 to -222 in one or both alleles of HBG1 and all or a portion of nucleotides -114 to -102 in one or both HBG1 alleles.

In those embodiments wherein a DNA repair mechanism-mediated (e.g., NHEJ- or HDR-mediated) deletion is used to delete one or more nucleotides from HBG1, HBG2, or HBG1 and HBG2 regulatory elements, the deletions may be identical to those observed in naturally occurring HPFH mutations, i.e., the deletion may consist of nucleotides -114 to -102 of HBG1 or HBG2, or nucleotides -225 to -222 of HBG1. In other embodiments, the DNA repair mechanism-mediated (e.g., the NHEJ- or HDR-mediated) deletion results in removal of only a portion of these nucleotides, e.g., deletion of 12 or fewer nucleotides falling within -114 to -102 of HBG1 or HBG2 or three or fewer nucleotides falling within -225 to -222 of HBG1. In certain embodiments, one or more nucleotides may be knocked out on either side of the naturally occurring HPFH mutation deletion boundaries (i.e., outside of -114 to -102 or -225 to -222) in addition to all or a portion of the nucleotides within the naturally occurring deletion boundaries.

In certain embodiments, the methods provided herein utilize a DNA repair mechanism-mediated (e.g., NHEJ- or HDR-mediated) insertion to insert one or more nucleotides into the region spanning nucleotides -114 to -102 of an HBG1 regulatory region, HBG2 regulatory region, or both HBG1 and HBG2 regulatory regions, or the region spanning nucleotides -225 to -222 of an HBG1 regulatory region, in order to disrupt a repressor binding site.

In certain embodiments, the methods provided herein utilize a DNA repair mechanism (e.g., HDR) to generate single nucleotide alterations (i.e., non-deletion mutants) corresponding to naturally occurring mutations associated with HPFH. For example, in

certain embodiments the methods utilize a DNA repair mechanism (e.g., HDR) to generate a single nucleotide alteration in an HBG1 regulatory region that corresponds to a naturally occurring mutation associated with HPFH, including for example c.-114 C>T; c.-117 G>A; c.-158 C>T; c.-167 C>T; c.-170 G>A; c.-175 T>G; c.-175 T>C; c.-195 C>G; c.-196 C>T; c.-198 T>C; c.-201 C>T; c.-251 T>C; or c.-499 T>A. In other embodiments, a DNA repair mechanism (e.g., HDR) is utilized to generate a single nucleotide alteration in an HBG2 regulatory region that corresponds to a naturally occurring mutation associated with HPFH, including for example c.-109 G>T; c.-114 C>A; c.-114 C>T; c.-157 C>T; c.-158 C>T; c.-167 C>T; c.-167 C>A; c.-175 T>C; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-255 C>G; c.-309 A>G; c.-369 C>G; c.-567 T>G.

In certain embodiments, a DNA repair mechanism (e.g., HDR) is utilized to generate a single nucleotide alteration in an HBG1 regulatory region corresponding to a naturally occurring HPFH mutation found in an HBG2 regulatory region but not an HBG1 regulatory region. Such alterations include, for example, c.-109 G>T; c.-114 C>A; c.-157 C>T; c.-167 C>A; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-255 C>G; c.-309 A>G; c.-369 C>G; or c.-567 T>G.

Likewise, in certain embodiments a DNA repair mechanism (e.g., HDR) is utilized to generate a single nucleotide alteration in an HBG2 regulatory region corresponding to a naturally occurring HPFH mutation found in an HBG1 regulatory region but not an HBG2 regulatory region. Such alterations include, for example, c.-117 G>A; c.-170 G>A; c.-175 T>G; c.-195 C>G; c.-196 C>T; c.-198 T>C; c.-201 C>T; c.-251 T>C; or c.-499 T>A.

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-114 C>T into an HBG1 and/or HBG2 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-158 C>T (i.e., rs7482144 or XmnI-HBG2 variant) into an HBG1 and/or HBG2 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-167 C>T into an HBG1 and/or HBG2 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-175 T>C (i.e., T→C substitution at position c.-175 in a conserved octamer [ATGCAAAT] sequence) into an HBG1 regulatory region by a DNA repair mechanism (e.g., HDR). This variant, which is associated with 40% HbF, has been shown to

abolish the ability of a ubiquitous octamer binding nuclear protein to bind the HBG promoter fragment, while simultaneously increased the ability of two erythroid specific proteins to bind the same fragment by 3-5 fold (Mantovani 1988).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-175 T>C into an HBG2 regulatory region by a DNA repair mechanism (e.g., HDR). This variant is associated with 20-30% HbF expression.

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-117 G>A into an HBG1 regulatory region by a DNA repair mechanism (e.g., HDR). This variant, referred to as "Greek type," is the most common nondeletion HPFH mutant and maps two nucleotides upstream from the distal CCAAT box (Waber 1986). HBG1 c.-117 G>A greatly decreases binding of erythroid-specific factors, but not of the ubiquitous protein, to the CCAAT box region fragment, and is associated with 10-20% HbF (Mantovani 1988). The mutation is thought to interfere with binding of nuclear factor E (NF-E), which is likely to play a role in repression of γ -globin transcription in adult erythroid cells (Superti-Furga 1988). In other embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-117 G>A into an HBG2 regulatory region, creating a non-naturally occurring HPFH variant.

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-170 G>A into an HBG1 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-175 T>G into an HBG1 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-195 C>G into an HBG1 regulatory region.

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-196 C>T into an HBG1 regulatory region by a DNA repair mechanism (e.g., HDR). This variant is associated with 10-20% HbF.

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-198 T>C into an HBG1 regulatory region by a DNA repair mechanism (e.g., HDR). This variant is associated with 18-21% HbF.

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-201 C>T into an HBG1 regulatory region.

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-251 T>C into an HBG1 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-499 T>A into an HBG1 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-109 G>T ("Hellenic mutation") into an HBG2 regulatory region by a DNA repair mechanism (e.g., HDR). This mutation is located at the 3' end of the HBG2 CCAAT box in the promoter region (Chassanidis 2009).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-114 C>A into an HBG2 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-157 C>T into an HBG2 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-167 C>A into an HBG2 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-202 C>G into an HBG2 regulatory region by a DNA repair mechanism (e.g., HDR). This variant is associated with 15-25% HbF expression.

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-211 C>T into an HBG2 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-228 T>C into an HBG2 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-255 C>G into an HBG2 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-309 A>G into an HBG2 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-369 C>G into an HBG2 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise inserting the non-deletion HPFH variant c.-567 T>G into an HBG2 regulatory region by a DNA repair mechanism (e.g., HDR).

In certain embodiments, the methods provided herein comprise deletion, disruption, or mutation of a BCL11a core binding motif (i.e., GGCCGG) located at position c.-56 relative to HBG1 and/or HBG2 and/or at another location in a γ -globin gene regulatory region.

In certain embodiments, the methods provided herein comprise altering one or more nucleotides in a GATA (e.g., GATA1) motif. In certain of these embodiments, a DNA repair mechanism (e.g., HDR) is used to insert a T>C mutation into the HBG1 GATA binding motif within the sequence AAATATCTGT, resulting in the altered sequence AAACATCTGT. This naturally occurring T>C HPFH mutation is associated with 40% HbF.

In certain embodiments, the methods provided herein utilize one or more DNA repair mechanism (e.g., both NHEJ and HDR) approaches. For example, in certain embodiments, the methods utilize NHEJ-mediated deletion, e.g., introduction of 13 bp del c.-114 to -102 into one or both alleles of HBG1 and/or HBG2 and/or 4 bp del c.-225 to -222 into one or both alleles of HBG1, in combination with HDR-mediated single nucleotide alteration, e.g., introduction of one or more of c.-109 G>T; c.-114 C>A; c.-114 C>T; c.-117 G>A; c.-157 C>T; c.-158 C>T; c.-167 C>T; c.-167 C>A; c.-170 G>A; c.-175 T>C; c.-175 T>G; c.-195 C>G; c.-196 C>T; c.-198 T>C; c.-201 C>T; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-251 T>C; c.-255 C>G; c.-309 A>G; c.-369 C>G; c.-499 T>A; or c.-567 T>G into one or both alleles of HBG1 and/or HBG2.

In certain embodiments, the methods utilize HDR-mediated deletion, e.g., introduction of 13 bp del c.-114 to -102 into one or both alleles of HBG1 and/or HBG2 and/or 4 bp del c.-225 to -222 into one or both alleles of HBG1, in combination with HDR-mediated single nucleotide alteration, e.g., introduction of one or more of c.-109 G>T; c.-114 C>A; c.-114 C>T; c.-117 G>A; c.-157 C>T; c.-158 C>T; c.-167 C>T; c.-167 C>A; c.-170 G>A; c.-175 T>C; c.-175 T>G; c.-195 C>G; c.-196 C>T; c.-198 T>C; c.-201 C>T; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-251 T>C; c.-255 C>G; c.-309 A>G; c.-369 C>G; c.-499 T>A; or c.-567 T>G into one or both alleles of HBG1 and/or HBG2.

While not wishing to be bound by theory, introduction of 4 bp del c.-225 to -222 into the HBG1 gene regulatory region reverses the normal ratio of 70% γ^A -globin (γ -globin product of the HBG1 gene) to 30% γ^G -globin (γ -globin product of the HBG2 gene), so that γ -globin is produced as approximately 30% γ^A -globin and 70% γ^G -globin. While not wishing to be bound by theory, reversal of γ^G -globin and γ^A -globin ratio results in increased production of γ^G -globin in a subject. While not wishing to be bound by theory, concomitant introduction of 4 bp del c.-225 to -222 into the HBG1 gene regulatory region and 13 bp del c.-114 to -102 into the HBG2 gene regulatory region leads to increased transcriptional activity of HBG2, increased production of γ^G -globin, and increased HbF in a subject. While not wishing to be bound by theory, concomitant introduction of (a) 4 bp del c.-225 to -222 into the HBG1 gene regulatory region, e.g., by NHEJ- or HDR-mediated deletion, and (b) a non-deletion HPFH variant, e.g., by HDR, e.g., c.-109 G>T; c.-114 C>T; c.-114 C>A; c.-157 C>T; c.-158 C>T; c.-167 C>T; c.-167 C>A; c.-175 T>C; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-255 C>G; c.-309 A>G; c.-369 C>G; c.-567 T>G, into the HBG2 gene regulatory region leads to increased transcriptional activity of HBG2, increased production of γ^G -globin and increased HbF in a subject.

While not wishing to be bound by theory, introduction of 4 bp del c.-225 to -222 into the HBG2 gene regulatory region may decrease the production of γ^G -globin (γ -globin product of the HBG2 gene) relative to production of γ^A -globin (γ -globin product of the HBG1 gene), so that more γ^A -globin is produced than by γ^G -globin. While not wishing to be bound by theory, concomitant introduction of 4 bp del c.-225 to -222 into the HBG2 gene regulatory region and 13 bp del c.-114 to -102 into the HBG1 gene regulatory region may lead to increased transcriptional activity of HBG1, increased production of γ^A -globin and increased HbF in a subject. While not wishing to be bound by theory, concomitant introduction of (a) 4 bp del c.-225 to -222 into the HBG2 gene regulatory region, e.g., by NHEJ- or HDR-mediated deletion, and (b) a non-deletion HPFH variant, e.g., by HDR, e.g., c.-114 C>T; c.-117 G>A; c.-158 C>T; c.-167 C>T; c.-170 G>A; c.-175 T>G; c.-175 T>C; c.-195 C>G; c.-196 C>T; c.-198 T>C; c.-201 C>T; c.-251 T>C; or c.-499 T>A, into the HBG1 gene regulatory region may lead to increased transcriptional activity of HBG1, increased production of γ^A -globin, and increased HbF in a subject.

While not wishing to be bound by theory, concomitant introduction of (a) 13 bp del c.-114 to -102 into the HBG1 gene regulatory region, e.g., by NHEJ- or HDR-mediated deletion, and (b) a non-deletion HPFH variant, e.g., by HDR, e.g., c.-109 G>T; c.-114 C>T;

c.-114 C>A; c.-157 C>T; c.-158 C>T; c.-167 C>A; c.-167 C>T; c.-175 T>C; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-255 C>G; c.-309 A>G; c.-369 C>G; or c.-567 T>G, into the HBG2 gene regulatory region leads to increased transcriptional activity of HBG2, increased production of γ^G -globin, and increased HbF in a subject.

While not wishing to be bound by theory, concomitant introduction of (a) 13 bp del c.-114 to -102 into the HBG2 gene regulatory region, e.g., by NHEJ- or HDR-mediated deletion, and (b) a non-deletion HPFH variant, e.g., by HDR, e.g., c.-114 C>T; c.-117 G>A; c.-158 C>T; c.-167 C>T; c.-170 G>A; c.-175 T>C; c.-175 T>G; c.-195 C>G; c.-196 C>T; c.-198 T>C; c.-201 C>T; c.-251 T>C; or c.-499 T>A, into the HBG1 gene regulatory region leads to increased transcriptional activity of HBG1, increased production of γ^A -globin and increased HbF in a subject.

Concomitant (a) BCL11A knockdown by siRNA and (b) SOX6 knockdown by siRNA leads to increased expression of HBG1 and HBG2 (Xu 2010). In certain embodiments, the methods provided herein comprise disrupting the action of BCL11A, SOX6, or BCL11A and SOX6 on the expression of HBG1 and HBG2 using a DNA repair mechanism (e.g., HDR, NHEJ, or NHEJ and HDR) modification of the HBG1 and HBG2 promoter regions and the erythroid-specific enhancer of BCL11A, alone or in parallel. In certain embodiments, the methods provided herein comprise decreasing BCL11A expression by disrupting the function of its intronic erythroid-specific enhancer by NHEJ and HDR and simultaneously inducing HPFH mutations for a synergistic effect on the production of HbF.

The embodiments described herein may be used in all classes of vertebrate including, but not limited to, primates, mice, rats, rabbits, pigs, dogs, and cats.

Timing and Subject Selection

Initiation of treatment using the methods disclosed herein may occur prior to disease onset, for example in a subject who has been deemed at risk of developing a β -hemoglobinopathy (e.g., SCD, β -Thal) based on genetic testing, familial history, or other factors, but who has not yet displayed any manifestations or symptoms of the disease. In certain of these embodiments, treatment may be initiated prior to naturally occurring globin switching, i.e., prior to the transition from predominantly HbF to predominantly HbA. In other embodiments, treatment may be initiated after naturally occurring globin switching has occurred.

In certain embodiments, treatment is initiated after disease onset, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 16, 24, 36, or 48 or more months after onset of SCD or β -Thal or one or more symptoms associated therewith. In certain of these embodiments, treatment is initiated at an early stage of disease progression, e.g., when a subject has displayed only minor symptoms or only a subset of symptoms. Exemplary symptoms include, but are not limited to, anemia, diarrhea, fever, failure to thrive, sickle cell crises, vaso-occlusive crises, aplastic crises, and acute chest syndrome anemia, vaso-occlusion, hepatomegaly, thrombosis, pulmonary embolus, stroke, leg ulcer, cardiomyopathy, cardiac arrhythmia, splenomegaly, delayed bone growth and/or puberty, and evidence of extramedullary erythropoiesis. In other embodiments, treatment is initiated well after disease onset or at a more advanced stage of disease progression, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 16, 24, 36, or 48 or more months after onset of SCD or β -Thal. While not wishing to be bound by theory, it is believed that this treatment will be effective if subjects present well into the course of illness.

In certain embodiments, the methods provided herein prevent or slow the development of one or more symptoms associated with the disease being treated. In certain embodiments, the methods provided herein result in prevention or delay of disease progression as compared to a subject who has not received the therapy. In certain embodiments, the methods provided herein result in the disease being cured entirely.

In certain embodiments, the methods provided herein are performed on a one-time basis. In other embodiments, the methods provided herein utilize multi-dose therapy.

In certain embodiments, a subject being treated using the methods provided herein is transfusion-dependent.

In certain embodiments, the methods provided herein comprise altering expression of one or more γ -globin genes (e.g., HBG1, HBG2) using CRISPR/Cas-mediated genome editing in a cell *in vivo*. In other embodiments, the methods provided herein comprise altering expression of one or more γ -globin genes using CRISPR/Cas-mediated genome editing in a cell *ex vivo*, then transplanting the cell into a subject. In certain of these embodiments, the cell is originally from the subject. In certain embodiments, the cell undergoing alteration is an adult erythroid cell. In other embodiments, the cell is a hematopoietic stem cell (HSC).

In certain embodiments, the methods provided herein comprise delivery to a cell of one or more gRNA molecules and one or more Cas9 polypeptides or nucleic acid sequences encoding a Cas9 polypeptide. In certain embodiments, the methods further comprise delivery of one or more nucleic acids, e.g., HDR donor templates.

In certain embodiments, one or more of these components (i.e., one or more gRNA molecules, one or more Cas9 polypeptides or nucleic acid sequences encoding a Cas9 polypeptide, and one or more nucleic acids, e.g., HDR donor templates) are delivered using one or more AAV vectors, lentiviral vectors, nanoparticles, or a combination thereof.

In certain embodiments, the methods provided herein are performed on a subject who has one or more mutations in an HBB gene, including one or more mutations associated with a β -hemoglobinopathy such as SCD or β -Thal. Examples of such mutations include, but are not limited to, c.17A>T, c.-136C>G, c.92+1G>A, c.92+6T>C, c.93-21G>A, c.118C>T, c.316-106C>G, c.25_26delAA, c.27_28insG, c.92+5G>C, c.118C>T, c.135delC, c.315+1G>A, c.-78A>G, c.52A>T, c.59A>G, c.92+5G>C, c.124_127delTTCT, c.316-197C>T, c.-78A>G, c.52A>T, c.124_127delTTCT, c.316-197C>T, c.-138C>T, c.-79A>G, c.92+5G>C, c.75T>A, c.316-2A>G, and c.316-2A>C.

NHEJ-mediated introduction of an indel within a γ -globin gene regulatory element

In certain embodiments, the methods provided herein utilize NHEJ-mediated insertions or deletions to disrupt all or a portion of a γ -globin gene regulatory element in order to increase expression of the γ -globin gene (e.g., HBG1, HBG2, or HBG1 and HBG2).

In certain embodiments, methods provided herein that utilize NHEJ comprise deletion or disruption of all or a portion of an HBG1 or HBG2 silencer element via NHEJ, resulting in inactivation of the silencer and a subsequent increase in HBG1 and/or HBG2 expression. In certain embodiments, NHEJ-mediated deletion results in removal of all or a part of c.-114 to -102 or -225 to -222 in one or both alleles of HBG1 and/or removal of all or a part of c.-114 to -102 in one or both alleles of HBG2. In certain of these embodiments, one or more nucleotides 5' or 3' to these regions are also deleted.

In certain embodiments, methods provided herein that utilize NHEJ comprise introduction of one or more breaks (e.g., single strand breaks or double strand breaks) within a γ -globin gene regulatory region, and in certain of these embodiments the one or more breaks are located sufficiently close to an HBG target position that a break-induced indel could be reasonably expected to span all or part of the HBG target position.

In certain embodiments, the targeting domain of a first gRNA molecule is configured to provide a cleavage event, e.g., a double strand break or a single strand break, sufficiently close to an HBG target position to allow NHEJ-mediated insertion or deletion at the HBG target position. In certain embodiments, the gRNA targeting domain is configured such that a cleavage event, e.g., a double strand or single strand break, is positioned within 1, 2, 3, 4, 5,

10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of an HBG target position. The break, e.g., a double strand or single strand break, can be positioned upstream or downstream of an HBG target position.

In certain embodiments, a second gRNA molecule comprising a second targeting domain is configured to provide a cleavage event, e.g., a double strand break or a single strand break, sufficiently close to an HBG target position to allow NHEJ-mediated insertion or deletion at the HBG target position, either alone or in combination with the break positioned by said first gRNA molecule. In certain embodiments, the targeting domains of the first and second gRNA molecules are configured such that a cleavage event, e.g., a double strand or single strand break, is positioned, independently for each of the gRNA molecules, within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position. In certain embodiments, the breaks, e.g., double strand or single strand breaks, are positioned on either side of a nucleotide of an HBG target position. In other embodiments, the breaks, e.g., double strand or single strand breaks, are both positioned on one side, e.g., upstream or downstream, of a nucleotide of an HBG target position.

In certain embodiments, a single strand break is accompanied by an additional single strand break, positioned by a second gRNA molecule, as discussed below. For example, the gRNA targeting domains may be configured such that a cleavage event, e.g., two single strand breaks, is positioned within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of an HBG target position. In certain embodiments, the first and second gRNA molecules are configured such that, when guiding a Cas9 nickase, a single strand break will be accompanied by an additional single strand break, positioned by a second gRNA, sufficiently close to one another to result in alteration of the HBG target position. In certain embodiments, the first and second gRNA molecules are configured such that a single strand break positioned by said second gRNA is within 10, 20, 30, 40, or 50 nucleotides of the break positioned by said first gRNA molecule, e.g., when the Cas9 is a nickase. In certain embodiments, the two gRNA molecules are configured to position cuts at the same position, or within a few nucleotides of one another, on different strands, e.g., essentially mimicking a double strand break.

In certain embodiments, a double strand break can be accompanied by an additional double strand break, positioned by a second gRNA molecule, as is discussed below. For example, the targeting domain of a first gRNA molecule is configured such that a double strand break is positioned upstream of HBG target position, e.g., within 1, 2, 3, 4, 5, 10, 15,

20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position; and the targeting domain of a second gRNA molecule is configured such that a double strand break is positioned downstream of the HBG target position, e.g., within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position.

In certain embodiments, a double strand break can be accompanied by two additional single strand breaks, positioned by a second gRNA molecule and a third gRNA molecule. For example, the targeting domain of a first gRNA molecule is configured such that a double strand break is positioned upstream of the HBG target position, e.g., within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position; and the targeting domains of a second and third gRNA molecule are configured such that two single strand breaks are positioned downstream of the HBG target position, e.g., within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position. In certain embodiments, the targeting domains of the first, second and third gRNA molecules are configured such that a cleavage event, e.g., a double strand or single strand break, is positioned, independently for each of the gRNA molecules.

In certain embodiments, a first and second single strand break can be accompanied by two additional single strand breaks positioned by a third and fourth gRNA molecule. For example, the targeting domains of a first and second gRNA molecule are configured such that two single strand breaks are positioned upstream of the HBG target position, e.g., within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position; and the targeting domains of a third and fourth gRNA molecule are configured such that two single strand breaks are positioned downstream of the HBG target position, e.g., within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position.

In certain embodiments, the methods provided herein comprise introducing an NHEJ-mediated deletion of a genomic sequence including an HBG target position. In certain embodiments, the methods comprise introduction of two double strand breaks, one 5' to and the other 3' to (i.e., flanking) the HBG target position. Two gRNAs, e.g., unimolecular (or chimeric) or modular gRNA molecules, are configured to position the two double strand breaks on opposite sides of the HBG target position. In certain embodiments, the first double strand break is positioned upstream of the mutation, and the second double strand break is positioned downstream of the mutation. In certain embodiments, the two double strand

breaks are positioned to remove all or a portion of HBG1 c.-114 to -102, HBG1 4 bp del -225 to -222. In an embodiment, the breaks (i.e., the two double strand breaks) are positioned to avoid unwanted target chromosome elements, such as repeat elements, e.g., an *Alu* repeat, or the endogenous splice sites.

In other embodiments, the methods comprise the introduction of two sets of breaks, one double strand break and a pair of single strand breaks. The two sets flank the HBG target position, i.e., one set is 5' to and the other is 3' to the HBG target position. Two gRNAs, e.g., unimolecular (or chimeric) or modular gRNA molecules, are configured to position the two sets of breaks (either the double strand break or the pair of single strand breaks) on opposite sides of the HBG target position. In certain embodiments, the breaks (i.e., the two sets of breaks (either the double strand break or the pair of single strand breaks)) are positioned to avoid unwanted target chromosome elements, such as repeat elements, e.g., an *Alu* repeat, or the endogenous splice sites.

In other embodiments, the methods comprise the introduction of two pairs of single strand breaks, one 5' to and the other 3' to (i.e., flanking) the HBG target position. Two gRNAs, e.g., unimolecular (or chimeric) or modular gRNA molecules, are configured to position the two sets of breaks on opposite sides of the HBG target position. In certain embodiments, the breaks (i.e., the two pairs of single strand breaks) are positioned to avoid unwanted target chromosome elements, such as repeat elements, e.g., an *Alu* repeat, or the endogenous splice sites.

HDR-mediated introduction of a sequence alteration in a γ -globin gene regulatory element

In certain embodiments, the methods provided herein utilize HDR to modify one or more nucleotides in a γ -globin gene regulatory element in order to increase expression of the γ -globin gene (e.g., HBG1, HBG2, or HBG1 and HBG2). In certain of these embodiments, HDR is utilized to incorporate one or more nucleotide modifications corresponding to naturally occurring mutations associated with HPFH. For example, in certain embodiments HDR is used to incorporate one or more of the following single nucleotide alterations into an HBG1 regulatory region: c.-114 C>T; c.-117 G>A; c.-158 C>T; c.-167 C>T; c.-170 G>A; c.-175 T>C; c.-175 T>G; c.-195 C>G; c.-196 C>T; c.-198 T>C; c.-201 C>T; c.-251 T>C; or c.-499 T>A. In other embodiments, HDR is used to incorporate one or more of the following single nucleotide alterations into an HBG2 regulatory region: c.-109 G>T; c.-114 C>A; c.-114 C>T; c.-157 C>T; c.-158 C>T; c.-167 C>T; c.-167 C>A; c.-175 T>C; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-255 C>G; c.-309 A>G; c.-369 C>G; c.-567 T>G.

In certain embodiments, the methods provided herein utilize HDR-mediated alteration (e.g., insertions or deletions) to disrupt all or a portion of a γ -globin gene regulatory element in order to increase expression of the γ -globin gene (e.g., HBG1, HBG2, or HBG1 and HBG2).

In certain embodiments, methods provided herein that utilize HDR comprise deletion or disruption of all or a portion of an HBG1 or HBG2 silencer element via HDR, resulting in inactivation of the silencer and a subsequent increase in HBG1 and/or HBG2 expression. In certain embodiments, HDR-mediated deletion results in removal of all or a part of c.-114 to -102 or -225 to -222 in one or both alleles of HBG1 and/or removal of all or a part of c.-114 to -102 in one or both alleles of HBG2. In certain of these embodiments, one or more nucleotides 5' or 3' to these regions are also deleted.

In certain embodiments, methods provided herein that utilize HDR comprise introduction of one or more breaks (e.g., single strand breaks or double strand breaks) within a γ -globin gene regulatory region, and in certain of these embodiments the one or more breaks are located sufficiently close to an HBG target position that a break-induced alteration could be reasonably expected to span all or part of the HBG target position.

In certain embodiments, HDR-mediated alteration may include the use of a template nucleic acid.

In certain embodiments, an HDR-mediated genetic alteration is incorporated into one γ -globin gene allele (e.g., one allele of HBG1 and/or HBG2). In another embodiment, the genetic alteration is incorporated into both alleles (e.g., both alleles of HBG1 and/or HBG2). In either situation, the treated subject exhibits increased γ -globin gene expression (e.g., HBG1, HBG2, or HBG1 and HBG2 expression).

In certain embodiments, methods provided herein that utilize HDR comprise introduction of one or more breaks (e.g., single strand breaks or double strand breaks) sufficiently close to (e.g., either 5' or 3' to) an HBG target position to allow for an alteration associated with HDR at the target position.

In certain embodiments, the targeting domain of a first gRNA molecule is configured to provide a cleavage event, e.g., a double strand break or a single strand break, sufficiently close to an HBG target position to allow for an alteration associated with HDR at the target position. In certain embodiments, the gRNA targeting domain is configured such that a cleavage event, e.g., a double strand or single strand break, is positioned within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or

500 nucleotides of an HBG target position. The break, e.g., a double strand or single strand break, can be positioned upstream or downstream of an HBG target position.

In certain embodiments, a second, third, and/or fourth gRNA molecule is configured to provide a cleavage event, e.g., a double strand break or a single strand break, sufficiently close to (e.g., either 5' or 3' to) an HBG target position to allow for an alteration associated with HDR at the target position. In certain embodiments, the gRNA targeting domain is configured such that a cleavage event, e.g., a double strand or single strand break, is positioned within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450 or 500 nucleotides of an HBG target position. The break, e.g., a double strand or single strand break, can be positioned upstream or downstream of the target position.

In certain embodiments, a single strand break is accompanied by an additional single strand break, positioned by a second, third and/or fourth gRNA molecule. For example, the gRNA targeting domains may be configured such that a cleavage event, e.g., the two single strand breaks, is positioned within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of an HBG target position. In certain embodiments, the first and second gRNA molecules are configured such that when guiding a Cas9 nickase, a single strand break will be accompanied by an additional single strand break positioned by a second gRNA sufficiently close to the first strand break to result in alteration of the HBG target position. In certain embodiments, the first and second gRNA molecules are configured such that a single strand break positioned by said second gRNA is within 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides of the break positioned by said first gRNA molecule, e.g., when the Cas9 is a nickase. In certain embodiments, the two gRNA molecules are configured to position cuts at the same position, or within a few nucleotides of one another, on different strands, e.g., essentially mimicking a double strand break.

In certain embodiments, a double strand break can be accompanied by an additional double strand break, positioned by a second, third and/or fourth gRNA molecule. For example, the targeting domain of a first gRNA molecule may be configured such that a double strand break is positioned upstream of the HBG target position, e.g., within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position; and the targeting domain of a second gRNA molecule may be configured such that a double strand break is positioned downstream from the HBG

target position, e.g., within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position.

In certain embodiments, a double strand break can be accompanied by two additional single strand breaks, positioned by a second and third gRNA molecule. For example, the targeting domain of a first gRNA molecule may be configured such that a double strand break is positioned upstream of the HBG target position, e.g., within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position; and the targeting domains of a second and third gRNA molecule may be configured such that two single strand breaks are positioned downstream of the target position, e.g., within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position. In certain embodiments, the targeting domains of the first, second and third gRNA molecules are configured such that a cleavage event, e.g., a double strand or single strand break, is positioned independently for each of the gRNA molecules.

In certain embodiments, first and second single strand breaks can be accompanied by two additional single strand breaks positioned by a third gRNA molecule and a fourth gRNA molecule. For example, the targeting domains of a first and second gRNA molecule may be configured such that two single strand breaks are positioned upstream of an HBG target position, e.g., within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position; and the targeting domains of a third and fourth gRNA molecule may be configured such that two single strand breaks are positioned downstream of the HBG target position, e.g., within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of the target position.

Guide RNA (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) (e.g., chimeric), or modular (comprising more than one, and typically two, separate RNA molecules). The gRNA molecules provided herein comprise a targeting domain comprising, consisting of, or consisting essentially of a nucleic acid sequence fully or partially complementary to a target domain. In certain embodiments, the gRNA molecule further comprises one or more additional domains, including for example a first complementarity

domain, a linking domain, a second complementarity domain, a proximal domain, a tail domain, and a 5' extension domain. Each of these domains is discussed in detail below. In certain embodiments, one or more of the domains in the gRNA molecule comprises a nucleotide sequence identical to or sharing sequence homology with a naturally occurring sequence, e.g., from *S. pyogenes*, *S. aureus*, or *S. thermophilus*.

Several exemplary gRNA structures are provided in **Figs. 1A-1I**. With regard to the three-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-1I** and other depictions provided herein. **Fig. 7** illustrates gRNA domain nomenclature using the gRNA sequence of SEQ ID NO:42, which contains one hairpin loop in the tracrRNA-derived region. In certain embodiments, a gRNA may contain more than one (e.g., two, three, or more) hairpin loops in this region (see, e.g., **Figs. 1H-1I**).

In certain embodiments, a unimolecular, or chimeric, gRNA comprises, preferably from 5' to 3':

- a targeting domain complementary to a target domain in a γ -globin gene regulatory region, e.g., a targeting domain from any of SEQ ID NOs:251-901;
- a first complementarity domain;
- a linking domain;
- a second complementarity domain (which is complementary to the first complementarity domain);
- a proximal domain; and
- optionally, a tail domain.

In certain embodiments, a modular gRNA comprises:

- a first strand comprising, preferably from 5' to 3':
 - a targeting domain complementary to a target domain in in a γ -globin gene regulatory region, e.g., a targeting domain from any of SEQ ID NOs:251-901; and
 - a first complementarity domain; and
- a second strand, comprising, preferably from 5' to 3':
 - optionally, a 5' extension domain;
 - a second complementarity domain;
 - a proximal domain; and
 - optionally, a tail domain.

Targeting domain

The targeting domain (sometimes referred to alternatively as the guide sequence or complementarity region) comprises, consists of, or consists essentially of a nucleic acid sequence that is complementary or partially complementary to a target nucleic acid sequence in a γ -globin gene regulatory region. The nucleic acid sequence in a γ -globin gene regulatory region to which all or a portion of the targeting domain is complementary or partially complementary is referred to herein as the target domain. In certain embodiments, the target domain comprises an HBG target position. In other embodiments, an HBG target position lies outside (i.e., upstream or downstream of) the target domain. In certain embodiments, the target domain is located entirely within a γ -globin gene regulatory region, e.g., in a regulatory element associated with a γ -globin gene or a regulatory element associated with a gene encoding a repressor of γ -globin gene expression. In other embodiments, all or part of the target domain is located outside of γ -globin gene regulatory region, e.g., in an HBG1 or HBG2 coding region, exon, or intron.

Methods for selecting targeting domains are known in the art (see, e.g., Fu 2014; Sternberg 2014). Examples of suitable targeting domains for use in the methods, compositions, and kits described herein include those set forth in SEQ ID NOs:251-901.

The strand of the target nucleic acid comprising the target domain is referred to herein as the complementary strand because it is complementary to the targeting domain sequence. Since the targeting domain is part of a gRNA molecule, it comprises the base uracil (U) rather than thymine (T); conversely, any DNA molecule encoding the gRNA molecule will comprise thymine rather than uracil. In a targeting domain/target domain pair, the uracil bases in the targeting domain will pair with the adenine bases in the target domain. In certain embodiments, the degree of complementarity between the targeting domain and target domain is sufficient to allow targeting of a Cas9 molecule to the target nucleic acid.

In certain embodiments, the targeting domain comprises a core domain and an optional secondary domain. In certain of these embodiments, the core domain is located 3' to the secondary domain, and in certain of these embodiments the core domain is located at or near the 3' end of the targeting domain. In certain of these embodiments, the core domain consists of or consists essentially of about 8 to about 13 nucleotides at the 3' end of the targeting domain. In certain embodiments, only the core domain is complementary or partially complementary to the corresponding portion of the target domain, and in certain of these embodiments the core domain is fully complementary to the corresponding portion of the target domain. In other embodiments, the secondary domain is also complementary or

partially complementary to a portion of the target domain. In certain embodiments, the core domain is complementary or partially complementary to a core domain target in the target domain, while the secondary domain is complementary or partially complementary to a secondary domain target in the target domain. In certain embodiments, the core domain and secondary domain have the same degree of complementarity with their respective corresponding portions of the target domain. In other embodiments, the degree of complementarity between the core domain and its target and the degree of complementarity between the secondary domain and its target may differ. In certain of these embodiments, the core domain may have a higher degree of complementarity for its target than the secondary domain, whereas in other embodiments the secondary domain may have a higher degree of complementarity than the core domain.

In certain embodiments, the targeting domain and/or the core domain within the targeting domain is 3 to 100, 5 to 100, 10 to 100, or 20 to 100 nucleotides in length, and in certain of these embodiments the targeting domain or core domain is 3 to 15, 3 to 20, 5 to 20, 10 to 20, 15 to 20, 5 to 50, 10 to 50, or 20 to 50 nucleotides in length. In certain embodiments, the targeting domain and/or the core domain within the targeting domain is 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length. In certain embodiments, the targeting domain and/or the core domain within the targeting domain is 6 +/-2, 7 +/-2, 8 +/-2, 9 +/-2, 10 +/-2, 10 +/-4, 10 +/-5, 11 +/-2, 12 +/-2, 13 +/-2, 14 +/-2, 15 +/-2, or 16 +/-2, 20 +/-5, 30 +/-5, 40 +/-5, 50 +/-5, 60 +/-5, 70 +/-5, 80 +/-5, 90 +/-5, or 100 +/-5 nucleotides in length.

In certain embodiments wherein the targeting domain includes a core domain, the core domain is 3 to 20 nucleotides in length, and in certain of these embodiments the core domain 5 to 15 or 8 to 13 nucleotides in length. In certain embodiments wherein the targeting domain includes a secondary domain, the secondary domain is 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 nucleotides in length. In certain embodiments wherein the targeting domain comprises a core domain that is 8 to 13 nucleotides in length, the targeting domain is 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, or 16 nucleotides in length, and the secondary domain is 13 to 18, 12 to 17, 11 to 16, 10 to 15, 9 to 14, 8 to 13, 7 to 12, 6 to 11, 5 to 10, 4 to 9, or 3 to 8 nucleotides in length, respectively.

In certain embodiments, the targeting domain is fully complementary to the target domain. Likewise, where the targeting domain comprises a core domain and/or a secondary domain, in certain embodiments one or both of the core domain and the secondary domain are fully complementary to the corresponding portions of the target domain. In other

embodiments, the targeting domain is partially complementary to the target domain, and in certain of these embodiments where the targeting domain comprises a core domain and/or a secondary domain, one or both of the core domain and the secondary domain are partially complementary to the corresponding portions of the target domain. In certain of these embodiments, the nucleic acid sequence of the targeting domain, or the core domain or targeting domain within the targeting domain, is at least 80, 85, 90, or 95% complementary to the target domain or to the corresponding portion of the target domain. In certain embodiments, the targeting domain and/or the core or secondary domains within the targeting domain include one or more nucleotides that are not complementary with the target domain or a portion thereof, and in certain of these embodiments the targeting domain and/or the core or secondary domains within the targeting domain include 1, 2, 3, 4, 5, 6, 7, or 8 nucleotides that are not complementary with the target domain. In certain embodiments, the core domain includes 1, 2, 3, 4, or 5 nucleotides that are not complementary with the corresponding portion of the target domain. In certain embodiments wherein the targeting domain includes one or more nucleotides that are not complementary with the target domain, one or more of said non-complementary nucleotides are located within five nucleotides of the 5' or 3' end of the targeting domain. In certain of these embodiments, the targeting domain includes 1, 2, 3, 4, or 5 nucleotides within five nucleotides of its 5' end, 3' end, or both its 5' and 3' ends that are not complementary to the target domain. In certain embodiments wherein the targeting domain includes two or more nucleotides that are not complementary to the target domain, two or more of said non-complementary nucleotides are adjacent to one another, and in certain of these embodiments the two or more consecutive non-complementary nucleotides are located within five nucleotides of the 5' or 3' end of the targeting domain. In other embodiments, the two or more consecutive non-complementary nucleotides are both located more than five nucleotides from the 5' and 3' ends of the targeting domain.

In certain embodiments, the targeting domain, core domain, and/or secondary domain do not comprise any modifications. In other embodiments, the targeting domain, core domain, and/or secondary domain, or one or more nucleotides therein, have a modification, including but not limited to the modifications set forth below. In certain embodiments, one or more nucleotides of the targeting domain, core domain, and/or secondary domain may comprise a 2' modification (e.g., a modification at the 2' position on ribose), e.g., a 2-acetylation, e.g., a 2' methylation. In certain embodiments, the backbone of the targeting domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the targeting domain, core domain, and/or secondary domain

render the targeting domain and/or the gRNA comprising the targeting domain less susceptible to degradation or more bio-compatible, e.g., less immunogenic. In certain embodiments, the targeting domain and/or the core or secondary domains include 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments the targeting domain and/or core or secondary domains include 1, 2, 3, or 4 modifications within five nucleotides of their respective 5' ends and/or 1, 2, 3, or 4 modifications within five nucleotides of their respective 3' ends. In certain embodiments, the targeting domain and/or the core or secondary domains comprise modifications at two or more consecutive nucleotides.

In certain embodiments wherein the targeting domain includes core and secondary domains, the core and secondary domains contain the same number of modifications. In certain of these embodiments, both domains are free of modifications. In other embodiments, the core domain includes more modifications than the secondary domain, or vice versa.

In certain embodiments, modifications to one or more nucleotides in the targeting domain, including in the core or secondary domains, are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification using a system as set forth below. gRNAs having a candidate targeting domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated using a system as set forth below. 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 certain embodiments, all of the modified nucleotides are complementary to and capable of hybridizing to corresponding nucleotides present in the target domain. In another embodiment, 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.

Figs. 1A-II provide examples of the placement of the targeting domain within a gRNA molecule.

First and second complementarity domains

The first and second complementarity (sometimes referred to alternatively as the crRNA-derived hairpin sequence and tracrRNA-derived hairpin sequences, respectively) domains are fully or partially complementary to one another. In certain embodiments, the degree of complementarity is sufficient for the two domains to form a duplexed region under at least some physiological conditions. In certain embodiments, the degree of complementarity between the first and second complementarity domains, together with other

properties of the gRNA, is sufficient to allow targeting of a Cas9 molecule to a target nucleic acid. Examples of first and second complementarity domains are set forth in **Figs. 1A-1G**.

In certain embodiments (see, e.g., **Figs. 1A-1B**) the first and/or second complementarity domain includes one or more nucleotides that lack complementarity with the corresponding complementarity domain. In certain embodiments, the first and/or second complementarity domain includes 1, 2, 3, 4, 5, or 6 nucleotides that do not complement with the corresponding complementarity domain. For example, the second complementarity domain may contain 1, 2, 3, 4, 5, or 6 nucleotides that do not pair with corresponding nucleotides in the first complementarity domain. In certain embodiments, the nucleotides on the first or second complementarity domain that do not complement with the corresponding complementarity domain loop out from the duplex formed between the first and second complementarity domains. In certain of these embodiments, the unpaired loop-out is located on the second complementarity domain, and in certain of these embodiments the unpaired region begins 1, 2, 3, 4, 5, or 6 nucleotides from the 5' end of the second complementarity domain.

In certain embodiments, the first complementarity domain is 5 to 30, 5 to 25, 7 to 25, 5 to 24, 5 to 23, 7 to 22, 5 to 22, 5 to 21, 5 to 20, 7 to 18, 7 to 15, 9 to 16, or 10 to 14 nucleotides in length, and in certain of these embodiments the first complementarity 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 certain embodiments, the second complementarity domain is 5 to 27, 7 to 27, 7 to 25, 5 to 24, 5 to 23, 5 to 22, 5 to 21, 7 to 20, 5 to 20, 7 to 18, 7 to 17, 9 to 16, or 10 to 14 nucleotides in length, and in certain of these embodiments the second complementarity domain is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length. In certain embodiments, the first and second complementarity domains are each 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. In certain embodiments, the second complementarity domain is longer than the first complementarity domain, e.g., 2, 3, 4, 5, or 6 nucleotides longer.

In certain embodiments, the first and/or second complementarity domains each independently comprise three subdomains, which, in the 5' to 3' direction are: a 5' subdomain, a central subdomain, and a 3' subdomain. In certain embodiments, the 5' subdomain and 3' subdomain of the first complementarity domain are fully or partially complementary to the 3' subdomain and 5' subdomain, respectively, of the second complementarity domain.

In certain embodiments, the 5' subdomain of the first complementarity domain is 4 to 9 nucleotides in length, and in certain of these embodiments the 5' domain is 4, 5, 6, 7, 8, or 9 nucleotides in length. In certain embodiments, the 5' subdomain of the second complementarity domain is 3 to 25, 4 to 22, 4 to 18, or 4 to 10 nucleotides in length, and in certain of these embodiments the 5' domain is 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 certain embodiments, the central subdomain of the first complementarity domain is 1, 2, or 3 nucleotides in length. In certain embodiments, the central subdomain of the second complementarity domain is 1, 2, 3, 4, or 5 nucleotides in length. In certain embodiments, the 3' subdomain of the first complementarity domain is 3 to 25, 4 to 22, 4 to 18, or 4 to 10 nucleotides in length, and in certain of these embodiments the 3' subdomain is 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 certain embodiments, the 3' subdomain of the second complementarity domain is 4 to 9, e.g., 4, 5, 6, 7, 8, or 9 nucleotides in length.

The first and/or second complementarity domains can share homology with, or be derived from, naturally occurring or reference first and/or second complementarity domains. In certain of these embodiments, the first and/or second complementarity domains have at least 50%, 60%, 70%, 80%, 85%, 90%, or 95% homology with, or differ by no more than 1, 2, 3, 4, 5, or 6 nucleotides from, the naturally occurring or reference first and/or second complementarity domain. In certain of these embodiments, the first and/or second complementarity domains may have at least 50%, 60%, 70%, 80%, 85%, 90%, or 95% homology with a first and/or second complementarity domain from *S. pyogenes* or *S. aureus*.

In certain embodiments, the first and/or second complementarity domains do not comprise any modifications. In other embodiments, the first and/or second complementarity domains or one or more nucleotides therein have a modification, including but not limited to a modification set forth below. In certain embodiments, one or more nucleotides of the first and/or second complementarity domain may comprise a 2' modification (e.g., a modification at the 2' position on ribose), e.g., a 2-acetylation, e.g., a 2' methylation. In certain embodiments, the backbone of the targeting domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the first and/or second complementarity domain render the first and/or second complementarity domain and/or the gRNA comprising the first and/or second complementarity less susceptible to degradation or more bio-compatible, e.g., less immunogenic. In certain embodiments, the first and/or second complementarity domains each independently include 1, 2, 3, 4, 5, 6, 7, or

8 or more modifications, and in certain of these embodiments the first and/or second complementarity domains each independently include 1, 2, 3, or 4 modifications within five nucleotides of their respective 5' ends, 3' ends, or both their 5' and 3' ends. In other embodiments, the first and/or second complementarity domains each independently contain no modifications within five nucleotides of their respective 5' ends, 3' ends, or both their 5' and 3' ends. In certain embodiments, one or both of the first and second complementarity domains comprise modifications at two or more consecutive nucleotides.

In certain embodiments, modifications to one or more nucleotides in the first and/or second complementarity domains are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in a system as set forth below. gRNAs having a candidate first or second complementarity domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated in a system as set forth below. 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 certain embodiments, the duplexed region formed by the first and second complementarity domains is, for example, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or 22 bp in length, excluding any looped out or unpaired nucleotides.

In certain embodiments, the first and second complementarity domains, when duplexed, comprise 11 paired nucleotides (see, for e.g., gRNA of SEQ ID NO:48). In certain embodiments, the first and second complementarity domains, when duplexed, comprise 15 paired nucleotides (see, e.g., gRNA of SEQ ID NO:50). In certain embodiments, the first and second complementarity domains, when duplexed, comprise 16 paired nucleotides (see, e.g., gRNA of SEQ ID NO:51). In certain embodiments, the first and second complementarity domains, when duplexed, comprise 21 paired nucleotides (see, e.g., gRNA of SEQ ID NO:29).

In certain embodiments, one or more nucleotides are exchanged between the first and second complementarity domains to remove poly-U tracts. For example, nucleotides 23 and 48 or nucleotides 26 and 45 of the gRNA of SEQ ID NO:48 may be exchanged to generate the gRNA of SEQ ID NOs:49 or 31, respectively. Similarly, nucleotides 23 and 39 of the gRNA of SEQ ID NO:29 may be exchanged with nucleotides 50 and 68 to generate the gRNA of SEQ ID NO:30.

Linking domain

The linking domain is disposed between and serves to link the first and second complementarity domains in a unimolecular or chimeric gRNA. **Figs. 1B-1E** provide examples of linking domains. In certain embodiments, part of the linking domain is from a crRNA-derived region, and another part is from a tracrRNA-derived region.

In certain embodiments, the linking domain links the first and second complementarity domains covalently. In certain of these embodiments, the linking domain consists of or comprises a covalent bond. In other embodiments, the linking domain links the first and second complementarity domains non-covalently. In certain embodiments, the linking domain is ten or fewer nucleotides in length, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides. In other embodiments, the linking domain is greater than 10 nucleotides in length, e.g., 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 or more nucleotides. In certain embodiments, the linking domain is 2 to 50, 2 to 40, 2 to 30, 2 to 20, 2 to 10, 2 to 5, 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, 10 to 15, 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 certain embodiments, the linking domain is 10 +/-5, 20 +/-5, 20 +/-10, 30 +/-5, 30 +/-10, 40 +/-5, 40 +/-10, 50 +/-5, 50 +/-10, 60 +/-5, 60 +/-10, 70 +/-5, 70 +/-10, 80 +/-5, 80 +/-10, 90 +/-5, 90 +/-10, 100 +/-5, or 100 +/-10 nucleotides in length.

In certain embodiments, the linking 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 certain embodiments, the linking domain has at least 50%, 60%, 70%, 80%, 90%, or 95% homology with or differs by no more than 1, 2, 3, 4, 5, or 6 nucleotides from a linking domain disclosed herein, e.g., the linking domains of **Figs. 1B-1E**.

In certain embodiments, the linking domain does not comprise any modifications. In other embodiments, the linking domain or one or more nucleotides therein have a modification, including but not limited to the modifications set forth below. In certain embodiments, one or more nucleotides of the linking domain may comprise a 2' modification (e.g., a modification at the 2' position on ribose), e.g., a 2-acetylation, e.g., a 2' methylation. In certain embodiments, the backbone of the linking domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the linking domain render the linking domain and/or the gRNA comprising the linking domain less susceptible to degradation or more bio-compatible, e.g., less immunogenic. In certain embodiments, the linking domain includes 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments the linking domain includes 1, 2, 3, or 4 modifications within

five nucleotides of its 5' and/or 3' end. In certain embodiments, the linking domain comprises modifications at two or more consecutive nucleotides.

In certain embodiments, modifications to one or more nucleotides in the linking domain are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in a system as set forth below. gRNAs having a candidate linking domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated in a system as set forth below. The 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 certain embodiments, the linking domain comprises a duplexed region, typically adjacent to 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 certain of these embodiments, the duplexed region of the linking region is 10+/-5, 15+/-5, 20+/-5, 20+/-10, or 30+/-5 bp in length. In certain embodiments, the duplexed region of the linking domain is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 bp in length. In certain embodiments, the sequences forming the duplexed region of the linking domain are fully complementary. In other embodiments, one or both of the sequences forming the duplexed region contain one or more nucleotides (e.g., 1, 2, 3, 4, 5, 6, 7, or 8 nucleotides) that are not complementary with the other duplex sequence.

5' extension domain

In certain embodiments, a modular gRNA as disclosed herein comprises a 5' extension domain, i.e., one or more additional nucleotides 5' to the second complementarity domain (see, e.g., **Fig. 1A**). In certain embodiments, the 5' extension domain is 2 to 10 or more, 2 to 9, 2 to 8, 2 to 7, 2 to 6, 2 to 5, or 2 to 4 nucleotides in length, and in certain of these embodiments the 5' extension domain is 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more nucleotides in length.

In certain embodiments, the 5' extension domain nucleotides do not comprise modifications, e.g., modifications of the type provided below. However, in certain embodiments, 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) as set forth below. In certain embodiments, a nucleotide of the 5' extension domain can comprise a 2' modification (e.g., a

modification at the 2' position on ribose), e.g., a 2-acetylation, e.g., a 2' methylation, or other modification(s) as set forth below.

In certain embodiments, the 5' extension domain can comprise as many as 1, 2, 3, 4, 5, 6, 7, or 8 modifications. In certain embodiments, 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 certain embodiments, 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 certain 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 certain embodiments, 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. In certain embodiments, 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 a system as set forth below. gRNAs having a candidate 5' extension domain having a selected length, sequence, degree of complementarity, or degree of modification, can be evaluated in a system as set forth below. 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.

In certain embodiments, 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**.

Proximal domain

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

In certain embodiments, the proximal domain is 5 to 20 or more nucleotides in length, e.g., 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length. In certain of these embodiments, 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 certain embodiments, the proximal domain is 5 to 20, 7, to 18, 9 to 16, or 10 to 14 nucleotides in length.

In certain embodiments, the proximal domain can share homology with or be derived from a naturally occurring proximal domain. In certain of these embodiments, the proximal domain has at least 50%, 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 proximal domain disclosed herein, e.g., an *S. pyogenes*, *S. aureus*, or *S. thermophilus* proximal domain, including those set forth in **Figs. 1A-1G**.

In certain embodiments, the proximal domain does not comprise any modifications. In other embodiments, the proximal domain or one or more nucleotides therein have a modification, including but not limited to the modifications set forth in herein. In certain embodiments, one or more nucleotides of the proximal domain may comprise a 2' modification (e.g., a modification at the 2' position on ribose), e.g., a 2-acetylation, e.g., a 2' methylation. In certain embodiments, the backbone of the proximal domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the proximal domain render the proximal domain and/or the gRNA comprising the proximal domain less susceptible to degradation or more bio-compatible, e.g., less immunogenic. In certain embodiments, the proximal domain includes 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments the proximal domain includes 1, 2, 3, or 4 modifications within five nucleotides of its 5' and/or 3' end. In certain embodiments, the proximal domain comprises modifications at two or more consecutive nucleotides.

In certain embodiments, modifications to one or more nucleotides in the proximal domain are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in a system as set forth below. gRNAs having a candidate proximal domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated in a system as set forth below. 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.

Tail domain

A broad spectrum of tail domains are suitable for use in the gRNA molecules disclosed herein. **Figs. 1A** and **1C-1G** provide examples of such tail domains.

In certain embodiments, the tail domain is absent. In other embodiments, the tail domain is 1 to 100 or more nucleotides in length, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotides in length. In certain embodiments, the tail domain is 1 to 5, 1 to 10, 1 to 15, 1 to 20, 1 to 50, 10 to 100, 20 to 100, 10 to 90, 20 to 90, 10 to 80, 20 to 80, 10 to 70, 20 to 70, 10 to 60, 20 to 60, 10 to 50, 20 to 50, 10 to 40, 20 to 40, 10 to 30, 20 to 30, 20 to 25, 10 to 20, or 10 to 15 nucleotides in length. In certain embodiments, the tail domain is 5 +/-5, 10 +/-5, 20 +/-10, 20 +/-5, 25 +/-10, 30 +/-10, 30 +/-5, 40 +/-10, 40 +/-5, 50 +/-10, 50 +/-5, 60 +/-10, 60 +/-5, 70 +/-10, 70 +/-5, 80 +/-10, 80 +/-5, 90 +/-10, 90 +/-5, 100 +/-10, or 100 +/-5 nucleotides in length.

In certain embodiments, the tail domain can share homology with or be derived from a naturally occurring tail domain or the 5' end of a naturally occurring tail domain. In certain of these embodiments, the proximal domain has at least 50%, 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 naturally occurring tail domain disclosed herein, e.g., an *S. pyogenes*, *S. aureus*, or *S. thermophilus* tail domain, including those set forth in **Figs. 1A** and **1C-1G**.

In certain embodiments, the tail domain includes sequences that are complementary to each other and which, under at least some physiological conditions, form a duplexed region. In certain of these embodiments, the tail domain comprises a tail duplex domain which can form a tail duplexed region. In certain embodiments, the tail duplexed region is 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 bp in length. In certain embodiments, the tail domain comprises a single stranded domain 3' to the tail duplex domain that does not form a duplex. In certain of these embodiments, the single stranded domain is 3 to 10 nucleotides in length, e.g., 3, 4, 5, 6, 7, 8, 9, 10, or 4 to 6 nucleotides in length.

In certain embodiments, the tail domain does not comprise any modifications. In other embodiments, the tail domain or one or more nucleotides therein have a modification, including but not limited to the modifications set forth herein. In certain embodiments, one or more nucleotides of the tail domain may comprise a 2' modification (e.g., a modification at the 2' position on ribose), e.g., a 2-acetylation, e.g., a 2' methylation. In certain

embodiments, the backbone of the tail domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the tail domain render the tail domain and/or the gRNA comprising the tail domain less susceptible to degradation or more bio-compatible, e.g., less immunogenic. In certain embodiments, the tail domain includes 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments the tail domain includes 1, 2, 3, or 4 modifications within five nucleotides of its 5' and/or 3' end. In certain embodiments, the tail domain comprises modifications at two or more consecutive nucleotides.

In certain embodiments, modifications to one or more nucleotides in the tail domain are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification as set forth below. gRNAs having a candidate tail domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated using a system as set forth below. 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 certain embodiments, 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 an H1 promoter is used for transcription, these nucleotides may be the sequence UUUU. When alternate pol-III promoters are used, these nucleotides may be various numbers of uracil bases depending on, e.g., the termination signal of the pol-III promoter, or they may include alternate bases.

In certain embodiments, the proximal and tail domain taken together comprise, consist of, or consist essentially of the sequence set forth in SEQ ID NOs:32, 33, 34, 35, 36, or 37.

Exemplary unimolecular/chimeric gRNAs

In certain embodiments, a unimolecular or chimeric gRNA as disclosed herein has the structure: 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 certain embodiments 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 certain embodiments has at least 50, 60, 70, 80, 85, 90, or 95% homology with a reference second complementarity domain disclosed herein;

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

In certain embodiments, a unimolecular gRNA as disclosed herein comprises, preferably from 5' to 3':

a targeting domain, e.g., comprising 10-50 nucleotides;

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;

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;

(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.

In certain embodiments, the sequence from (a), (b), and/or (c) has at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% homology with the corresponding sequence of a naturally occurring gRNA, or with a gRNA described herein.

In certain embodiments, 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 certain embodiments, 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 certain embodiments, 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 are complementary to the corresponding nucleotides of the first complementarity domain.

In certain embodiments, the targeting domain consists of, consists essentially of, or comprises 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) complementary or partially complementary to the target domain or a portion thereof, e.g., the targeting domain is 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length. In certain of these embodiments, the targeting domain is complementary to the target domain over the entire length of the targeting domain, the entire length of the target domain, or both.

In certain embodiments, a unimolecular or chimeric gRNA molecule disclosed herein (comprising a targeting domain, a first complementarity domain, a linking domain, a second complementarity domain, a proximal domain and, optionally, a tail domain) comprises the nucleotide sequence set forth in SEQ ID NO:42, wherein the targeting domain is listed as 20 N's (residues 1-20) but may range in length from 16 to 26 nucleotides, and wherein the final six residues (residues 97-102) represent a termination signal for the U6 promoter but may be absent or fewer in number. In certain embodiments, the unimolecular, or chimeric, gRNA molecule is a *S. pyogenes* gRNA molecule.

In certain embodiments, a unimolecular or chimeric gRNA molecule disclosed herein (comprising a targeting domain, a first complementarity domain, a linking domain, a second complementarity domain, a proximal domain and, optionally, a tail domain) comprises the nucleotide sequence set forth in SEQ ID NO:38, wherein the targeting domain is listed as 20 Ns (residues 1-20) but may range in length from 16 to 26 nucleotides, and wherein the final six residues (residues 97-102) represent a termination signal for the U6 promoter but may be absent or fewer in number. In certain embodiments, the unimolecular or chimeric gRNA molecule is an *S. aureus* gRNA molecule.

The sequences and structures of exemplary chimeric gRNAs are also shown in **Figs. 1H-1I.**

Exemplary modular gRNAs

In certain embodiments, a modular gRNA disclosed herein 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':
 - optionally a 5' extension domain;
 - a second complementarity domain;
 - 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;
- (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.

In certain embodiments, 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 certain embodiments, 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 certain embodiments, 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 certain embodiments, 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 certain embodiments, 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 certain embodiments, the targeting domain consists of, consists essentially of, or comprises 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) complementary to the target domain or a portion thereof. In certain of these embodiments, the targeting domain is complementary to the target domain over the entire length of the targeting domain, the entire length of the target domain, or both.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially 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 certain embodiments of these embodiments, (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; and/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.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially 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 certain of these embodiments, (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; and/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.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially 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 certain of these embodiments, (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; and/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.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially 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 certain of these embodiments, (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; and/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.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially 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 certain of these embodiments, (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; and/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.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially 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 certain of these embodiments, (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; and/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.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially 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 certain of these embodiments, (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; and/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.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 23 nucleotides (e.g., 23 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 23 nucleotides in length. In certain of these embodiments, (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; and/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.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially 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 certain of these embodiments, (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; and/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.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially 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 certain of these embodiments, (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; and/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.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 26 nucleotides (e.g., 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 26 nucleotides in length. In certain of these embodiments, (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; and/or 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.

gRNA delivery

In certain embodiments of the methods provided herein, the methods comprise delivery of one or more (e.g., two, three, or four) gRNA molecules as described herein. In certain of these embodiments, the gRNA molecules are delivered by intravenous injection, intramuscular injection, subcutaneous injection, or inhalation.

Methods for designing gRNAs

Methods for selecting, designing, and validating targeting domains for use in the gRNAs described herein are provided. Exemplary targeting domains for incorporation into gRNAs are also provided herein.

Methods for selection and validation of target sequences as well as off-target analyses have been described previously (see, e.g., Mali 2013; Hsu 2013; Fu 2014; Heigwer 2014; Bae 2014; Xiao 2014). For example, a software tool can be used to optimize the choice of potential targeting domains corresponding to 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 targeting domain choice using *S. pyogenes* Cas9, 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 targeting domain is then ranked according to its total predicted off-target cleavage; the top-ranked targeting domains represent those that are likely to have the greatest on-target cleavage 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 targeting domains and gRNAs comprising those targeting domains can be functionally evaluated using methods known in the art and/or as set forth herein.

As a non-limiting example, targeting domains for use in gRNAs for use with *S. pyogenes* and *S. aureus* Cas9s were identified using a DNA sequence searching algorithm.

17-mer and 20-mer targeting domains were designed for *S. pyogenes* targets, while 18-mer, 19-mer, 20-mer, 21-mer, 22-mer, 23-mer, and 24-mer targeting domains were designed for *S. aureus* targets. gRNA design was carried out using custom gRNA design software based on the public tool cas-offinder (Bae 2014). This software scores guides after calculating their genome-wide 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 target sites adjacent to PAM sequences, the software also identifies all PAM adjacent sequences that differ by 1, 2, 3, or more than 3 nucleotides from the selected target sites. Genomic DNA sequences for HBG1 and HBG2 regulatory regions were 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, targeting domains 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., an NGG PAM for *S. pyogenes*, or an NNGRRT (SEQ ID NO:204) or NNGRRV (SEQ ID NO:205) PAM for *S. aureus*). 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 targeting domain 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.

Targeting domains were identified for both single-gRNA nuclease cleavage and for a dual-gRNA paired “nickase” strategy. Criteria for selecting targeting domains and the determination of which targeting domains can be used for the dual-gRNA paired “nickase” strategy is based on two considerations:

- (1) Targeting domain 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; and
- (2) An assumption that cleaving with dual nickase pairs will result in deletion of the entire intervening sequence at a reasonable frequency.

However, cleaving with dual nickase pairs can also result in indel mutations at the site of only one of the gRNAs. Candidate pair members can be tested for how efficiently they remove the entire sequence versus causing indel mutations at the target site of one targeting domain.

Targeting domains for use in deleting HBG1 c.-114 to -102

Targeting domains for use in gRNAs for deleting c.-114 to -102 of HBG1 in conjunction with the methods disclosed herein were identified and ranked into 4 tiers for *S. pyogenes* and *S. aureus*.

For *S. pyogenes*, tier 1 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1 c.-114 to -102), specifically within 400 bp of either end of the target site, (2) a high level of orthogonality, and (3) the presence of 5' G. Tier 2 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1 c.-114 to -102), specifically within 400 bp of either end of the target site, and (2) a high level of orthogonality. Tier 3 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1 c.-114 to -102), specifically within 400 bp of either end of the target site and (2) the presence of 5' G. Tier 4 targeting domains were selected based on distance upstream or downstream from either end of the target site (i.e., HBG1 c.-114 to -102), specifically within 400 bp of either end of the target site.

For *S. aureus*, tier 1 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1 c.-114 to -102), specifically within 400 bp of either end of the target site, (2) a high level of orthogonality, (3) the presence of 5' G, and (4) PAM having the sequence NNGRRT (SEQ ID NO:204). Tier 2 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1 c.-114 to -102), specifically within 400 bp of either end of the target site, (2) a high level of orthogonality, and (3) PAM having the sequence NNGRRT (SEQ ID NO:204). Tier 3 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1 c.-114 to -102), specifically within 400 bp of either end of the target site, and (2) PAM having the sequence NNGRRT (SEQ ID NO:204). Tier 4 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1 c.-114 to -102), specifically within

400 bp of either end of the target site, and (2) PAM having the sequence NNGRRV (SEQ ID NO:205).

Note that tiers are non-inclusive (each targeting domain is listed only once for the strategy). In certain instances, no targeting domain was identified based on the criteria of the particular tier. The identified targeting domains are summarized below in **Table 6**.

Table 6: Nucleotide sequences of *S. pyogenes* and *S. aureus* targeting domains

	<i>S. pyogenes</i>	<i>S. aureus</i>
Tier 1	SEQ ID NOs:251-256	SEQ ID NOs:367-376
Tier 2	SEQ ID NOs:257-274	SEQ ID NOs:343, 377-393
Tier 3	SEQ ID NOs:275-300	SEQ ID NOs:357, 365, 394-461
Tier 4	SEQ ID NOs:301-366	SEQ ID NOs:252-254, 256, 268, 272-274, 292, 295, 347, 348, 353, 360-362, 366, 598-759

Targeting domains for use in deleting HBG2 c.-114 to -102

Targeting domains for use in gRNAs for deleting c.-114 to -102 of HBG2 in conjunction with the methods disclosed herein were identified and ranked into 4 tiers for *S. pyogenes* and *S. aureus*.

For *S. pyogenes*, tier 1 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG2 c.-114 to -102), specifically within 400 bp of either end of the target site, (2) a high level of orthogonality, and (3) the presence of 5' G. Tier 2 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG2 c.-114 to -102), specifically within 400 bp of either end of the target site, and (2) a high level of orthogonality. Tier 3 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG2 c.-114 to -102), specifically within 400 bp of either end of the target site and (2) the presence of 5' G. Tier 4 targeting domains were selected based on distance upstream or downstream from either end of the target site (i.e., HBG2 c.-114 to -102), specifically within 400 bp of either end of the target site.

For *S. aureus*, tier 1 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG2 c.-114 to -102), specifically within 400 bp of either end of the target site, (2) a high level of orthogonality, (3) the presence of 5' G, and (4) PAM having the sequence NNGRRT (SEQ ID NO:204). Tier 2 targeting domains were selected based on (1) distance upstream or downstream from either

end of the target site (i.e., HBG2 c.-114 to -102), specifically within 400 bp of either end of the target site, (2) a high level of orthogonality, and (3) PAM having the sequence NNGRRT (SEQ ID NO:204). Tier 3 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG2 c.-114 to -102), specifically within 400 bp of either end of the target site, and (2) PAM having the sequence NNGRRT (SEQ ID NO:204). Tier 4 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG2 c.-114 to -102), specifically within 400 bp of either end of the target site, and (2) PAM having the sequence NNGRRV (SEQ ID NO:205).

Note that tiers are non-inclusive (each targeting domain is listed only once for the strategy). In certain instances, no targeting domain was identified based on the criteria of the particular tier. The identified targeting domains are summarized below in **Table 7**.

Table 7: Nucleotide sequences of *S. pyogenes* and *S. aureus* targeting domains

	<i>S. pyogenes</i>	<i>S. aureus</i>
Tier 1	SEQ ID NOs:760-764	SEQ ID NOs:784-791
Tier 2	SEQ ID NOs:765-781	SEQ ID NOs:778, 792-803
Tier 3	SEQ ID NOs:275-281, 283-300	SEQ ID NOs:357, 365, 394-461
Tier 4	SEQ ID NOs:301-311, 313-342, 344-348, 350-366, 782, 783	SEQ ID NOs:292, 295, 347, 348, 353, 360-362, 366, 462-468, 476-481, 489-587, 601-607, 614-620, 640-666, 674-679, 687-693, 708-714, 733-753, 762-764, 775, 779-781, 804-901

In certain embodiments, two or more (e.g., three or four) gRNA molecules are used with one Cas9 molecule. In another embodiment, when two or more (e.g., three or four) gRNAs are used with two or more Cas9 molecules, at least one Cas9 molecule is from a different species than the other Cas9 molecule(s). For example, when two gRNA molecules are used with two Cas9 molecules, one Cas9 molecule can be from one species and the other Cas9 molecule can be from a different species. Both Cas9 species are used to generate a single or double-strand break, as desired.

Any of the targeting domains in the tables described herein can be used with a Cas9 molecule that generates a single strand break (i.e., *S. pyogenes* or *S. aureus* Cas9 nickase) or

with a Cas9 molecule that generates a double strand break (i.e., *S. pyogenes* or *S. aureus* Cas9 nuclease).

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

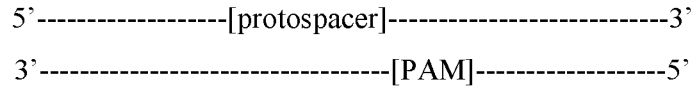
It is contemplated herein that any upstream gRNA described herein may be paired with any downstream gRNA described herein. When an upstream gRNA designed for use with one species of Cas9 is paired with a downstream gRNA designed for use from a different species of Cas9, both Cas9 species are used to generate a single or double-strand break, as desired.

RNA-guided nucleases

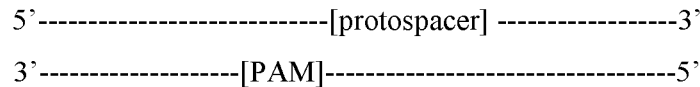
RNA-guided nucleases according to the present disclosure include, but are not limited to, naturally-occurring Class 2 CRISPR nucleases such as Cas9, and Cpf1, as well as other nucleases derived or obtained therefrom. In functional terms, RNA-guided nucleases are defined as those nucleases that: (a) interact with (e.g., complex with) a gRNA; and (b) together with the gRNA, associate with, and optionally cleave or modify, a target region of a DNA that includes (i) a sequence complementary to the targeting domain of the gRNA and, optionally, (ii) a PAM. RNA-guided nucleases can be defined, in broad terms, by their PAM specificity and cleavage activity, even though variations may exist between individual RNA-guided nucleases that share the same PAM specificity or cleavage activity. Skilled artisans will appreciate that some aspects of the present disclosure relate to systems, methods and compositions that can be implemented using any suitable RNA-guided nuclease having a certain PAM specificity and/or cleavage activity. For this reason, unless otherwise specified, the term RNA-guided nuclease should be understood as a generic term, and not limited to any particular type (e.g. Cas9 vs. Cpf1), species (e.g. *S. pyogenes* vs. *S. aureus*) or variation (e.g., full-length vs. truncated or split; naturally-occurring PAM specificity vs. engineered PAM specificity, etc.) of RNA-guided nuclease.

The PAM sequence takes its name from its sequential relationship to the “protospacer” sequence that is complementary to gRNA targeting domains (or “spacers”). Together with protospacer sequences, PAM sequences define target regions or sequences for specific RNA-guided nuclease / gRNA combinations.

Various RNA-guided nucleases may require different sequential relationships between PAMs and protospacers. In general, Cas9s recognize PAM sequences that are 3' of the protospacer as visualized relative to the top or complementary strand:



Cpf1, on the other hand, generally recognizes PAM sequences that are 5' of the protospacer:



In addition to recognizing specific sequential orientations of PAMs and protospacers, RNA-guided nucleases can also recognize specific PAM sequences. *S. aureus* Cas9, for instance, recognizes a PAM sequence of NNGRRT or NNGRRV, wherein the N residues are immediately 3' of the region recognized by the gRNA targeting domain. *S. pyogenes* Cas9 recognizes NGG PAM sequences. And *F. novicida* Cpf1 recognizes a TTN PAM sequence. PAM sequences have been identified for a variety of RNA-guided nucleases, and a strategy for identifying novel PAM sequences has been described by Shmakov 2015. It should also be noted that engineered RNA-guided nucleases can have PAM specificities that differ from the PAM specificities of reference molecules (for instance, in the case of an engineered RNA-guided nuclease, the reference molecule may be the naturally occurring variant from which the RNA-guided nuclease is derived, or the naturally occurring variant having the greatest amino acid sequence homology to the engineered RNA-guided nuclease).

In addition to their PAM specificity, RNA-guided nucleases can be characterized by their DNA cleavage activity: naturally-occurring RNA-guided nucleases typically form DSBs in target nucleic acids, but engineered variants have been produced that generate only SSBs (discussed above) Ran 2013, incorporated by reference herein), or that that do not cut at all.

Cas9 molecules

Cas9 molecules of a variety of species can be used in the methods and compositions described herein. While *S. pyogenes* and *S. aureus* 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. These include, for example, Cas9 molecules from *Acidovorax avenae*, *Actinobacillus pleuropneumoniae*, *Actinobacillus succinogenes*, *Actinobacillus suis*, *Actinomyces sp.*, *Cyclophilus 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*, *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., *Methylosinus trichosporium*, *Mobiluncus mulieris*, *Neisseria bacilliformis*, *Neisseria cinerea*, *Neisseria flavescens*, *Neisseria lactamica*, *Neisseria* sp., *Neisseria wadsworthii*, *Nitrosomonas* sp., *Parvibaculum lavamentivorans*, *Pasteurella multocida*, *Phascolarctobacterium succinatutens*, *Ralstonia syzygii*, *Rhodopseudomonas palustris*, *Rhodovulum* sp., *Simonsiella muelleri*, *Sphingomonas* sp., *Sporolactobacillus vineae*, *Staphylococcus lugdunensis*, *Streptococcus* sp., *Subdoligranulum* sp., *Tistrella mobilis*, *Treponema* sp., or *Verminephrobacter eiseniae*.

Cas9 domains

Crystal structures have been determined for two different naturally occurring bacterial Cas9 molecules (Jinek 2014) and for *S. pyogenes* Cas9 with a guide RNA (e.g., a synthetic fusion of crRNA and tracrRNA) (Nishimasu 2014; Anders 2014).

A naturally occurring Cas9 molecule comprises two lobes: a recognition (REC) lobe and a nuclease (NUC) lobe; each of which further comprise domains described herein. **Figs. 8A-8B** 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 previously (Nishimasu 2014). The numbering of the 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 *S. pyogenes* Cas9 (SEQ ID NO:2). 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 *S. pyogenes* Cas9 (SEQ ID NO:2). 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 *S. pyogenes* Cas9 (SEQ ID NO:2).

The NUC lobe comprises the RuvC domain, the HNH 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 (RuvCI, 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, respectively) at amino acids 1-59, 718-769, and 909-1098, respectively, of *S. pyogenes* Cas9 (SEQ ID NO:2). 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 *S. pyogenes* Cas9 (SEQ ID NO:2). The PI domain interacts with the PAM of the target nucleic acid molecule, and comprises amino acids 1099-1368 of *S. pyogenes* Cas9 (SEQ ID NO:2).

RuvC-like domain and HNH-like domain

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain and a RuvC-like domain, and in certain of these embodiments cleavage activity is dependent on the RuvC-like domain and the HNH-like domain. A Cas9 molecule or Cas9 polypeptide can comprise one or more of a RuvC-like domain and an HNH-like domain. In certain embodiments, a Cas9 molecule or Cas9 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 certain embodiments, 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 certain embodiments, 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 certain embodiments, 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, a Cas9 molecule or Cas9 polypeptide can comprise an N-terminal RuvC-like domain. Exemplary N-terminal RuvC-like domains are described below.

In certain embodiments, a Cas9 molecule or Cas9 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:20),

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 (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 certain embodiments, the N-terminal RuvC-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 certain embodiments, the N-terminal RuvC-like domain is cleavage competent. In other embodiments, the N-terminal RuvC-like domain is cleavage incompetent.

In certain embodiments, a Cas9 molecule or Cas9 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:21),

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);

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 (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).

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of SEQ ID NO:21 by as many as 1 but not more than 2, 3, 4, or 5 residues.

In certain embodiments, 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:22),

wherein

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).

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of SEQ ID NO:22 by as many as 1 but not more than, 2, 3, 4, or 5 residues.

In certain embodiments, the N-terminal RuvC-like domain comprises an amino acid sequence of Formula IV:

D-I-G-T-N-S-V-G-W-A-V-X (SEQ ID NO:23),

wherein

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 Cas9 molecule can comprise an N-terminal RuvC-like domain shown in **Figs. 2A-2G** (depicted as Y)).

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of SEQ ID NO:23 by as many as 1 but not more than, 2, 3, 4, or 5 residues.

In certain embodiments, 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**, 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** are present.

In certain embodiments, 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**, 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** are present.

Additional RuvC-like domains

In addition to the N-terminal RuvC-like domain, the Cas9 molecule or Cas9 polypeptide can comprise one or more additional RuvC-like domains. In certain embodiments, 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.

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

I-X₁-X₂-E-X₃-A-R-E (SEQ ID NO:15),

wherein

X₁ is V or H;

X₂ is I, L or V (e.g., I or V); and

X₃ is M or T.

In certain embodiments, the additional RuvC-like domain comprises an amino acid sequence of Formula VI:

I-V-X₂-E-M-A-R-E (SEQ ID NO:16),

wherein

X₂ is I, L or V (e.g., I or V) (e.g., the Cas9 molecule or Cas9 polypeptide can comprise an additional RuvC-like domain shown in **Fig. 2A-2G** (depicted as B)).

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

H-H-A-X₁-D-A-X₂-X₃ (SEQ ID NO:17),

wherein

X₁ is H or L;

X₂ is R or V; and

X₃ is E or V.

In certain embodiments, the additional RuvC-like domain comprises the amino acid sequence: H-H-A-H-D-A-Y-L (SEQ ID NO:18).

In certain embodiments, the additional RuvC-like domain differs from a sequence of SEQ ID NOs:15-18 by as many as 1 but not more than 2, 3, 4, or 5 residues.

In certain embodiments, the sequence flanking the N-terminal RuvC-like domain has the amino acid sequence of Formula VIII:

$K-X_1'-Y-X_2'-X_3'-X_4'-Z-T-D-X_9'-Y$ (SEQ ID NO:19),

wherein

X_1' is selected from K and P;

X_2' is selected from V, L, I, and F (e.g., V, I and L);

X_3' is selected from G, A and S (e.g., G);

X_4' is selected from L, I, V, and F (e.g., L);

X_9' is selected from D, E, N, and Q; and

Z is an N-terminal RuvC-like domain, e.g., as described above, e.g., having 5 to 20 amino acids.

HNH-like domains

In certain embodiments, an HNH-like domain cleaves a single stranded complementary domain, e.g., a complementary strand of a double stranded nucleic acid molecule. In certain embodiments, an HNH-like domain is at least 15, 20, or 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, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain having an amino acid sequence of Formula IX:

$X_1-X_2-X_3-H-X_4-X_5-P-X_6-X_7-X_8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-N-X^{16}-X^{17}-X^{18}-X^{19}-X_{20}-X_{21}-X_{22}-X_{23}-N$ (SEQ ID NO:25),

wherein

X_1 is selected from D, E, Q, and N (e.g., D and E);

X^2 is selected from L, I, R, Q, V, M, and K;

X_3 is selected from D and E;

X_4 is selected from I, V, T, A, and L (e.g., A, I, and V);

X_5 is selected from V, Y, I, L, F, and W (e.g., V, I, and L);

X_6 is selected from Q, H, R, K, Y, I, L, F, and W;

X_7 is selected from S, A, D, T, and K (e.g., S and A);

X_8 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;

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;

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

X₂₃ is selected from K, V, A, E, Y, I, C, L, S, T, G, K, M, D, and F.

In certain embodiments, a HNH-like domain differs from a sequence of SEQ ID NO:25 by at least one but not more than, 2, 3, 4, or 5 residues.

In certain embodiments, the HNH-like domain is cleavage competent. In other embodiments, the HNH-like domain is cleavage incompetent.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain comprising an amino acid sequence of Formula X:

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:26),

wherein

X₁ is selected from D and E;

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;

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;

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.

In certain embodiments, the HNH-like domain differs from a sequence of SEQ ID NO:26 by 1, 2, 3, 4, or 5 residues.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain comprising an amino acid sequence of Formula XI:

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:27),

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;

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;

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 certain embodiments, the HNH-like domain differs from a sequence of SEQ ID NO:27 by 1, 2, 3, 4, or 5 residues.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain having an amino acid sequence of Formula XII:

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:28),

wherein

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;

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

X₂₃ is selected from E, K, G, and N (e.g., the Cas9 molecule or Cas9 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:28 by as many as 1 but no more than 2, 3, 4, or 5 residues.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises the amino acid sequence of Formula XIII:

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:24),

wherein

X₁' is selected from K and R;

X₂' is selected from V and T;

X₃' is selected from G and D;

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;

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 certain embodiments, the Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence that differs from a sequence of SEQ ID NO:24 by as many as 1 but not more than 2, 3, 4, or 5 residues.

In certain embodiments, the HNH-like domain differs from a sequence of an HNH-like domain disclosed herein, e.g., in **Figs. 5A-5C**, by as many as 1 but not more than 2, 3, 4,

or 5 residues. In certain embodiments, 1 or both of the highly conserved residues identified in **Figs. 5A-5C** are present.

In certain embodiments, the HNH -like domain differs from a sequence of an HNH-like domain disclosed herein, e.g., in **Figs. 6A-6B**, by as many as 1 but not more than 2, 3, 4, or 5 residues. In an embodiment, 1, 2, or all 3 of the highly conserved residues identified in **Figs. 6A-6B** are present.

Cas9 Activities

In certain embodiments, 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 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 (an enzymatically active Cas9) molecule or eaCas9 polypeptide.

In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide comprises one or more of the following enzymatic activities:

- (1) 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;
- (2) 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;
- (3) endonuclease activity;
- (4) exonuclease activity; and
- (5) helicase activity, i.e., the ability to unwind the helical structure of a double stranded nucleic acid.

In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide cleaves both DNA strands and results in a double stranded break. In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide 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 domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with a RuvC domain. In an embodiment,

an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with an HNH domain and cleavage activity associated with a RuvC domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an active, or cleavage competent, HNH domain and an inactive, or cleavage incompetent, RuvC domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an inactive, or cleavage incompetent, HNH domain and an active, or cleavage competent, RuvC domain.

Targeting and PAMs

A Cas9 molecule or Cas9 polypeptide can interact with a gRNA molecule and, in concert with the gRNA molecule, localize to a site which comprises a target domain, and in certain embodiments, a PAM sequence.

In certain embodiments, 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 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, bp upstream from that sequence (see, e.g., Mali 2013). In an embodiment, an eaCas9 molecule of *S. thermophilus* recognizes the sequence motif NGGNG (SEQ ID NO:199) and/or NNAGAAW (W = A or T) (SEQ ID NO:200) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from these sequences (see, e.g., Horvath 2010; Deveau 2008). In an embodiment, an eaCas9 molecule of *S. mutans* recognizes the sequence motif NGG and/or NAAR (R = A or G) (SEQ ID NO:201) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5 bp, upstream from this sequence (see, e.g., Deveau 2008). In an embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRR (R = A or G) (SEQ ID NO:202) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In an embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRRN (R = A or G) (SEQ ID NO:203) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In an embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRRT (R = A or G) (SEQ ID NO:204) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In an embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRRV (R = A or G, V = A, G, or C)

(SEQ ID NO:205) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. The ability of a Cas9 molecule to recognize a PAM sequence can be determined, e.g., using a transformation assay as described previously (Jinek 2012). In each of the aforementioned embodiments (i.e., SEQ ID NOs:199-205), 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.

Exemplary naturally occurring Cas9 molecules have been described previously (see, e.g., Chylinski 2013). 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 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 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 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 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.

Exemplary naturally occurring Cas9 molecules include a Cas9 molecule of a cluster 1 bacterial family. Examples include a Cas9 molecule of: *S. aureus*, *S. pyogenes* (e.g., strains SF370, MGAS10270, MGAS10750, MGAS2096, MGAS315, MGAS5005, MGAS6180, MGAS9429, NZ131, SSI-1), *S. thermophilus* (e.g., strain LMD-9), *S. pseudoporcinus* (e.g., strain SPIN 20026), *S. mutans* (e.g., strains UA159, NN2025), *S. macacae* (e.g., strain NCTC11558), *S. gallolyticus* (e.g., strains UCN34, ATCC BAA-2069), *S. equines* (e.g., strains ATCC 9812, MGCS 124), *S. dysdalaiae* (e.g., strain GGS 124), *S. bovis* (e.g., strain ATCC 700338), *S. anginosus* (e.g., strain F0211), *S. agalactiae* (e.g., strains 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).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence:

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

identical to any Cas9 molecule sequence described herein, or to a naturally occurring Cas9 molecule sequence, e.g., a Cas9 molecule from a species listed herein (e.g., SEQ ID NOs:1, 2, 4-6, or 12) or described in Chylinski 2013. 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 (e.g., an endonuclease and/or exonuclease activity); a helicase activity; or the ability, together with a gRNA molecule, to localize to a target nucleic acid.

In certain embodiments, 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*, or *L. innocua*, and “-” indicates absent. 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 certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises the amino acid sequence of SEQ ID NO:2. In other embodiments, a Cas9 molecule or Cas9 polypeptide differs from the sequence of SEQ ID NO:2 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 (residues 1 to 180, or in the case of region 1' residues 120 to 180)

region 2 (residues 360 to 480);

region 3 (residues 660 to 720);

region 4 (residues 817 to 900); and

region 5 (residues 900 to 960).

In certain embodiments, 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 certain embodiments, regions 1-5 each independently have 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with the corresponding residues of a Cas9 molecule or Cas9 polypeptide described herein, e.g., a sequence from **Figs. 2A-2G** (SEQ ID NOs:1, 2, 4, 5, 14).

In certain embodiments, a Cas9 molecule or Cas9 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 **Fig. 2**; 52% of residues in the four Cas9 sequences in **Figs. 2A-2G** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes* (SEQ ID NO:2);

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* (SEQ ID NOs:2, 4, 1, and 5, respectively); or

is identical to amino acids 1-180 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 1':

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 **Fig. 2**

are conserved) of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively);

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* (SEQ ID NOs:2, 4, 1, and 5, respectively); or

is identical to amino acids 120-180 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 2:

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 **Fig. 2** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively);

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* (SEQ ID NOs:2, 4, 1, and 5, respectively); or

is identical to amino acids 360-480 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 3:

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 **Fig. 2** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively);

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* (SEQ ID NOs:2, 4, 1, and 5, respectively); or

is identical to amino acids 660-720 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively).

In certain embodiments, a Cas9 molecule or Cas9 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* (SEQ ID NOs:2, 4, 1, and 5, respectively);

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* (SEQ ID NOs:2, 4, 1, and 5, respectively); or

is identical to amino acids 817-900 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively).

In certain embodiments, a Cas9 molecule or Cas9 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* (SEQ ID NOs:2, 4, 1, and 5, respectively);

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* (SEQ ID NOs:2, 4, 1, and 5, respectively); or

is identical to amino acids 900-960 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively).

Engineered or altered Cas9

Cas9 molecules and Cas9 polypeptides described herein can possess any of a number of properties, including 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 certain embodiments, a Cas9 molecule or Cas9 polypeptide can include all or a subset of these properties. In a typical embodiment, 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 certain embodiments, 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 Cas9 activities. In certain embodiments, 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 certain embodiments, 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 certain embodiments, a mutation or mutations do not have a substantial effect on a Cas9 activity, e.g., a Cas9 activity described herein. In other embodiments, 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

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.

In certain embodiments, 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 certain embodiments, 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 NOs:24-28) 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:2, 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 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. aureus*, 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., a RuvC-like domain described herein, e.g., SEQ ID NOs:15-23). 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*, *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.

In certain embodiments, 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 domains, e.g., an N-terminal RuvC domain; an HNH domain; a region outside the RuvC domains and the HNH domain. In an embodiment, a mutation(s) is present in a RuvC domain. In an embodiment, a mutation(s) is present in an HNH domain. In an embodiment, mutations are present in both a RuvC domain and an HNH domain.

Exemplary mutations that may be made in the RuvC domain or HNH domain with reference to the *S. pyogenes* Cas9 sequence include: D10A, E762A, H840A, N854A, N863A, and/or D986A. Exemplary mutations that may be made in the RuvC domain with reference to the *S. aureus* Cas9 sequence include N580A (see, e.g., SEQ ID NO:11).

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. 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 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 can differ from naturally occurring Cas9 molecules, e.g., a Cas9 molecule of *S. aureus* or *S. pyogenes*, 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* or *S. pyogenes*); 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 naturally occurring Cas9 molecule (e.g., a Cas9 molecule of *S. aureus* or *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. In certain embodiments, the nickase is *S. aureus* Cas9-derived nickase comprising the sequence of SEQ ID NO:10 (D10A) or SEQ ID NO:11 (N580A) (Friedland 2015).

In an embodiment, the altered Cas9 molecule is an eaCas9 molecule comprising one or more of the following activities: cleavage activity 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 certain embodiments, 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**; and

the sequence corresponding to the residues identified by “*” in the consensus sequence disclosed in **Figs. 2A-2G** differs 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*, *S. thermophilus*, *S. mutans*, or *L. innocua* Cas9 molecule.

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising the amino acid sequence of *S. pyogenes* Cas9 disclosed in **Figs. 2A-2G** (SEQ ID NO:2) with one or more amino acids that differ from the sequence of *S. pyogenes* (e.g., substitutions) at one or more residues (e.g., 2, 3, 5, 10, 15, 20, 30, 50, 70, 80, 90, 100, or 200 amino acid residues) represented by an “*” in the consensus sequence disclosed in **Figs. 2A-2G** (SEQ ID NO:14).

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising the amino acid sequence of *S. thermophilus* Cas9 disclosed in **Figs. 2A-2G** (SEQ ID NO:4) with one or more amino acids that differ from the sequence of *S. thermophilus* (e.g., substitutions) at one or more residues (e.g., 2, 3, 5, 10, 15, 20, 30, 50, 70, 80, 90, 100, or 200 amino acid residues) represented by an “*” in the consensus sequence disclosed in **Figs. 2A-2G** (SEQ ID NO:14).

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising the amino acid sequence of *S. mutans* Cas9

disclosed in **Figs. 2A-2G** (SEQ ID NO:1) with one or more amino acids that differ from the sequence of *S. mutans* (e.g., substitutions) at one or more residues (e.g., 2, 3, 5, 10, 15, 20, 30, 50, 70, 80, 90, 100, or 200 amino acid residues) represented by an “*” in the consensus sequence disclosed in **Figs. 2A-2G** (SEQ ID NO:14).

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising the amino acid sequence of *L. innocua* Cas9 disclosed in **Figs. 2A-2G** (SEQ ID NO:5) with one or more amino acids that differ from the sequence of *L. innocua* (e.g., substitutions) at one or more residues (e.g., 2, 3, 5, 10, 15, 20, 30, 50, 70, 80, 90, 100, or 200 amino acid residues) represented by an “*” in the consensus sequence disclosed in **Figs. 2A-2G** (SEQ ID NO:14).

In certain embodiments, 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 a 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 with altered or no PAM recognition

Naturally occurring Cas9 molecules can recognize specific PAM sequences, for example the PAM recognition sequences described above for, e.g., *S. pyogenes*, *S. thermophilus*, *S. mutans*, and *S. aureus*.

In certain embodiments, 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 or Cas9 polypeptide recognizes in order to decrease off-target sites and/or improve specificity; or eliminate a PAM recognition requirement. In certain embodiments, 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/or increase specificity. In certain embodiments, 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 (see, e.g., Esvelt 2011). Candidate Cas9 molecules can be evaluated, e.g., by methods described below.

Size-optimized Cas9

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* or *S. pyogenes* 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 can comprise one or more deletions that do not substantially affect or decrease the activity of the resultant Cas9 molecules described herein. Activities that are retained in the Cas9 molecules 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;

- 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.

Activity of the Cas9 molecules 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. Naturally-occurring orthologous Cas9 molecules from various bacterial species, e.g., any one of those listed in **Table 1**, can be modeled onto the crystal structure of *S. pyogenes* Cas9 (Nishimasu 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.

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 have been described previously (see, e.g., Cong 2013; Wang 2013; Mali 2013; Jinek 2012).

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 chemically modified, e.g., as described herein. 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 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 sequences are known in the art.

An exemplary codon optimized nucleic acid sequence encoding a Cas9 molecule of *S. pyogenes* is set forth in SEQ ID NO:3. The corresponding amino acid sequence of an *S. pyogenes* Cas9 molecule is set forth in SEQ ID NO:2.

Exemplary codon optimized nucleic acid sequences encoding a Cas9 molecule of *S. aureus* are set forth in SEQ ID NOs:7-9. An amino acid sequence of an *S. aureus* Cas9 molecule is set forth in SEQ ID NO:6.

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 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) have been described previously (see, e.g., Haft 2005 and Makarova 2011). Exemplary Cas molecules (and Cas systems) are also shown in **Table 2**.

Table 2: Cas systems

Gene name [‡]	System type or subtype	Name from Haft 2005 [§]	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	<i>cas5a</i> , <i>cas5d</i> , <i>cas5e</i> , <i>cas5h</i> ,	3KG4	COG1688 (RAMP)	APE1234, BH0337, <i>devS</i> and <i>ygcI</i>

Gene name [‡]	System type or subtype	Name from Haft 2005 [§]	Structure of encoded protein (PDB accessions) [¶]	Families (and superfamily) of encoded protein ^{***}	Representatives
	• Subtype I-C • Subtype I-E	<i>cas5p</i> , <i>cas5t</i> and <i>cmx5</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)	y1727
<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>csb2</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>csb1</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
<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-F ^{††}	<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

Gene name [‡]	System type or subtype	Name from Haft 2005 [§]	Structure of encoded protein (PDB accessions) [¶]	Families (and superfamily) of encoded protein ^{***}	Representatives
<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
<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

Gene name [‡]	System type or subtype	Name from Haft 2005 [§]	Structure of encoded protein (PDB accessions) [¶]	Families (and superfamily) of encoded protein ^{***}	Representatives
<i>csf4</i>	• Type U	<i>csf4</i>	NA	NA	AFE_1037

Cpf1 molecules

The crystal structure of *Acidaminococcus* sp. Cpf1 in complex with crRNA and a double-stranded (ds) DNA target including a TTTN PAM sequence has been solved by Yamano 2016, incorporated by reference herein. Cpf1, like Cas9, has two lobes: a REC (recognition) lobe, and a NUC (nuclease) lobe. The REC lobe includes REC1 and REC2 domains, which lack similarity to any known protein structures. The NUC lobe, meanwhile, includes three RuvC domains (RuvC-I, -II and -III) and a BH domain. However, in contrast to Cas9, the Cpf1 REC lobe lacks an HNH domain, and includes other domains that also lack similarity to known protein structures: a structurally unique PI domain, three Wedge (WED) domains (WED-I, -II and -III), and a nuclease (Nuc) domain.

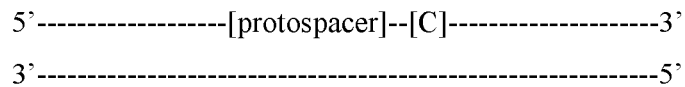
While Cas9 and Cpf1 share similarities in structure and function, it should be appreciated that certain Cpf1 activities are mediated by structural domains that are not analogous to any Cas9 domains. For instance, cleavage of the complementary strand of the target DNA appears to be mediated by the Nuc domain, which differs sequentially and spatially from the HNH domain of Cas9. Additionally, the non-targeting portion of Cpf1 gRNA (the handle) adopts a pseudoknot structure, rather than a stem loop structure formed by the repeat:anti-repeat duplex in Cas9 gRNAs.

Modifications of RNA-guided nucleases

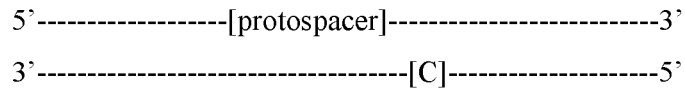
The RNA-guided nucleases described above have activities and properties that can be useful in a variety of applications, but the skilled artisan will appreciate that RNA-guided nucleases can also be modified in certain instances, to alter cleavage activity, PAM specificity, or other structural or functional features.

Turning first to modifications that alter cleavage activity, mutations that reduce or eliminate the activity of domains within the NUC lobe have been described above. Exemplary mutations that may be made in the RuvC domains, in the Cas9 HNH domain, or in the Cpf1 Nuc domain are described in Ran 2013 and Yamano 2016, as well as in Cotta-Ramusino 2016. In general, mutations that reduce or eliminate activity in one of the two nuclease domains result in RNA-guided nucleases with nickase activity, but it should be noted that the type of nickase activity varies depending on which domain is inactivated. As

one example, inactivation of a RuvC domain of a Cas9 will result in a nickase that cleaves the complementary or top strand as shown below (where C denotes the site of cleavage):



On the other hand, inactivation of a Cas9 HNH domain results in a nickase that cleaves the bottom or non-complementary strand:



Modifications of PAM specificity relative to naturally occurring Cas9 reference molecules has been described by Kleinstiver 2015a for both *S. pyogenes* and *S. aureus* (Kleinstiver 2015b). Kleinstiver et al. have also described modifications that improve the targeting fidelity of Cas9 (Kleinstiver 2016). Each of these references is incorporated by reference herein.

RNA-guided nucleases have been split into two or more parts, as described by Zetsche 2015, incorporated by reference, and by Fine 2015, incorporated by reference.

RNA-guided nucleases can be, in certain embodiments, size-optimized or truncated, for instance via one or more deletions that reduce the size of the nuclease while still retaining gRNA association, target and PAM recognition, and cleavage activities. In certain embodiments, RNA guided nucleases are bound, covalently or non-covalently, to another polypeptide, nucleotide, or other structure, optionally by means of a linker. Exemplary bound nucleases and linkers are described by Guilinger 2014, which is incorporated by reference for all purposes herein.

RNA-guided nucleases also optionally include a tag, such as, but not limited to, a nuclear localization signal to facilitate movement of RNA-guided nuclease protein into the nucleus. In certain embodiments, the RNA-guided nuclease can incorporate C- and/or N-terminal nuclear localization signals. Nuclear localization sequences are known in the art and are described in Maeder 2015 and elsewhere.

The foregoing list of modifications is intended to be exemplary in nature, and the skilled artisan will appreciate, in view of the instant disclosure, that other modifications may be possible or desirable in certain applications. For brevity, therefore, exemplary systems, methods and compositions of the present disclosure are presented with reference to particular RNA-guided nucleases, but it should be understood that the RNA-guided nucleases used may be modified in ways that do not alter their operating principles. Such modifications are within the scope of the present disclosure.

Nucleic acids encoding RNA-guided nucleases

Nucleic acids encoding RNA-guided nucleases, e.g., Cas9, Cpf1 or functional fragments thereof, are provided herein. Exemplary nucleic acids encoding RNA-guided nucleases have been described previously (see, e.g., Cong 2013; Wang 2013; Mali 2013; Jinek 2012).

In some cases, a nucleic acid encoding an RNA-guided nuclease can be a synthetic nucleic acid sequence. For example, the synthetic nucleic acid molecule can be chemically modified. In certain embodiments, an mRNA encoding an RNA-guided nuclease will have one or more (e.g., all) of the following properties: it can be capped; polyadenylated; and substituted with 5-methylcytidine and/or pseudouridine.

Synthetic nucleic acid sequences can also be codon optimized, e.g., at least one non-common codon or less-common codon has been replaced by a common 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. Examples of codon optimized Cas9 coding sequences are presented in Cotta-Ramusino 2016.

In addition, or alternatively, a nucleic acid encoding an RNA-guided nuclease may comprise a nuclear localization sequence (NLS). Nuclear localization sequences are known in the art.

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 have been described previously (Jinek 2012).

Binding and cleavage assay: testing the endonuclease activity of Cas9 molecules

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, a 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 minutes 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, while 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 - ^{32} P]-ATP in 1X T4 polynucleotide kinase reaction buffer at 37°C for 30 minutes, 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 minutes, followed by slow cooling to room temperature. For cleavage assays, gRNA molecules are annealed by heating to 95°C for 30 seconds, 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 hour 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 minutes. 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 molecules to target DNA

Exemplary methods for evaluating the binding of Cas9 molecules to target DNA have been described previously (Jinek 2012).

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 minutes, and slow 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 minutes at 37°C. Polynucleotide kinase is heat denatured at 65°C for 20 minutes, and unincorporated radiolabel 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 molecules are 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 hour at 37°C and resolved at 4°C on an 8% native polyacrylamide 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 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 can be 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.

To determine the best solution conditions for forming RNP complexes, a 2 μ M solution of Cas9 is made in water with 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 minutes and brief centrifugation to 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°C increase in temperature every 10 seconds.

The second assay consists of mixing various concentrations of gRNA with 2 μ M Cas9 in optimal buffer from assay 1 above, and incubating at room temperature for 10 minutes in a 384 well plate. An equal volume of optimal buffer and 10x SYPRO Orange® (Life Technologies cat#S-6650) is added and the plate is sealed with Microseal® B adhesive (MSB-1001). Following brief centrifugation to 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°C increase in temperature every 10 seconds.

NHEJ approaches for gene targeting

In certain embodiments of the methods provided herein, NHEJ-mediated deletion is used to delete all or a portion of a γ -globin gene (e.g., HBG1, HBG2) negative regulatory element (e.g., silencer). As described herein, nuclease-induced NHEJ can be used to knock out all or a portion of a regulatory element in a target-specific manner. In other embodiments, NHEJ-mediated insertion is used to insert a sequence into a γ -globin gene negative regulatory element, resulting in inactivation of the regulatory element.

While not wishing to be bound by theory, it is believed that, in certain embodiments, 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; they are most commonly in the 1-50 bp range, but 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 a regulatory region of interest can be used to disrupt or delete a target regulatory element.

Placement of double strand or single strand breaks relative to the target position

In certain embodiments 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 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 certain embodiments 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 single-strand breaks to provide for NHEJ repair a nucleotide of the target position. In certain embodiments, 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 certain embodiments, 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 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 certain embodiments, the gRNAs are configured to place a single strand break on either side of a nucleotide of the target position.

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 certain embodiments, two gRNAs, e.g., independently, unimolecular (or chimeric) or modular gRNA, are configured to position a double-strand break on both sides of a target position. In other embodiments, 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 still other embodiments, four gRNAs, e.g., independently, 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).

HDR repair, HDR-mediated knock-in, knock-out, or deletion, and template nucleic acids

In certain embodiments of the methods provided herein, HDR-mediated sequence alteration is used to alter (e.g., delete, disrupt, or modify) the sequence of one or more nucleotides in a γ -globin gene (e.g., HBG1, HBG2) regulatory region using an exogenously provided template nucleic acid (also referred to herein as a donor construct). While not wishing to be bound by theory, it is believed that HDR-mediated alteration of an HBG target position within a γ -globin gene regulatory region occurs by HDR with an exogenously provided donor template or template nucleic acid. For example, the donor construct or template nucleic acid provides for alteration of an HBG target position. It is contemplated that a plasmid donor can be used as a template for homologous recombination. It is further contemplated that a single stranded donor template can be used as a template for alteration of the HBG target position by alternate methods of HDR (e.g., single strand annealing) between the target sequence and the donor template. Donor template-effected alteration of an HBG target position depends on cleavage by a Cas9 molecule. Cleavage by Cas9 can comprise a double-strand break or two single-strand breaks.

In certain embodiments of the methods provided herein, HDR-mediated alteration is used to knock out or delete all or a portion of a γ -globin gene (e.g., HBG1, HBG2) negative

regulatory element (e.g., silencer). As described herein, HDR can be used to knock out or delete all or a portion of a regulatory element in a target-specific manner.

In other embodiments, HDR-mediated sequence alteration is used to alter the sequence of one or more nucleotides in a γ -globin gene (e.g., HBG1, HBG2) regulatory region without using an exogenously provided template nucleic acid. While not wishing to be bound by theory, it is believed that alteration of an HBG target position occurs by HDR with an endogenous genomic donor sequence. For example, the endogenous genomic donor sequence provides for alteration of the HBG target position. It is contemplated that in an embodiment the endogenous genomic donor sequence is located on the same chromosome as the target sequence. It is further contemplated that in another embodiment the endogenous genomic donor sequence is located on a different chromosome from the target sequence. Alteration of an HBG target position by endogenous genomic donor sequence depends on cleavage by a Cas9 molecule. Cleavage by Cas9 can comprise a double-strand break or two single-strand breaks.

In certain embodiments of the methods provided herein, HDR-mediated alteration is used to alter a single nucleotide in a γ -globin gene regulatory region. These embodiments may utilize either one double-strand break or two single-strand breaks. In certain embodiments, a single nucleotide alteration is incorporated using (1) one double-strand break, (2) two single-strand breaks, (3) two double-strand breaks with a break occurring on each side of the target position, (4) one double-strand break and two single strand breaks with the double strand break and two single strand breaks occurring on each side of the target position, (5) four single-strand breaks with a pair of single-strand breaks occurring on each side of the target position, or (6) one single-strand break.

In certain embodiments wherein a single-stranded template nucleic acid is used, the target position can be altered by alternative HDR.

In certain embodiments of the methods provided herein, HDR-mediated alteration is used to introduce an alteration (e.g., deletion) of one or more nucleotides in a γ -globin gene regulatory region. In certain embodiments, the γ -globin gene regulatory region may be a HBG target position. In certain embodiments, the alteration (e.g., deletion) may be introduced at a target site within the HBG target position. In certain embodiments, the alteration (e.g., deletion) may be selected from one or more of HBG1 13 bp del c.-114 to -102, HBG1 4 bp del c.-225 to -222, and HBG1 13 bp del c.-114 to -102. In certain embodiments, the target site may be selected from one or more of HBG1 c.-114 to -102 (e.g., nucleotides 2824-2836 of SEQ ID NO:902 (HBG1)), HBG1 c.-225 to -222 (e.g., nucleotides

2716-2719 of SEQ ID NO:902 (HBG1)), and HBG2 c.-114 to -102 (e.g., nucleotides 2748-2760 of SEQ ID NO:903 (HBG2)).

Donor template-effected alteration of an HBG target position depends on cleavage by a Cas9 molecule. Cleavage by Cas9 can comprise a nick, a double-strand break, or two single-strand breaks, e.g., one on each strand of the target nucleic acid. After introduction of the breaks on the target nucleic acid, resection occurs at the break ends resulting in single stranded overhanging DNA regions.

In canonical HDR, a double-stranded donor template is introduced, comprising homologous sequence to the target nucleic acid that will either be directly incorporated into the target nucleic acid or used as a template to change the sequence of the target nucleic acid. After resection at the break, repair can progress by different pathways, e.g., by the double Holliday junction model (or double-strand break repair, DSB, pathway) or the synthesis-dependent strand annealing (SDSA) pathway. In the double Holliday junction model, strand invasion by the two single stranded overhangs of the target nucleic acid to the homologous sequences in the donor template occurs, resulting in the formation of an intermediate with two Holliday junctions. The junctions migrate as new DNA is synthesized from the ends of the invading strand to fill the gap resulting from the resection. The end of the newly synthesized DNA is ligated to the resected end, and the junctions are resolved, resulting in alteration of the target nucleic acid, e.g., incorporation of an HPFH mutant sequence of the donor template at the corresponding HBG target position. Crossover with the donor template may occur upon resolution of the junctions. In the SDSA pathway, only one single stranded overhang invades the donor template and new DNA is synthesized from the end of the invading strand to fill the gap resulting from resection. The newly synthesized DNA then anneals to the remaining single stranded overhang, new DNA is synthesized to fill in the gap, and the strands are ligated to produce the altered DNA duplex.

In alternative HDR, a single strand donor template, e.g., template nucleic acid, is introduced. A nick, single-strand break, or double-strand break at the target nucleic acid, for altering a desired HBG target position, is mediated by a Cas9 molecule, e.g., described herein, and resection at the break occurs to reveal single stranded overhangs. Incorporation of the sequence of the template nucleic acid to alter an HBG target position typically occurs by the SDSA pathway, as described above.

Additional details on template nucleic acids are provided in Section IV entitled "Template nucleic acids" in International Application PCT/US2014/057905.

In certain embodiments, double-strand cleavage is effected by a Cas9 molecule having cleavage activity associated with an HNH-like domain and cleavage activity associated with a RuvC-like domain, e.g., an N-terminal RuvC-like domain, e.g., a wild-type Cas9. Such embodiments require only a single gRNA.

In certain embodiments, one single-strand break, or nick, is effected by a Cas9 molecule having nickase activity, e.g., a Cas9 nickase as described herein. A nicked target nucleic acid can be a substrate for alt-HDR.

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 usually 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 certain embodiments, 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 (see, e.g., SEQ ID NO:10). D10A inactivates RuvC; therefore, the Cas9 nickase has (only) HNH activity and will cut on the strand to which the gRNA hybridizes (e.g., 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 (e.g., the strand that has the NGG PAM and whose sequence is identical to the gRNA). In other embodiments, a Cas9 molecule having an N863 mutation, e.g., the N863A mutation, mutation can be used as a nickase. N863A 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 certain embodiments 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 about 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 2013).

In certain embodiments, a single nick can be used to induce HDR, e.g., alt-HDR. It is contemplated herein that a single nick can be used to increase the ratio of HR to NHEJ at a given cleavage site. 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 other embodiments, 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.

Placement of double-strand or single-strand breaks relative to the target position

A double-strand break or single-strand break in one of the strands should be sufficiently close to an HBG target position that an alteration is produced in the desired region, e.g., incorporation of an HPFH mutation. In certain embodiments, the distance is not more than 50, 100, 200, 300, 350, or 400 nucleotides from the HBG target position. While not wishing to be bound by theory, in certain embodiments it is believed that the break should be sufficiently close to the HBG target position that the target position is within the region that is subject to exonuclease-mediated removal during end resection. If the distance between the HBG target position and a break is too great, the sequence desired to be altered may not be included in the end resection and, therefore, may not be altered, as donor sequence, either exogenously provided donor sequence or endogenous genomic donor sequence, in some embodiments is only used to alter sequence within the end resection region.

In certain embodiments, the methods described herein introduce one or more breaks near a γ -globin gene regulatory region(s), e.g., enhancer region(s), e.g., silencer region(s), e.g., promoter region(s) of the HGB1 and/or HGB2 gene(s). In certain of these embodiments, two or more breaks are introduced that flank at least a portion of the regulatory region(s), e.g., enhancer region(s), e.g., silencer region(s), of the HGB1 and/or HGB2 gene(s). The two or more breaks remove (e.g., delete) a genomic sequence including at least a portion of the γ -globin gene regulatory region(s), e.g., enhancer region(s), e.g., silencer region(s), of the HGB1 and/or HGB2 gene(s). All methods described herein result in altering the regulatory region(s), e.g., enhancer region(s), e.g., silencer region(s), of the HGB1 and/or HGB2 gene(s).

In certain embodiments, the gRNA targeting domain is configured such that a cleavage event, e.g., a double strand or single strand break, is positioned within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, or 200 nucleotides of the region

desired to be altered, e.g., a mutation. The break, e.g., a double-strand or single-strand break, can be positioned upstream or downstream of the region desired to be altered, e.g., a mutation. In some embodiments, a break is positioned within the region desired to be altered, e.g., within a region defined by at least two mutant nucleotides. In some embodiments, a break is positioned immediately adjacent to the region desired to be altered, e.g., immediately upstream or downstream of a mutation.

In certain embodiments, a single-strand break is accompanied by an additional single-strand break, positioned by a second gRNA molecule, as discussed below. For example, the targeting domains bind configured such that a cleavage event, e.g., the two single strand breaks, are positioned within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, or 200 nucleotides of an HBG target position. In an embodiment, the first and second gRNA molecules are configured such that, when guiding a Cas9 nickase, a single-strand break will be accompanied by an additional single strand break, positioned by a second gRNA, sufficiently close to one another to result in alteration of the desired region. In an embodiment, the first and second gRNA molecules are configured such that a single-strand break positioned by said second gRNA is within 10, 20, 30, 40, or 50 nucleotides of the break positioned by said first gRNA molecule, e.g., when the Cas9 is a nickase. In an embodiment, the two gRNA molecules are configured to position cuts at the same position, or within a few nucleotides of one another, on different strands, e.g., essentially mimicking a double-strand break.

In certain embodiments in which a gRNA (unimolecular (or chimeric) or modular gRNA) and Cas9 nuclease induce a double-strand break for the purpose of inducing HDR-mediated sequence alteration, the cleavage site is 0-200 bp (e.g., 0 to 175, 0 to 150, 0 to 125, 0 to 100, 0 to 75, 0 to 50, 0 to 25, 25 to 200, 25 to 175, 25 to 150, 25 to 125, 25 to 100, 25 to 75, 25 to 50, 50 to 200, 50 to 175, 50 to 150, 50 to 125, 50 to 100, 50 to 75, 75 to 200, 75 to 175, 75 to 150, 75 to 125, 75 to 100 bp) away from the HBG target position. In certain embodiments, the cleavage site is 0-100 bp (e.g., 0 to 75, 0 to 50, 0 to 25, 25 to 100, 25 to 75, 25 to 50, 50 to 100, 50 to 75 or 75 to 100 bp) away from the HBG target position.

In certain embodiments, one can promote HDR by using nickases to generate a break with overhangs. While not wishing to be bound by theory, the single stranded nature of the overhangs can enhance the likelihood of a cell repairing the break by HDR as opposed to, e.g., NHEJ. Specifically, in some embodiments, HDR is promoted by selecting a first gRNA that targets a first nickase to a first target sequence, and a second gRNA that targets a second

nickase to a second target sequence which is on the opposite DNA strand from the first target sequence and offset from the first nick.

In certain embodiments, the targeting domain of a gRNA molecule is configured to position a cleavage event sufficiently far from a preselected nucleotide that the nucleotide is not altered. In certain embodiments, the targeting domain of a gRNA molecule is configured to position an intronic cleavage event sufficiently far from an intron/exon border, or naturally occurring splice signal, to avoid alteration of the exonic sequence or unwanted splicing events. The gRNA molecule may be a first, second, third and/or fourth gRNA molecule, as described herein.

Placement of a first break and a second break relative to each other

In certain embodiments, a double-strand break can be accompanied by an additional double-strand break, positioned by a second gRNA molecule, as is discussed below.

In certain embodiments, a double-strand break can be accompanied by two additional single-strand breaks, positioned by a second gRNA molecule and a third gRNA molecule.

In certain embodiments, first and second single-strand breaks can be accompanied by two additional single-strand breaks positioned by a third gRNA molecule and a fourth gRNA molecule.

When two or more gRNAs are used to position two or more cleavage events, e.g., double-strand or single-strand breaks, in a target nucleic acid, it is contemplated that the two or more cleavage events may be made by the same or different Cas9 proteins. For example, when two gRNAs are used to position two double-strand breaks, a single Cas9 nuclease may be used to create both double-strand breaks. When two or more gRNAs are used to position two or more single-strand breaks (nicks), a single Cas9 nickase may be used to create the two or more nicks. When two or more gRNAs are used to position at least one double-strand break and at least one single-strand break, two Cas9 proteins may be used, e.g., one Cas9 nuclease and one Cas9 nickase. It is contemplated that when two or more Cas9 proteins are used that the two or more Cas9 proteins may be delivered sequentially to control specificity of a double-strand versus a single-strand break at the desired position in the target nucleic acid.

In some embodiments, the targeting domain of the first 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 configured such that the PAMs are oriented outward.

In certain embodiments, two gRNA are selected to direct Cas9-mediated cleavage at two positions that are a preselected distance from each other. In certain embodiments, the two points of cleavage are on opposite strands of the target nucleic acid. In some embodiments, the two cleavage points form a blunt ended break, and in other embodiments, they are offset so that the DNA ends comprise one or two overhangs (e.g., one or more 5' overhangs and/or one or more 3' overhangs). In some embodiments, each cleavage event is a nick. In certain embodiments, the nicks are close enough together that they form a break that is recognized by the double stranded break machinery (as opposed to being recognized by, e.g., the SSB machinery). In certain embodiments, the nicks are far enough apart that they create an overhang that is a substrate for HDR, i.e., the placement of the breaks mimics a DNA substrate that has experienced some resection. For instance, in some embodiments the nicks are spaced to create an overhang that is a substrate for processive resection. In some embodiments, the two breaks are spaced within 25-65 nucleotides of each other. The two breaks may be, e.g., about 25, 30, 35, 40, 45, 50, 55, 60, or 65 nucleotides of each other. The two breaks may be, e.g., at least about 25, 30, 35, 40, 45, 50, 55, 60, or 65 nucleotides of each other. The two breaks may be, e.g., at most about 30, 35, 40, 45, 50, 55, 60, or 65 nucleotides of each other. In certain embodiments, the two breaks are about 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, or 60-65 nucleotides of each other.

In some embodiments, the break that mimics a resected break comprises a 3' overhang (e.g., generated by a DSB and a nick, where the nick leaves a 3' overhang), a 5' overhang (e.g., generated by a DSB and a nick, where the nick leaves a 5' overhang), a 3' and a 5' overhang (e.g., generated by three cuts), two 3' overhangs (e.g., generated by two nicks that are offset from each other), or two 5' overhangs (e.g., generated by two nicks that are offset from each other).

In certain embodiments in which two gRNAs (independently, unimolecular (or chimeric) or modular gRNA) complexing with Cas9 nickases induce two single-strand breaks for the purpose of inducing HDR-mediated alteration, the closer nick is between 0-200 bp (e.g., 0 to 175, 0 to 150, 0 to 125, 0 to 100, 0 to 75, 0 to 50, 0 to 25, 25 to 200, 25 to 175, 25 to 150, 25 to 125, 25 to 100, 25 to 75, 25 to 50, 50 to 200, 50 to 175, 50 to 150, 50 to 125, 50 to 100, 50 to 75, 75 to 200, 75 to 175, 75 to 150, 75 to 125, or 75 to 100 bp) away from the HBG target position and the two nicks will ideally be within 25-65 bp of each other (e.g., 25 to 50, 25 to 45, 25 to 40, 25 to 35, 25 to 30, 30 to 55, 30 to 50, 30 to 45, 30 to 40, 30 to 35, 35 to 55, 35 to 50, 35 to 45, 35 to 40, 40 to 55, 40 to 50, 40 to 45 bp, 45 to 50 bp, 50 to 55 bp, 55 to 60 bp, or 60 to 65 bp) and no more than 100 bp away from each other (e.g., no more than

90, 80, 70, 60, 50, 40, 30, 20, 10, or 5 bp away from each other). In certain embodiments, the cleavage site is between 0-100 bp (e.g., 0 to 75, 0 to 50, 0 to 25, 25 to 100, 25 to 75, 25 to 50, 50 to 100, 50 to 75, or 75 to 100 bp) away from the HBG target position.

In some embodiments, two gRNAs, e.g., independently, unimolecular (or chimeric) or modular gRNA, are configured to position a double-strand break on both sides of a target position. In other embodiments, 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 other embodiments, four gRNAs, e.g., independently, 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 HBG 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, in certain embodiments, within 25-65 bp of each other (e.g., between 25 to 55, 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, 40 to 45 bp, 45 to 50 bp, 50 to 55 bp, 55 to 60 bp, or 60 to 65 bp) and no more than 100 bp away from each other (e.g., no more than 90, 80, 70, 60, 50, 40, 30, or 20 or 10 bp).

When two gRNAs are used to target Cas9 molecules to breaks, different combinations of Cas9 molecules are envisioned. In some embodiments, a first gRNA is used to target a first Cas9 molecule to a first target position, and a second gRNA is used to target a second Cas9 molecule to a second target position. In some embodiments, the first Cas9 molecule creates a nick on the first strand of the target nucleic acid, and the second Cas9 molecule creates a nick on the opposite strand, resulting in a double-strand break (e.g., a blunt ended cut or a cut with overhangs).

Different combinations of nickases can be chosen to target one single-strand break to one strand and a second single-strand break to the opposite strand. When choosing a combination, one can take into account that there are nickases having one active RuvC-like domain, and nickases having one active HNH domain. In certain embodiments, a RuvC-like domain cleaves the non-complementary strand of the target nucleic acid molecule. In certain embodiments, an HNH-like domain cleaves a single stranded complementary domain, e.g., a complementary strand of a double stranded nucleic acid molecule. Generally, if both Cas9

molecules have the same active domain (e.g., both have an active RuvC domain or both have an active HNH domain), one will choose two gRNAs that bind to opposite strands of the target. In more detail, in some embodiments a first gRNA is complementary with a first strand of the target nucleic acid and binds a nickase having an active RuvC-like domain and causes that nickase to cleave the strand that is non-complementary to that first gRNA, i.e., a second strand of the target nucleic acid; and a second gRNA is complementary with a second strand of the target nucleic acid and binds a nickase having an active RuvC-like domain and causes that nickase to cleave the strand that is non-complementary to that second gRNA, i.e., the first strand of the target nucleic acid. Conversely, in some embodiments, a first gRNA is complementary with a first strand of the target nucleic acid and binds a nickase having an active HNH domain and causes that nickase to cleave the strand that is complementary to that first gRNA, i.e., a first strand of the target nucleic acid; and a second gRNA is complementary with a second strand of the target nucleic acid and binds a nickase having an active HNH domain and causes that nickase to cleave the strand that is complementary to that second gRNA, i.e., the second strand of the target nucleic acid. In another arrangement, if one Cas9 molecule has an active RuvC-like domain and the other Cas9 molecule has an active HNH domain, the gRNAs for both Cas9 molecules can be complementary to the same strand of the target nucleic acid, so that the Cas9 molecule with the active RuvC-like domain will cleave the non-complementary strand and the Cas9 molecule with the HNH domain will cleave the complementary strand, resulting in a double stranded break.

Homology arms of the donor template

A 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 or LINE repeats.

Exemplary homology arm lengths include at least 50, 100, 250, 500, 750, 1000, 2000, 3000, 4000, or 5000 nucleotides. In some embodiments, the homology arm length is 50-100, 100-250, 250-500, 500-750, 750-1000, 1000-2000, 2000-3000, 3000-4000, or 4000-5000 nucleotides.

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 (e.g., delete, disrupt, or modify) the structure of an HBG target position. In certain embodiments,

the HBG 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. Alternatively, the HBG target position may comprise one or more nucleotides that are altered by a template nucleic acid. In certain embodiments, an alteration (e.g., deletion) may be introduced at a target site within the HBG target position. In certain embodiments, the alteration (e.g., deletion) may be selected from one or more of HBG1 13 bp del c.-114 to -102, HBG1 4 bp del c.-225 to -222, and HBG1 13 bp del c.-114 to -102. In certain embodiments, the target site may be selected from one or more of HBG1 c.-114 to -102 (e.g., nucleotides 2824-2836 of SEQ ID NO:902 (HBG1)), HBG1 c.-225 to -222 (e.g., nucleotides 2716-2719 of SEQ ID NO:902 (HBG1)), and HBG2 c.-114 to -102 (e.g., nucleotides 2748-2760 of SEQ ID NO:903 (HBG2)).

In certain embodiments, the target nucleic acid is modified to have some or all of the sequence of the template nucleic acid, typically at or near cleavage site(s). In certain embodiments, the template nucleic acid is single stranded. In other embodiments, the template nucleic acid is double stranded. In certain embodiments, the template nucleic acid is DNA, e.g., double stranded DNA. In other embodiments, 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 certain embodiments, the template nucleic acid is excised from a vector backbone *in vivo*, e.g., it is flanked by gRNA recognition sequences. In certain embodiments, the template nucleic acid comprises endogenous genomic sequence.

In certain embodiments, the template nucleic acid alters the structure of the target position by participating in an HDR event. In certain embodiments, the template nucleic acid alters the sequence of the target position. In certain embodiments, the template nucleic acid results in the incorporation of a modified, or non-naturally occurring base into the target nucleic acid.

In certain embodiments, the template nucleic acid results in a deletion of one or more nucleotides of the target nucleic acid. In certain embodiments, the template nucleic acid results in deletion of one or more nucleotides of a HBG target position. In certain embodiments, an alteration (e.g., deletion) may be introduced at a target site within the HBG target position. In certain embodiments, the alteration (e.g., deletion) may be selected from one or more of HBG1 13 bp del c.-114 to -102, HBG1 4 bp del c.-225 to -222, and HBG1 13 bp del c.-114 to -102. In certain embodiments, the target site may be selected from one or more of HBG1 c.-114 to -102 (e.g., nucleotides 2824-2836 of SEQ ID NO:902 (HBG1)),

HBG1 c.-225 to -222 (e.g., nucleotides 2716-2719 of SEQ ID NO:902 (HBG1)), and HBG2 c.-114 to -102 (e.g., nucleotides 2748-2760 of SEQ ID NO:903 (HBG2)).

Typically, the template sequence undergoes a breakage mediated or catalyzed recombination with the target sequence. In certain embodiments, 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 certain embodiments, 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 second site on the target sequence that is cleaved in a second Cas9 mediated event.

A template nucleic acid having homology with an HBG target position in a γ -globin gene regulatory region can be used to alter the structure of the regulatory region. For example, a template nucleic acid having homology with the region 5' and 3' of an HBG target position in a γ -globin gene regulatory region can be used to delete one or more nucleotides of an HBG target position.

A template nucleic acid typically comprises the following components:

[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. The homology arms are regions that are homologous to regions of DNA within or near (e.g., flanking or adjoining) a target nucleic acid to be cleaved. In certain embodiments, the homology arms flank the most distal cleavage sites.

In certain embodiments, a template nucleic acid may be used to remove (e.g., delete) a genomic sequence including at least a portion of the γ -globin gene regulatory region(s), e.g., enhancer region(s), e.g., silencer region(s), of the HGB1 and/or HGB2 gene(s). In certain embodiments, a template nucleic acid may be used to delete one or more nucleotides of an HBG target position, i.e., introduce an alteration (e.g., deletion) into an HBG target position. In certain embodiments, an alteration (e.g., deletion) may be introduced at a target site within the HBG target position. In certain embodiments, the alteration (e.g., deletion) may be selected from one or more of HBG1 13 bp del c.-114 to -102, HBG1 4 bp del c.-225 to -222, and HBG1 13 bp del c.-114 to -102. In certain embodiments, the target site may be selected from one or more of HBG1 c.-114 to -102 (e.g., nucleotides 2824-2836 of SEQ ID NO:902 (HBG1)), HBG1 c.-225 to -222 (e.g., nucleotides 2716-2719 of SEQ ID NO:902 (HBG1)), and HBG2 c.-114 to -102 (e.g., nucleotides 2748-2760 of SEQ ID NO:903 (HBG2)).

Replacement sequences in donor templates have been described elsewhere, including in Cotta-Ramusino 2016, which is incorporated by reference herein. A replacement sequence can be any suitable length. In certain embodiments, a replacement sequence may include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 or more sequence modifications relative to the naturally-occurring sequence within a cell in which editing is desired.

In certain embodiments, where the desired repair outcome is a deletion of the target nucleic acid, a replacement sequence may be 0 nucleotides or 0 bp. In certain embodiments, the template nucleic acid omits the sequence that is homologous to the target nucleic acid sequence to be deleted. If the replacement sequence is 0 nucleotides or 0 bp, then the sequence of the target nucleic acid that is positioned between where the 5' homology arm and 3' homology arm anneal to the template nucleic acid will be deleted.

In certain embodiments, the 3' end of the 5' homology arm is the position next to the 5' end of the replacement sequence. In certain embodiments, the 5' homology arm can extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, or 5000 nucleotides 5' from the 5' end of the replacement sequence. In certain embodiments, when the replacement sequence is 0 nucleotides or 0 bp, the 3' end of the 5' homology arm is the position next to the 5' end of the 3' homology arm. In certain embodiments where the replacement sequence is 0 nucleotides or 0 bp, the 5' homology arm can extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, or 5000 nucleotides 5' from the 5' end of the 3' homology arm.

In certain embodiments, 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, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, or 5000 nucleotides 3' from the 3' end of the replacement sequence. In certain embodiments where the replacement sequence is 0 nucleotides or 0 bp, the 5' end of the 3' homology arm is the position next to the 3' end of the 5' homology arm. In an embodiment, the 3' homology arm can extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, or 5000 nucleotides 3' from the 3' end of the 5' homology arm.

In certain embodiments, to alter one or more nucleotides at an HBG target position, the homology arms, e.g., the 5' and 3' homology arms, may each comprise about 1000 bp of sequence flanking the most distal gRNAs (e.g., 1000 bp of sequence on either side of the HBG target position).

It is contemplated herein that one or both homology arms may be shortened to avoid including certain sequence repeat elements, e.g., Alu repeats or 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.

It is contemplated herein that template nucleic acids for altering the sequence of an HBG target position may be designed for use as a single-stranded oligonucleotide, e.g., a single-stranded oligodeoxynucleotide (ssODN). When using a ssODN, 5' and 3' homology arms may range up to about 200 nucleotides 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 some embodiments, a longer homology arm is made by a method other than chemical synthesis, e.g., by denaturing a long double stranded nucleic acid and purifying one of the strands, e.g., by affinity for a strand-specific sequence anchored to a solid substrate.

While not wishing to be bound by theory, in certain embodiments alt-HDR proceeds more efficiently when the template nucleic acid has extended homology 5' to the nick (i.e., in the 5' direction of the nicked strand) or target site (i.e., in the 5' direction of the target site). Accordingly, in some embodiments, the template nucleic acid has a longer homology arm and a shorter homology arm, wherein the longer homology arm can anneal 5' of the nick or target site. In some embodiments, the arm that can anneal 5' to the nick or target site is at least 25, 50, 75, 100, 125, 150, 175, or 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, or 5000 nucleotides from the nick or target site or the 5' or 3' end of the replacement sequence. In some embodiments, the arm that can anneal 5' to the nick or target site is at least 10%, 20%, 30%, 40%, or 50% longer than the arm that can anneal 3' to the nick or target site. In some embodiments, the arm that can anneal 5' to the nick or target site is at least 2x, 3x, 4x, or 5x longer than the arm that can anneal 3' to the nick or target site. Depending on whether a ssDNA template can anneal to the intact strand or the nicked strand, the homology arm that anneals 5' to the nick or target site may be at the 5' end of the ssDNA template or the 3' end of the ssDNA template, respectively.

Similarly, in some embodiments, the template nucleic acid has a 5' homology arm, a replacement sequence, and a 3' homology arm, such that the template nucleic acid has extended homology to the 5' of the nick. For example, the 5' homology arm and 3' homology arm may be substantially the same length, but the replacement sequence may

extend farther 5' of the nick than 3' of the nick. In some embodiments, the replacement sequence extends at least 10%, 20%, 30%, 40%, 50%, 2x, 3x, 4x, or 5x further to the 5' end of the nick than the 3' end of the nick.

While not wishing to be bound by theory, in some embodiments alt-HDR proceeds more efficiently when the template nucleic acid is centered on the nick or target site. Accordingly, in some embodiments, the template nucleic acid has two homology arms that are essentially the same size. For instance, the first homology arm of a template nucleic acid may have a length that is within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% of the second homology arm of the template nucleic acid.

Similarly, in some embodiments, the template nucleic acid has a 5' homology arm, a replacement sequence, and a 3' homology arm, such that the template nucleic acid extends substantially the same distance on either side of the nick or target site. For example, the homology arms may have different lengths, but the replacement sequence may be selected to compensate for this. For example, the replacement sequence may extend further 5' from the nick than it does 3' of the nick, but the homology arm 5' of the nick is shorter than the homology arm 3' of the nick, to compensate. The converse is also possible, e.g., that the replacement sequence may extend further 3' from the nick than it does 5' of the nick, but the homology arm 3' of the nick is shorter than the homology arm 5' of the nick, to compensate.

Exemplary template nucleic acids

In certain embodiments, the template nucleic acid is double stranded. In other embodiments, the template nucleic acid is single stranded. In certain embodiments, the template nucleic acid comprises a single stranded portion and a double stranded portion. In certain embodiments, the template nucleic acid comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 65 to 85, or 70 to 80 bp, homology on either side of the nick, target site, and/or replacement sequence. In certain embodiments, the template nucleic acid comprises about 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 bp homology 5' of the nick, target site, or replacement sequence, 3' of the nick, target site, or replacement sequence, or both 5' and 3' of the nick, target site, or replacement sequences.

In certain embodiments, the template nucleic acid comprises about 150 to 200 bp, e.g., 155 to 195, 160 to 190, 165 to 185, or 170 to 180 bp, homology 3' of the nick, target site, and/or replacement sequence. In certain embodiments, the template nucleic acid comprises about 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, or 200 bp homology 3' of the nick, target site, or replacement sequence. In certain embodiments, the template nucleic

acid comprises less than about 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, or 10 bp homology 5' of the nick, target site, or replacement sequence.

In certain embodiments, the template nucleic acid comprises about 150 to 200 bp, e.g., 155 to 195, 160 to 190, 165 to 185, or 170 to 180 bp, homology 5' of the nick, target site, and/or replacement sequence. In certain embodiments, the template nucleic acid comprises about 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, or 200 bp homology 5' of the nick, target site, or replacement sequence. In certain embodiments, the template nucleic acid comprises less than about 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, or 10 bp homology 3' of the nick, target site, or replacement sequence.

In certain 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 that may be used to delete one or more nucleotides of a HBG target position.

The template nucleic acid may comprise a replacement sequence. In some embodiments, the template nucleic acid comprises a 5' homology arm. In other embodiments, the template nucleic acid comprises a 3' homology arm.

The template nucleic acid may comprise a 5' homology arm, replacement sequence that is 0 nucleotides or 0 bp, and a 3' homology arm.

In certain embodiments, the template nucleic acid is linear double stranded DNA. The length may be, e.g., about 150-200 bp, e.g., about 150, 160, 170, 180, 190, or 200 bp. The length may be, e.g., at least 150, 160, 170, 180, 190, or 200 bp. In some embodiments, the length is no greater than 150, 160, 170, 180, 190, or 200 bp. In some embodiments, a double stranded template nucleic acid has a length of about 160 bp, e.g., about 155-165, 150-170, 140-180, 130-190, 120-200, 110-210, 100-220, 90-230, or 80-240 bp.

The template nucleic acid can be linear single stranded DNA. In certain embodiments, the template nucleic acid is (i) linear single stranded DNA that can anneal to the nicked strand of the target nucleic acid, (ii) linear single stranded DNA that can anneal to the intact strand of the target nucleic acid, (iii) linear single stranded DNA that can anneal to the plus strand of the target nucleic acid, (iv) linear single stranded DNA that can anneal to the minus strand of the target nucleic acid, or more than one of the preceding. The length may be, e.g., about 150-200 nucleotides, e.g., about 150, 160, 170, 180, 190, or 200 nucleotides. The length may be, e.g., at least 150, 160, 170, 180, 190, or 200 nucleotides. In

some embodiments, the length is no greater than 150, 160, 170, 180, 190, or 200 nucleotides. In some embodiments, a single stranded template nucleic acid has a length of about 160 nucleotides, e.g., about 155-165, 150-170, 140-180, 130-190, 120-200, 110-210, 100-220, 90-230, or 80-240 nucleotides.

In some embodiments, the template nucleic acid is circular double stranded DNA, e.g., a plasmid. In some embodiments, the template nucleic acid comprises about 500 to 1000 bp of homology on either side of the replacement sequence, target site, and/or the nick. In some embodiments, the template nucleic acid comprises about 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 bp of homology 5' of the nick, target site, or replacement sequence, 3' of the nick, target site, or replacement sequence, or both 5' and 3' of the nick, target site, or replacement sequence. In some embodiments, the template nucleic acid comprises at least 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 bp of homology 5' of the nick, target site, or replacement sequence, 3' of the nick, target site, or replacement sequence, or both 5' and 3' of the nick, target site, or replacement sequence. In some embodiments, the template nucleic acid comprises no more than 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 bp of homology 5' of the nick, target site, or replacement sequence, 3' of the nick, target site, or replacement sequence, or both 5' and 3' of the nick, target site, or replacement sequence.

In certain embodiments, 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, while 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.

In some embodiments, the template nucleic acid is an adenovirus vector, e.g., an AAV vector, e.g., a ssDNA molecule of a length and sequence that allows it to be packaged in an AAV capsid. The vector may be, e.g., less than 5 kb and may contain an ITR sequence that promotes packaging into the capsid. The vector may be integration-deficient. In some embodiments, the template nucleic acid comprises about 150 to 1000 nucleotides of homology on either side of the replacement sequence, target site, and/or the nick. In some embodiments, the template nucleic acid comprises about 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' of the nick, target site, or replacement sequence, 3' of the nick, target site, or replacement sequence, or both 5' and 3' of the nick, target site, or replacement sequence. In some embodiments, the template nucleic acid

comprises at least 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' of the nick, target site, or replacement sequence, 3' of the nick, target site, or replacement sequence, or both 5' and 3' of the nick, target site, or replacement sequence. In some embodiments, the template nucleic acid comprises at most 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' of the nick, target site, or replacement sequence, 3' of the nick, target site, or replacement sequence, or both 5' and 3' of the nick, target site, or replacement sequence.

In some embodiments, the template nucleic acid is a lentiviral vector, e.g., an IDLV (integration deficiency lentivirus). In some embodiments, the template nucleic acid comprises about 500 to 1000 bp of homology on either side of the replacement sequence, target site, and/or the nick. In some embodiments, the template nucleic acid comprises about 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 bp of homology 5' of the nick, target site, or replacement sequence, 3' of the nick, target site, or replacement sequence, or both 5' and 3' of the nick, target site, or replacement sequence. In some embodiments, the template nucleic acid comprises at least 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 bp of homology 5' of the nick, target site, or replacement sequence, 3' of the nick, target site, or replacement sequence, or both 5' and 3' of the nick or replacement sequence. In some embodiments, the template nucleic acid comprises no more than 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 bp of homology 5' of the nick, target site, or replacement sequence, 3' of the nick, target site, or replacement sequence, or both 5' and 3' of the nick, target site, or replacement sequence.

In an embodiment, the template nucleic acid comprises one or more mutations, e.g., silent mutations, that prevent Cas9 from recognizing and cleaving the template nucleic acid. The template nucleic acid may comprise, e.g., at least 1, 2, 3, 4, 5, 10, 20, or 30 silent mutations relative to the corresponding sequence in the genome of the cell to be altered. In certain embodiments, the template nucleic acid comprises at most 2, 3, 4, 5, 10, 20, 30, or 50 silent mutations relative to the corresponding sequence in the genome of the cell to be altered. In an embodiment, the cDNA comprises one or more mutations, e.g., silent mutations that prevent Cas9 from recognizing and cleaving the template nucleic acid. The template nucleic acid may comprise, e.g., at least 1, 2, 3, 4, 5, 10, 20, or 30 silent mutations relative to the corresponding sequence in the genome of the cell to be altered. In certain embodiments, the template nucleic acid comprises at most 2, 3, 4, 5, 10, 20, 30, or 50 silent mutations relative to the corresponding sequence in the genome of the cell to be altered.

In certain embodiments of the methods provided herein, HDR-mediated alteration is used to introduce an alteration (e.g., deletion) of one or more nucleotides in a γ -globin gene regulatory region. In certain embodiments, the γ -globin gene regulatory region may be a HBG target position. In certain embodiments, an alteration (e.g., deletion) may be introduced at a target site within the HBG target position. In certain embodiments, the alteration (e.g., deletion) may be selected from one or more of HBG1 13 bp del c.-114 to -102, HBG1 4 bp del c.-225 to -222, and HBG1 13 bp del c.-114 to -102. In certain embodiments, the target site may be selected from one or more of HBG1 c.-114 to -102 (e.g., nucleotides 2824-2836 of SEQ ID NO:902 (HBG1)), HBG1 c.-225 to -222 (e.g., nucleotides 2716-2719 of SEQ ID NO:902 (HBG1)), and HBG2 c.-114 to -102 (e.g., nucleotides 2748-2760 of SEQ ID NO:903 (HBG2)).

In certain embodiments, a template nucleic acid for introducing an alteration (e.g., deletion) at a target site within an HBG target position (i.e., an HBG1 or HBG2 regulatory region) comprises, from the 5' to 3' direction, a 5' homology arm, a replacement sequence, and a 3' homology arm, wherein the replacement sequence is 0 nucleotides or 0 bp. In certain embodiments, the template nucleic acid may be a single stranded oligodeoxynucleotide (ssODN). In certain embodiments, the 5' homology arm may be any of the 5' homology arms described herein. In certain embodiments, the 3' homology arms may be any of the 3' homology arms described herein. In certain embodiments, an alteration (e.g., deletion) may be introduced at a target site within the HBG target position. In certain embodiments, the alteration (e.g., deletion) may be selected from one or more of HBG1 13 bp del c.-114 to -102, HBG1 4 bp del c.-225 to -222, and HBG1 13 bp del c.-114 to -102. In certain embodiments, the target site may be selected from one or more of HBG1 c.-114 to -102 (e.g., nucleotides 2824-2836 of SEQ ID NO:902 (HBG1)), HBG1 c.-225 to -222 (e.g., nucleotides 2716-2719 of SEQ ID NO:902 (HBG1)), and HBG2 c.-114 to -102 (e.g., nucleotides 2748-2760 of SEQ ID NO:903 (HBG2)).

For example, a template nucleic acid for introducing the alteration HBG1 13 bp del c.-114 to -102 at the target site HBG1 c.-114 to -102 (e.g., nucleotides 2824-2836 of SEQ ID NO:902 (HBG1)) may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm, where the replacement sequence is 0 nucleotides or 0 bp. In certain embodiments, the 5' homology arm comprises about 200 nucleotides in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the target site HBG1 c.-114 to -102 (e.g., nucleotides 2824-2836 of SEQ ID

NO:902 (HBG1)). In certain embodiments, the 5' homology arm comprises, consists essentially of, or consists of SEQ ID NO:904 (ssODN1 5' homology arm). In certain embodiments, the 5' homology arm comprises, consists essentially of, or consists of SEQ ID NO:907 (PhTx ssODN1 5' homology arm). In certain embodiments, the 3' homology arm comprises about 200 nucleotides in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 3' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the target site HBG1 c.-114 to -102 (e.g., nucleotides 2824-2836 of SEQ ID NO:902 (HBG1)). In certain embodiments, the 3' homology arm comprises, consists essentially of, or consists of SEQ ID NO:905 (ssODN1 3' homology arm). In certain embodiments, the 3' homology arm comprises, consists essentially of, or consists of SEQ ID NO:908 (PhTx ssODN1 3' homology arm). In certain embodiments, the template nucleic acid comprises, consists essentially of, or consists of SEQ ID NO:906. In certain embodiments, the template nucleic acid comprises, consists essentially of, or consists of SEQ ID NO:909 (PhTx ssODN1).

In another example, a template nucleic acid for introducing the alteration HBG2 13 bp del c.-114 to -102 at the target site HBG2 c.-114 to -102 (e.g., nucleotides 2748-2760 of SEQ ID NO:903 (HBG2)) may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm, where the replacement sequence is 0 nucleotides or 0 bp. In certain embodiments, the 5' homology arm comprises about 200 nucleotides in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the target site HBG2 c.-114 to -102 (e.g., nucleotides 2748-2760 of SEQ ID NO:903 (HBG2)). In certain embodiments, the 5' homology arm comprises, consists essentially of, or consists of SEQ ID NO:904 (ssODN1 5' homology arm). In certain embodiments, the 5' homology arm comprises, consists essentially of, or consists of SEQ ID NO:907 (PhTx ssODN1 5' homology arm). In certain embodiments, the 3' homology arm comprises about 200 nucleotides in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 3' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the target site HBG2 c.-114 to -102 (e.g., nucleotides 2748-2760 of SEQ ID NO:903 (HBG2)). In certain embodiments, the 3' homology arm comprises, consists essentially of, or consists of SEQ ID NO:905 (ssODN1 3' homology arm). In certain embodiments, the 3' homology arm comprises, consists essentially of, or consists of SEQ ID NO:908 (PhTx ssODN1 3' homology arm). In certain embodiments, the template nucleic acid comprises, consists essentially of, or

consists of SEQ ID NO:906. In certain embodiments, the template nucleic acid comprises, consists essentially of, or consists of SEQ ID NO:909 (PhTx ssODN1).

In another example, a template nucleic acid for introducing the alteration HBG1 4 bp del c.-225 to -222 at the target site HBG1 c.-225 to -222 (e.g., nucleotides 2716-2719 of SEQ ID NO:902 (HBG1)) may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm, where the replacement sequence is 0 nucleotides or 0 bp. In certain embodiments, the 5' homology arm comprises about 200 nucleotides in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the target site HBG1 c.-225 to -222 (e.g., nucleotides 2716-2719 of SEQ ID NO:902 (HBG1)). In certain embodiments, the 3' homology arm comprises about 200 nucleotides in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 3' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the target site HBG1 c.-225 to -222 (e.g., nucleotides 2716-2719 of SEQ ID NO:902 (HBG1)).

In certain embodiments, the 5' homology arm comprises a 5' phosphorothionate (PhTx) modification. In certain embodiments, the 3' homology arm comprises a 3' PhTx modification. In certain embodiments, the template nucleic acid comprises a 5' and 3' PhTx modification.

In certain embodiments, a template nucleic acid for altering a single nucleotide in a γ -globin gene (e.g., HBG1, HBG2) regulatory region comprises, from the 5' to 3' direction, a 5' homology arm, a replacement sequence, and a 3' homology arm, wherein the replacement is designed to incorporate the single nucleotide alteration. For example, where the alteration being incorporated is HBG1 c.-114 C>T; c.-158 C>T; c.-167 C>T; c.-196 C>T; or c.-201 C>T or HBG2 c.-109 G>T; c.-114 C>T; c.-157 C>T; c.-158 C>T; c.-167 C>T; c.-211 C>T, the replacement sequence may comprise the single nucleotide T and, optionally, one more nucleotides on one or both sides of this T. Similarly, where the alteration being incorporated is HBG1 c.-117 G>A; c.-170 G>A; or c.-499 T>A or HBG2 c.-114 C>A or c.-167 C>A, the replacement sequence may comprise the single nucleotide A and, optionally, one more nucleotides on one or both sides of this A; where the alteration being incorporated is HBG1 c.-175 T>G or c.-195 C>G or HBG2 c.-202 C>G; c.-255 C>G; c.-309 A>G; c.-369 C>G; or c.-567 T>G, the replacement sequence may comprise the single nucleotide G and, optionally, one more nucleotides on one or both sides of this G; and where the alteration being incorporated is HBG1 c.-175 T>C; c.-198 T>C; or c.-251 T>C or HBG2 c.-175 T>C or c.-

228 T>C, the replacement sequence may comprise the single nucleotide C and, optionally, one more nucleotides on one or both sides of this C.

In certain embodiments, the 5' and 3' homology arms each comprise a length of sequence flanking the nucleotides corresponding to the replacement sequence. In certain embodiments, a template nucleic acid comprises a replacement sequence flanked by a 5' homology arm and a 3' homology arm each independently comprising 10 or more, 20 or more, 50 or more, 100 or more, 150 or more, 200 or more, 250 or more, 300 or more, 350 or more, 400 or more, 450 or more, 500 or more, 550 or more, 600 or more, 650 or more, 700 or more, 750 or more, 800 or more, 850 or more, 900 or more, 1000 or more, 1100 or more, 1200 or more, 1300 or more, 1400 or more, 1500 or more, 1600 or more, 1700 or more, 1800 or more, 1900 or more, or 2000 or more nucleotides. In certain embodiments, a template nucleic acid comprises a replacement sequence flanked by a 5' homology arm and a 3' homology arm each independently comprising at least 50, 100, or 150 nucleotides, but not long enough to include a repeated element. In certain embodiments, a template nucleic acid comprises a replacement sequence flanked by a 5' homology arm and a 3' homology arm each independently comprising 5 to 100, 10 to 150, or 20 to 150 nucleotides. In certain embodiments, the replacement sequence optionally comprises a promoter and/or polyA signal.

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, 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 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 a target nucleic acid sequence. Instead, the complementary repeat sequence is utilized.

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 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.

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 b, DNA polymerase d, DNA polymerase e, PCNA, LIG1, PNK, PNKP, APE1, APTX, APLF, TDP1, LIG3, FEN1, CtIP, MRN, and ERCC1.

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 sub-sequentially filled in by DNA polymerase and finally sealed with ligase. A more complete picture is given in Li 2008, 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 d, RPA, HMGB1, RFC, and DNA ligase I.

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 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 desired nucleotide (e.g., HPFH mutant) 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 Marteiijn 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 β and Rev1.

Error-free postreplication repair (PRR) is another pathway for repairing a single stranded break left after a defective replication event.

Examples of gRNAs in genome editing methods

gRNA molecules as described herein can be used with Cas9 molecules that generate a double strand break or a single strand break to alter the sequence of a target nucleic acid, e.g.,

a target position or target genetic signature. gRNA molecules useful in these methods are described below.

In certain embodiments, 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, 200, 250, 300, 350, 400, 450, or 500 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 16 nucleotides, e.g., a targeting domain of (i) 16, (ii) 17, (iii) 18, (iv) 19, (v) 20, (vi) 21, (vii) 22, (viii) 23, (ix) 24, (x) 25, or (xi) 26 nucleotides; and

(c)(i) 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, e.g., at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides from a naturally occurring *S. pyogenes* or *S. aureus* tail and proximal domain, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(ii) 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, e.g., at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides from the corresponding sequence of a naturally occurring *S. pyogenes* or *S. aureus* gRNA, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides therefrom;

(c)(iii) 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, e.g., at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides from the corresponding sequence of a naturally occurring *S. pyogenes* or *S. aureus* gRNA, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides therefrom;

(c)(iv) the tail domain is at least 10, 15, 20, 25, 30, 35 or 40 nucleotides in length, e.g., it comprises at least 10, 15, 20, 25, 30, 35 or 40 nucleotides from a naturally occurring *S. pyogenes* or *S. aureus* tail domain, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom; or

(c)(v) 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. aureus* tail domain.

In certain embodiments, the gRNA is configured such that it comprises properties a and b(i); a and b(ii); a and b(iii); a and b(iv); a and b(v); a and b(vi); a and b(vii); a and b(viii); a and b(ix); a and b(x); a and b(xi); a and c; a, b, and c; a(i), b(i), and c(i); a(i), b(i), and c(ii); a(i), b(ii), and c(i); a(i), b(ii), and c(ii); a(i), b(iii), and c(i); a(i), b(iii), and c(ii); a(i), b(iv), and c(i); a(i), b(iv), and c(ii); a(i), b(v), and c(i); a(i), b(v), and c(ii); a(i), b(vi), and c(i); a(i), b(vi), and c(ii); a(i), b(vii), and c(i); a(i), b(vii), and c(ii); a(i), b(viii), and c(i); a(i), b(viii), and c(ii); a(i), b(ix), and c(i); a(i), b(ix), and c(ii); a(i), b(x), and c(i); a(i), b(x), and c(ii); a(i), b(xi), or c(i); a(i), b(xi), and c(ii).

In certain embodiments, the gRNA, e.g., a chimeric gRNA, is configured such that it 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, 200, 250, 300, 350, 400, 450, or 500 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 16 nucleotides, e.g., a targeting domain of (i) 16, (ii) 17, (iii) 18, (iv) 19, (v) 20, (vi) 21, (vii) 22, (viii) 23, (ix) 24, (x) 25, or (xi) 26 nucleotides; and

(c)(i) 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, e.g., at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides from a naturally occurring *S. pyogenes* or *S. aureus* tail and proximal domain, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides therefrom;

(c)(ii) 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, e.g., at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides from the corresponding sequence of a naturally occurring *S. pyogenes*, or *S. aureus* gRNA, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides therefrom;

(c)(iii) 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, e.g., at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides from the corresponding sequence of a naturally occurring *S. pyogenes* or *S. aureus* gRNA, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(iv) the tail domain is at least 10, 15, 20, 25, 30, 35 or 40 nucleotides in length, e.g., it comprises at least 10, 15, 20, 25, 30, 35 or 40 nucleotides from a naturally occurring *S. pyogenes* or *S. aureus* tail domain, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides therefrom; or

(c)(v) the tail domain comprises 15, 20, 25, 30, 35, or 40 nucleotides or all of the corresponding portions of a naturally occurring tail domain, e.g., a naturally occurring *S. pyogenes* or *S. aureus* tail domain.

In certain embodiments, the gRNA is configured such that it comprises properties a and b(i); a and b(ii); a and b(iii); a and b(iv); a and b(v); a and b(vi); a and b(vii); a and b(viii); a and b(ix); a and b(x); a and b(xi); a and c; a, b, and c; a(i), b(i), and c(i); a(i), b(i), and c(ii); a(i), b(ii), and c(i); a(i), b(ii), and c(ii); a(i), b(iii), and c(i); a(i), b(iii), and c(ii); a(i), b(iv), and c(i); a(i), b(iv), and c(ii); a(i), b(v), and c(i); a(i), b(v), and c(ii); a(i), b(vi), and c(i); a(i), b(vi), and c(ii); a(i), b(vii), and c(i); a(i), b(vii), and c(ii); a(i), b(viii), and c(i); a(i), b(viii), and c(ii); a(i), b(ix), and c(i); a(i), b(ix), and c(ii); a(i), b(x), and c(i); a(i), b(x), and c(ii); a(i), b(xi), and c(i); a(i), b(xi), and c(ii).

In certain embodiments, 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 a mutation at 840, e.g., the 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 N863, e.g., the N863A mutation.

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, 200, 250, 300, 350, 400, 450, or 500 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 16 nucleotides, e.g., a targeting domain of (i) 16, (ii) 17, (iii) 18, (iv) 19, (v) 20, (vi) 21, (vii) 22, (viii) 23, (ix) 24, (x) 25, or (xi) 26 nucleotides;

(c)(i) 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, e.g., at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides from a naturally occurring *S. pyogenes* or *S. aureus* tail and proximal domain, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(ii) 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, e.g., at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides from the corresponding sequence of a naturally occurring *S. pyogenes* or *S. aureus* gRNA, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(iii) 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, e.g., at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides from the corresponding sequence of a naturally occurring *S. pyogenes* or *S. aureus* gRNA, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides therefrom;

(c)(iv) the tail domain is at least 10, 15, 20, 25, 30, 35 or 40 nucleotides in length, e.g., it comprises at least 10, 15, 20, 25, 30, 35 or 40 nucleotides from a naturally occurring *S. pyogenes* or *S. aureus* tail domain; or, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom; or

(c)(v) the tail domain comprises 15, 20, 25, 30, 35, or 40 nucleotides or all of the corresponding portions of a naturally occurring tail domain, e.g., a naturally occurring *S. pyogenes* or *S. aureus* 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 certain embodiments, one or both of the gRNAs is configured such that it comprises properties a and b(i); a and b(ii); a and b(iii); a and b(iv); a and b(v); a and b(vi); a and b(vii); a and b(viii); a and b(ix); a and b(x); a and b(xi); a and c; a, b, and c; a(i), b(i), and c(i); a(i), b(i), and c(ii); a(i), b(i), c, and d; a(i), b(i), c, and e; a(i), b(i), c, d, and e; a(i), b(ii), and c(i); a(i), b(ii), and c(ii); a(i), b(ii), c, and d; a(i), b(ii), c, and e; a(i), b(ii), c, d, and e; a(i), b(iii), and c(i); a(i), b(iii), and c(ii); a(i), b(iii), c, and d; a(i), b(iii), c, and e; a(i), b(iii), c, d,

and e; a(i), b(iv), and c(i); a(i), b(iv), and c(ii); a(i), b(iv), c, and d; a(i), b(iv), c, and e; a(i), b(iv), c, d, and e; a(i), b(v), and c(i); a(i), b(v), and c(ii); a(i), b(v), c, and d; a(i), b(v), c, and e; a(i), b(v), c, d, and e; a(i), b(vi), and c(i); a(i), b(vi), and c(ii); a(i), b(vi), c, and d; a(i), b(vi), c, and e; a(i), b(vi), c, d, and e; a(i), b(vii), and c(i); a(i), b(vii), and c(ii); a(i), b(vii), c, and d; a(i), b(vii), c, and e; a(i), b(vii), c, d, and e; a(i), b(viii), and c(i); a(i), b(viii), and c(ii); a(i), b(viii), c, and d; a(i), b(viii), c, and e; a(i), b(viii), c, d, and e; a(i), b(ix), and c(i); a(i), b(ix), and c(ii); a(i), b(ix), c, and d; a(i), b(ix), c, and e; a(i), b(ix), c, d, and e; a(i), b(x), and c(i); a(i), b(x), and c(ii); a(i), b(x), c, and d; a(i), b(x), c, and e; a(i), b(x), c, d, and e; a(i), b(xi), and c(i); a(i), b(xi), and c(ii); a(i), b(xi), c, and d; a(i), b(xi), c, and e; a(i), b(xi), c, d, and e.

In certain embodiments, 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.

In certain embodiments, 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., the H840A mutation.

In certain embodiments, 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 N863, e.g., the N863A mutation.

Target cells

Cas9 molecules and gRNA molecules, e.g., a Cas9 molecule/gRNA molecule complex, can be used to alter (e.g., introduce a mutation or deletion in) a target nucleic acid, e.g., a γ -globin gene (e.g., HBG1, HBG2) regulatory region, in a wide variety of cells. In certain embodiments, alteration of a target nucleic acid in a target cell may be performed *in vitro*, *ex vivo* or *in vivo*.

The Cas9 and gRNA molecules described herein can be delivered to a target cell. In certain embodiments, the target cell is an erythroid cell, e.g., an erythroblast. In certain embodiments, erythroid cells are preferentially targeted, e.g., at least about 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the targeted cells are erythroid cells. For example, in the case of *in vivo* delivery, erythroid cells are preferentially targeted, and if cells are treated *ex vivo* and returned to the subject, erythroid cells are preferentially modified.

In certain embodiments, the target cell is a circulating blood cell, e.g., a reticulocyte, megakaryocyte erythroid progenitor (MEP) cell, myeloid progenitor cell (CMP/GMP),

lymphoid progenitor (LP) cell, hematopoietic stem/progenitor cell (HSC), or endothelial cell (EC). In certain embodiments, the target cell is a bone marrow cell (e.g., a reticulocyte, an erythroid cell (e.g., erythroblast), an MEP cell, myeloid progenitor cell (CMP/GMP), LP cell, erythroid progenitor (EP) cell, HSC, multipotent progenitor (MPP) cell, endothelial cell (EC), hemogenic endothelial (HE) cell, or mesenchymal stem cell). In certain embodiments, the target cell is a myeloid progenitor cell (e.g., a common myeloid progenitor (CMP) cell or granulocyte macrophage progenitor (GMP) cell). In certain embodiments, the target cell is a lymphoid progenitor cell, e.g., a common lymphoid progenitor (CLP) cell. In certain embodiments, the target cell is an erythroid progenitor cell (e.g., an MEP cell). In certain embodiments, the target cell is a hematopoietic stem/progenitor cell (e.g., a long term HSC (LT-HSC), short term HSC (ST-HSC), MPP cell, or lineage restricted progenitor (LRP) cell). In certain embodiments, the target cell is a $CD34^+$ cell, $CD34^+CD90^+$ cell, $CD34^+CD38^-$ cell, $CD34^+CD90^+CD49f^+CD38^-CD45RA^-$ cell, $CD105^+$ cell, $CD31^+$, or $CD133^+$ cell, or a $CD34^+CD90^+CD133^+$ cell. In certain embodiments, the target cell is an umbilical cord blood $CD34^+$ HSPC, umbilical cord venous endothelial cell, umbilical cord arterial endothelial cell, amniotic fluid $CD34^+$ cell, amniotic fluid endothelial cell, placental endothelial cell, or placental hematopoietic $CD34^+$ cell. In certain embodiments, the target cell is a mobilized peripheral blood hematopoietic $CD34^+$ cell (after the patient is treated with a mobilization agent, e.g., G-CSF or Plerixafor). In certain embodiments, the target cell is a peripheral blood endothelial cell.

In certain embodiments, a target cell is manipulated *ex vivo* by editing a γ -globin gene regulatory region, then the target cell is administered to the subject. Sources of target cells for *ex vivo* manipulation may include, for example, the subject's blood, bone marrow, or cord blood. Other sources of target cells for *ex vivo* manipulation may include, for example, heterologous donor blood, cord blood, or bone marrow. In certain embodiments, an erythrocyte is removed from the subject, manipulated *ex vivo* as described above, and returned to the subject. In certain embodiments, a hematopoietic stem cell is removed from the subject, manipulated *ex vivo* as described above, and returned to the subject. In certain embodiments, an erythroid progenitor cell is removed from a subject, manipulated *ex vivo* as described above, and returned to the subject. In certain embodiments, a myeloid progenitor cell is removed from a subject, manipulated *ex vivo* as described above, and returned to the subject. In certain embodiments, a multipotent progenitor (MPP) cell is removed from a subject, manipulated *ex vivo* as described above, and returned to the subject. In certain embodiments, a hematopoietic stem/progenitor cell (HSC) is removed from a subject,

manipulated *ex vivo* as described above, and returned to the subject. In certain embodiments, a CD34⁺ HSC is removed from a subject, manipulated *ex vivo* as described above, and returned to the subject.

In certain embodiments, modified HSCs generated *ex vivo* are administered to a subject without myeloablative pre-conditioning. In other embodiments, the modified HSCs are administered after mild myeloblastic conditioning such that, following engraftment, some of the hematopoietic cells are derived from the modified HSCs. In still other embodiments, the modified HSCs are administered after full myeloblastic conditioning such that, following engraftment, 100% of the hematopoietic cells are derived from the modified HSCs.

A suitable cell can also include a stem cell such as, e.g., an embryonic stem cell, induced pluripotent stem cell, hematopoietic stem cell, or hemogenic endothelial (HE) cell (precursor to both hematopoietic stem cells and endothelial cells). In certain embodiments, the cell is an induced pluripotent stem (iPS) cell or a cell derived from an iPS cell, e.g., an iPS cell generated from the subject, modified using the methods disclosed herein, and differentiated into a clinically relevant cell such as e.g., an erythrocyte. In certain embodiments, AAV is used to transduce the target cells.

In certain embodiments, stem cells used for gene editing as described herein may be prepared for use in accordance with the methods described in the examples in Gori 2016 at, e.g., pages 219-223, 223-224, 227-231, 231-236, 235-238, 240-241, 242-244, which is incorporated herein by reference. Stem cells can be cultured and expanded in any manner that is suitable and known by the skilled artisan.

Cells produced by the methods described herein may be used immediately. Alternatively, the cells may be frozen (e.g., in liquid nitrogen) and stored for later use. The cells will usually be frozen in 10% dimethylsulfoxide (DMSO), 50% serum, 40% buffered medium, or some other such solution as is commonly used in the art to preserve cells at such freezing temperature, and thawed in such a manner as commonly known in the art for thawing frozen cultured cells. Cells may also be thermostabilized for prolonged storage at 4°C.

Delivery, formulations, and routes of administration

Genome editing system components, e.g., RNA-guided nuclease molecules, e.g., a Cas9 molecule, gRNA molecule (e.g., a Cas9 molecule/gRNA molecule complex), and a donor template nucleic acid, or all three, can be delivered, formulated, or administered in a variety of forms, see, e.g., **Tables 3 and 4**.

In certain embodiments, one Cas9 molecule and two or more (e.g., 2, 3, 4, or more) different gRNA molecules are delivered, e.g., by an AAV vector. In certain embodiments, the sequence encoding the Cas9 molecule and the sequence(s) encoding the two or more (e.g., 2, 3, 4, or more) different gRNA molecules are present on the same nucleic acid molecule, e.g., an AAV vector. 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, SFFV, EFS, EF-1a, PGK, CAG, and CBH promoters or a blood cell specific promoter. In an embodiment, the promoter is a constitutive promoter. In another embodiment, the promoter is a tissue specific promoter. Useful promoters for gRNAs include T7.H1, EF-1a, U6, U1, and tRNA 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, the sequence encoding a Cas9 molecule comprises at least two nuclear localization signals. In an embodiment, a promoter for a Cas9 molecule or a gRNA molecule can be, independently, inducible, tissue specific, or cell specific.

Table 3 provides examples of how the components can be formulated, delivered, or administered.

Table 3

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		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 <i>in vitro</i> transcribed or synthesized RNA. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	RNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is transcribed from DNA, and a gRNA is provided as <i>in vitro</i> 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 <i>in vitro</i> transcribed mRNA, and a gRNA is provided as <i>in vitro</i> 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 <i>in vitro</i> 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 <i>in vitro</i> 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 4 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 4

Delivery Vector/Mode		Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Physical (e.g., electroporation, particle gun, Calcium Phosphate transfection, cell compression or squeezing)		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</i>	Attenuated Bacteria	YES	Transient	NO	Nucleic Acids

<i>Delivery Vehicles</i>	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 an RNA-guided nuclease and or one or more gRNA molecules

Nucleic acids encoding RNA-guided nucleases, e.g., Cas9 molecules (e.g., eaCas9 molecules), gRNA molecules, a donor template nucleic acid, or any combination (e.g., two or all) thereof 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, as well as donor template nucleic acids can be delivered by, e.g., vectors (e.g., viral or non-viral vectors), non-vector based methods (e.g., using naked DNA or DNA complexes), or a combination thereof.

Nucleic acids 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., erythrocytes, HSCs). Donor template molecules can likewise be conjugated to molecules (e.g., N-acetylgalactosamine) promoting uptake by the target cells (e.g., erythrocytes, HSCs).

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a vector (e.g., viral vector/virus or plasmid).

Vectors can comprise a sequence that encodes a Cas9 molecule and/or a gRNA molecule and/or a donor template with high homology to the region (e.g., target sequence) being targeted. In certain embodiments, the donor template comprises all or part of a target sequence. Exemplary donor templates are a repair template, e.g., a gene correction template, or a gene mutation template, e.g., point mutation (e.g., single nucleotide (nt) substitution) template). A vector can also comprise a sequence encoding a signal peptide (e.g., for nuclear localization, nucleolar localization, or 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.

One or more regulatory/control elements, e.g., promoters, enhancers, introns, polyadenylation signals, Kozak consensus sequences, or 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 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 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.

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 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.

In an embodiment, the viral vector recognizes a specific cell type or tissue. 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(s) of one or more viral envelope glycoproteins to incorporate a targeting ligand such as a peptide ligand, a single chain antibody, or 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., a ligand-receptor, monoclonal antibody, avidin-biotin and chemical conjugation).

In some embodiments, the Cas9- and/or gRNA-encoding nucleic acid sequence 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 nucleic acid sequence is delivered by a recombinant lentivirus. In an embodiment, the donor template nucleic acid is delivered by a recombinant retrovirus. For example, the lentivirus is replication-defective, e.g., does not comprise one or more genes required for viral replication.

In an embodiment, the Cas9- and/or gRNA-encoding nucleic acid sequence is delivered by a recombinant lentivirus. In an embodiment, the donor template nucleic acid 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 nucleic acid sequence is delivered by a recombinant adenovirus. In an embodiment, the donor template nucleic acid 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 nucleic acid sequence is delivered by a recombinant AAV. In an embodiment, the donor template nucleic acid is delivered by a recombinant AAV. In some embodiments, the AAV does not incorporate its genome into that of a host cell, e.g., a target cell as describe herein. 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 an embodiment, an AAV capsid that can be used in the methods described herein is a capsid sequence from serotype AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV.rh8, AAV.rh10, AAV.rh32/33, AAV.rh43, AAV.rh64R1, or AAV7m8.

In an embodiment, the Cas9- and/or gRNA-encoding DNA is delivered in a re-engineered AAV capsid, e.g., with 50% or greater, e.g., 60% or greater, 70% or greater, 80% or greater, 90% or greater, or 95% or greater, sequence homology with a capsid sequence from serotypes AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV.rh8, AAV.rh10, AAV.rh32/33, AAV.rh43, or AAV.rh64R1.

In an embodiment, the Cas9- and/or gRNA-encoding DNA is delivered by a chimeric AAV capsid. In an embodiment, the donor template nucleic acid is delivered by a chimeric AAV capsid. Exemplary chimeric AAV capsids include, but are not limited to, AAV9i1, AAV2i8, AAV-DJ, AAV2G9, AAV2i8G9, or AAV8G9.

In an embodiment, 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. In an embodiment, the hybrid virus is hybrid of an AAV (e.g., of any AAV serotype), with a Bocavirus, B19 virus, porcine AAV, goose AAV, feline AAV, canine AAV, or MVM.

A packaging cell is used to form a virus particle that is capable of infecting a target cell. Exemplary packaging cells include 293 cells, which can package adenovirus, and ψ 2 or PA317 cells, 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, e.g., 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 can be supplied in *trans* by the packaging cell line and/or plasmid containing E2A, E4, and VA genes from adenovirus, and plasmid encoding *Rep* and *Cap* genes from AAV, as described in "Triple Transfection Protocol." 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. In certain embodiments, the viral DNA is packaged in a producer cell line, which contains E1A and/or E1B genes from adenovirus. The cell line is also infected with adenovirus as a helper. The helper virus (e.g., adenovirus or HSV) or helper plasmid promotes replication of the AAV vector and expression of AAV genes from the helper plasmid with ITRs. 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 certain embodiments, the viral vector is capable 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, single chain antibody, or 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 certain embodiments, the viral vector achieves cell type specific expression. For example, a tissue-specific promoter can be constructed to restrict expression of the transgene (Cas9 and gRNA) to 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 nuclear envelope (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, transient cell compression or squeezing (see, e.g., Lee 2012), gene gun, sonoporation, magnetofection, lipid-mediated transfection, dendrimers, inorganic nanoparticles, calcium phosphates, or a combination thereof.

In an embodiment, delivery via electroporation comprises mixing the cells with the Cas9-and/or gRNA-encoding DNA in a cartridge, chamber or cuvette and applying one or more electrical impulses of defined duration and amplitude. In an embodiment, delivery via electroporation is performed using a system in which cells are mixed with the Cas9-and/or gRNA-encoding DNA in a vessel connected to a device (e.g., a pump) which feeds the mixture into a cartridge, chamber or cuvette wherein one or more electrical impulses of defined duration and amplitude are applied, after which the cells are delivered to a second vessel.

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a combination of a vector and a non-vector based method. In an embodiment, the donor template nucleic acid 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 certain embodiments, the delivery vehicle is a non-viral vector, and in certain of these embodiments 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 1**.

Table 1: 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
Diioctadecylamidoglycylspermine	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-propanaminium bromide	MDRIE	Cationic
Dimyristooxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3β-[<i>N</i> -(<i>N,N'</i> -Dimethylaminoethane)-carbonyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Dioleoyl-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic
Dimethyloctadecylammonium bromide	DDAB	Cationic
Diioctadecylamidoglycylspermidin	DSL	Cationic
rac-[(2,3-Dioctadecyloxypropyl)(2-hydroxyethyl)]-dimethylammonium chloride	CLIP-1	Cationic
rac-[2(2,3-Dihexadecyloxypropyl-oxy methyloxy)ethyl]trimethylammonium bromide	CLIP-6	Cationic

Ethyl dimyristoyl phosphatidylcholine	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- <i>sn</i> -glycero-3-ethylphosphocholine	DSEPC	Cationic
<i>N</i> -Palmitoyl D-erythro-sphingosyl carbamoyl-spermine	CCS	Cationic
<i>N-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-Cholesteryl oxycarbonyl-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 5**.

Table 5: 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(<i>N</i> -ethyl-4-vinylpyridinium bromide)	
Poly(phosphazene)s	PPZ
Poly(phosphoester)s	PPE
Poly(phosphoramidate)s	PPA
Poly(<i>N</i> -2-hydroxypropylmethacrylamide)	pHPMA
Poly (2-(dimethylamino)ethyl methacrylate)	pDMAEMA
Poly(2-aminoethyl propylene phosphate)	PPE-EA
Chitosan	
Galactosylated chitosan	
<i>N</i> -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 an RNA-guided nuclease

RNA encoding RNA-guided nucleases, e.g., Cas9 molecules, and/or gRNA molecules, can be delivered into cells, e.g., target cells described herein, 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, transient cell compression or squeezing (see, e.g., Lee 2012), lipid-mediated transfection, peptide-mediated delivery, or a combination thereof. Cas9-encoding and/or gRNA-encoding RNA can be conjugated to molecules promoting uptake by the target cells (e.g., target cells described herein).

In an embodiment, delivery via electroporation comprises mixing the cells with the RNA encoding Cas9 molecules and/or gRNA molecules, with or without donor template nucleic acid molecules, in a cartridge, chamber or cuvette and applying one or more electrical impulses of defined duration and amplitude. In an embodiment, delivery via electroporation is performed using a system in which cells are mixed with the RNA encoding Cas9 molecules and/or gRNA molecules, with or without donor template nucleic acid molecules in a vessel connected to a device (e.g., a pump) which feeds the mixture into a cartridge, chamber or cuvette wherein one or more electrical impulses of defined duration and amplitude are

applied, after which the cells are delivered to a second vessel. Cas9-encoding and/or gRNA-encoding RNA can be conjugated to molecules to promote uptake by the target cells (e.g., target cells described herein).

Delivery of RNA-guided nucleases

RNA-guided nucleases, e.g., Cas9 molecules, 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, transient cell compression or squeezing (see, e.g., Lee 2012), lipid-mediated transfection, peptide-mediated delivery, or a combination thereof. Delivery can be accompanied by DNA encoding a gRNA or by a gRNA. Cas9 protein can be conjugated to molecules promoting uptake by the target cells (e.g., target cells described herein).

In an embodiment, delivery via electroporation comprises mixing the cells with the Cas9 molecules and/or gRNA molecules, with or without donor nucleic acid, in a cartridge, chamber or cuvette and applying one or more electrical impulses of defined duration and amplitude. In an embodiment, delivery via electroporation is performed using a system in which cells are mixed with the Cas9 molecules and/or gRNA molecules, with or without donor nucleic acid in a vessel connected to a device (e.g., a pump) which feeds the mixture into a cartridge, chamber or cuvette wherein one or more electrical impulses of defined duration and amplitude are applied, after which the cells are delivered to a second vessel. Cas9-encoding and/or gRNA-encoding RNA can be conjugated to molecules to promote uptake by the target cells (e.g., target cells described herein).

Route of administration of genome editing system components

Systemic modes of administration include oral and parenteral routes. Parenteral routes include, by way of example, intravenous, intramarrow, intrarterial, intramuscular, intradermal, subcutaneous, intranasal, and intraperitoneal routes. Components administered systemically may be modified or formulated to target, e.g., HSCs, hematopoietic stem/progenitor cells, or erythroid progenitors or precursor cells.

Local modes of administration include, by way of example, intramarrow injection into the trabecular bone or intrafemoral injection into the marrow space, and infusion into the portal vein. In an embodiment, significantly smaller amounts of the components (compared with systemic approaches) may exert an effect when administered locally (for example, directly into the bone marrow) 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 therapeutically effective amounts of a component are administered systemically.

Administration may be provided as a periodic bolus (for example, intravenously) or as continuous infusion from an internal reservoir or from an external reservoir (for example, from an intravenous bag or implantable pump). Components may be administered locally, for example, by continuous release from a sustained release drug delivery device.

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, with the choice of the appropriate system depending on the required rate of release for 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; 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 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, 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. 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 and can be loaded with components described herein.

Bi-modal or differential delivery of genome editing system components

Separate delivery of Cas system components, 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.

In certain embodiments, 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 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., AAV or lentivirus, delivery.

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 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 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, 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.

In certain embodiments, the first pharmacodynamic or pharmacokinetic property, e.g., distribution, persistence or exposure, is more limited than the second pharmacodynamic or pharmacokinetic property.

In certain embodiments, the first mode of delivery is selected to optimize, e.g., minimize, a pharmacodynamic or pharmacokinetic property, e.g., distribution, persistence or exposure.

In certain embodiments, the second mode of delivery is selected to optimize, e.g., maximize, a pharmacodynamic or pharmacokinetic property, e.g., distribution, persistence or exposure.

In certain embodiments, 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.

In certain embodiments, the second mode of delivery comprises a relatively transient element, e.g., an RNA or protein.

In certain embodiments, 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/or 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 certain embodiments, 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 of Cas system components

In certain embodiments, Cas system components described in **Table 3** 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 4**.

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 chimeric) 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 an embodiment, the modified nucleic acids comprise one, two, three or more modified nucleotides. In an embodiment, 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 an embodiment, 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 an embodiment, 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 an embodiment, 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 an embodiment, 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

Phosphate group

In an embodiment, 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 an embodiment, the modification of the phosphate backbone can include alterations that result in either an uncharged linker or a charged linker with unsymmetrical charge distribution.

Examples of modified phosphate groups include, phosphorothioate, phosphoroselenates, borano phosphates, borano phosphate esters, hydrogen phosphonates, phosphoroamidates, alkyl or aryl phosphonates and phosphotriesters. In an embodiment, 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), BR₃ (wherein R can be, e.g., hydrogen, alkyl, or aryl), C (e.g., an alkyl group, an aryl group, and the like), H, NR₂ (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 center. The stereogenic phosphorous atom can possess either the “R” configuration (herein Rp) or the “S” configuration (herein Sp).

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 an embodiment, modifications to one or both non-

bridging oxygens can 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 phosphoramidates), sulfur (bridged phosphorothioates) and carbon (bridged methylenephosphonates). The 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 an embodiment, the charged phosphate group can be replaced by a neutral moiety.

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 an embodiment, 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 an embodiment, 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 an embodiment, 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 an embodiment, 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, heteroarylamino, diheteroarylamino, or amino acid); $NH(CH_2CH_2NH)_nCH_2CH_2$ -amino (wherein amino can be, e.g., as described herein), - $NHC(O)R$ (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 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. 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 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 an embodiment, 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) or threose nucleic acid (TNA, where ribose is replaced with α -L-threofuranosyl- (3'→2')).

Modifications on the nucleobase

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 an embodiment, the nucleobase can include, for example, naturally-occurring and synthetic derivatives of a base.

Uracil

In an embodiment, 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 (mm^5U), 5-methylaminomethyl-2-thio-uridine (mm^5s2U), 5-methylaminomethyl-2-seleno-uridine (mm^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 (inm^5s2U), α -thio-uridine, 2'-O-methyl-uridine (Um), 5,2'-O-dimethyl-uridine (m^5Um), 2'-O-methyl-pseudouridine (ψm), 2-thio-2'-O-methyl-uridine ($s2Um$), 5-methoxycarbonylmethyl-2'-O-methyl-uridine (mcm^5Um), 5-carbamoylmethyl-2'-O-methyl-uridine (ncm^5Um), 5-carboxymethylaminomethyl-2'-O-methyl-uridine ($cmnm^5Um$), 3,2'-O-dimethyl-uridine (m^3Um), 5-(isopentenylaminomethyl)-2'-O-methyl-uridine (inm^5Um), 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

In an embodiment, 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 (m^3C), N4-acetyl-cytidine (act), 5-formyl-cytidine (f^5C), N4-methyl-cytidine (m^4C), 5-methyl-cytidine (m^5C), 5-halo-cytidine (e.g., 5-iodo-cytidine), 5-hydroxymethyl-cytidine (hm^5C), 1-methyl-pseudoisocytidine, pyrrolo-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 (k^2C), α -thio-cytidine, 2'-O-methyl-cytidine (Cm), 5,2'-O-dimethyl-cytidine (m^5Cm), N4-acetyl-2'-O-methyl-cytidine (ac^4Cm), N4,2'-O-dimethyl-cytidine (m^4Cm), 5-formyl-2'-O-methyl-cytidine (f^5Cm), N4,N4,2'-O-trimethyl-cytidine (m^4_2Cm), 1-thio-cytidine, 2'-F-ara-cytidine, 2'-F-cytidine, and 2'-OH-ara-cytidine.

Adenine

In an embodiment, 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-adenosine, 7-deaza-8-aza-adenosine, 7-deaza-2-amino-purine, 7-deaza-8-aza-2-amino-purine, 7-deaza-2,6-diaminopurine, 7-deaza-8-aza-2,6-diaminopurine, 1-methyl-adenosine (m^1A), 2-methyl-adenosine (m^2A), N6-methyl-adenosine (m^6A), 2-methylthio-N6-methyl-adenosine (ms^2m^6A), 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 (ms^2io^6A), N6-glycylcarbamoyl-adenosine (g^6A), N6-threonylcarbamoyl-adenosine (t^6A), N6-methyl-N6-threonylcarbamoyl-adenosine (m^6t^6A), 2-methylthio-N6-threonylcarbamoyl-adenosine (ms^2g^6A), N6,N6-dimethyl-adenosine (m^6_2A), N6-hydroxynorvalylcarbamoyl-adenosine (hn^6A), 2-methylthio-N6-hydroxynorvalylcarbamoyl-adenosine (ms^2hn^6A), N6-acetyl-adenosine (ac^6A), 7-methyl-adenosine, 2-methylthio-adenosine, 2-methoxy-adenosine, 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-pentaoxonadecyl)-adenosine.

Guanine

In an embodiment, 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 ($preQ_0$), 7-aminomethyl-7-deaza-guanosine ($preQ_1$), 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-

methyl-6- 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^2 Gm), N2,N2-dimethyl-2'-O-methyl-guanosine (m^2_2 Gm), 1-methyl-2'-O-methyl-guanosine (m' Gm), N2,7-dimethyl-2'-O-methyl-guanosine (m^2_2 7Gm), 2'-O-methyl-inosine (Im), 1,2'-O-dimethyl-inosine (m' Im), 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.

Exemplary Modified gRNAs

In some embodiments, the modified nucleic acids can be modified gRNAs. It is to be understood that any of the gRNAs described herein can be modified in accordance with this section, including any gRNA that comprises a targeting domain from SEQ ID NOs:251-901.

As discussed above, transiently expressed or delivered nucleic acids can be prone to degradation by, e.g., cellular nucleases. Accordingly, in one aspect the modified gRNAs described herein can contain one or more modified nucleosides or nucleotides which introduce stability toward nucleases. While not wishing to be bound by theory it is also believed that certain modified gRNAs described herein can exhibit a reduced innate immune response when introduced into a population of cells, particularly the cells of the present invention. As noted above, 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.

While some of the exemplary modification discussed in this section may be included at any position within the gRNA sequence, in some embodiments, a gRNA comprises a modification at or near its 5' end (e.g., within 1-10, 1-5, or 1-2 nucleotides of its 5' end). In some embodiments, a gRNA comprises a modification at or near its 3' end (e.g., within 1-10, 1-5, or 1-2 nucleotides of its 3' end). In some embodiments, a gRNA comprises both a modification at or near its 5' end and a modification at or near its 3' end.

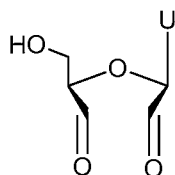
In an embodiment, the 5' end of a gRNA is modified by the inclusion of a eukaryotic mRNA cap structure or cap analog (e.g., a $G(5')ppp(5')G$ cap analog, a $m^7G(5')ppp(5')G$ cap analog, or a 3'-O-Me- $m^7G(5')ppp(5')G$ anti reverse cap analog (ARCA)). The cap or cap analog can be included during either chemical synthesis or *in vitro* transcription of the gRNA.

In an embodiment, an *in vitro* transcribed gRNA is modified by treatment with a phosphatase (e.g., calf intestinal alkaline phosphatase) to remove the 5' triphosphate group.

In an embodiment, the 3' end of a gRNA is modified by the addition of one or more (e.g., 25-200) adenine (A) residues. The polyA tract can be contained in the nucleic acid (e.g., plasmid, PCR product, viral genome) encoding the gRNA, or can be added to the gRNA during chemical synthesis, or following *in vitro* transcription using a polyadenosine polymerase (e.g., *E. coli* Poly(A)Polymerase).

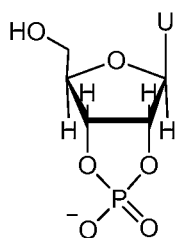
In an embodiment, *in vitro* transcribed gRNA contains both a 5' cap structure or cap analog and a 3' polyA tract. In an embodiment, an *in vitro* transcribed gRNA is modified by treatment with a phosphatase (e.g., calf intestinal alkaline phosphatase) to remove the 5' triphosphate group and comprises a 3' polyA tract.

In some embodiments, gRNAs can be modified at a 3' terminal U ribose. For example, the two terminal hydroxyl groups of the U 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 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 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, sugar-modified ribonucleotides can be incorporated into the gRNA, 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, aralkyl, heteroaryl or sugar), amino (wherein amino can be, e.g., NH₂; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, diheteroarylamino, 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, one or more of the nucleotides of the gRNA can each independently be a modified or unmodified nucleotide including, but not limited to 2'-sugar modified, such as, 2'-O-methyl, 2'-O-methoxyethyl, or 2'-Fluoro modified including, e.g., 2'-F or 2'-O-methyl, adenosine (A), 2'-F or 2'-O-methyl, cytidine (C), 2'-F or 2'-O-methyl, uridine (U), 2'-F or 2'-O-methyl, thymidine (T), 2'-F or 2'-O-methyl, guanosine (G), 2'-O-methoxyethyl-5-methyluridine (Teo), 2'-O-methoxyethyladenosine (Aeo), 2'-O-methoxyethyl-5-methylcytidine (m5Ceo), and any combinations thereof.

In some embodiments, a gRNA can include "locked" nucleic acids (LNA) in which the 2' OH-group can be connected, e.g., by a C1-6 alkylene or C1-6 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., NH₂; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino) and aminoalkoxy or O(CH₂)_n-amino (wherein amino can be, e.g., NH₂; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino).

In some embodiments, a gRNA can include a modified nucleotide which is multicyclic (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), or threose nucleic acid (TNA, where ribose is replaced with α -L-threofuranosyl-(3'→2')).

Generally, gRNA molecules include the sugar group ribose, which is a 5-membered ring having an oxygen. Exemplary modified gRNAs can include, without 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). Although the majority of sugar analog alterations are localized to the 2' position, other sites are amenable to modification, including the 4' position. In an embodiment, a gRNA comprises a 4'-S, 4'-Se or a 4'-C-aminomethyl-2'-O-Me modification.

In some embodiments, deaza nucleotides, e.g., 7-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, one or more or all of the nucleotides in 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 an mRNA, and down-regulate gene expression. While not wishing to be bound by theory, it is believed that this down regulation occurs by either reducing nucleic acid molecule stability or 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.

EXAMPLES

The following Examples are merely illustrative and are not intended to limit the scope or content of the invention in any way.

Example 1: Screening of *S. pyogenes* gRNAs for use in inserting 13 bp del c.-114 to -102 into HBG1 and HBG2 regulatory regions

S. pyogenes gRNAs targeting a 26 nt fragment spanning and including the 13 nucleotides at c.-114 to -102 of HBG1 (e.g., nucleotides 2824-2836 of SEQ ID NO:902 (HBG1)), resulting in the alteration HBG1 13 bp del c.-114 to -102) and HBG2 (e.g., nucleotides 2748-2760 of SEQ ID NO:903 (HBG2), resulting in the alteration HBG1 13 bp del c.-114 to -102) were designed as set forth herein. After gRNAs were designed *in silico* and tiered, a subset of the gRNAs were selected and screened for activity and specificity in human K562 cells. The gRNAs selected for screening contain the targeting domain sequences as set forth in **Table 8**. The DNA encoding a U6 promoter and each of the *S. pyogenes* gRNAs were co-electroporated (Amaza Nucleofector) with plasmid DNA encoding

S. pyogenes Cas9 into human K562 cells. The experimental conditions were generally in accordance with those known in the art (e.g., Gori 2016, which is hereby incorporated by reference herein). Three days after electroporation, gDNA was extracted from K562 cells and then the HBG1 and HBG2 loci were PCR amplified from the gDNA. Gene editing was evaluated in the PCR products by T7E1 endonuclease assay analysis. Of the ten sgRNAs screened, eight cut in both the HBG1 and HBG2 targeted regions in the promoter sequences (**Fig. 10A**).

The HBG1 and HBG2 PCR products for the K562 cells that were targeted with the eight active sgRNAs were then analyzed by DNA sequencing analysis and scored for insertions and deletions detected. The deletions were subdivided into precise 13 nt deletions at the HPFH site, HPFH inclusive and proximal small deletions (18-26 nt), 12 nt deletions (i.e., partial deletion) of the HPFH target site, >26 nt deletions that span a portion of the HPFH target site, and other deletions, e.g., deletions proximal to but outside the HPFH target site. Seven of the eight sgRNAs targeted deletion of the 13 nt (HPFH mutation induction) (**Fig. 10B**) for HBG1. At least five of the eight sgRNAs also supported targeted deletion of the 13 nt (HPFH mutation induction) in HBG2 promoter region (**Fig. 10C**). Note that DNA sequence results for HBG2 in cells treated with HBG Sp34 sgRNA were not available. These data indicate that Cas9 and sgRNA support precise induction of the 13 bp del c.-114 to -102 HPFH mutation. Figs. **10D-10F** depict examples of the types of deletions observed in target sequences in HBG1. The gRNA used in each particular example is shown in black, and the other gRNAs not targeted in each panel's example are shown in white.

Table 8: Select list of gRNAs for screening in K562 cells

gRNA ID	Targeting domain sequence (RNA)	Targeting domain sequence (DNA)	Targeting domain sequence plus 3 NT <i>S. pyogenes</i> PAM sequence (NGG) (RNA)	Targeting domain sequence plus 3 NT <i>S. pyogenes</i> PAM sequence (NGG) (DNA)	Sense	Guide category
Sp9	GGCUAUUGGU CAAGGCA (SEQ ID NO:277)	GGCTATTGGTC AAGGCA (SEQ ID NO:910)	GGCUAUUGGU CAAGGCAAGG (SEQ ID NO:920)	GGCTATTGGTC AAGGCAAGG (SEQ ID NO:930)	Anti-sense	spy17
Sp36	CAAGGCUAUU GGUCAAGGCA (SEQ ID NO:338)	CAAGGCTATT GGTCAAGGCA (SEQ ID NO:911)	CAAGGCUAUU GGUCAAGGCA AGG (SEQ ID NO:921)	CAAGGCTATT GGTCAAGGCA AGG (SEQ ID NO:931)	Anti-sense	spy20
Sp40	UGCCUUGUCA AGGCUAU (SEQ ID NO:327)	TGCCTTGTC GGCTAT (SEQ ID NO:912)	UGCCUUGUCA AGGCUAUUGG (SEQ ID NO:922)	TGCCTTGTC GGCTATTGG (SEQ ID NO:932)	Anti-sense	spy17
Sp42	GUUUGCCUUG UCAAGGCUAU (SEQ ID NO:299)	GTTTGCCTTGT CAAGGCTAT (SEQ ID NO:913)	GUUUGCCUUG UCAAGGCUAU UGG (SEQ ID NO:923)	GTTTGCCTTGT CAAGGCTATT GG (SEQ ID NO:933)	Anti-sense	spy20

Sp38	GACCAAUAGC CUUGACA (SEQ ID NO:276)	GACCAATAGC CTTGACA (SEQ ID NO:914)	GACCAAUAGC CUUGACAAGG (SEQ ID NO:924)	GACCAATAGC CTTGACAAGG (SEQ ID NO:934)	Sense	spy17
Sp37	CUUGACCAAU AGCCUUGACA (SEQ ID NO:333)	CTTGACCAATA GCCTTGACA (SEQ ID NO:915)	CUUGACCAAU AGCCUUGACA AGG (SEQ ID NO:925)	CTTGACCAATA GCCTTGACAA GG (SEQ ID NO:935)	Sense	spy20
Sp43	GUCAAGGCUA UUGGUCA (SEQ ID NO:278)	GTCAAGGCTA TTGGTCA (SEQ ID NO:916)	GUCAAGGCUA UUGGUCAAGG (SEQ ID NO:926)	GTCAAGGCTA TTGGTCAAGG (SEQ ID NO:936)	Anti- sense	spy17
Sp35	CUUGUCAAGG CUAUUGGUCA (SEQ ID NO:339)	CTTGTCAAGGC TATTGGTCA (SEQ ID NO:917)	CUUGUCAAGG CUAUUGGUCA AGG (SEQ ID NO:927)	CTTGTCAAGGC TATTGGTCAAG G (SEQ ID NO:937)	Anti- sense	spy20
Sp41	UCAAGUUUGC CUUGUCA (SEQ ID NO:310)	TCAAGTTTGCC TTGTCA (SEQ ID NO:918)	UCAAGUUUGC CUUGUCAAGG (SEQ ID NO:928)	TCAAGTTTGCC TTGTCAAGG (SEQ ID NO:938)	Anti- sense	spy17
Sp34	UGGUCAAGUU UGCCUUGUCA (SEQ ID NO:340)	TGGTCAAGTTT GCCTTGTC (SEQ ID NO:919)	UGGUCAAGUU UGCCUUGUCA AGG (SEQ ID NO:929)	TGGTCAAGTTT GCCTTGTCAG G (SEQ ID NO:939)	Anti- sense	spy20

Example 2: Cas9 RNP containing gRNA targeting the HPFH mutation supports gene editing in human hematopoietic stem/progenitor cells

Human cord blood (CB) CD34⁺ cells were prestimulated with human cytokines (stem cell factor (SCF), thrombopoietin (TPO), Flt3 ligand (FL)) and small molecules (prostaglandin E2 (PGE2), StemRegenin 1 (SR1)) for two days. The experimental conditions were generally in accordance with the methods provided in Gori 2016 at pages 240-241, which is hereby incorporated by reference herein. The CB CD34⁺ cells were electroporated (Amaza Nucleofector) with *S. pyogenes* Cas9 RNP containing (e.g., 5' ARCA capped and 3' polyA (20A) tail) sgRNAs (Table 8) that target the HBG1 and HBG2 regulatory regions. Three days after electroporation, gDNA was extracted from the RNP-treated CB CD34⁺ cells, and gene editing was analyzed by T7E1 assay and DNA sequencing.

Of the RNPs containing different gRNAs tested in CB CD34⁺ cells, only Sp37 gRNA (comprising SEQ ID NO:333) resulted in detectable editing at the target site in the *HBG1* and *HBG2* promoters as determined by T7E1 analysis of indels in *HBG1* and *HBG2* specific PCR products amplified from gDNA extracted from electroporated CB CD34⁺ cells from a three cord blood donors (**Fig. 11A**). The average level of editing detected in cells electroporated with Cas9 protein complexed to Sp37 was 5 ± 2 % indels at *HBG1* and 3 ± 1 % indels detected at *HBG2* (three separate experiments, and CB donors).

Next, three *S. pyogenes* gRNAs whose target sites are within the *HBG* promoter (Sp35 (comprising SEQ ID NO:339), Sp36 (comprising SEQ ID NO:338), Sp37 (comprising

SEQ ID NO:333)) were complexed to wild-type *S. pyogenes* Cas9 protein to form ribonucleoprotein complexes. These HBG targeted RNPS were electroporated into CB CD34⁺ cells (n=3 donors) and adult mobilized peripheral blood (mPB) CD34⁺ cell donors (n=3 donors). CB CD34⁺ cells were prepared in accordance with the methods described above and as in Gori 2016 at pages 240-241. Adult mPB CD34⁺ cells were prepared in substantially the same way as the CB CD34⁺ cells, except without the addition of SR1. Approximately three days after Cas9 RNP delivery, the level of insertions/deletions at the target site was analyzed by T7E1 endonuclease analysis of the *HBG2* PCR products amplified from genomic DNA extracted from the samples. Each of these RNPs supported only low level gene editing in both the CB and adult CD34⁺ cells across three donors and three separate experiments (**Fig. 11B**).

In order to increase gene editing at the target site and to increase the occurrence of the 13 bp deletion at the target site, single strand deoxynucleotide donor repair templates (ssODNs) that encode 87 bp and 89 bp of homology on the 5' and 3' side of the targeted deletion site of *HBG1* and *HBG2* were generated. The construct ssODN1 (SEQ ID NO:906, **Table 9**), including the 5' and 3' homology arm, was designed to 'encode' the 13 bp deletion with sequence homology arms engineered flanking this absent sequence to create a perfect deletion. The 5' homology arm (SEQ ID NO:904, **Table 9**) includes nucleotides homologous to the sequence 5' of c.-114 to -102 of *HBG1* and *HBG2* (i.e., nucleotides homologous to the sequence 5' of nucleotides 2824-2836 of SEQ ID NO:902 (*HBG1*) and nucleotides homologous to the sequence 5' of nucleotides 2748-2760 of SEQ ID NO:903 (*HBG2*)). The 3' homology arm (SEQ ID NO:905, **Table 9**) includes nucleotides homologous to the 3' region of c.-114 to -102 of *HBG1* and *HBG2* (i.e., nucleotides homologous to the sequence 3' of nucleotides 2824-2836 of SEQ ID NO:902 (*HBG1*) and nucleotides homologous to the sequence 3' of nucleotides 2748-2760 of SEQ ID NO:903 (*HBG2*)). The ssODN1 construct was modified at the ends to contain phosphothioates (PhTx) at the 5' and 3' ends (SEQ ID NO:909, **Table 9**) to form PhTx ssODN1.

Table 9: Single strand deoxynucleotide donor repair templates (ssODN)

ssODN ID	SEQ ID NO	Sequence
ssODN1 5' homology arm	904	GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGGCCCCCTG GCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTT
ssODN1 3' homology arm	905	GTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGGACCG TTTCAGACAGATATTTCATTGAGATAGTGTGGGGAAGGGG
ssODN1	906	GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGGCCCCCTG GCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAG GCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGGACCGTTTCAG ACAGATATTTGCATTGAGATAGTGTGGGGAAGGGG

PhTx ssODN1 5' homology arm	907	*GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGCCCCCTG GCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTT
PhTx ssODN1 3' homology arm	908	<u>GTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGGACCG</u> <u>TTTCAGACAGATATTTGCATTGAGATAGTGTGGGGAAGGGG*</u>
PhTx ssODN1	909	*GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGCCCCCTG GCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAG GCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGGACCGTTTCAG ACAGATATTTGCATTGAGATAGTGTGGGGAAGGGG*
The homology arms flanking the deletion are indicated by bold [5' homology arm] and underline [3' homology arm]). Note the absence of the 13 bp sequence in ssODN1 and PhTx ssODN1. *Represents modification by phosphothioate.		

CB CD34⁺ cells were prepared in accordance with the methods described above and as in Gori 2016 at pages 240-241. The ssODNs (*i.e.*, ssODN1 and PhTx ssODN1) were co-delivered with RNP targeting *HBG* containing the Sp37 gRNA (*HBG* Sp37 RNP) or *HBG* Sp35 (*HBG* Sp35 RNP) to CB CD34⁺ cells.

Co-delivery of the ssODN1 and PhTx ssODN1 donor templates encoding the 13bp deletion with RNP containing Sp35 gRNA (*i.e.*, *HBG* Sp35 RNP) or RNP containing Sp37 gRNA (*i.e.*, *HBG* Sp37 RNP) led to a 6-fold and 5-fold increase in gene editing of the target site, respectively, as determined by T7E1 analysis of the *HBG2* PCR product (**Fig. 11C**). DNA sequencing analysis (Sanger sequencing) of the *HBG2* PCR product indicated that 20% gene editing in cells that were treated with *HBG* Sp37 RNP and PhTx ssODN1, with 15% deletions and 5% insertions (**Fig. 11C, lower left panel**). Further analysis of the specific type and size of deletions at the target site revealed that ¾ of 75% of the total deletions detected contained the HPFH 13 bp deletion (which included deletion of the CAAT box in the proximal promoter), the absence of which is associated with elevation of HbF expression (**Fig. 11C, lower right panel**). The remaining ¼ of deletions were partial deletions that did not span the full 13 bp deletion. These data indicate that co-delivery of a homologous ssODN that is engineered to have a deletion supported precise gene editing (deletion) at *HBG* in human CD34⁺ cells.

Example 3: Screening of *S. pyogenes* gRNAs delivered to K562 cells as ribonucleoprotein complexes for use in causing 13 bp del c.-114 to -102 into *HBG1* and *HBG2* regulatory regions

Guide RNAs screened by electroporation of Cas9 and gRNA DNA into K562 cells as described in Example 1 (**Fig. 10**) were *in vitro* transcribed and then complexed with *S. pyogenes* Wt Cas9 protein to form ribonucleoprotein complexes (RNPs). To compare the activity of these RNPs to what was observed by DNA delivery of Cas9 and gRNAs into K562

cells (*i.e.*, Example 1) and by RNP delivery to human CD34⁺ cells (*i.e.*, Example 2), here RNPs were delivered to K562 cells by electroporation (Amaxa Nucleofector). The gRNAs complexed to *S. pyogenes* Cas9 protein were modified gRNAs ((*e.g.*, 5' ARCA capped and 3' polyA (20A) tail; **Table 8**) and target the *HBG1* and *HBG2* regulatory regions.

Three days after electroporation, gDNA was extracted from K562 cells and then the *HBG1* and *HBG2* promoter regions amplified by PCR, followed by T7E1 analysis of the PCR products. (**Fig. 12A**). Eight out of nine RNPs supported a high percentage of NHEJ. Sp37 RNP, the only gRNA shown to be active in human CD34⁺ cells (<10% editing in CD34⁺ cells), was highly active in K562 cells, with >60% indels detected at both *HBG1* and *HBG2* (**Fig. 12A**). The other gRNA that targets the HPFH deletion mutation site, Sp35, supported 43% editing at *HBG1* and *HBG2* (**Fig. 12A**).

DNA sequencing analysis was performed on a subset of PCR products from the gDNA from cells that were treated with Cas9 complexed to gRNAs closest to the targeted HPFH site. DNA sequences were scored for insertions and deletions detected. The deletions were subdivided into precise 13 nt deletions at the HPFH site, HPFH inclusive and proximal small deletions (18-26 nt), 12 nt deletions (*i.e.*, partial deletion) of the HPFH target site, >26 nt deletions that span a portion of the HPFH target site, and other deletions, *e.g.*, deletions proximal to but outside the HPFH target site. The 13 nt deletion was detected in cells treated with RNP complexed to gRNAs Sp35 and 37 (HPFH mutation induction) (**Fig. 12B**) for *HBG1*/*HBG2*. These data indicate that Cas9 and sgRNAs (Sp35 and Sp37) delivered to hematopoietic cells as ribonucleoprotein complexes caused the c.-114 to -102 HPFH mutation.

Example 4: Cas9 RNP targeting the HPFH mutation supports gene editing in human adult mobilized peripheral blood hematopoietic stem/progenitor cells with increased *HBG* expression in erythroblast progeny

To determine whether editing *HBG* with Cas9 RNP complexed to Sp37 gRNA or Sp35 gRNA (*i.e.*, the gRNAs that target the 13bp deletion that is associated with HPFH) in the promoter of *HBG* supports an increase in *HBG* expression in erythroid progeny of edited CD34⁺ cells, human adult CD34⁺ cells from mobilized peripheral blood (mPB) were electroporated with the RNPs. Briefly, mPB CD34⁺ cells were prestimulated for 2 days with human cytokines and PGE2 in StemSpan Serum-Free Expansion Medium (SFEM) and then electroporated with Cas9 protein precomplexed to Sp35 and Sp37, respectively. *See* Gori 2016. T7E1 analysis of *HBG* PCR product indicated ~3% indels detected for mPB CD34⁺

cells treated with RNP complexed to Sp37 while no editing was detected for cells that were treated with RNP complexed to Sp35 (**Fig. 13A**).

In order to increase gene editing at the target site and to increase the occurrence of the 13 bp deletion at the target site, PhTx ssODN1 was co-delivered with the precomplexed RNP targeting *HBG* containing the Sp37 gRNA. Co-delivery of the PhTx ssODN1 donor encoding the 13bp deletion led to a nearly 2-fold increase in gene editing of the target site (**Fig. 13A**).

To determine whether editing *HBG* increases production of fetal hemoglobin in erythroid progeny of edited adult CD34⁺ cells, the cells were differentiated into erythroblasts by culture for up to 18 days in the presence of human cytokines (erythropoietin, SCF, IL3), human plasma (Octoplas), and other supplements (hydrocortisone, heparin, transferrin). Over the time course of differentiation, mRNA was collected to evaluate *HBG* gene expression in the erythroid progeny of RNP treated mPB CD34⁺ cells and donor matched negative (untreated) controls. By day seven of differentiation, erythroblast progeny of human CD34⁺ cells that were treated with *HBG* Sp37 RNP and 13bp HPFH deletion encoding ssODN1 (~5% indels detected in gDNA from the bulk cell population by T7E1 analysis) exhibited a 2-fold increase in *HBG* mRNA production (**Fig. 13B**). Furthermore, the erythroblasts differentiated from RNP treated CD34⁺ cells maintained the kinetics of differentiation observed for donor matched untreated control cells as determined by flow analysis for acquisition of erythroid phenotype (% Glycophorin A⁺ cells) (**Fig. 14A**). Importantly, CD34⁺ cells that were electroporated with *HBG* Sp37 RNP and ssODN1 maintained their *ex vivo* hematopoietic activity (*i.e.*, no difference in the quantity or diversity of erythroid and myeloid colonies compared to untreated donor matched CD34⁺ cell negative control), as determined in hematopoietic colony forming cell (CFC) assays (**Fig. 14B**). These data indicate that targeted disruption of *HBG1/HBG2* proximal promoter region supported an increase in *HBG* expression in erythroid progeny of RNP treated adult hematopoietic stem/progenitor cells without altering differentiation potential.

SEQUENCES

Genome editing system components according to the present disclosure (including without limitation, RNA-guided nucleases, guide RNAs, donor template nucleic acids, nucleic acids encoding nucleases or guide RNAs, and portions or fragments of any of the foregoing), are exemplified by the nucleotide and amino acid sequences presented in the

Sequence Listing. The sequences presented in the Sequence Listing are not intended to be limiting, but rather illustrative of certain principles of genome editing systems and their component parts, which, in combination with the instant disclosure, will inform those of skill in the art about additional implementations and modifications that are within the scope of this disclosure. A list of the sequences presented is provided in the following **Table 10**.

Table 10: Sequences presented in the Sequence Listing:

SEQ ID NOS:	DESCRIPTION
1-2, 4-6, 12, 14	Cas9 polypeptides
3, 7-11, 13	Cas9 coding sequences
15-23, 52-123	Cas9 RuvC-like domains
24-28, 124-198	Cas9 HNH-like domains
29-31, 38-51	Full-length modular and unimolecular gRNAs
32-37	gRNA proximal and tail domains
199-205	PAM sequences
251-901	gRNA targeting domains (RNA)– see Tables 6-8
910-919	gRNA targeting domains (DNA)– see Table 8
920-929	gRNA targeting domains plus PAM (NGG) (RNA) – see Table 8
930-939	gRNA targeting domains plus PAM (NGG) (DNA) – see Table 8
902, 903	Human HBG1, 2 promoter sequences including HPFH deletion site
904-909	Oligonucleotide donor sequences and homology arms – see Table 9

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 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 herein. Such equivalents are intended to be encompassed by the following claims.

References

- Ahern et al., Br J Haematol 25(4):437-444 (1973)
- Akinbami Hemoglobin 40:64-65 (2016)
- Aliyu et al. Am J Hematol 83:63-70 (2008)
- Anders et al. Nature 513(7519):569-573 (2014)
- Angastiniotis & Modell Ann N Y Acad Sci 850:251-269 (1998)
- Bae et al. Bioinformatics 30(10):1473-1475 (2014)
- Barbosa et al. Braz J Med Bio Res 43(8):705-711 (2010)
- Bouva Hematologica 91(1):129-132 (2006)
- Brousseau Am J Hematol 85(1):77-78 (2010)
- Caldecott Nat Rev Genet 9(8):619-631 (2008)
- Chassanidis Ann Hematol 88(6):549-555 (2009)
- Chylinski et al. RNA Biol 10(5):726-737 (2013)
- Cong et al. Science 399(6121):819-823 (2013)
- Costa et al., Cad Saude Publica 18(5):1469-1471 (2002)
- Cotta-Ramusino et al., Int'l Patent Publication No. WO 2016/073990 (2016)
- Fine et al. Sci Rep. 5:10777 (2015)
- Friedland et al. Genome Biol 16:257 (2015)
- Fu et al. Nat Biotechnol 32:279-284 (2014)
- Gori et al., Int'l Patent Publication No. WO 2016/182959 A1 (2016)
- Guilinger et al. Nat Biotechnol 32:577-582 (2014)
- Jinek et al. Science 337(6096):816-821 (2012)
- Jinek et al. Science 343(6176):1247997 (2014)
- Kleinstiver et al. Nature 523(7561):481-485 (2015a)
- Kleinstiver et al. Nat Biotechnol 33(12):1293-1298 (2015b)
- Kleinstiver et al. Nature 529(7587):490-495 (2016)
- Lee et al. Nano Lett 12(12):6322-6327 (2012)
- Lewis "Medical-Surgical Nursing: Assessment and Management of Clinical Problems" (2014)

- Li Cell Res 18(1):85-98 (2008)
- Maeder et al. Int'l Patent Publication No. WO 2015/138510 (2015)
- Mali et al. Science 339(6121):823-826 (2013)
- Mantovani et al. Nucleic Acids Res 16(16):7783-7797 (1988)
- Marteijn et al. Nat Rev Mol Cell Biol 15(7):465-481 (2014)
- Nishimasu et al. Cell 156(5):935-949 (2014)
- Ran et al. Cell 154(6):1380-1389 (2013)
- Shmakov et al. Molecular Cell 60(3):385-397 (2015)
- Sternberg et al. Nature 507(7490):62-67 (2014)
- Superti-Furga et al. EMBO J 7(10):3099-3107 (1988)
- Thein Hum Mol Genet 18(R2):R216-223 (2009)
- Waber et al. Blood 67(2):551-554 (1986)
- Wang et al. Cell 153(4):910-918 (2013)
- Xu et al. Genes Dev 24(8):783-798 (2010)
- Yamano et al. Cell 165(4): 949-962 (2016)
- Zetsche et al. Nat Biotechnol 33(2):139-42 (2015)

CLAIMS

1. A method of increasing the level of fetal hemoglobin in a human cell by genome editing, the method comprising the step of introducing into the human cell a ribonucleoprotein (RNP) complex comprising:
 - (a) an RNA-guided nuclease; and
 - (b) a guide RNA (gRNA) comprising a targeting domain comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:310, SEQ ID NO:333, SEQ ID NO:338, SEQ ID NO:339, SEQ ID NO:340, SEQ ID NO:370, SEQ ID NO:371, SEQ ID NO:372, and SEQ ID NO:379.
2. The method of claim 1, wherein the step of introducing into the human cell further comprises introducing into the human cell a single stranded oligodeoxynucleotide (ssODN).
3. The method according to claim 1 or 2, wherein the step of introducing into the human cell is performed via electroporation.
4. The method according to any one of claims 1-3, wherein the step of introducing into the human cell occurs *in vitro* or *ex vivo*.
5. A cell generated by the method of any one of claims 1-4.
6. A cell modified using a ribonucleoprotein (RNP) complex comprising:
 - (a) an RNA-guided nuclease protein; and
 - (b) a guide RNA (gRNA) comprising a targeting domain comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:310, SEQ ID NO:333, SEQ ID NO:338, SEQ ID NO:339, SEQ ID NO:340, SEQ ID NO:370, SEQ ID NO:371, SEQ ID NO:372, and SEQ ID NO:379.
7. The method according to any one of claims 1-4 or the cell according to claim 5 or 6, wherein the cell is capable of differentiating into an erythroblast, erythrocyte, or a precursor of an erythroblast or erythrocyte.
8. The method according to claim 7 or the cell according to claim 7, wherein the cell is a CD34+ cell.
9. A genome editing system comprising a ribonucleoprotein (RNP) complex comprising:
 - (a) an RNA-guided nuclease; and

(b) a guide RNA (gRNA) comprising a targeting domain comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:310, SEQ ID NO:333, SEQ ID NO:338, SEQ ID NO:339, SEQ ID NO:340, SEQ ID NO:370, SEQ ID NO:371, SEQ ID NO:372, and SEQ ID NO:379.

10. The method according to any one of claims 1-4 or 7-8; the cell according to any one of claims 5-8; or the genome editing system according to claim 9; wherein the gRNA comprises one or more modifications selected from the group consisting of a 2'-acetylation, a 2'-methylation, and a phosphorothioate modification.

11. The method according to any one of claims 1-4, 7-8, or 10; the cell according to any one of claims 5-8 or 10; or the genome editing system according to any one of claims 9 or 10; wherein the RNA-guided nuclease is a Cas9 endonuclease protein.

12. The method according to claim 11; the cell according to claim 11; or the genome editing system according to claim 11; wherein the Cas9 endonuclease protein is selected from the group consisting of a *S. pyogenes*, *S. aureus*, and *S. thermophilus* Cas9 endonuclease protein.

Editas Medicine, Inc.

Patent Attorneys for the Applicant/Nominated Person

SPRUSON & FERGUSON

[illegible]

Fig. 1B

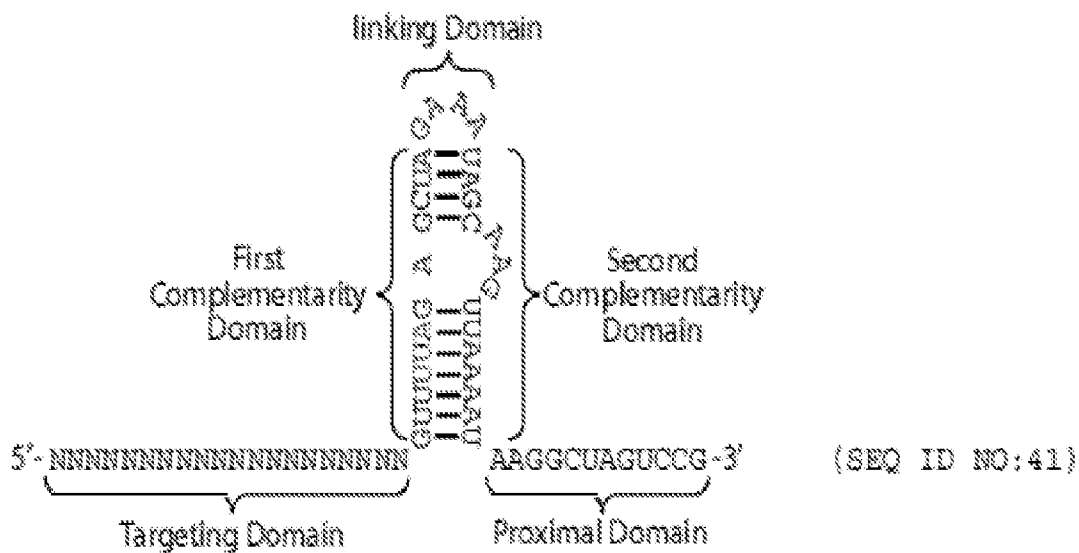


Fig. 1C

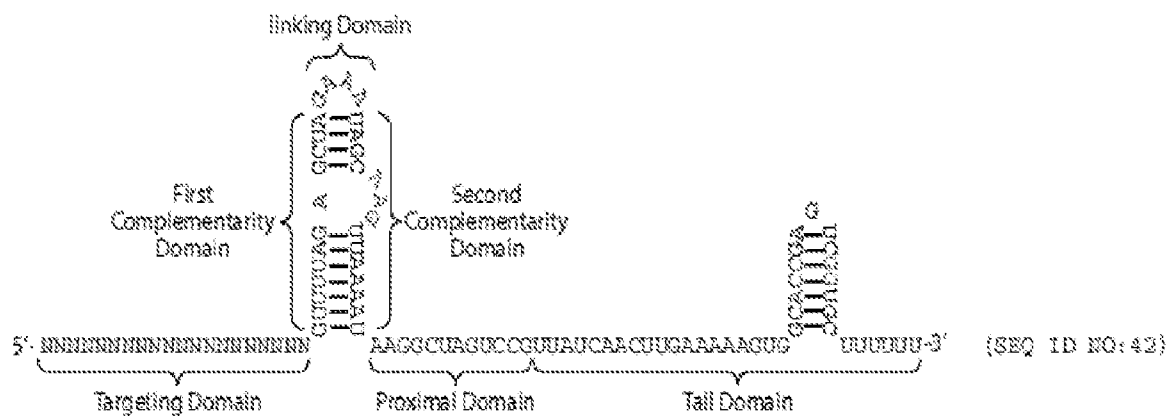


Fig. 1D

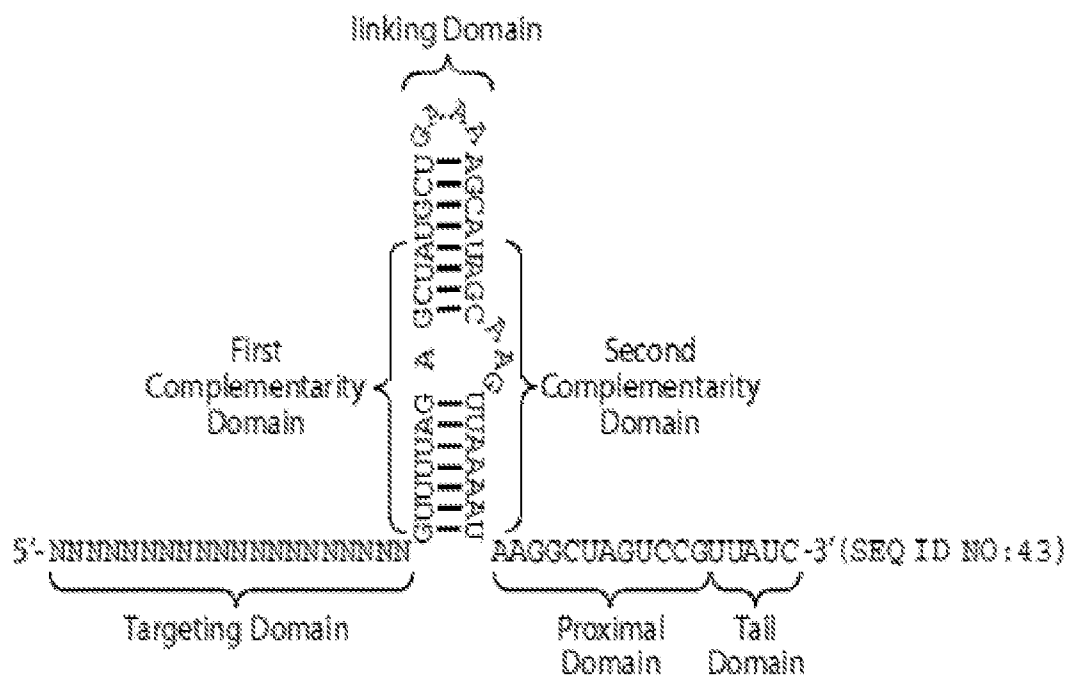


Fig. 1E

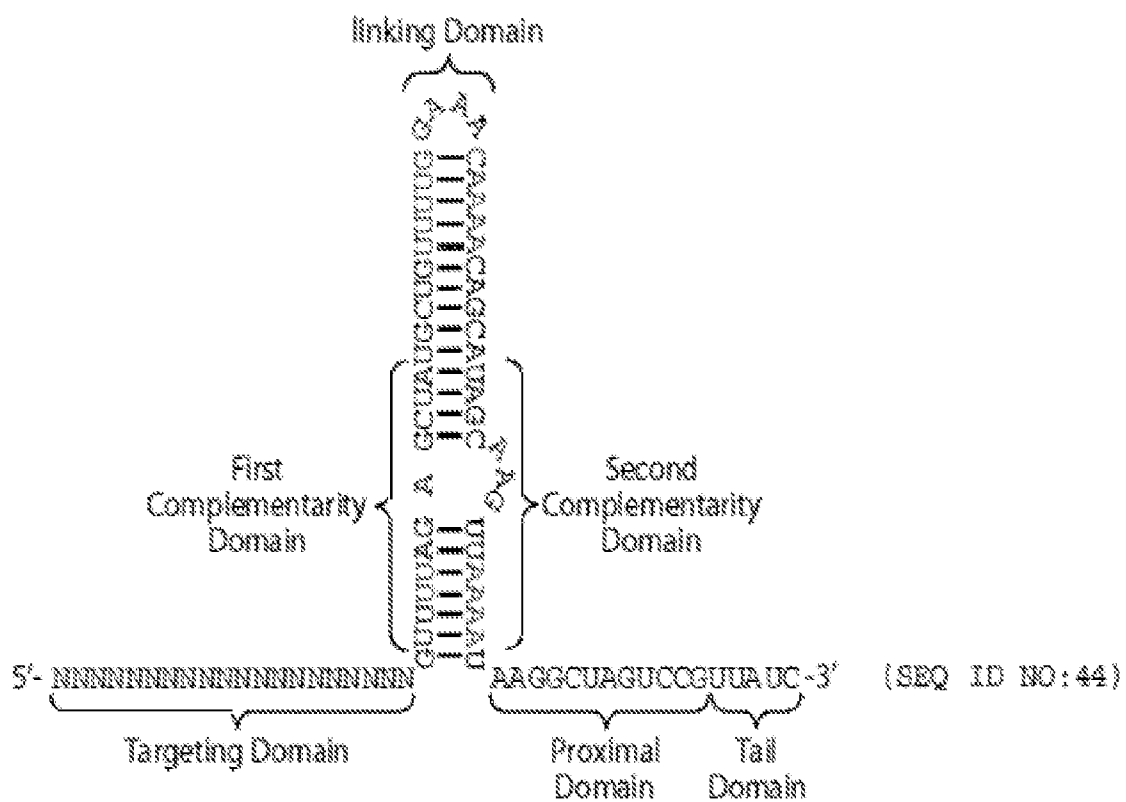
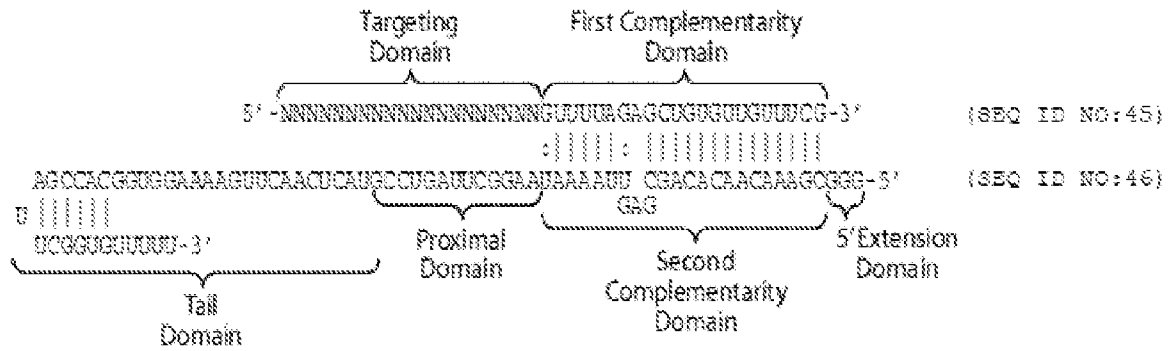


Fig. 1F

Streptococcus thermophilus

a) Structure (See, e.g., Karvelis et al. RNA Biology, 2013; 10(5): 841-851)



Signature

SUBSTITUTE SHEET (RULE 26)

Fig. 1H

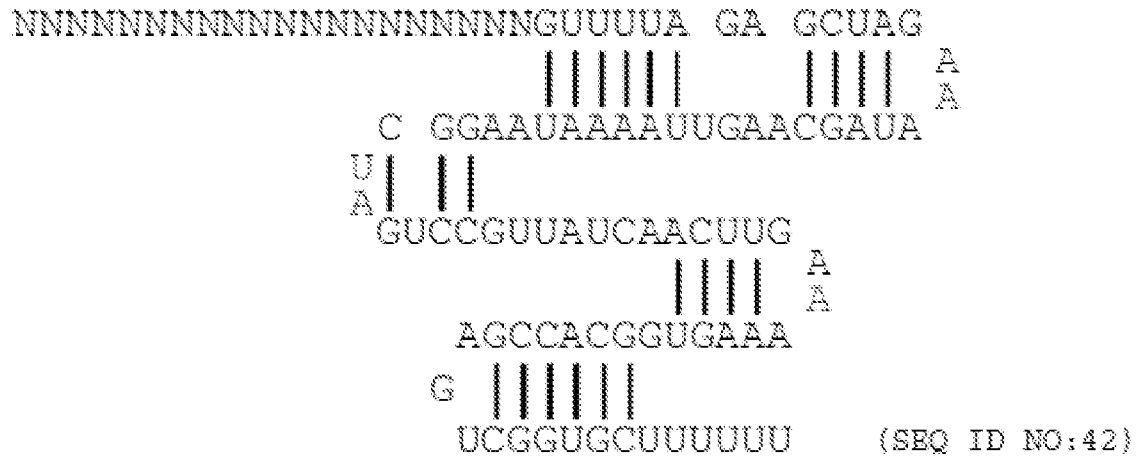
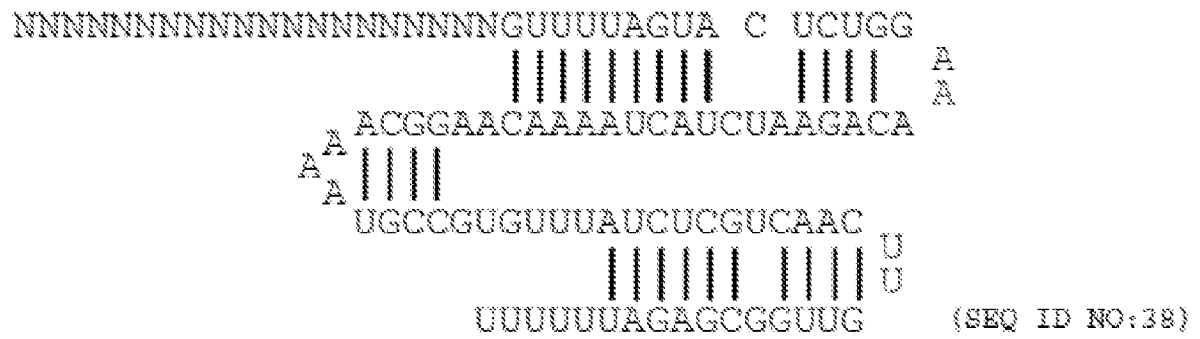


Fig. 11



CLUSTAL format alignment by MAFFT (v7.058b)



[illegible]

[illegible]

[illegible]

32

5

[illegible]

CS2

SM GNSDK-LIPRTKKKFTWDTKKYGGFDSPIVAYSILVIADIENGKSEKLLKTVKALVGVTIM
SP RNSDK-LIARKKD---WDFKKYGGFDSPTVAYSVLVAKVEKGEKSKLLKSVKELLGITIM
ST PMSNENLVGAKKEY---LDPKKYGGYAGISNSFTVLVXGTIEKGAKKKITNVLSFQGISIL
LI GNSSK-LIPKRTN---WDFMKYGGLDSPIMAYAVVI--EYAKGKN-KLVFEKKILRVTIM
*: *: * *: **** . : : : : : ** . *: : : :
Motif: -NS*-L*-K-----D--KYGG-----*****-----KG---K*-----*---*I*

SM EKMTFEEDPVAFLERKGYRQVQRENI IKLPKYSLSFKLENGRKRLLAS-----ARELQK
SP ERSSFEKNPIDFLEAKGYKEVKKDLI IKLPKYSLSFELENGRKRMLAS-----AGELQK
ST DRINVRKDKLNFLENGYKDI--ELI IELPKYSLSFELSDGSRRLASILSTNNKRGEIHK
LI ERKAFEKDEKAFLEEQGYRQP--KVLAKLPKYTLTYECEGRRRMLAS-----ANEAQK
: : : : : ** : **: : . : : *****: : : . : * : * : * : *
Motif: **---*---*---*FL--*GY*---*---*LPKY*L*---*G-*R*LAS-----E-*E

SM GNEIVLPNHLGTLTYHAKNIHKV-----DEPKHLDYVDKHKDESKELLDVVSNSFSKKYT
SP GNEIALPSKYVNFYLYASHYEKLGKSPEDNEQKQL-FVEQKHLYLDELIEQISEFSKRVI
ST GNQIFLSQKRFVKLLYHAKRISNT-----INENHREKYVENHKKKEFEELFYIIEFNENYV
LI GNQQVLPNHLVTLHLHAANCEVS-----DGKSLDYIESNREMFALLAHVSEFAKRYT
*: *: * . : : : *: * . : : : : : : : : : : * : : : : * : :
Motif: GN*--L--*---*L*-A-----*---*---*---*---*E*---*---*F*---

SM LAEGNLEKI KELYAQINGEDLKELASSFI-----HLLTFTAIGAPATFKFFDKNIDR
SP LADANLDKVL SAYNKHHRDKPIREQAENI I-----HLFTLTNLGAPAPFKYFDITIDR
ST GAKKNGKLLNSAFQSWQHHSIDELCSSFIGPTGSEKGLFELTSRGSAADFEPLGVKIPR
LI LAEAMLNKINQLFEQNKKEGDIKAI AQSFV-----DLMAFNAMGAPASFKFFETTIER
* . * . : . : . : : : : : : : : : : : * : : . * : * * : : : . * *
Motif: -A--N-----*---*---*---*---*---*---*---*---*L*---*---*G*-A-F*---*---*I-R

Fig. 2G

SM	KR~YTSTTEILNATLIHQSI TGLYETRIDLNKLGGD	(SEQ ID NO:1)
SP	KR~YTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD	(SEQ ID NO:2)
ST	YRDYTFSSLLK DATLIHQSV TGLYETRIDLAKLGGC	(SEQ ID NO:4)
LI	KR~YNNLKELLNSTI IYQSI TGLYESRKRLD----D	(SEQ ID NO:5)
	* * . . : : * : * : * : * : * : * : *	
Motif:	-R-Y-----*~**T*I*QS*TGLYE*R--L-----	(SEQ ID NO:14)

Fig. 3A

Alignment of the N terminal RuvC-like Domains disclosed in Chylinski et al.
(excluding sequence outliers).
(CLUSTAL format alignment by MAFFT (v7.058b))

1,12	DIGTNSVGVAVT	(SEQ ID NO:120)
3,20	DVGTVSVGVAVT	(SEQ ID NO:121)
15	DMGTNSVGVAVT	(SEQ ID NO:122)
4	DVGTSSVGVAVT	(SEQ ID NO:123)
7	DIGTASVGVAVT	(SEQ ID NO:52)
6	DVGTCVSVGVAVT	(SEQ ID NO:53)
9	DIGTNSVGVAVV	(SEQ ID NO:54)
10	DIGTNSVGVAVI	(SEQ ID NO:55)
11	DIGTNSVGVAVL	(SEQ ID NO:56)
42	DLGTNSIGWAVV	(SEQ ID NO:57)
48	DLGTNSIGWAI-	(SEQ ID NO:58)
43	DLGTNSIGWALV	(SEQ ID NO:59)
2	DIGTNSVGVAVT	(SEQ ID NO:60)
14	DIGTNSVGVAVT	(SEQ ID NO:61)
5	DMGTGSLGVAVT	(SEQ ID NO:62)
16	DIGTSSVGVAAI	(SEQ ID NO:63)
8	DLGTCSVGVAVV	(SEQ ID NO:64)
23	DLGVGSVGVAVV	(SEQ ID NO:65)
23	DLGIASIGWAI I	(SEQ ID NO:66)
24	DLGIASVGVAVV	(SEQ ID NO:67)
25	DLGVASVGVWIV	(SEQ ID NO:68)
26	DIGIASVGVWIV	(SEQ ID NO:69)
28	DLGISSVGVWIV	(SEQ ID NO:70)
32	DIGIASVGVWIV	(SEQ ID NO:71)
33	DVGIGSIGWAVI	(SEQ ID NO:72)
39	DLGVGSIGFAIV	(SEQ ID NO:73)
34	DIGYASIGWAVI	(SEQ ID NO:74)
47	DTGTNSLGVAVV	(SEQ ID NO:75)
50	DLGTNSIGWCLL	(SEQ ID NO:76)
49	DIGTDSLGVAVF	(SEQ ID NO:77)
18	DIGSNSIGFAVV	(SEQ ID NO:78)
41	DLGVGSIGVAVA	(SEQ ID NO:79)
45	DLGIASCGWGVV	(SEQ ID NO:80)

Fig. 3B

21	DLGIASVGWCLT	(SEQ ID NO:81)
27	DIGIGSVGVGIL	(SEQ ID NO:82)
29	DIGITSVGYGLI	(SEQ ID NO:83)
30	DIGITSVGFCII	(SEQ ID NO:84)
31	DVGITSTGYAVL	(SEQ ID NO:85)
40	DLGITSFGYAIL	(SEQ ID NO:86)
17	DIGNASVGWSAF	(SEQ ID NO:87)
19	DVGITNSCGWVAM	(SEQ ID NO:88)
35	DVGERSIGLAAV	(SEQ ID NO:89)
36	DVGLNSVGLAAV	(SEQ ID NO:90)
37	DVGLMSVGLAAI	(SEQ ID NO:91)
38	DVGTF SVGLAAI	(SEQ ID NO:92)
13	DIGTG SVGYACM	(SEQ ID NO:92)
44	DLGTT SIGFAHI	(SEQ ID NO:94)
46	DLGTN SIGSSVR	(SEQ ID NO:95)
	* * * *	

Fig. 4A

Alignment of the N terminal RuvC-like Domains disclosed in Chylinski et al.
(CLUSTAL format alignment by MAFFT (v7.058b))

1,12	D---IGTNSVGWAVT	(SEQ ID NO:120)
3,20	D---VGTNSVGWAVT	(SEQ ID NO:121)
15	D---MGTNSVGWAVT	(SEQ ID NO:122)
4	D---VGTSSVGWAVT	(SEQ ID NO:123)
7	D---IGTASVGWAVT	(SEQ ID NO:52)
6	D---VGTGSVGWAVT	(SEQ ID NO:53)
9	D---IGTNSVGWAVV	(SEQ ID NO:54)
10	D---IGTNSVGWAVI	(SEQ ID NO:55)
52	D---IGTNSIGWAVI	(SEQ ID NO:96)
11	D---IGTNSVGWAVL	(SEQ ID NO:56)
42	D---LGTNSIGWAVV	(SEQ ID NO:57)
48	D---LGTNSIGWAI-	(SEQ ID NO:58)
43	D---LGTNSIGWALV	(SEQ ID NO:59)
2	D---IGTNSVGWCVT	(SEQ ID NO:60)
14	D---IGTNSVGWAVT	(SEQ ID NO:61)
5	D---MGTGSLGWAVT	(SEQ ID NO:62)
16	D---IGTSSVGWAAI	(SEQ ID NO:63)
8	D---LGTGSVGWAVV	(SEQ ID NO:64)
22	D---LGVGSVGWAVV	(SEQ ID NO:65)
23	D---LGIASIGWAI I	(SEQ ID NO:66)
24	D---LGIASVGWAVV	(SEQ ID NO:67)
68	D---LGIASVGWAVV	(SEQ ID NO:97)
25	D---LGVASVGWAI V	(SEQ ID NO:68)
26	D---IGIASVGWAIL	(SEQ ID NO:69)
66	D---IGIASVGWAVL	(SEQ ID NO:98)
59	D---IGIASIGWAVI	(SEQ ID NO:99)
61	D---IGIASVGWAI I	(SEQ ID NO:100)
64	D---VGIASVGWAVI	(SEQ ID NO:101)
62	D---IGIASVGWAL-	(SEQ ID NO:102)
67	D---IGIASVGWAMV	(SEQ ID NO:103)
32	D---IGIASVGWSVI	(SEQ ID NO:71)
28	D---LGISSVGWSVI	(SEQ ID NO:70)
63	D---IGITSVGWAVI	(SEQ ID NO:104)

Fig. 4B

33	D----VGIGSIGWAVI	(SEQ ID NO:72)
57	D----LGISSLGWAIV	(SEQ ID NO:105)
39	D----LGVGSIGFAIV	(SEQ ID NO:73)
34	D----IGYASIGWAVI	(SEQ ID NO:74)
50	D----LGTNSIGWCLL	(SEQ ID NO:76)
54	D----LGTNSIGWGLL	(SEQ ID NO:106)
47	D----TGTNSLGWAIV	(SEQ ID NO:75)
49	D----IGTDSLGVAVF	(SEQ ID NO:77)
51	D----LGSTSLGWAIF	(SEQ ID NO:107)
58	D----IGISSIGWAFS	(SEQ ID NO:108)
21	D----LGIASVGVWCLT	(SEQ ID NO:81)
45	D----LGIASCGWGVV	(SEQ ID NO:80)
18	D----IGSNSIGFAVV	(SEQ ID NO:78)
65	D----IGTTSIGFVSI	(SEQ ID NO:109)
29	D----IGITSVGYGLI	(SEQ ID NO:83)
30	D----IGITSVGFGII	(SEQ ID NO:84)
44	D----LGTTSIGFAHI	(SEQ ID NO:94)
27	D----IGICSVGVGIL	(SEQ ID NO:82)
41	D----LGVGSIGVAVA	(SEQ ID NO:79)
31	D----VGITSTGYAVL	(SEQ ID NO:85)
40	D----LGITSPGYAIL	(SEQ ID NO:86)
53	D----IGTSSIGWWLY	(SEQ ID NO:110)
55	D----LGSNSLGVFVT	(SEQ ID NO:111)
56	D----LGANSLGVFVV	(SEQ ID NO:112)
17	D----IGNASVGWSAF	(SEQ ID NO:87)
19	D----VGTNSCGWVAM	(SEQ ID NO:88)
35	D----VGERSIGLAAV	(SEQ ID NO:89)
36	D----VGLNSVGLAAV	(SEQ ID NO:90)
37	D----VGLMSVGLAAI	(SEQ ID NO:91)
38	D----VGTFSVGLAAI	(SEQ ID NO:92)
13	D----IGTGSVGYACM	(SEQ ID NO:93)
46	D----LGTNSIGSSVR	(SEQ ID NO:95)
60	DIGLRIGITSCGWSI-	(SEQ ID NO:113)
69	D----MGAKYTG VFYA	(SEQ ID NO:114)
73	D----LGGKNTGFFSF	(SEQ ID NO:115)
74	D----LGVKNTGVFSA	(SEQ ID NO:116)
70	D----LGAKPTGVALY	(SEQ ID NO:117)
71	D----LGGKFTGVCLS	(SEQ ID NO:118)
72	D----LGGTYTGTFIT	(SEQ ID NO:119)

Fig. 5A

Alignment of the HNH-like Domains disclosed in Chylinski et al.
(CLUSTAL format alignment by MAPPT (v7.058b))

1	YDIDHIYPRS-LTKD-----DSF-DNLVLCERTAN	(SEQ ID NO:196)
2	-DIDHIYPRSKVIKD-----DSF-DNLVLVLKNE	(SEQ ID NO:197)
3	-DRDHIYPQS-KIKD-----DSI-DNLVLVNKTYN	(SEQ ID NO:198)
4	-DIDHIYPRS-KIKD-----DSI-TNRVLVEKDIN	(SEQ ID NO:195)
6	-DIDHIYPQS-KIKD-----DSI-SNRVLVCSSCN	(SEQ ID NO:124)
5	-DIDHIYPQS-KTMD-----DSL-NNRVLVKKKNYN	(SEQ ID NO:125)
7	-DQDHIYPKS-KIYD-----DSL-ENRVLVKKNLN	(SEQ ID NO:126)
8	-QIDHIVPQS-LVKD-----DSF-DNRVLVVPSEN	(SEQ ID NO:127)
9	-DIDHIIPQA-PIKD-----NSI-DNRVLVTSSKEN	(SEQ ID NO:128)
12	-DIDHIIPQA-FLKD-----NSI-DNKVLVSSASN	(SEQ ID NO:129)
16	-DIDHIIPQA-YTKD-----NSL-DNRVLVSNITN	(SEQ ID NO:130)
11	-DIDHIVPQS-FITD-----NSI-DNLVLVTSSAGN	(SEQ ID NO:131)
10	-DVDHIVPQS-FLKD-----DSI-DNKVLTRSDKN	(SEQ ID NO:132)
14	-NIDHIYPQS-MVKD-----DSL-DNKVLVQSEIN	(SEQ ID NO:133)
18	-DIDHILPQS-LIKD-----DSL-DNRVLVNATIN	(SEQ ID NO:134)
19	-DIDHILPQS-PIKD-----DSL-ENRVLVKKAVN	(SEQ ID NO:135)
13	-EVDHIIPRS-PIKD-----DSI-DNKVLVIKKMN	(SEQ ID NO:136)
15	-EVDHIIPRS-YIKD-----DSF-ENKVLVYREEN	(SEQ ID NO:137)
17	-DIDHIIPQA-VTON-----DSI-DNRVLVARAEN	(SEQ ID NO:138)
22	-EIDHIIPYS-ISFD-----DSS-SNKLLVLAESN	(SEQ ID NO:139)
24	-EIDHIIPYS-LCFD-----DSS-ANKVLVHKQSN	(SEQ ID NO:140)
32	-DIDHIIPYS-RSMD-----DSY-SNKVLVLSCEN	(SEQ ID NO:141)
63	-DIDHIIPYS-KSMD-----DSF-NNKVLCLABEN	(SEQ ID NO:142)
59	-EIDHIIPYS-RSFD-----DSY-MNKVLVFTKQN	(SEQ ID NO:143)
65	-QIDHIIPYS-RSMD-----DSY-MNKVLVLTDEN	(SEQ ID NO:144)
64	-EIDHIIPFS-RSFD-----DSL-SNKILVLGSEN	(SEQ ID NO:145)
68	-EIDHALPFS-RTWD-----DSF-NNKVLVLGSEN	(SEQ ID NO:146)
69	-EIDHALPFS-RTWD-----DSF-NNKVLVLASEN	(SEQ ID NO:147)
28	-EIDHIIPIS-ISLD-----DSI-NNKVLVLSEAN	(SEQ ID NO:148)
30	-EVDHIIPIS-ISLD-----DSI-TNKVLVTHREN	(SEQ ID NO:149)
62	-QVDHALPYS-RSYD-----DSK-NNKVLVLTHEN	(SEQ ID NO:150)
27	-EVDHILPLS-ITFD-----DSL-ANKVLVYATAN	(SEQ ID NO:151)
26	-EIDHIIPRS-ISFD-----DAR-SNKVLVYRSEN	(SEQ ID NO:152)

Fig. 5B

29	-EVDHIIPRS-VSFD-----NSY-HNKVLVKQSEN	(SEQ ID NO:153)
67	-DIDHILPYS-ITFD-----DSF-HNKVLVTSQEN	(SEQ ID NO:154)
58	-EIDHILPRS-RSAD-----DSF-ANKVLCLARAN	(SEQ ID NO:155)
51	-EIEHLLPFS-LTLD-----DSM-ANKFVCFRQAN	(SEQ ID NO:156)
55	-DIDHILPFS-VSLD-----DSA-ANKVVCLREAN	(SEQ ID NO:157)
57	-DIDHLIPFS-ISWD-----DSA-ANKVVCMBRYAN	(SEQ ID NO:158)
56	-DIDHILPVA-MTLD-----DSP-ANKIICMBRYAN	(SEQ ID NO:159)
54	-DVDHILPYS-RTLD-----DSP-PNRTLCCLREAN	(SEQ ID NO:160)
52	-EIEHILPFS-RTLD-----DSL-NNRTVAMRRAN	(SEQ ID NO:161)
31	-EVDHIIPYS-ISWD-----DSY-TNKVLTSACKN	(SEQ ID NO:162)
45	-QVDHILPWS-RFGD-----DSY-LNKTLCCTARSN	(SEQ ID NO:163)
53	-QVDHILPFS-KTLD-----DSF-ANKVLAQHGAN	(SEQ ID NO:164)
60	-QIDHAFPLS-RSLD-----DSQ-SNKVLCCLTSSN	(SEQ ID NO:165)
21	-DIDHIVPRS-ISFD-----DSF-SNLVIVNKLGN	(SEQ ID NO:166)
23	-EIEHIIPYS-MSYD-----NSQ-ANKILTEKAEN	(SEQ ID NO:167)
25	-RIDHVIPYS-KSAD-----DSW-FNKLLVKKSTN	(SEQ ID NO:168)
49	-EMDHILPYS-RSLD-----NGW-HNRVLVHGKGN	(SEQ ID NO:169)
33	-EVDHIVPYS-LILD-----NTI-NNKALVYAEEN	(SEQ ID NO:170)
42	-EIEHVIPOQS-LYFD-----DSF-SNKVICEAEVN	(SEQ ID NO:171)
43	-DIEHIIPQA-RLFD-----DSF-SNKTLEARSVN	(SEQ ID NO:172)
44	-EIEHIVPKA-RVFD-----DSF-SNKTLTFRHRIN	(SEQ ID NO:173)
20	-DKDHIIPOQS-MKKD-----DSI-INNLVLVYNKNAN	(SEQ ID NO:174)
66	-EVEHIWPRS-RSFD-----NSP-RNKTLCCKDGN	(SEQ ID NO:175)
61	-IVNHIIPYN-RSFD-----DTY-HNRVLTLTETK	(SEQ ID NO:176)
46	-DMEHTIPKS-ISFD-----NSD-QNLTLCESYIN	(SEQ ID NO:177)
47	-DIEHTIPRS-AGGD-----STK-MNLTLCSSRFN	(SEQ ID NO:178)
48	-DIEHTIPRS-ISGD-----NSQ-MNKTLCCLKFN	(SEQ ID NO:179)
50	-DIDHVIPLA-RGGR-----DSL-DNMVLCQSDAN	(SEQ ID NO:180)
39	-DIEHLFPIA-ESED-----NGR-NNLVISHSACN	(SEQ ID NO:181)
41	-DVDHIFPRD-DTAD-----NSY-GNKVVAHRQCN	(SEQ ID NO:182)
40	-DIEHIVPQS-LGGL-----STD-YNTIVTLKSVN	(SEQ ID NO:183)
35	-ELDHIVPRT-DGGS-----NRH-ENLAITCGACN	(SEQ ID NO:184)
36	-EMDHIVPRKGVGST-----NTR-TNFAAVCABCN	(SEQ ID NO:185)
37	-EMDHIVPRKGVGST-----NTR-VNLAAACAACN	(SEQ ID NO:186)
38	-EMDHIVPRAGQGST-----NTR-ENLVAVCHRCN	(SEQ ID NO:187)
70	-EIDHILPRS-LIKIARGIVFNAE-PNLIYASSRGN	(SEQ ID NO:188)
71	-EIDHIIPRS-LTGRTKTVFNSE-ANLIYCCKGN	(SEQ ID NO:189)
73	-EIDHIIPRS-LTLKKSESIYNSE-VNLIFVSAQGN	(SEQ ID NO:190)

Fig. 5C

72	-EIDHIYPRS-LSKKHFGVIFNSE-VNLIYCSSLQGN	(SEQ ID NO:191)
74	-EIDHILPRS-HTLKITYGTVFNDE-GNLIYVHQKCN	(SEQ ID NO:192)
75	-ELDHIIPRS-HKKY---GTLNDE-ANLICVTRGDN	(SEQ ID NO:193)
34	-ELEHIVPHS-FRQS-----NAL-SSLVLTWPGVN	(SEQ ID NO:194)
	1 * * . . 5	

Fig. 6A

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:196)
2      -DIDHIYPRSKVIKDDS-FDNLVLVLKNE      (SEQ ID NO:197)
3      -DEDHIYPQS-KIKDDS-IDNLVLVNKTYN      (SEQ ID NO:198)
4      -DIDHIYPRS-KIKDDS-ITNRVLVEKDIN      (SEQ ID NO:195)
5      -DIDHIYPQS-KIKDDS-IGNRVLCSSCN      (SEQ ID NO:124)
6      -DIDHIYPQS-KTMDDS-LNNRVLVKKNYN      (SEQ ID NO:125)
7      -DQDHIYPKS-KIYDDS-LENRVLVKKNLN      (SEQ ID NO:126)
8      -QIDHIVPQS-LVKDDS-FDNRVLVVPSEN      (SEQ ID NO:127)
9      -DIDHIIPQA-FIKDMS-IDNRLVLTSSKEN      (SEQ ID NO:128)
10     -DIDHIIPQA-FLKDNS-IDNKVLVSSASN      (SEQ ID NO:129)
11     -DIDHIIPQA-YTKDNS-LDNRVLVSNITN      (SEQ ID NO:130)
12     -DIDHIVPQS-FITDNS-IDNLVLTSSAGN      (SEQ ID NO:131)
13     -DVDHIVPQS-FLKDDS-IDNKVLTRSDKN      (SEQ ID NO:132)
14     -NIDHIYPQS-MVKDDS-LENKVLVQSEIN      (SEQ ID NO:133)
15     -DIDHILPQS-LIKDDS-LDNRVLVNATIN      (SEQ ID NO:134)
16     -DIDHILPQS-FIKDDS-LENRVLVKKAVN      (SEQ ID NO:135)
17     -EVDHIFPRS-FIKDDS-IDNKVLVKKQMN      (SEQ ID NO:136)
18     -EVDHIIPRS-YIKDDS-FENKVLVYREEN      (SEQ ID NO:137)
19     -DIDHIIPQA-VTQDNS-IDNRVLVARAEN      (SEQ ID NO:138)
20     -DIDHIVPRS-ISFDDS-FSNLVIYNKLEN      (SEQ ID NO:166)
21     -EIDHIIPYS-ISFDDS-SSNKLVLVLAESN      (SEQ ID NO:139)
22     -EIDHIIPYS-LCFDDS-SANKVLVHKQSN      (SEQ ID NO:140)
23     -EIDHIIPIS-ISLDDS-INNKVLVLSEAN      (SEQ ID NO:148)
24     -EVDHIIPIS-ISLDDS-ITNKVLVTHREN      (SEQ ID NO:149)
25     -EVDHILPLS-ITFDDS-LANKVLVYATAN      (SEQ ID NO:151)
26     -EIDHIIPRS-ISFDDA-RSNKVLVYRSEN      (SEQ ID NO:152)
27     -EVDHIIPRS-VSEFDS-YHNKVLVQSEN      (SEQ ID NO:153)
28     -EVDHIIPYS-ISWDDS-YTNKVLTSACKN      (SEQ ID NO:162)
29     -DIDHIIPYS-RSMDDS-YSNKVLVLSCEN      (SEQ ID NO:141)
30     -EIEHIIPYS-MSYDNS-QANKILTEKAEN      (SEQ ID NO:167)
31     -EVDHIVPYS-LILDNT-INNKALVYAEEN      (SEQ ID NO:170)
32     -EIDHVIPYS-KSADDS-WFNKLLVKKSTN      (SEQ ID NO:168)
33     -EMDHILPYS-RELENG-WHNRVLVHGKDN      (SEQ ID NO:169)
34     -EIEHVIPQS-LYFDDS-FSNKVICEAEVN      (SEQ ID NO:171)
35     -DIEHIIPQA-RLFDDS-FSNKTLARSVN      (SEQ ID NO:172)

```

Fig. 6B

44	-EIEHIVPKA-RVFDDS-PSNKTTLTFHRIN	(SEQ ID NO:173)
20	-DKDHIIPQS-MKKDDSIINNVLVLVNKNAN	(SEQ ID NO:174)
45	-QVDHILPWS-RPGDDS-YLNKTLC TARSN	(SEQ ID NO:163)
50	-DIDHVIPLA-RGGRDS-LDNMVLCQSDAN	(SEQ ID NO:180)
46	-DMEHTIPKS-ISFDNS-DQNLTLCESYYN	(SEQ ID NO:177)
47	-DIEHTIPRS-ACGDST-KMNLTLCSSRFN	(SEQ ID NO:178)
48	-DIEHTIPRS-ISQDNS-QMNKTLC SLKFN	(SEQ ID NO:179)
39	-DIEHLFPIA-ESEDNG-RNNLVISHSACN	(SEQ ID NO:181)
41	-DVDHIFPRD-DTADNS-YCNKVVAHRQCN	(SEQ ID NO:182)
40	-DIEHIVPQS-LGGLST-DYNTIVTLKSVN	(SEQ ID NO:183)
35	-ELDHIVPRT-DGGSNR-HENLAITCGACN	(SEQ ID NO:184)
36	-EMDHIVPRKGVGSTNT-RTNFAAVCAECN	(SEQ ID NO:185)
37	-EMDHIVPRKGVGSTNT-RVNLAACAACN	(SEQ ID NO:186)
38	-EMDHIVPRAGQGSTNT-RENLVAVCHRCN	(SEQ ID NO:187)
34	-ELEHIVPHS-FRQSNA-LSSLVLTWPCVN	(SEQ ID NO:194)

Fig. 7

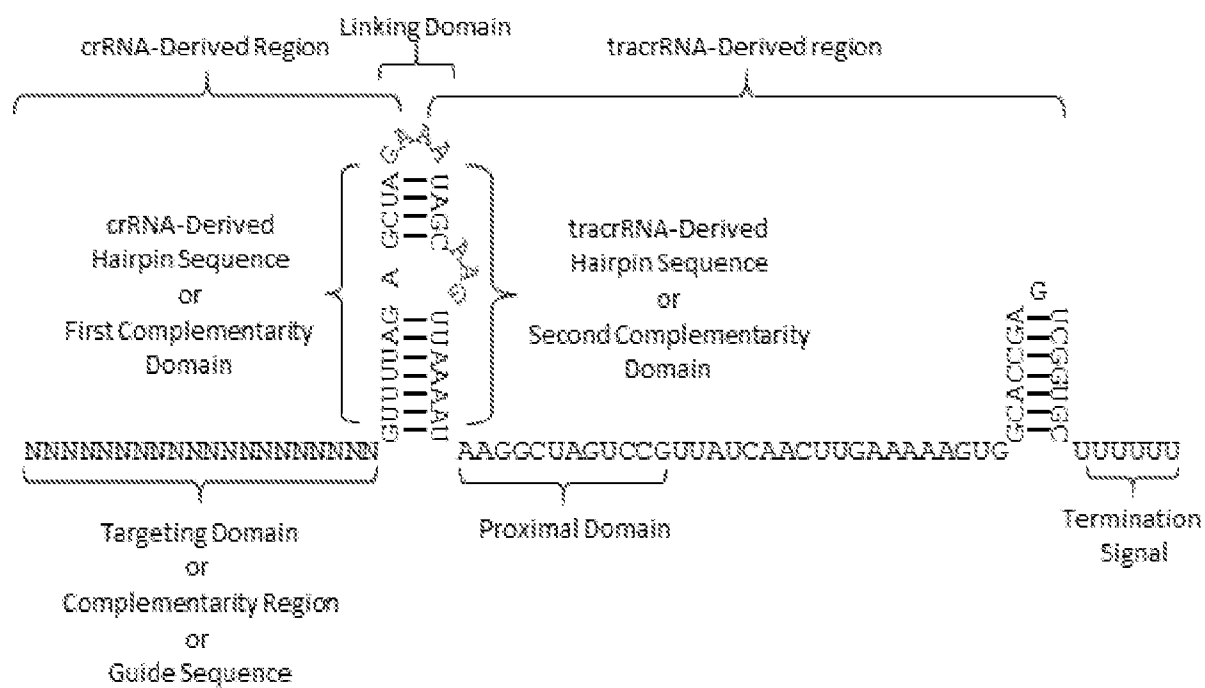


Fig. 8A

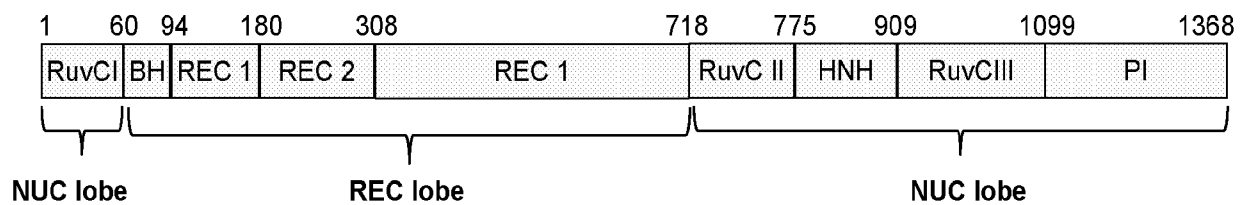


Fig. 8B

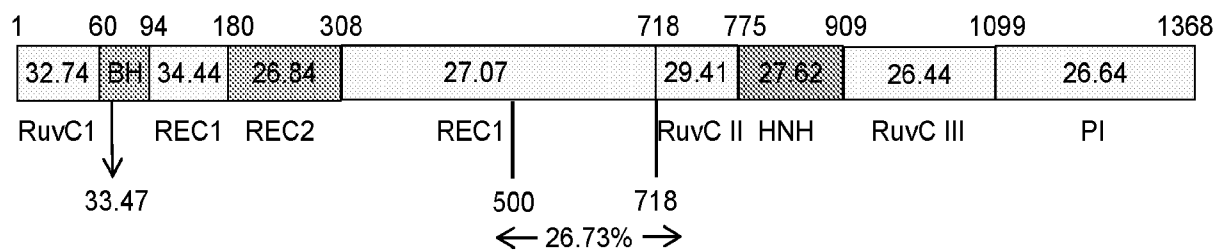


Fig. 9

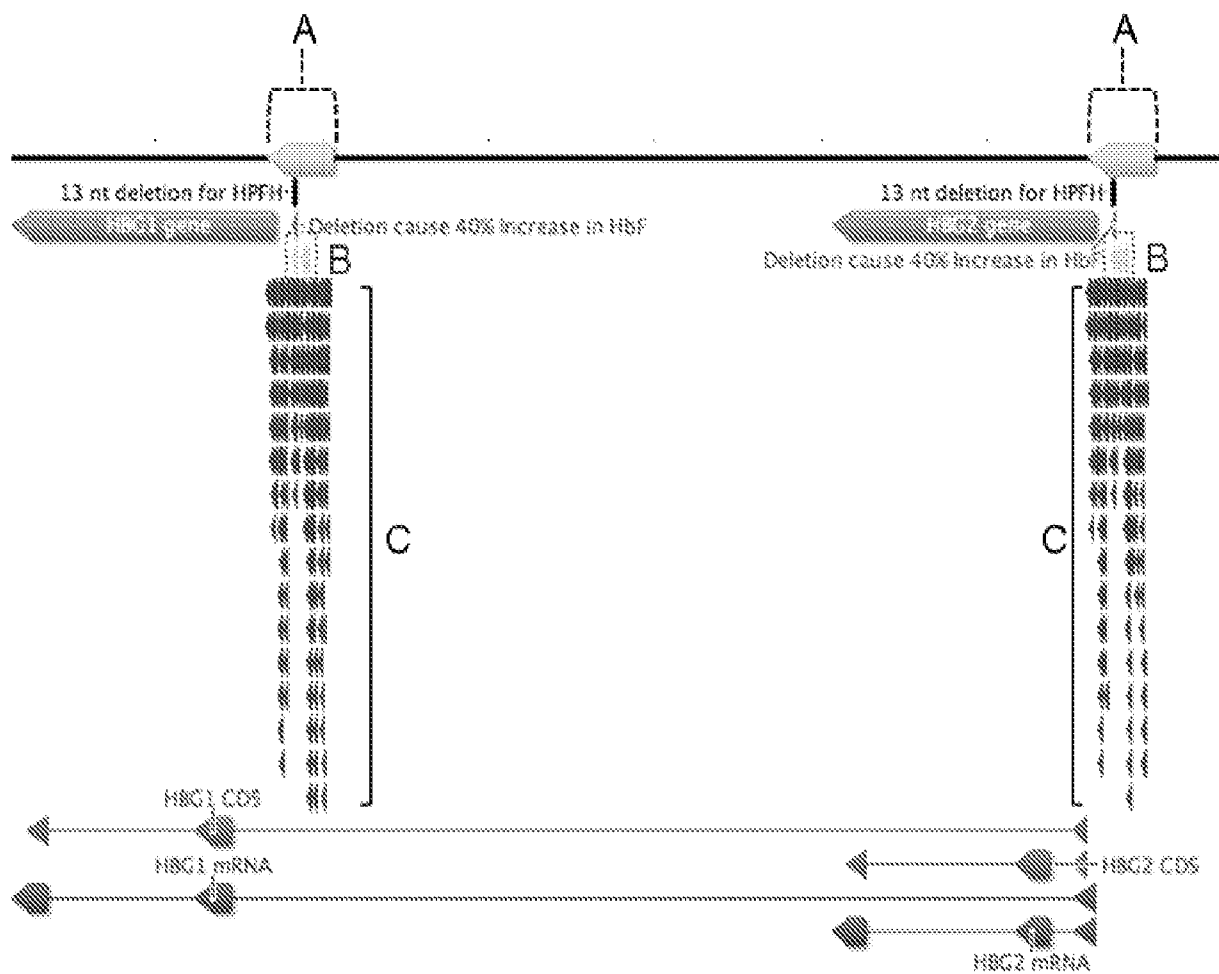


Fig. 10A

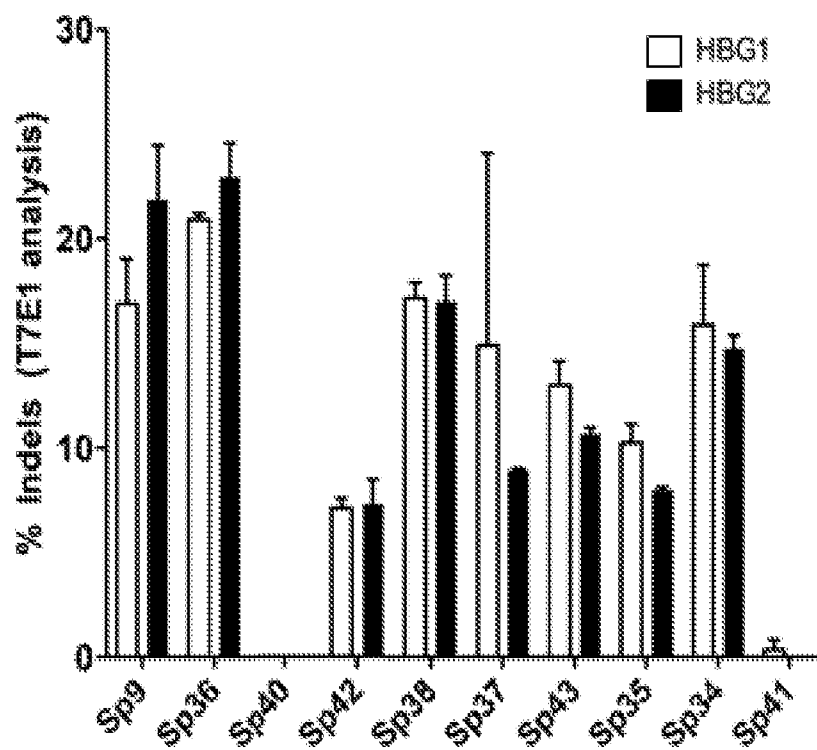


Fig. 10B

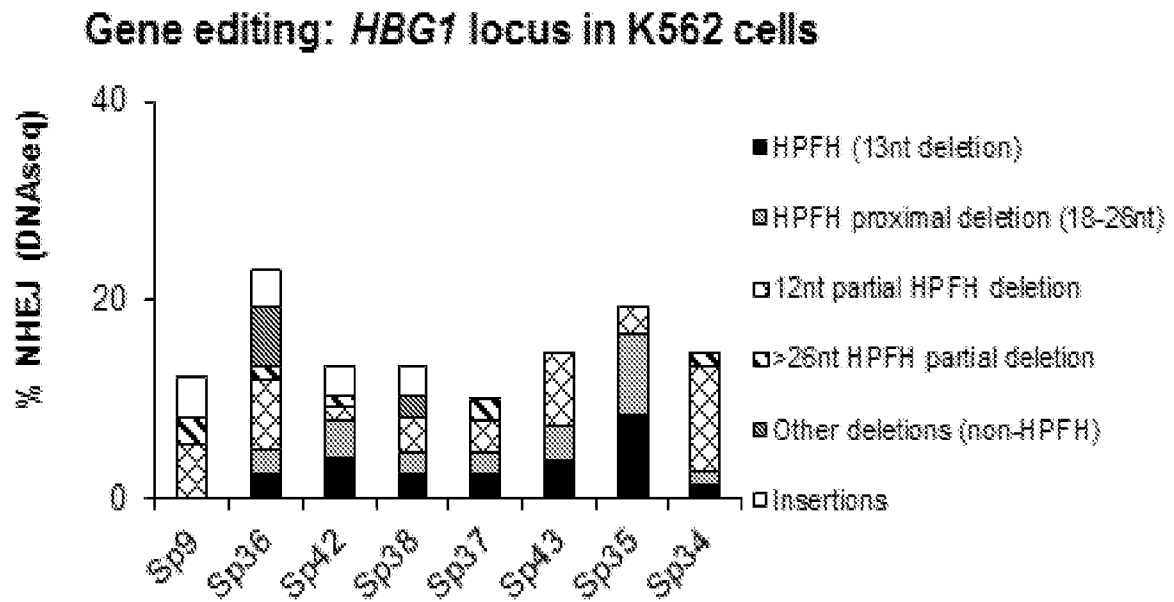


Fig. 10C

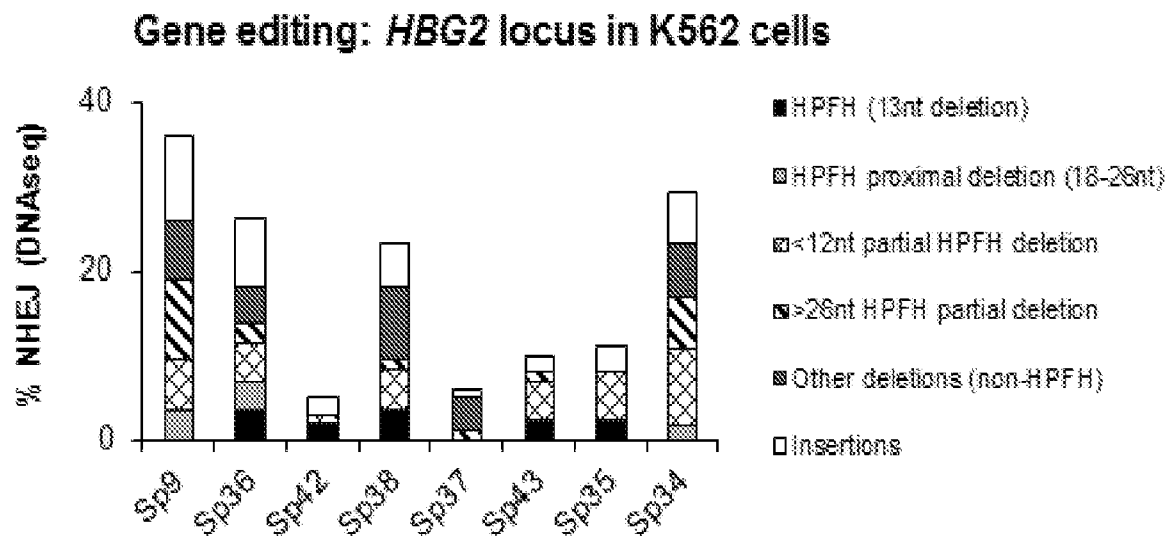


Fig. 10D

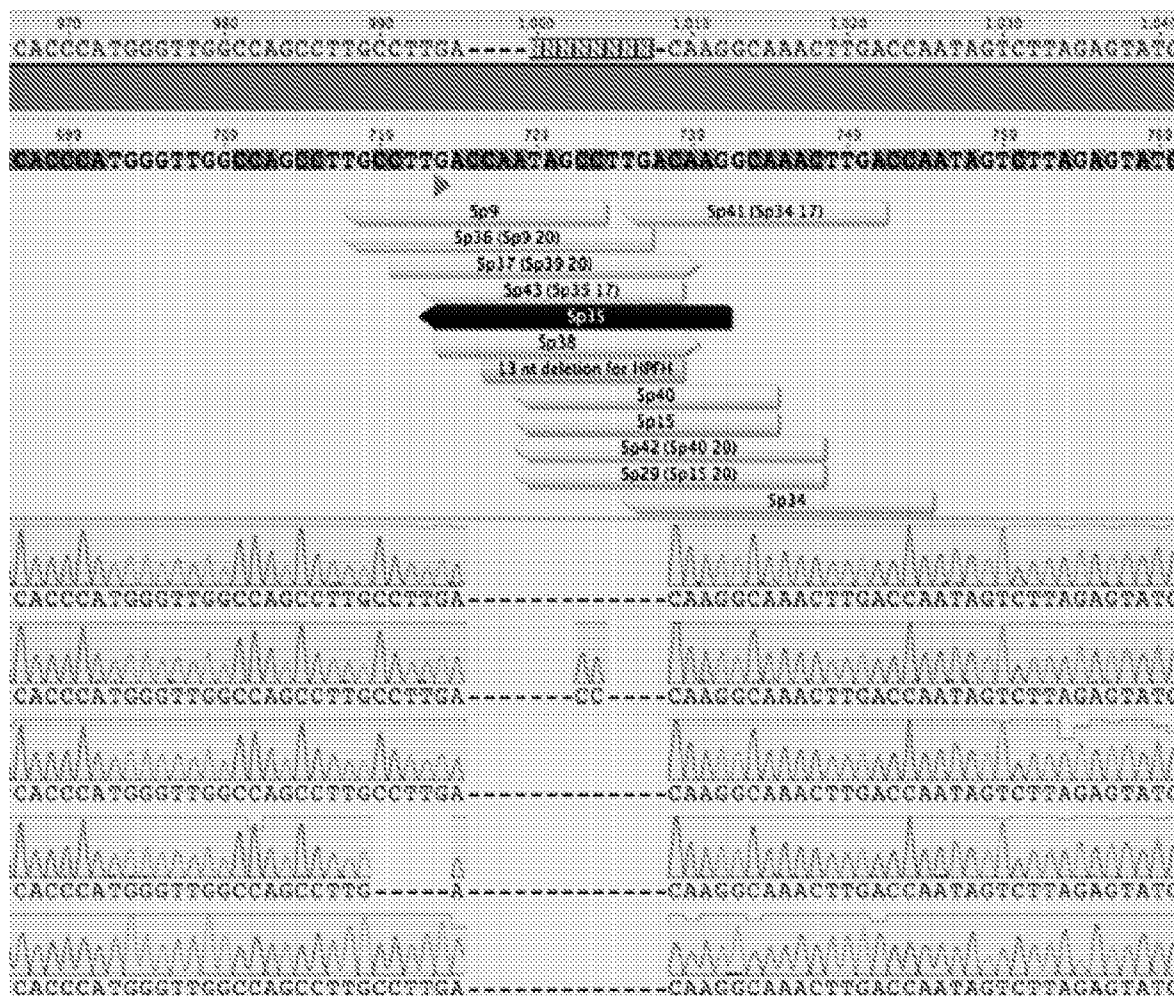


Fig. 10E

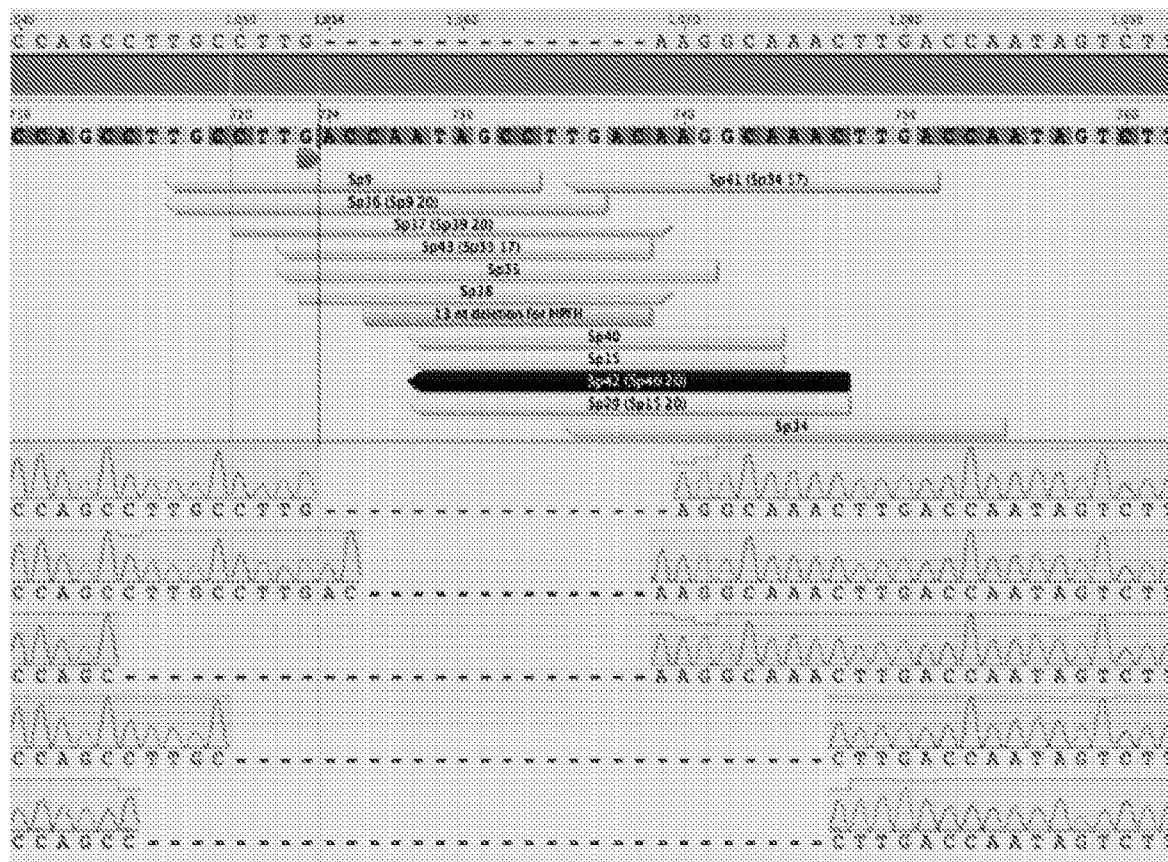


Fig. 10F

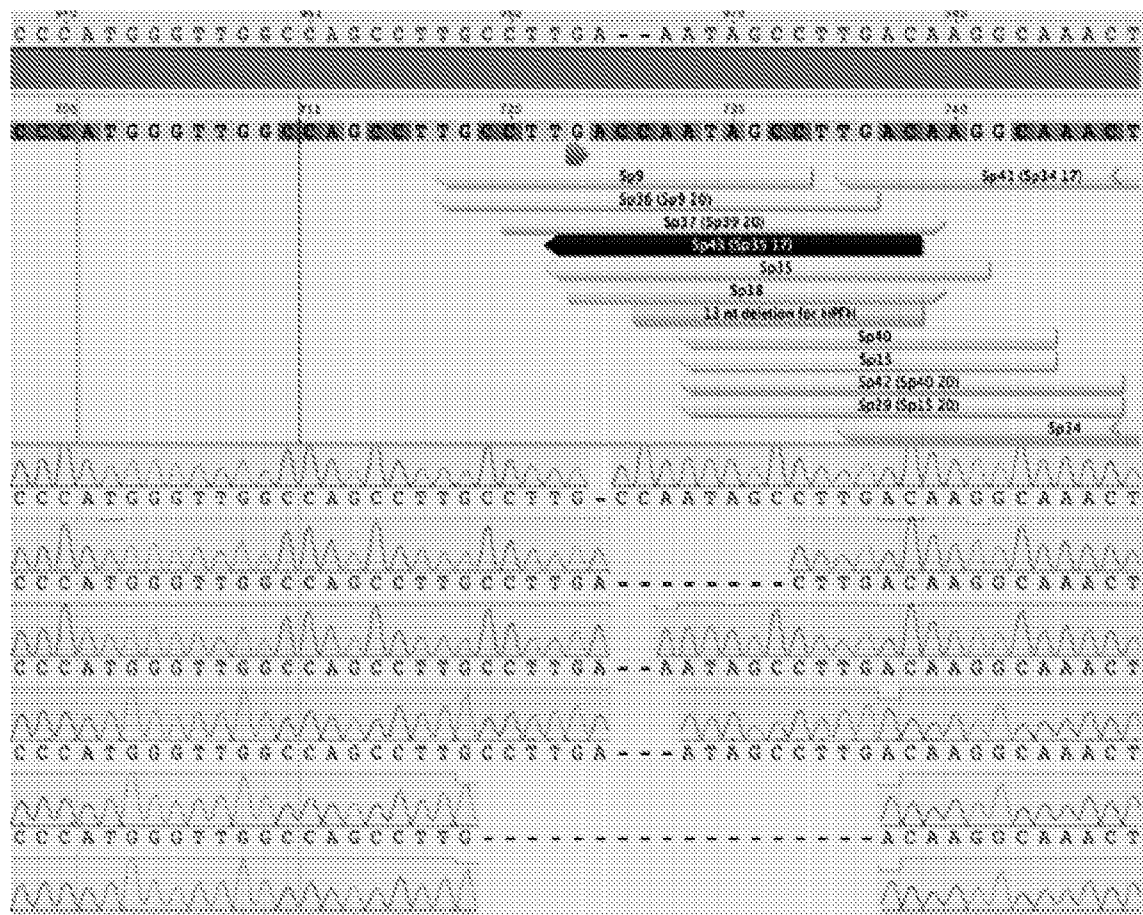


Fig. 11A

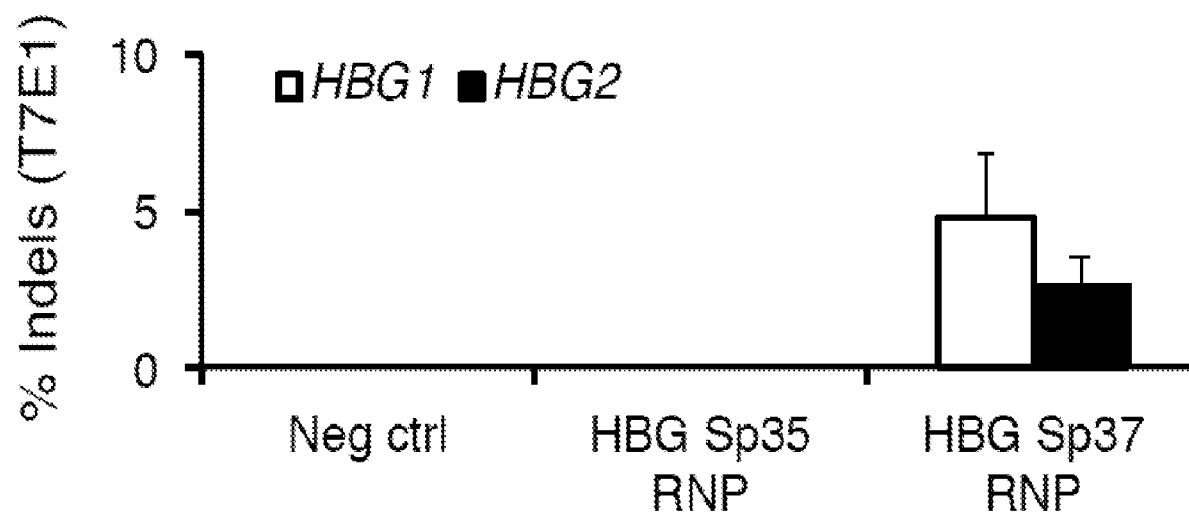
CB CD34⁺

Fig. 11B

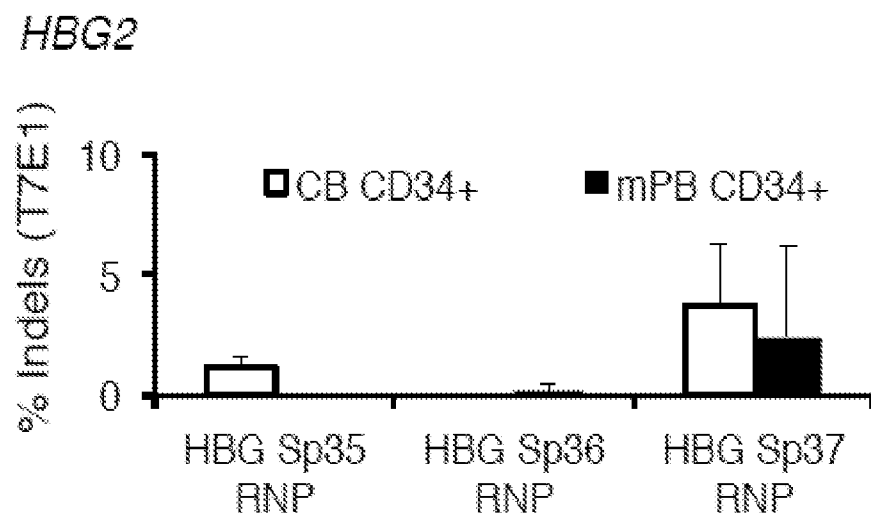


Fig. 11C

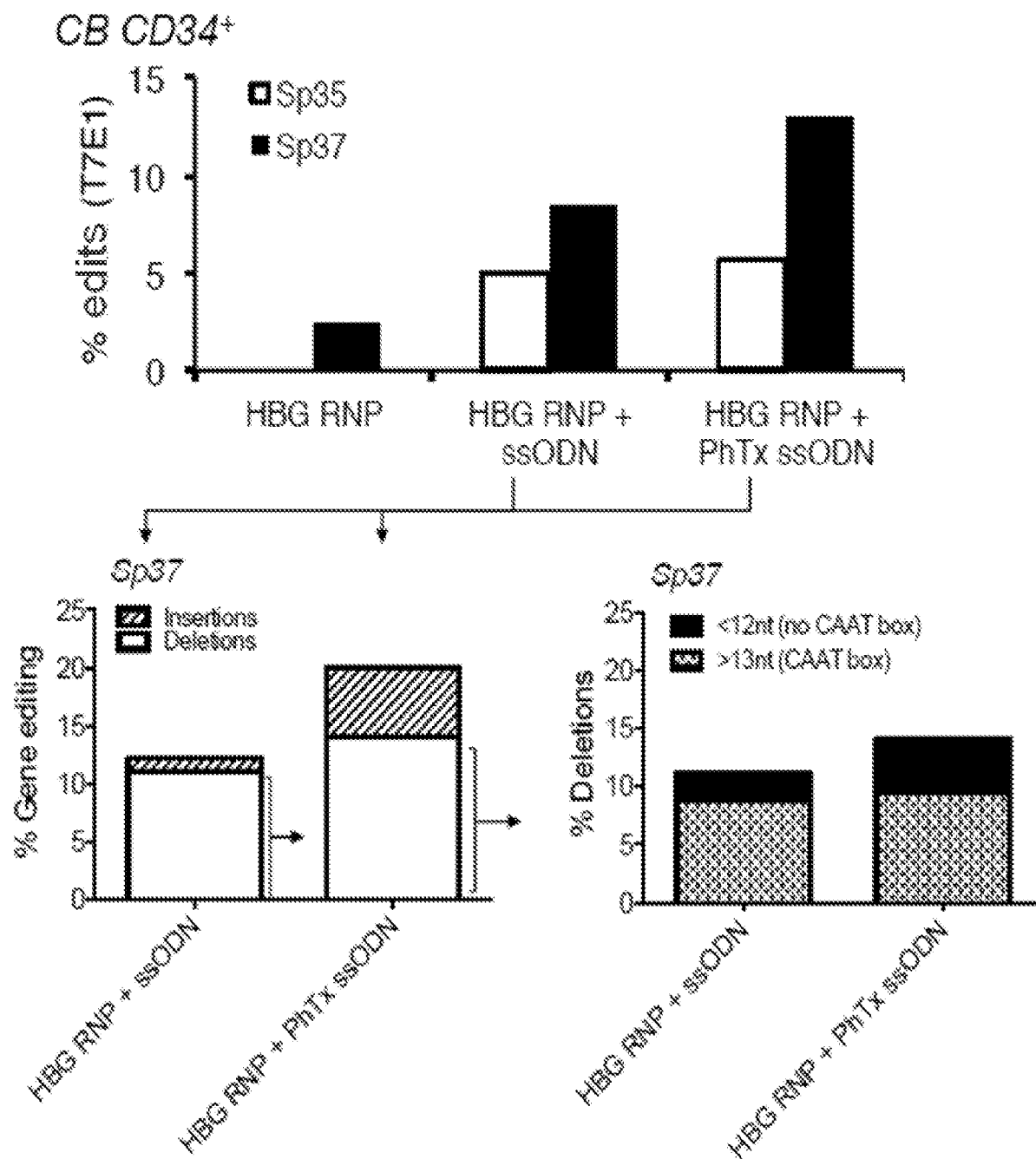


Fig. 12A

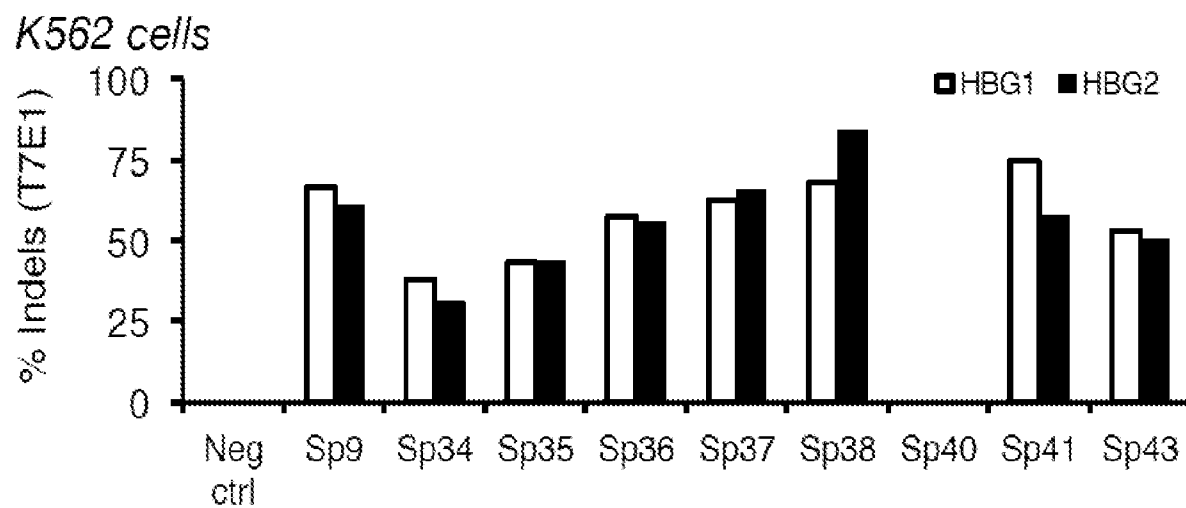


Fig. 12B

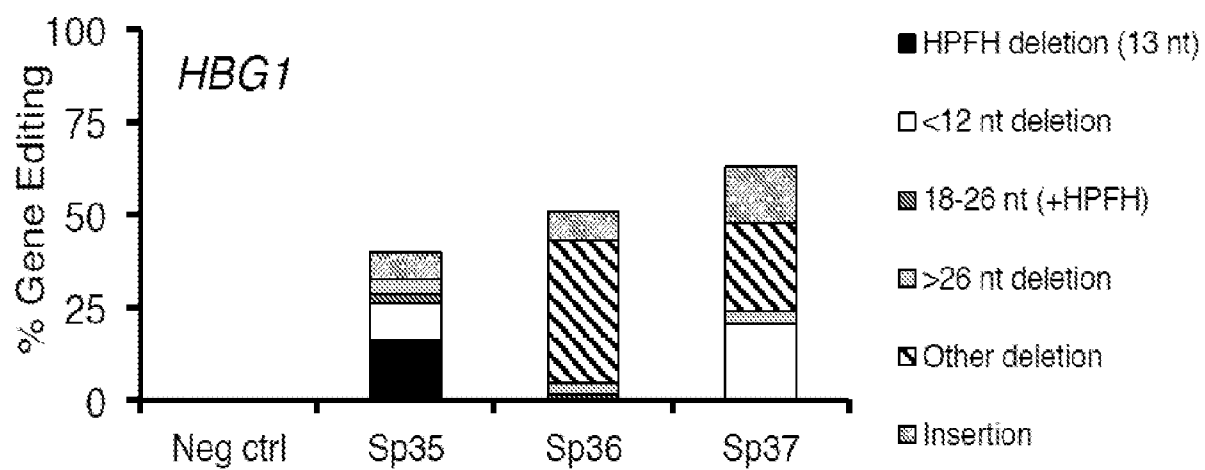


Fig. 12C

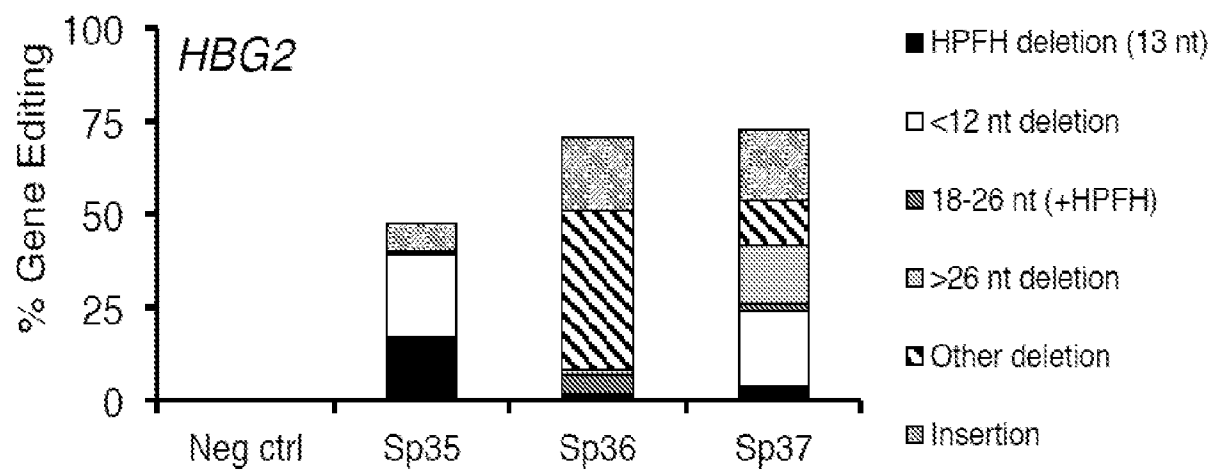


Fig. 13A

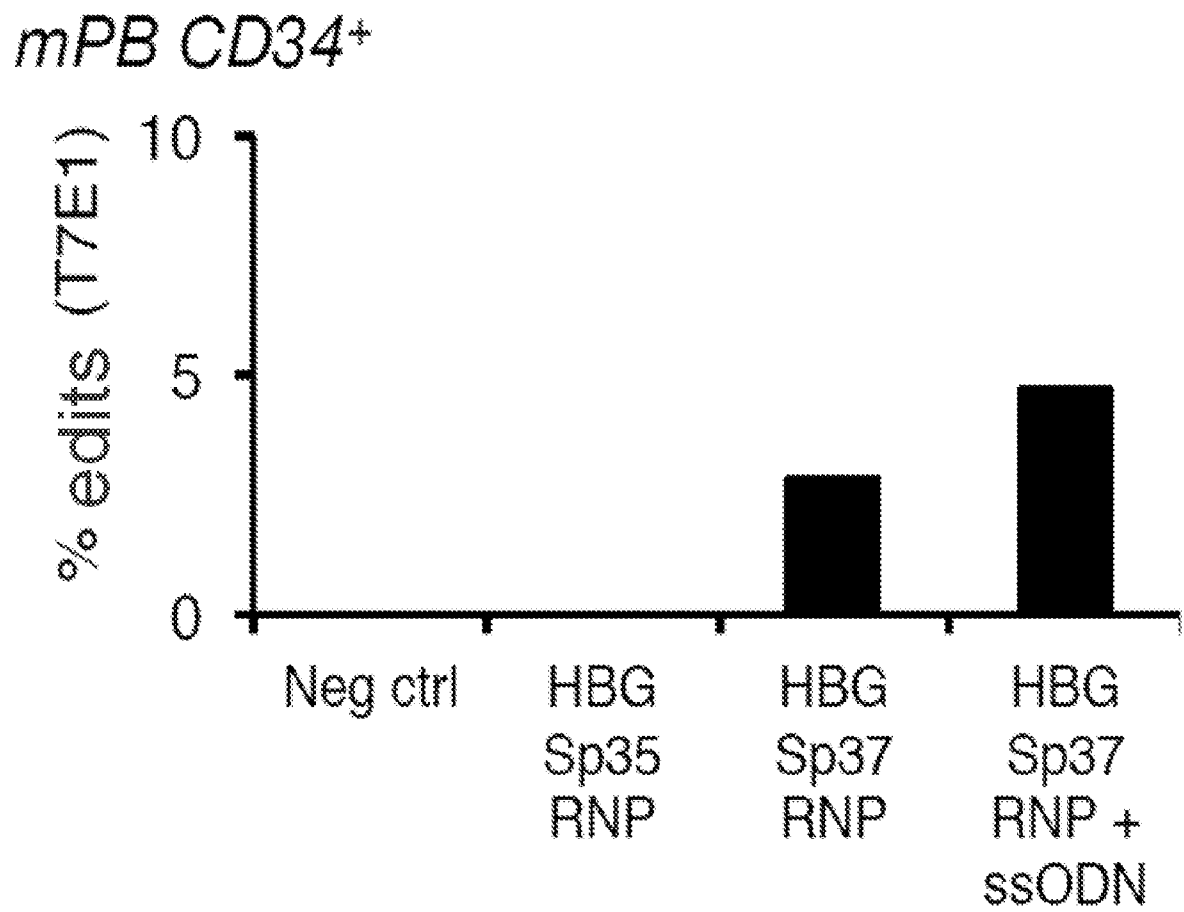


Fig. 13B

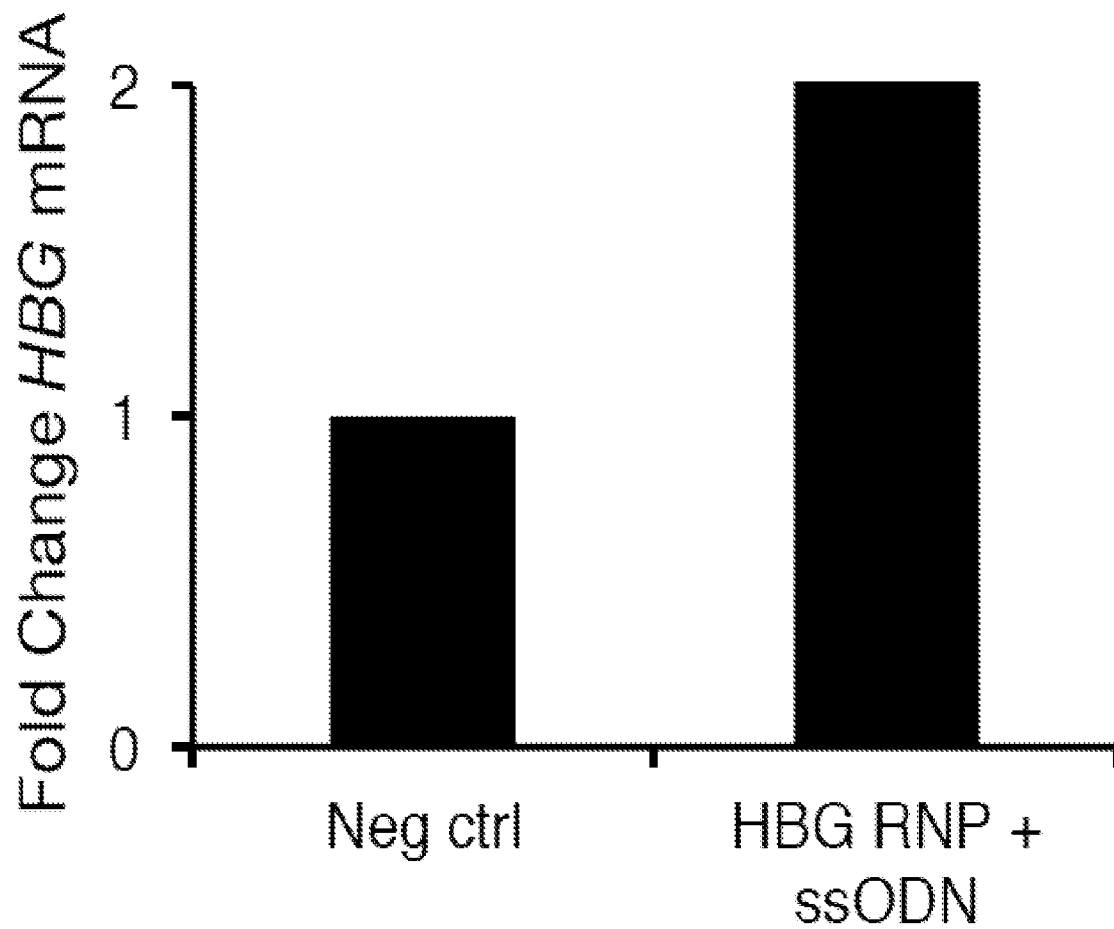
mPB CD34⁺ erythroblasts

Fig. 14A

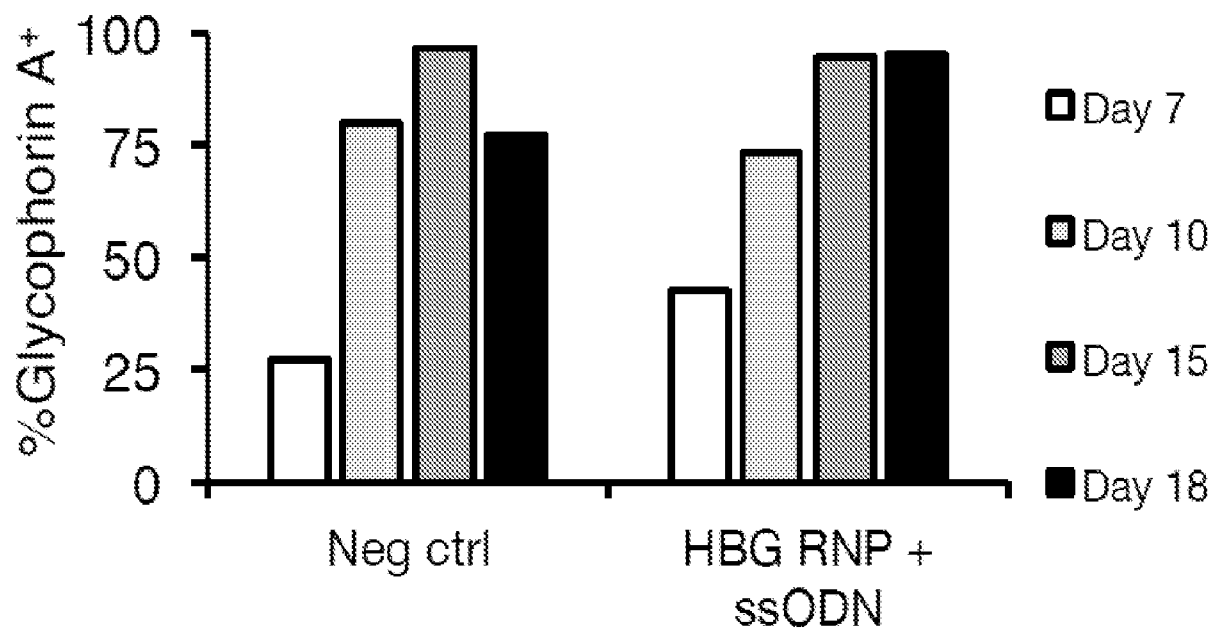
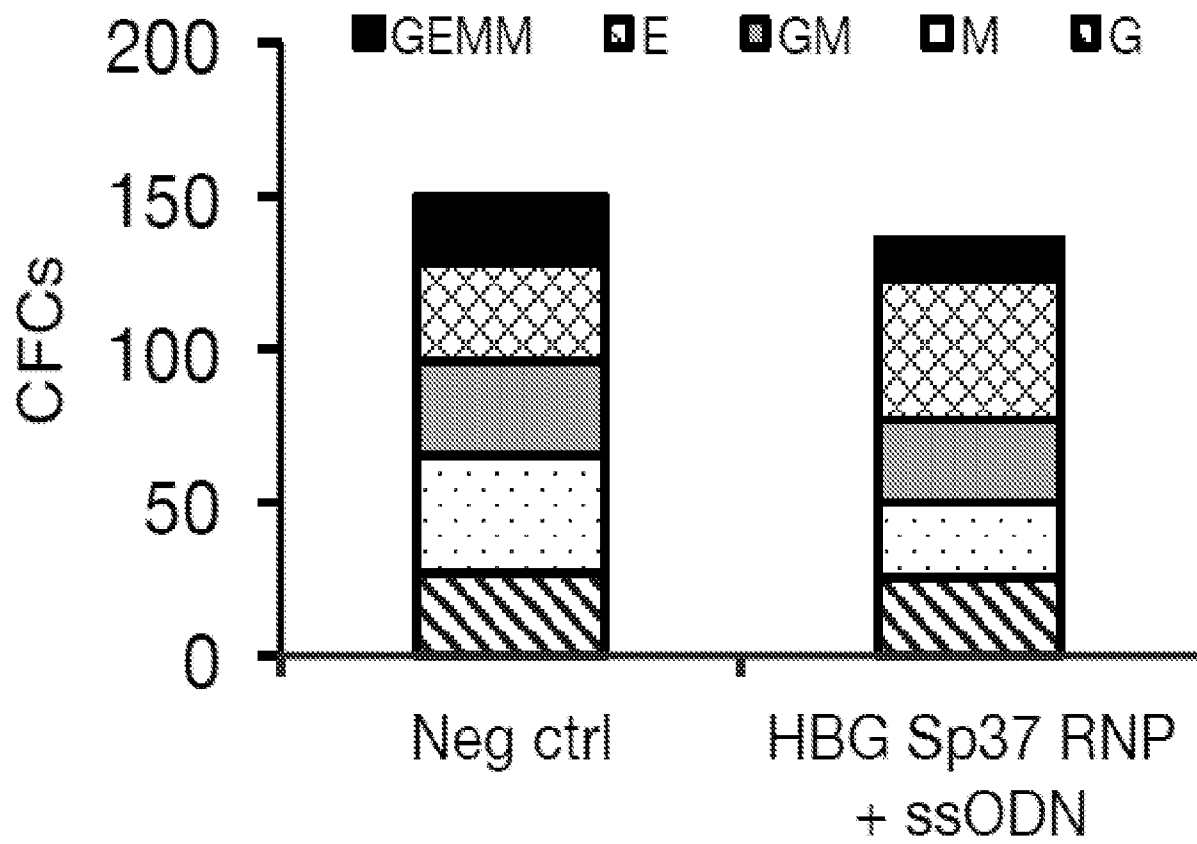


Fig. 14B



8009W000_SequenceListing.txt
SEQUENCE LISTING

<110> Editas Medicine
Gori, Jennifer L
Barrera, Luis A.

<120> CRISPR/Cas-Related Methods and Compositions for Treating Beta Hemoglobinopathies

<130> 118945.8009.W000 (EM076PCT)

<150> US 62/308,190
<151> 2016-03-14

<150> US 62/456,615
<151> 2017-02-08

<160> 939

<170> PatentIn version 3.5

<210> 1
<211> 1345
<212> PRT
<213> Streptococcus mutans

<220>
<221> misc_feature
<222> (10)..(21)
<223> N-terminal RuvC-like domain

<220>
<221> misc_feature
<222> (759)..(766)
<223> RuvC-like domain

<220>
<221> misc_feature
<222> (837)..(863)
<223> HNH-like domain

<220>
<221> misc_feature
<222> (982)..(989)
<223> RuvC-like domain

<400> 1

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

Gly Trp Ala Val Val Thr Asp Asp Tyr Lys Val Pro Ala Lys Lys Met
20 25 30

Lys Val Leu Gly Asn Thr Asp Lys Ser His Ile Glu Lys Asn Leu Leu
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Asn Thr Ala Glu Asp Arg Arg Leu
50 55 60

8009W000_SequenceListing.txt

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu
 65 70 75 80
 Tyr Leu Gln Glu Ile Phe Ser Glu Glu Met Gly Lys Val Asp Asp Ser
 85 90 95
 Phe Phe His Arg Leu Glu Asp Ser Phe Leu Val Thr Glu Asp Lys Arg
 100 105 110
 Gly Glu Arg His Pro Ile Phe Gly Asn Leu Glu Glu Glu Val Lys Tyr
 115 120 125
 His Glu Asn Phe Pro Thr Ile Tyr His Leu Arg Gln Tyr Leu Ala Asp
 130 135 140
 Asn Pro Glu Lys Val Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His
 145 150 155 160
 Ile Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Lys Phe Asp Thr
 165 170 175
 Arg Asn Asn Asp Val Gln Arg Leu Phe Gln Glu Phe Leu Ala Val Tyr
 180 185 190
 Asp Asn Thr Phe Glu Asn Ser Ser Leu Gln Glu Gln Asn Val Gln Val
 195 200 205
 Glu Glu Ile Leu Thr Asp Lys Ile Ser Lys Ser Ala Lys Lys Asp Arg
 210 215 220
 Val Leu Lys Leu Phe Pro Asn Glu Lys Ser Asn Gly Arg Phe Ala Glu
 225 230 235 240
 Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Lys Lys His Phe
 245 250 255
 Glu Leu Glu Glu Lys Ala Pro Leu Gln Phe Ser Lys Asp Thr Tyr Glu
 260 265 270
 Glu Glu Leu Glu Val Leu Leu Ala Gln Ile Gly Asp Asn Tyr Ala Glu
 275 280 285
 Leu Phe Leu Ser Ala Lys Lys Leu Tyr Asp Ser Ile Leu Leu Ser Gly
 290 295 300
 Ile Leu Thr Val Thr Asp Val Gly Thr Lys Ala Pro Leu Ser Ala Ser

8009W000_SequenceListing.txt

```

305          310          315          320
Met Ile Gln Arg Tyr Asn Glu His Gln Met Asp Leu Ala Gln Leu Lys
      325      330
Gln Phe Ile Arg Gln Lys Leu Ser Asp Lys Tyr Asn Glu Val Phe Ser
      340      345      350
Asp Val Ser Lys Asp Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn
      355      360      365
Gln Glu Ala Phe Tyr Lys Tyr Leu Lys Gly Leu Leu Asn Lys Ile Glu
      370      375      380
Gly Ser Gly Tyr Phe Leu Asp Lys Ile Glu Arg Glu Asp Phe Leu Arg
      385      390      395      400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
      405      410      415
Gln Glu Met Arg Ala Ile Ile Arg Arg Gln Ala Glu Phe Tyr Pro Phe
      420      425      430
Leu Ala Asp Asn Gln Asp Arg Ile Glu Lys Leu Leu Thr Phe Arg Ile
      435      440      445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Lys Ser Asp Phe Ala Trp
      450      455      460
Leu Ser Arg Lys Ser Ala Asp Lys Ile Thr Pro Trp Asn Phe Asp Glu
      465      470      475      480
Ile Val Asp Lys Glu Ser Ser Ala Glu Ala Phe Ile Asn Arg Met Thr
      485      490      495
Asn Tyr Asp Leu Tyr Leu Pro Asn Gln Lys Val Leu Pro Lys His Ser
      500      505      510
Leu Leu Tyr Glu Lys Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
      515      520      525
Tyr Lys Thr Glu Gln Gly Lys Thr Ala Phe Phe Asp Ala Asn Met Lys
      530      535      540
Gln Glu Ile Phe Asp Gly Val Phe Lys Val Tyr Arg Lys Val Thr Lys
      545      550      555      560

```

8009W000_SequenceListing.txt

Asp Lys Leu Met Asp Phe Leu Glu Lys Glu Phe Asp Glu Phe Arg Ile
 565 570 575
 Val Asp Leu Thr Gly Leu Asp Lys Glu Asn Lys Val Phe Asn Ala Ser
 580 585 590
 Tyr Gly Thr Tyr His Asp Leu Cys Lys Ile Leu Asp Lys Asp Phe Leu
 595 600 605
 Asp Asn Ser Lys Asn Glu Lys Ile Leu Glu Asp Ile Val Leu Thr Leu
 610 615 620
 Thr Leu Phe Glu Asp Arg Glu Met Ile Arg Lys Arg Leu Glu Asn Tyr
 625 630 635 640
 Ser Asp Leu Leu Thr Lys Glu Gln Val Lys Lys Leu Glu Arg Arg His
 645 650 655
 Tyr Thr Gly Trp Gly Arg Leu Ser Ala Glu Leu Ile His Gly Ile Arg
 660 665 670
 Asn Lys Glu Ser Arg Lys Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly
 675 680 685
 Asn Ser Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp Ala Leu Ser
 690 695 700
 Phe Lys Glu Glu Ile Ala Lys Ala Gln Val Ile Gly Glu Thr Asp Asn
 705 710 715 720
 Leu Asn Gln Val Val Ser Asp Ile Ala Gly Ser Pro Ala Ile Lys Lys
 725 730 735
 Gly Ile Leu Gln Ser Leu Lys Ile Val Asp Glu Leu Val Lys Ile Met
 740 745 750
 Gly His Gln Pro Glu Asn Ile Val Val Glu Met Ala Arg Glu Asn Gln
 755 760 765
 Phe Thr Asn Gln Gly Arg Arg Asn Ser Gln Gln Arg Leu Lys Gly Leu
 770 775 780
 Thr Asp Ser Ile Lys Glu Phe Gly Ser Gln Ile Leu Lys Glu His Pro
 785 790 795 800
 Val Glu Asn Ser Gln Leu Gln Asn Asp Arg Leu Phe Leu Tyr Tyr Leu
 805 810 815

8009W000_SequenceListing.txt

Gln Asn Gly Arg Asp Met Tyr Thr Gly Glu Glu Leu Asp Ile Asp Tyr
820 825 830

Leu Ser Gln Tyr Asp Ile Asp His Ile Ile Pro Gln Ala Phe Ile Lys
835 840 845

Asp Asn Ser Ile Asp Asn Arg Val Leu Thr Ser Ser Lys Glu Asn Arg
850 855 860

Gly Lys Ser Asp Asp Val Pro Ser Lys Asp Val Val Arg Lys Met Lys
865 870 875 880

Ser Tyr Trp Ser Lys Leu Leu Ser Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Thr Asp Asp Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Arg Ile Leu Asp Glu Arg Phe Asn Thr Glu Thr Asp
930 935 940

Glu Asn Asn Lys Lys Ile Arg Gln Val Lys Ile Val Thr Leu Lys Ser
945 950 955 960

Asn Leu Val Ser Asn Phe Arg Lys Glu Phe Glu Leu Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asp Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Ile Gly Lys Ala Leu Leu Gly Val Tyr Pro Gln Leu Glu Pro Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Pro His Phe His Gly His Lys Glu Asn Lys
1010 1015 1020

Ala Thr Ala Lys Lys Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe
1025 1030 1035

Lys Lys Asp Asp Val Arg Thr Asp Lys Asn Gly Glu Ile Ile Trp
1040 1045 1050

Lys Lys Asp Glu His Ile Ser Asn Ile Lys Lys Val Leu Ser Tyr
1055 1060 1065

8009W000_SequenceListing.txt

Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Val	Glu	Glu	Gln	Thr	Gly	Gly
	1070					1075					1080			
Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	Lys	Gly	Asn	Ser	Asp	Lys	Leu
	1085					1090					1095			
Ile	Pro	Arg	Lys	Thr	Lys	Lys	Phe	Tyr	Trp	Asp	Thr	Lys	Lys	Tyr
	1100					1105					1110			
Gly	Gly	Phe	Asp	Ser	Pro	Ile	Val	Ala	Tyr	Ser	Ile	Leu	Val	Ile
	1115					1120					1125			
Ala	Asp	Ile	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	Lys	Thr	Val	Lys
	1130					1135					1140			
Ala	Leu	Val	Gly	Val	Thr	Ile	Met	Glu	Lys	Met	Thr	Phe	Glu	Arg
	1145					1150					1155			
Asp	Pro	Val	Ala	Phe	Leu	Glu	Arg	Lys	Gly	Tyr	Arg	Asn	Val	Gln
	1160					1165					1170			
Glu	Glu	Asn	Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser	Leu	Phe	Lys	Leu
	1175					1180					1185			
Glu	Asn	Gly	Arg	Lys	Arg	Leu	Leu	Ala	Ser	Ala	Arg	Glu	Leu	Gln
	1190					1195					1200			
Lys	Gly	Asn	Glu	Ile	Val	Leu	Pro	Asn	His	Leu	Gly	Thr	Leu	Leu
	1205					1210					1215			
Tyr	His	Ala	Lys	Asn	Ile	His	Lys	Val	Asp	Glu	Pro	Lys	His	Leu
	1220					1225					1230			
Asp	Tyr	Val	Asp	Lys	His	Lys	Asp	Glu	Phe	Lys	Glu	Leu	Leu	Asp
	1235					1240					1245			
Val	Val	Ser	Asn	Phe	Ser	Lys	Lys	Tyr	Thr	Leu	Ala	Glu	Gly	Asn
	1250					1255					1260			
Leu	Glu	Lys	Ile	Lys	Glu	Leu	Tyr	Ala	Gln	Asn	Asn	Gly	Glu	Asp
	1265					1270					1275			
Leu	Lys	Glu	Leu	Ala	Ser	Ser	Phe	Ile	Asn	Leu	Leu	Thr	Phe	Thr
	1280					1285					1290			
Ala	Ile	Gly	Ala	Pro	Ala	Thr	Phe	Lys	Phe	Phe	Asp	Lys	Asn	Ile

8009W000_SequenceListing.txt

1295

1300

1305

Asp Arg Lys Arg Tyr Thr Ser Thr Thr Glu Ile Leu Asn Ala Thr
 1310 1315 1320

Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
 1325 1330 1335

Leu Asn Lys Leu Gly Gly Asp
 1340 1345

<210> 2
 <211> 1368
 <212> PRT
 <213> Streptococcus pyogenes

<220>
 <221> misc_feature
 <222> (10)..(21)
 <223> N-terminal RuvC-like domain

<220>
 <221> misc_feature
 <222> (759)..(766)
 <223> RuvC-like domain

<220>
 <221> misc_feature
 <222> (837)..(863)
 <223> HNH-like domain

<220>
 <221> misc_feature
 <222> (982)..(989)
 <223> RuvC-like domain

<400> 2

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
 1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
 20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
 35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
 50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
 65 70 75 80

8009W000_SequenceListing.txt

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
 85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
 100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
 115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
 130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
 145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
 165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
 180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
 195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
 210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
 225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
 260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
 275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
 290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
 305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 325 330 335

8009W000_SequenceListing.txt

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

8009W000_SequenceListing.txt

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

8009W000_SequenceListing.txt

835 840 845
 Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
 850 855 860
 Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
 865 870 875 880
 Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
 885 890 895
 Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
 900 905 910
 Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
 915 920 925
 Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 930 935 940
 Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
 945 950 955 960
 Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
 965 970 975
 Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
 980 985 990
 Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
 995 1000 1005
 Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
 1010 1015 1020
 Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
 1025 1030 1035
 Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
 1040 1045 1050
 Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
 1055 1060 1065
 Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
 1070 1075 1080

8009W000_SequenceListing.txt

Arg	Lys	Val	Leu	Ser	Met	Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Thr
	1085					1090					1095			
Glu	Val	Gln	Thr	Gly	Gly	Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	Lys
	1100					1105					1110			
Arg	Asn	Ser	Asp	Lys	Leu	Ile	Ala	Arg	Lys	Lys	Asp	Trp	Asp	Pro
	1115					1120					1125			
Lys	Lys	Tyr	Gly	Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala	Tyr	Ser	Val
	1130					1135					1140			
Leu	Val	Val	Ala	Lys	Val	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	Lys
	1145					1150					1155			
Ser	Val	Lys	Glu	Leu	Leu	Gly	Ile	Thr	Ile	Met	Glu	Arg	Ser	Ser
	1160					1165					1170			
Phe	Glu	Lys	Asn	Pro	Ile	Asp	Phe	Leu	Glu	Ala	Lys	Gly	Tyr	Lys
	1175					1180					1185			
Glu	Val	Lys	Lys	Asp	Leu	Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser	Leu
	1190					1195					1200			
Phe	Glu	Leu	Glu	Asn	Gly	Arg	Lys	Arg	Met	Leu	Ala	Ser	Ala	Gly
	1205					1210					1215			
Glu	Leu	Gln	Lys	Gly	Asn	Glu	Leu	Ala	Leu	Pro	Ser	Lys	Tyr	Val
	1220					1225					1230			
Asn	Phe	Leu	Tyr	Leu	Ala	Ser	His	Tyr	Glu	Lys	Leu	Lys	Gly	Ser
	1235					1240					1245			
Pro	Glu	Asp	Asn	Glu	Gln	Lys	Gln	Leu	Phe	Val	Glu	Gln	His	Lys
	1250					1255					1260			
His	Tyr	Leu	Asp	Glu	Ile	Ile	Glu	Gln	Ile	Ser	Glu	Phe	Ser	Lys
	1265					1270					1275			
Arg	Val	Ile	Leu	Ala	Asp	Ala	Asn	Leu	Asp	Lys	Val	Leu	Ser	Ala
	1280					1285					1290			
Tyr	Asn	Lys	His	Arg	Asp	Lys	Pro	Ile	Arg	Glu	Gln	Ala	Glu	Asn
	1295					1300					1305			
Ile	Ile	His	Leu	Phe	Thr	Leu	Thr	Asn	Leu	Gly	Ala	Pro	Ala	Ala
	1310					1315					1320			

8009W000_SequenceListing.txt

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
 1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
 1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
 1355 1360 1365

<210> 3
 <211> 4107
 <212> DNA
 <213> Streptococcus pyogenes

<400> 3
 atggataaaa agtacagcat cgggctggac atcgggtacaa actcagtggg gtgggcccgtg 60
 attacggacg agtacaaggt accctccaaa aaattttaaag tgctgggtaa cacggacaga 120
 cactctataa agaaaaatct tattggagcc ttgctgttcg actcaggcga gacagccgaa 180
 gccacaaggt tgaagcggac cgccaggagg cggatatacca ggagaaagaa ccgcatatgc 240
 tacctgcaag aaatcttcag taacgagatg gcaaagggtg acgatatgctt tttccatcgc 300
 ctggaagaat cctttcttgt tgaggaagac aagaagcacg aacggcaccc catctttggc 360
 aatattgtcg acgaagtggc atatcacgaa aagtaccgga ctatctacca cctcaggaag 420
 aagctggtgg actctaccga taaggcggac ctcagactta tttatttggc actcgcccac 480
 atgattaaat ttagaggaca tttcttgatc gagggcgacc tgaacccgga caacagtgc 540
 gtcgataagc tgttcatcca acttgtgcag acctacaatc aactgttcga agaaaaccct 600
 ataaatgctt caggagtcga cgctaaagca atcctgtccg cgcgccctctc aaaatctaga 660
 agacttgaga atctgattgc tcagttgccc ggggaaaaaga aaaatggatt gtttggcaac 720
 ctgatcgccc tcagtctcgg actgacccca aatttcaaaa gtaacttcga cctggccgaa 780
 gacgctaagc tccagctgtc caaggacaca tacgatgacg acctcgacaa tctgctggcc 840
 cagattgggg atcagtacgc cgatctcttt ttggcagcaa agaacctgtc cgacgccatc 900
 ctgttgagcg atatcttgag agtgaacacc gaaattacta aagcaccct tagcgcattc 960
 atgatcaagc ggtacgacga gcatcatcag gatctgaccc tgctgaaggc tcttgtgagg 1020
 caacagctcc ccgaaaaata caaggaaatc ttctttgacc agagcaaaaaa cggctacgct 1080
 ggctatatag atggtggggc cagtcaggag gaattctata aattcatcaa gccattctc 1140
 gagaaaatgg acggcacaga ggagttgctg gtcaaactta acagggagga cctgctgcgg 1200
 aagcagcgga cctttgacaa cgggtctatc cccaccaga ttcattctggg cgaactgcac 1260
 gcaatcctga ggaggcagga ggatttttat ccttttctta aagataaccg cgagaaaata 1320

8009W000_SequenceListing.txt

gaaaagattc ttacattcag gatcccggtac tacgtggggac ctctcgccccg gggcaattca	1380
cggtttgcct ggatgacaag gaagtcagag gagactatta caccttggaa cttcgaagaa	1440
gtggtggaca aggggtgcatc tgcccagtct ttcacgagc ggatgacaaa ttttgacaag	1500
aacctcccta atgagaaggt gctgccccaaa cattctctgc tctacgagta ctttaccgtc	1560
tacaatgaac tgactaaagt caagtacgtc accgagggaa tgaggaagcc ggcattcctt	1620
agtggagaac agaagaaggc gattgtagac ctgttgttca agaccaacag gaaggtgact	1680
gtgaagcaac ttaaagaaga ctactttaag aagatcgaat gttttgacag tgtggaaatt	1740
tcaggggttg aagaccgctt caatgctgca ttggggactt accatgatct tctcaagatc	1800
ataaaggaca aagacttcct ggacaacgaa gaaaatgagg atattctcga agacatcgtc	1860
ctcaccctga ccctgttcga agacagggaa atgatagaag agcgcttgaa aacctatgcc	1920
cacctcttcg acgataaagt tatgaagcag ctgaagcgca ggagatacac aggatgggga	1980
agattgtcaa ggaagctgat caatggaatt agggataaac agagtggcaa gaccatactg	2040
gatttcctca aatctgatgg cttcgccaat aggaacttca tgcaactgat tcacgatgac	2100
tctcttacct tcaaggagga cattcaaaaag gctcagggtga gcgggcaggg agactccctt	2160
catgaacaca tcgcaaatTTT ggcagggttcc cccgctatta aaaagggcat ccttcaaact	2220
gtcaagggtg tggatgaatt ggtcaaggta atgggcagac ataagccaga aaatattgtg	2280
atcgagatgg cccgcgaaaa ccagaccaca cagaagggcc agaaaaatag tagagagcgg	2340
atgaagagga tcgaggaggg catcaaagag ctgggatctc agattctcaa agaacacccc	2400
gtagaaaaca cacagctgca gaacgaaaaa ttgtacttgt actatctgca gaacggcaga	2460
gacatgtacg tcgaccaaga acttgatatt aatagactgt ccgactatga cgtagaccat	2520
atcgtgcccc agtccttcct gaaggacgac tccattgata acaaagtctt gacaagaagc	2580
gacaagaaca ggggtaaaag tgataatgtg cctagcgagg aggtggtgaa aaaaatgaag	2640
aactactggc gacagctgct taatgcaaag ctcatcac aacggaagtt cgataatctg	2700
acgaaagcag agagaggtgg cttgtctgag ttggacaagg cagggtttat taagcggcag	2760
ctggtggaaa ctaggcagat cacaagcac gtggcgaga ttttggacag ccgatgaac	2820
acaaaatacg acgaaaatga taaactgata cgagaggtca aagttatcac gctgaaaagc	2880
aagctggtgt ccgattttcg gaaagacttc cagttctaca aagttcgca gattaataac	2940
taccatcatg ctcacgatgc gtacctgaac gctgtgtgcg ggaccgcctt gataaagaag	3000
tacccaaagc tggaatccga gttcgtatac ggggattaca aagtgtacga tgtgagggaa	3060
atgatagcca agtccgagca ggagattgga aaggccacag ctaagtactt cttttattct	3120
aacatcatga atttttttaa gacggaaatt accctggcca acggagagat cagaaagcgg	3180
ccccttatag agacaaatgg tgaaacaggt gaaatcgtct gggataaggg cagggatttc	3240

8009W000_SequenceListing.txt

```

gctactgtga ggaaggtgct gagtatgcca caggtaaata tcgtgaaaaa aaccgaagta 3300
cagaccggag gattttccaa ggaaagcatt ttgcctaaaa gaaactcaga caagctcatc 3360
gcccgcgaaga aagattggga ccctaagaaa tacgggggat ttgactcacc caccgtagcc 3420
tattctgtgc tgggtgtagc taaggtggaa aaaggaaagt ctaagaagct gaagtccgtg 3480
aaggaactct tgggaatcac tatcatggaa agatcatcct ttgaaaagaa ccctatcgat 3540
ttcctggagg ctaagggtta caaggaggtc aagaaagacc tcatcattaa actgccaaaa 3600
tactctctct tcgagctgga aaatggcagg aagagaatgt tggccagcgc cggagagctg 3660
caaaagggaa acgagcttgc tctgccctcc aaatatgtta attttctcta tctcgcttcc 3720
cactatgaaa agctgaaagg gtctcccgaa gataacgagc agaagcagct gttcgtcgaa 3780
cagcacaagc actatctgga tgaaataatc gaacaaataa gcgagttcag caaaaggggt 3840
atcctggcgg atgctaattt ggacaaaagta ctgtctgctt ataacaagca ccgggataag 3900
cctattaggg aacaagccga gaatataatt cacctcttta cactcacgaa tctcggagcc 3960
cccgccgcct tcaaatactt tgatacgact atcgaccgga aacggtatac cagtacaaa 4020
gaggtcctcg atgccaccct catccaccag tcaattactg gcctgtacga aacacggatc 4080
gacctctctc aactgggcgg cgactag 4107

```

```

<210> 4
<211> 1388
<212> PRT
<213> Streptococcus thermophilus

```

```

<220>
<221> misc_feature
<222> (10)..(21)
<223> N-terminal RuvC-like domain

```

```

<220>
<221> misc_feature
<222> (760)..(767)
<223> RuvC-like domain

```

```

<220>
<221> misc_feature
<222> (844)..(870)
<223> HNH-like domain

```

```

<220>
<221> misc_feature
<222> (989)..(996)
<223> RuvC-like domain

```

```

<400> 4

```

```

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

```

8009W000_SequenceListing.txt

Gly Trp Ala Val Thr Thr Asp Asn Tyr Lys Val Pro Ser Lys Lys Met
20 25 30

Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile Lys Lys Asn Leu Leu
35 40 45

Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala Glu Gly Arg Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala Thr Leu Asp Asp Ala
85 90 95

Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val Pro Asp Asp Lys Arg
100 105 110

Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val Glu Glu Lys Ala Tyr
115 120 125

His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg Lys Tyr Leu Ala Asp
130 135 140

Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu Gly Glu Phe Asn Ser
165 170 175

Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp Phe Leu Asp Thr Tyr
180 185 190

Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu Asn Ser Lys Gln Leu
195 200 205

Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu Glu Lys Lys Asp Arg
210 215 220

Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser Gly Ile Phe Ser Glu
225 230 235 240

Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Arg Lys Cys Phe
245 250 255

Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser Lys Glu Ser Tyr Asp
260 265 270

8009W000_SequenceListing.txt

Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly Asp Asp Tyr Ser Asp
 275 280 285
 Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala Ile Leu Leu Ser Gly
 290 295 300
 Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala Pro Leu Ser Ser Ala
 305 310 315 320
 Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp Leu Ala Leu Leu Lys
 325 330 335
 Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr Asn Glu Val Phe Lys
 340 345 350
 Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn
 355 360 365
 Gln Glu Asp Phe Tyr Val Tyr Leu Lys Lys Leu Leu Ala Glu Phe Glu
 370 375 380
 Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg Glu Asp Phe Leu Arg
 385 390 395 400
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro Tyr Gln Ile His Leu
 405 410 415
 Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala Lys Phe Tyr Pro Phe
 420 425 430
 Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Asp Phe Ala Trp
 450 455 460
 Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro Trp Asn Phe Glu Asp
 465 470 475 480
 Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe Ile Asn Arg Met Thr
 485 490 495
 Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro Lys His Ser
 500 505 510
 Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu Leu Thr Lys Val Arg

8009W000_SequenceListing.txt

515

520

525

Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe Leu Asp Ser Lys Gln
 530 535 540

Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp Lys Arg Lys Val Thr
 545 550 555 560

Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile Tyr Gly Tyr Asp Gly
 565 570 575

Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn Ser Ser Leu Ser Thr
 580 585 590

Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys Glu Phe Leu Asp Asp
 595 600 605

Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile His Thr Leu Thr Ile
 610 615 620

Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu Ser Lys Phe Glu Asn
 625 630 635 640

Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser Arg Arg His Tyr Thr
 645 650 655

Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn Gly Ile Arg Asp Glu
 660 665 670

Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Ile Ser
 675 680 685

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ala Leu Ser Phe Lys
 690 695 700

Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp Glu Asp Lys Gly Asn
 705 710 715 720

Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser Pro Ala Ile Lys Lys
 725 730 735

Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu Leu Val Lys Val Met
 740 745 750

Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu Met Ala Arg Glu Asn
 755 760 765

8009W000_SequenceListing.txt

Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln Gln Arg Leu Lys Arg
770 775 780

Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys Ile Leu Lys Glu Asn
785 790 795 800

Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn Ala Leu Gln Asn Asp
805 810 815

Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly
820 825 830

Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr Asp Ile Asp His Ile
835 840 845

Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile Asp Asn Lys Val Leu
850 855 860

Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp Asp Val Pro Ser Leu
865 870 875 880

Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr Gln Leu Leu Lys Ser
885 890 895

Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
900 905 910

Gly Gly Leu Ser Pro Glu Asp Lys Ala Gly Phe Ile Gln Arg Gln Leu
915 920 925

Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Arg Leu Leu Asp Glu
930 935 940

Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg Ala Val Arg Thr Val
945 950 955 960

Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser Gln Phe Arg Lys Asp
965 970 975

Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp Phe His His Ala His
980 985 990

Asp Ala Tyr Leu Asn Ala Val Val Ala Ser Ala Leu Leu Lys Lys Tyr
995 1000 1005

Pro Lys Leu Glu Pro Glu Phe Val Tyr Gly Asp Tyr Pro Lys Tyr
1010 1015 1020

8009W000_SequenceListing.txt

Asn	Ser	Phe	Arg	Glu	Arg	Lys	Ser	Ala	Thr	Glu	Lys	Val	Tyr	Phe
1025						1030					1035			
Tyr	Ser	Asn	Ile	Met	Asn	Ile	Phe	Lys	Lys	Ser	Ile	Ser	Leu	Ala
1040						1045					1050			
Asp	Gly	Arg	Val	Ile	Glu	Arg	Pro	Leu	Ile	Glu	Val	Asn	Glu	Glu
1055						1060					1065			
Thr	Gly	Glu	Ser	Val	Trp	Asn	Lys	Glu	Ser	Asp	Leu	Ala	Thr	Val
1070						1075					1080			
Arg	Arg	Val	Leu	Ser	Tyr	Pro	Gln	Val	Asn	Val	Val	Lys	Lys	Val
1085						1090					1095			
Glu	Glu	Gln	Asn	His	Gly	Leu	Asp	Arg	Gly	Lys	Pro	Lys	Gly	Leu
1100						1105					1110			
Phe	Asn	Ala	Asn	Leu	Ser	Ser	Lys	Pro	Lys	Pro	Asn	Ser	Asn	Glu
1115						1120					1125			
Asn	Leu	Val	Gly	Ala	Lys	Glu	Tyr	Leu	Asp	Pro	Lys	Lys	Tyr	Gly
1130						1135					1140			
Gly	Tyr	Ala	Gly	Ile	Ser	Asn	Ser	Phe	Thr	Val	Leu	Val	Lys	Gly
1145						1150					1155			
Thr	Ile	Glu	Lys	Gly	Ala	Lys	Lys	Lys	Ile	Thr	Asn	Val	Leu	Glu
1160						1165					1170			
Phe	Gln	Gly	Ile	Ser	Ile	Leu	Asp	Arg	Ile	Asn	Tyr	Arg	Lys	Asp
1175						1180					1185			
Lys	Leu	Asn	Phe	Leu	Leu	Glu	Lys	Gly	Tyr	Lys	Asp	Ile	Glu	Leu
1190						1195					1200			
Ile	Ile	Glu	Leu	Pro	Lys	Tyr	Ser	Leu	Phe	Glu	Leu	Ser	Asp	Gly
1205						1210					1215			
Ser	Arg	Arg	Met	Leu	Ala	Ser	Ile	Leu	Ser	Thr	Asn	Asn	Lys	Arg
1220						1225					1230			
Gly	Glu	Ile	His	Lys	Gly	Asn	Gln	Ile	Phe	Leu	Ser	Gln	Lys	Phe
1235						1240					1245			
Val	Lys	Leu	Leu	Tyr	His	Ala	Lys	Arg	Ile	Ser	Asn	Thr	Ile	Asn
1250						1255					1260			

8009W000_SequenceListing.txt

Glu Asn His Arg Lys Tyr Val Glu Asn His Lys Lys Glu Phe Glu
1265 1270 1275

Glu Leu Phe Tyr Tyr Ile Leu Glu Phe Asn Glu Asn Tyr Val Gly
1280 1285 1290

Ala Lys Lys Asn Gly Lys Leu Leu Asn Ser Ala Phe Gln Ser Trp
1295 1300 1305

Gln Asn His Ser Ile Asp Glu Leu Cys Ser Ser Phe Ile Gly Pro
1310 1315 1320

Thr Gly Ser Glu Arg Lys Gly Leu Phe Glu Leu Thr Ser Arg Gly
1325 1330 1335

Ser Ala Ala Asp Phe Glu Phe Leu Gly Val Lys Ile Pro Arg Tyr
1340 1345 1350

Arg Asp Tyr Thr Pro Ser Ser Leu Leu Lys Asp Ala Thr Leu Ile
1355 1360 1365

His Gln Ser Val Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ala
1370 1375 1380

Lys Leu Gly Glu Gly
1385

<210> 5
<211> 1334
<212> PRT
<213> Listeria innocua

<220>
<221> misc_feature
<222> (10)..(21)
<223> N-terminal RuvC-like domain

<220>
<221> misc_feature
<222> (762)..(769)
<223> RuvC-like domain

<220>
<221> misc_feature
<222> (840)..(866)
<223> HNH-like domain

<220>
<221> misc_feature
<222> (985)..(992)
<223> RuvC-like domain

8009W000_SequenceListing.txt

<400> 5

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

Gly Trp Ala Val Leu Thr Asp Gln Tyr Asp Leu Val Lys Arg Lys Met
 20 25 30

Lys Ile Ala Gly Asp Ser Glu Lys Lys Gln Ile Lys Lys Asn Phe Trp
 35 40 45

Gly Val Arg Leu Phe Asp Glu Gly Gln Thr Ala Ala Asp Arg Arg Met
 50 55 60

Ala Arg Thr Ala Arg Arg Arg Ile Glu Arg Arg Arg Asn Arg Ile Ser
 65 70 75 80

Tyr Leu Gln Gly Ile Phe Ala Glu Glu Met Ser Lys Thr Asp Ala Asn
 85 90 95

Phe Phe Cys Arg Leu Ser Asp Ser Phe Tyr Val Asp Asn Glu Lys Arg
 100 105 110

Asn Ser Arg His Pro Phe Phe Ala Thr Ile Glu Glu Glu Val Glu Tyr
 115 120 125

His Lys Asn Tyr Pro Thr Ile Tyr His Leu Arg Glu Glu Leu Val Asn
 130 135 140

Ser Ser Glu Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His
 145 150 155 160

Ile Ile Lys Tyr Arg Gly Asn Phe Leu Ile Glu Gly Ala Leu Asp Thr
 165 170 175

Gln Asn Thr Ser Val Asp Gly Ile Tyr Lys Gln Phe Ile Gln Thr Tyr
 180 185 190

Asn Gln Val Phe Ala Ser Gly Ile Glu Asp Gly Ser Leu Lys Lys Leu
 195 200 205

Glu Asp Asn Lys Asp Val Ala Lys Ile Leu Val Glu Lys Val Thr Arg
 210 215 220

Lys Glu Lys Leu Glu Arg Ile Leu Lys Leu Tyr Pro Gly Glu Lys Ser
 225 230 235 240

8009W000_SequenceListing.txt

Ala Gly Met Phe Ala Gln Phe Ile Ser Leu Ile Val Gly Ser Lys Gly
245 250 255

Asn Phe Gln Lys Pro Phe Asp Leu Ile Glu Lys Ser Asp Ile Glu Cys
260 265 270

Ala Lys Asp Ser Tyr Glu Glu Asp Leu Glu Ser Leu Leu Ala Leu Ile
275 280 285

Gly Asp Glu Tyr Ala Glu Leu Phe Val Ala Ala Lys Asn Ala Tyr Ser
290 295 300

Ala Val Val Leu Ser Ser Ile Ile Thr Val Ala Glu Thr Glu Thr Asn
305 310 315 320

Ala Lys Leu Ser Ala Ser Met Ile Glu Arg Phe Asp Thr His Glu Glu
325 330 335

Asp Leu Gly Glu Leu Lys Ala Phe Ile Lys Leu His Leu Pro Lys His
340 345 350

Tyr Glu Glu Ile Phe Ser Asn Thr Glu Lys His Gly Tyr Ala Gly Tyr
355 360 365

Ile Asp Gly Lys Thr Lys Gln Ala Asp Phe Tyr Lys Tyr Met Lys Met
370 375 380

Thr Leu Glu Asn Ile Glu Gly Ala Asp Tyr Phe Ile Ala Lys Ile Glu
385 390 395 400

Lys Glu Asn Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ala Ile
405 410 415

Pro His Gln Leu His Leu Glu Glu Leu Glu Ala Ile Leu His Gln Gln
420 425 430

Ala Lys Tyr Tyr Pro Phe Leu Lys Glu Asn Tyr Asp Lys Ile Lys Ser
435 440 445

Leu Val Thr Phe Arg Ile Pro Tyr Phe Val Gly Pro Leu Ala Asn Gly
450 455 460

Gln Ser Glu Phe Ala Trp Leu Thr Arg Lys Ala Asp Gly Glu Ile Arg
465 470 475 480

Pro Trp Asn Ile Glu Glu Lys Val Asp Phe Gly Lys Ser Ala Val Asp
485 490 495

8009W000_SequenceListing.txt

Phe Ile Glu Lys Met Thr Asn Lys Asp Thr Tyr Leu Pro Lys Glu Asn
 500 505 510
 Val Leu Pro Lys His Ser Leu Cys Tyr Gln Lys Tyr Leu Val Tyr Asn
 515 520 525
 Glu Leu Thr Lys Val Arg Tyr Ile Asn Asp Gln Gly Lys Thr Ser Tyr
 530 535 540
 Phe Ser Gly Gln Glu Lys Glu Gln Ile Phe Asn Asp Leu Phe Lys Gln
 545 550 555 560
 Lys Arg Lys Val Lys Lys Lys Asp Leu Glu Leu Phe Leu Arg Asn Met
 565 570 575
 Ser His Val Glu Ser Pro Thr Ile Glu Gly Leu Glu Asp Ser Phe Asn
 580 585 590
 Ser Ser Tyr Ser Thr Tyr His Asp Leu Leu Lys Val Gly Ile Lys Gln
 595 600 605
 Glu Ile Leu Asp Asn Pro Val Asn Thr Glu Met Leu Glu Asn Ile Val
 610 615 620
 Lys Ile Leu Thr Val Phe Glu Asp Lys Arg Met Ile Lys Glu Gln Leu
 625 630 635 640
 Gln Gln Phe Ser Asp Val Leu Asp Gly Val Val Leu Lys Lys Leu Glu
 645 650 655
 Arg Arg His Tyr Thr Gly Trp Gly Arg Leu Ser Ala Lys Leu Leu Met
 660 665 670
 Gly Ile Arg Asp Lys Gln Ser His Leu Thr Ile Leu Asp Tyr Leu Met
 675 680 685
 Asn Asp Asp Gly Leu Asn Arg Asn Leu Met Gln Leu Ile Asn Asp Ser
 690 695 700
 Asn Leu Ser Phe Lys Ser Ile Ile Glu Lys Glu Gln Val Thr Thr Ala
 705 710 715 720
 Asp Lys Asp Ile Gln Ser Ile Val Ala Asp Leu Ala Gly Ser Pro Ala
 725 730 735
 Ile Lys Lys Gly Ile Leu Gln Ser Leu Lys Ile Val Asp Glu Leu Val
 740 745 750

8009W000_SequenceListing.txt

Ser Val Met Gly Tyr Pro Pro Gln Thr Ile Val Val Glu Met Ala Arg
755 760 765

Glu Asn Gln Thr Thr Gly Lys Gly Lys Asn Asn Ser Arg Pro Arg Tyr
770 775 780

Lys Ser Leu Glu Lys Ala Ile Lys Glu Phe Gly Ser Gln Ile Leu Lys
785 790 795 800

Glu His Pro Thr Asp Asn Gln Glu Leu Arg Asn Asn Arg Leu Tyr Leu
805 810 815

Tyr Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly Gln Asp Leu Asp
820 825 830

Ile His Asn Leu Ser Asn Tyr Asp Ile Asp His Ile Val Pro Gln Ser
835 840 845

Phe Ile Thr Asp Asn Ser Ile Asp Asn Leu Val Leu Thr Ser Ser Ala
850 855 860

Gly Asn Arg Glu Lys Gly Asp Asp Val Pro Pro Leu Glu Ile Val Arg
865 870 875 880

Lys Arg Lys Val Phe Trp Glu Lys Leu Tyr Gln Gly Asn Leu Met Ser
885 890 895

Lys Arg Lys Phe Asp Tyr Leu Thr Lys Ala Glu Arg Gly Gly Leu Thr
900 905 910

Glu Ala Asp Lys Ala Arg Phe Ile His Arg Gln Leu Val Glu Thr Arg
915 920 925

Gln Ile Thr Lys Asn Val Ala Asn Ile Leu His Gln Arg Phe Asn Tyr
930 935 940

Glu Lys Asp Asp His Gly Asn Thr Met Lys Gln Val Arg Ile Val Thr
945 950 955 960

Leu Lys Ser Ala Leu Val Ser Gln Phe Arg Lys Gln Phe Gln Leu Tyr
965 970 975

Lys Val Arg Asp Val Asn Asp Tyr His His Ala His Asp Ala Tyr Leu
980 985 990

Asn Gly Val Val Ala Asn Thr Leu Leu Lys Val Tyr Pro Gln Leu Glu

8009W000_SequenceListing.txt

995

1000

1005

Pro	Glu	Phe	Val	Tyr	Gly	Asp	Tyr	His	Gln	Phe	Asp	Trp	Phe	Lys
	1010					1015					1020			
Ala	Asn	Lys	Ala	Thr	Ala	Lys	Lys	Gln	Phe	Tyr	Thr	Asn	Ile	Met
	1025					1030					1035			
Leu	Phe	Phe	Ala	Gln	Lys	Asp	Arg	Ile	Ile	Asp	Glu	Asn	Gly	Glu
	1040					1045					1050			
Ile	Leu	Trp	Asp	Lys	Lys	Tyr	Leu	Asp	Thr	Val	Lys	Lys	Val	Met
	1055					1060					1065			
Ser	Tyr	Arg	Gln	Met	Asn	Ile	Val	Lys	Lys	Thr	Glu	Ile	Gln	Lys
	1070					1075					1080			
Gly	Glu	Phe	Ser	Lys	Ala	Thr	Ile	Lys	Pro	Lys	Gly	Asn	Ser	Ser
	1085					1090					1095			
Lys	Leu	Ile	Pro	Arg	Lys	Thr	Asn	Trp	Asp	Pro	Met	Lys	Tyr	Gly
	1100					1105					1110			
Gly	Leu	Asp	Ser	Pro	Asn	Met	Ala	Tyr	Ala	Val	Val	Ile	Glu	Tyr
	1115					1120					1125			
Ala	Lys	Gly	Lys	Asn	Lys	Leu	Val	Phe	Glu	Lys	Lys	Ile	Ile	Arg
	1130					1135					1140			
Val	Thr	Ile	Met	Glu	Arg	Lys	Ala	Phe	Glu	Lys	Asp	Glu	Lys	Ala
	1145					1150					1155			
Phe	Leu	Glu	Glu	Gln	Gly	Tyr	Arg	Gln	Pro	Lys	Val	Leu	Ala	Lys
	1160					1165					1170			
Leu	Pro	Lys	Tyr	Thr	Leu	Tyr	Glu	Cys	Glu	Glu	Gly	Arg	Arg	Arg
	1175					1180					1185			
Met	Leu	Ala	Ser	Ala	Asn	Glu	Ala	Gln	Lys	Gly	Asn	Gln	Gln	Val
	1190					1195					1200			
Leu	Pro	Asn	His	Leu	Val	Thr	Leu	Leu	His	His	Ala	Ala	Asn	Cys
	1205					1210					1215			
Glu	Val	Ser	Asp	Gly	Lys	Ser	Leu	Asp	Tyr	Ile	Glu	Ser	Asn	Arg
	1220					1225					1230			

8009W000_SequenceListing.txt

Glu Met Phe Ala Glu Leu Leu Ala His Val Ser Glu Phe Ala Lys
 1235 1240 1245

Arg Tyr Thr Leu Ala Glu Ala Asn Leu Asn Lys Ile Asn Gln Leu
 1250 1255 1260

Phe Glu Gln Asn Lys Glu Gly Asp Ile Lys Ala Ile Ala Gln Ser
 1265 1270 1275

Phe Val Asp Leu Met Ala Phe Asn Ala Met Gly Ala Pro Ala Ser
 1280 1285 1290

Phe Lys Phe Phe Glu Thr Thr Ile Glu Arg Lys Arg Tyr Asn Asn
 1295 1300 1305

Leu Lys Glu Leu Leu Asn Ser Thr Ile Ile Tyr Gln Ser Ile Thr
 1310 1315 1320

Gly Leu Tyr Glu Ser Arg Lys Arg Leu Asp Asp
 1325 1330

<210> 6
 <211> 1053
 <212> PRT
 <213> Staphylococcus aureus

<400> 6

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

Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
 20 25 30

Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
 35 40 45

Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile
 50 55 60

Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His
 65 70 75 80

Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu
 85 90 95

Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu
 100 105 110

8009W000_SequenceListing.txt

Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr
 115 120 125
 Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala
 130 135 140
 Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys
 145 150 155 160
 Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr
 165 170 175
 Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln
 180 185 190
 Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg
 195 200 205
 Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys
 210 215 220
 Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe
 225 230 235 240
 Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr
 245 250 255
 Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn
 260 265 270
 Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe
 275 280 285
 Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu
 290 295 300
 Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys
 305 310 315 320
 Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr
 325 330 335
 Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala
 340 345 350
 Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu
 355 360 365

8009W000_SequenceListing.txt

Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser
 370 375 380
 Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile
 385 390 395 400
 Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala
 405 410 415
 Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln
 420 425 430
 Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro
 435 440 445
 Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile
 450 455 460
 Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg
 465 470 475 480
 Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys
 485 490 495
 Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr
 500 505 510
 Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp
 515 520 525
 Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu
 530 535 540
 Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro
 545 550 555 560
 Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys
 565 570 575
 Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu
 580 585 590
 Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile
 595 600 605
 Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu
 610 615 620

8009W000_SequenceListing.txt

Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp
 625 630 635 640
 Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu
 645 650 655
 Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys
 660 665 670
 Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp
 675 680 685
 Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp
 690 695 700
 Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys
 705 710 715 720
 Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys
 725 730 735
 Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu
 740 745 750
 Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp
 755 760 765
 Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile
 770 775 780
 Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu
 785 790 795 800
 Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu
 805 810 815
 Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His
 820 825 830
 Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly
 835 840 845
 Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr
 850 855 860
 Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile

8009W000_SequenceListing.txt

865 870 875 880
 Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp
 885 890 895
 Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr
 900 905 910
 Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val
 915 920 925
 Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser
 930 935 940
 Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala
 945 950 955 960
 Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly
 965 970 975
 Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile
 980 985 990
 Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met
 995 1000 1005
 Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys
 1010 1015 1020
 Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu
 1025 1030 1035
 Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
 1040 1045 1050

<210> 7
 <211> 3159
 <212> DNA
 <213> Staphylococcus aureus

<400> 7
 atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg gtatgggatt 60
 attgactatg aaacaaggga cgtgatcgac gcaggcgtca gactgttcaa ggaggccaac 120
 gtggaaaaca atgaggggacg gagaagcaag aggggagcca ggcgcctgaa acgacggaga 180
 aggcacagaa tccagaggggt gaagaaaactg ctgttcgatt acaacctgct gaccgaccat 240
 tctgagctga gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg 300

8009W000_SequenceListing.txt

tcagaggaag	agttttccgc	agctctgctg	cacctggcta	agcgccgagg	agtgcataac	360
gtcaatgagg	tggaagagga	caccggcaac	gagctgtcta	caaaggaaca	gatctcacgc	420
aatagcaaag	ctctggaaga	gaagtatgtc	gcagagctgc	agctggaacg	gctgaagaaa	480
gatggcgagg	tgagaggggtc	aattaatagg	ttcaagacaa	gcgactacgt	caaagaagcc	540
aagcagctgc	tgaaagtgca	gaaggcttac	caccagctgg	atcagagctt	catcgatact	600
tatatcgacc	tgctggagac	tcggagaacc	tactatgagg	gaccaggaga	agggagcccc	660
ttcggatgga	aagacatcaa	ggaatggtag	gagatgctga	tgggacattg	cacctatfff	720
ccagaagagc	tgagaagcgt	caagtacgct	tataacgcag	atctgtacaa	cgccctgaat	780
gacctgaaca	acctggtcat	caccagggat	gaaaacgaga	aactggaata	ctatgagaag	840
ttccagatca	tcgaaaacgt	gtttaagcag	aagaaaaagc	ctacactgaa	acagattgct	900
aaggagatcc	tgggtcaacga	agaggacatc	aagggtacc	gggtgacaag	cactggaaaa	960
ccagagttca	ccaatctgaa	agtgtatcac	gatattaagg	acatcacagc	acggaaagaa	1020
atcattgaga	acgccgaact	gctggatcag	attgctaaga	tcctgactat	ctaccagagc	1080
tccgaggaca	tccaggaaga	gctgactaac	ctgaacagcg	agctgaccca	ggaagagatc	1140
gaacagatta	gtaatctgaa	gggttacacc	ggaacacaca	acctgtccct	gaaagctatc	1200
aatctgattc	tggatgagct	gtggcataca	aacgacaatc	agattgcaat	ctttaaccgg	1260
ctgaagctgg	tccccaaaaa	ggtggacctg	agtcagcaga	aagagatccc	aaccacactg	1320
gtggacgatt	tcattctgtc	acccgtgggtc	aagcggagct	tcattccagag	catcaaagtg	1380
atcaacgcca	tcattcaagaa	gtacggcctg	cccaatgata	tcattatcga	gctggctagg	1440
gagaagaaca	gcaaggacgc	acagaagatg	atcaatgaga	tgcagaaacg	aaaccggcag	1500
accaatgaac	gcattgaaga	gattatccga	actaccggga	aagagaacgc	aaagtacctg	1560
attgaaaaaa	tcaagctgca	cgatatgcag	gagggaaaagt	gtctgtattc	tctggaggcc	1620
atccccctgg	aggacctgct	gaacaatcca	ttcaactacg	aggtcgatca	tattatcccc	1680
agaagcgtgt	ccttcgacaa	ttcctttaac	aacaagggtc	tggtaagca	ggaagagaac	1740
tctaaaaagg	gcaataggac	tcctttccag	tacctgtcta	gttcagattc	caagatctct	1800
tacgaaacct	ttaaaaagca	cattctgaat	ctggccaaag	gaaagggccg	catcagcaag	1860
acaaaaaagg	agtacctgct	ggaagagcgg	gacatcaaca	gattctccgt	ccagaaggat	1920
tttattaacc	ggaatctggt	ggacacaaga	tacgtacttc	gcggcctgat	gaatctgctg	1980
cgatcctatt	tccgggtgaa	caatctggat	gtgaaaagtca	agtccatcaa	cggcgggttc	2040
acatcttttc	tgaggcgcaa	atggaagttt	aaaaaggagc	gcaacaaagg	gtacaagcac	2100
catgccgaag	atgctctgat	tatcgcaaat	gccgacttca	tctttaagga	gtggaaaaag	2160
ctggacaaaag	ccaagaaaagt	gatggagaac	cagatgtttcg	aagagaagca	ggccgaatct	2220

8009W000_SequenceListing.txt

atgcccga	aa tcgagacaga	acaggagtac	aaggagattt	tcatcactcc	tcaccagatc	2280
aagcatatca	aggatttcaa	ggactacaag	tactctcacc	gggtggataa	aaagcccaac	2340
agagagctga	tcaatgacac	cctgtatagt	acaagaaaag	acgataaggg	gaataccctg	2400
attgtgaaca	atctgaacgg	actgtacgac	aaagataatg	acaagctgaa	aaagctgac	2460
aacaaaagtc	ccgagaagct	gctgatgtac	caccatgac	ctcagacata	tcagaaactg	2520
aagctgatta	tggagcagta	cggcgacgag	aagaacccac	tgtataagta	ctatgaagag	2580
actgggaact	acctgaccaa	gtatagcaaa	aaggataatg	gccccgtgat	caagaagac	2640
aagtactatg	ggaacaagct	gaatgcccat	ctggacatca	cagacgatta	ccctaacagt	2700
cgcaacaagg	tggatcaagct	gtcactgaag	ccatacagat	tcgatgtcta	tctggacaac	2760
ggcgtgtata	aatttgtgac	tgtcaagaat	ctggatgtca	tcaaaaagga	gaactactat	2820
gaagtgaata	gcaagtgtca	cgaagaggct	aaaaagctga	aaaagattag	caaccaggca	2880
gagttcatcg	cctcctttta	caacaacgac	ctgattaaga	tcaatggcga	actgtatagg	2940
gtcatcgggg	tgaacaatga	tctgctgaac	cgcattgaag	tgaatatgat	tgacatcact	3000
taccgagagt	atctggaaaa	catgaatgat	aagcgccccc	ctcgaattat	caaaacaatt	3060
gcctctaaga	ctcagagtat	caaaaagtac	tcaaccgaca	ttctgggaaa	cctgtatgag	3120
gtgaagagca	aaaagcacc	tcagattatc	aaaaagggc			3159

<210> 8

<211> 3159

<212> DNA

<213> Staphylococcus aureus

<400> 8

atgaagcgga	actacatcct	gggcctggac	atcggcatca	ccagcgtggg	ctacggcatc	60
atcgactacg	agacacggga	cgtgatcgat	gccggcgtgc	ggctgttcaa	agaggccaac	120
gtggaaaaaca	acgagggcag	gcggagcaag	agaggcgcca	gaaggctgaa	gcggcggagg	180
cggcatagaa	tccagagagt	gaagaagctg	ctgttcgact	acaacctgct	gaccgaccac	240
agcgagctga	gcggcatcaa	cccctacgag	gccagagtga	agggcctgag	ccagaagctg	300
agcgaggaag	agttctctgc	cgccctgctg	cacctggcca	agagaagagg	cgtgcacaac	360
gtgaacgagg	tggaagagga	caccggcaac	gagctgtcca	caaagagca	gatcagccgg	420
aacagcaagg	ccctggaaga	gaaatacgtg	gccgaactgc	agctggaacg	gctgaagaaa	480
gacggcgaag	tgcggggcag	catcaacaga	ttcaagacca	gcgactacgt	gaaagaagcc	540
aaacagctgc	tgaaggtgca	gaaggcctac	caccagctgg	accagagctt	catcgacacc	600
tacatcgacc	tgctggaaac	ccggcggacc	tactatgagg	gacctggcga	gggcagcccc	660
ttcggctgga	aggacatcaa	agaatggtac	gagatgctga	tgggccactg	cacctacttc	720

8009W000_SequenceListing.txt

cccgaggaac	tgcggagcgt	gaagtacgcc	tacaacgccg	acctgtacaa	cgccctgaac	780
gacctgaaca	atctcgtgat	caccagggac	gagaacgaga	agctggaata	ttacgagaag	840
ttccagatca	tcgagaacgt	gttcaagcag	aagaagaagc	ccaccctgaa	gcagatcgcc	900
aaagaaatcc	tcgtgaacga	agaggatatt	aagggctaca	gagtgaccag	caccggcaag	960
cccgagttca	ccaacctgaa	ggtgtaccac	gacatcaagg	acattaccgc	ccggaaagag	1020
attattgaga	acgccgagct	gctggatcag	attgccaaga	tcctgaccat	ctaccagagc	1080
agcgaggaca	tccaggaaga	actgaccaat	ctgaactccg	agctgacca	ggaagagatc	1140
gagcagatct	ctaattctgaa	gggctatacc	ggcaccacaca	acctgagcct	gaaggccatc	1200
aacctgatcc	tggacgagct	gtggcacacc	aacgacaacc	agatcgctat	cttcaaccgg	1260
ctgaagctgg	tgccaagaa	ggtggacctg	tcccagcaga	aagagatccc	caccaccctg	1320
gtggacgact	tcattctgag	ccccgtcgtg	aagagaagct	tcattccagag	catcaaagtg	1380
atcaacgcca	tcattcaagaa	gtacggcctg	ccaacgaca	tcattatcga	gctggcccg	1440
gagaagaact	ccaaggacgc	ccagaaaatg	atcaacgaga	tgagaagcg	gaaccggcag	1500
accaacgagc	ggatcgagga	aatcatccgg	accaccggca	aagagaacgc	caagtacctg	1560
atcgagaaga	tcaagctgca	cgacatgcag	gaaggcaagt	gcctgtacag	cctggaagcc	1620
atccctctgg	aagatctgct	gaacaacccc	ttcaactatg	aggtggacca	catcatcccc	1680
agaagcgtgt	ccttcgacaa	cagcttcaac	aacaagggtg	tcgtgaagca	ggaagaaaac	1740
agcaagaagg	gcaaccggac	cccattccag	tacctgagca	gcagcgacag	caagatcagc	1800
tacgaaacct	tcaagaagca	catcctgaat	ctggccaagg	gcaagggcag	aatcagcaag	1860
accaagaaaag	agtatctgct	ggaagaacgg	gacatcaaca	ggttctccgt	gcagaaagac	1920
ttcatcaacc	ggaacctggt	ggataccaga	tacgccacca	gaggcctgat	gaacctgctg	1980
cggagctact	tcagagtga	caacctggac	gtgaaagtga	agtccatcaa	tggcggcttc	2040
accagctttc	tgcggcggaa	gtggaagttt	aagaaagagc	ggaacaaggg	gtacaagcac	2100
cacgccgagg	acgccctgat	cattgccaac	gccgatttca	tcttcaaaga	gtggaagaaa	2160
ctggacaagg	ccaaaaaagt	gatggaaaac	cagatgttcg	aggaaaagca	ggccgagagc	2220
atgcccgaga	tcgaaaccga	gcaggagtac	aaagagatct	tcattacccc	ccaccagatc	2280
aagcacatta	aggacttcaa	ggactacaag	tacagccacc	gggtggacaa	gaagccta	2340
agagagctga	ttaacgacac	cctgtactcc	acccggaagg	acgacaaggg	caacaccctg	2400
atcgtgaaca	atctgaacgg	cctgtacgac	aaggacaatg	acaagctgaa	aaagctgatc	2460
aacaagagcc	ccgaaaagct	gctgatgtac	caccacgacc	cccagacct	ccagaaactg	2520
aagctgatta	tggaacagta	cggcgacgag	aagaatcccc	tgtacaagta	ctacgaggaa	2580

8009W000_SequenceListing.txt

accgggaact	acctgaccaa	gtactccaaa	aaggacaacg	gccccgtgat	caagaagatt	2640
aagtattacg	gcaacaaact	gaacgcccac	ctggacatca	ccgacgacta	ccccaacagc	2700
agaacaagg	tcgtgaagct	gtccctgaag	ccctacagat	tcgacgtgta	cctggacaat	2760
ggcgtgtaca	agttcgtgac	cgtgaagaat	ctggatgtga	tcaaaaaaga	aaactactac	2820
gaagtgaata	gcaagtgcta	tgaggaagct	aagaagctga	agaagatcag	caaccaggcc	2880
gagtttatcg	cctccttcta	caacaacgat	ctgatcaaga	tcaacggcga	gctgtataga	2940
gtgatcggcg	tgaacaacga	cctgctgaac	cggatcgaag	tgaacatgat	cgacatcacc	3000
taccgcgagt	acctggaaaa	catgaacgac	aagaggcccc	ccaggatcat	taagacaatc	3060
gcctccaaga	cccagagcat	taagaagtac	agcacagaca	ttctgggcaa	cctgtatgaa	3120
gtgaaatcta	agaagcaccc	tcagatcatc	aaaaagggc			3159

<210> 9

<211> 3159

<212> DNA

<213> Staphylococcus aureus

<400> 9

atgaagcgca	actacatcct	cggactggac	atcggcatta	cctccgtggg	atacggcatc	60
atcgattacg	aaactagggg	tgtgatcgac	gctggagtca	ggctgttcaa	agaggcgaac	120
gtggagaaca	acgagggggc	gcgctcaaag	agggggggcc	gccggctgaa	gcgccggccg	180
agacatagaa	tccagcgcg	gaagaagctg	ctgttcgact	acaaccttct	gaccgaccac	240
tccgaacttt	ccggcatcaa	cccatatgag	gctagagtga	agggattgtc	ccaaaagctg	300
tccgaggaag	agttctccgc	cgcggttgct	cacctcgcca	agcgcagggg	agtgacacat	360
gtgaacgaag	tggaagaaga	taccggaaac	gagctgtcca	ccaaggagca	gatcagccgg	420
aactccaagg	ccctggaaga	gaaatacgtg	gcggaactgc	aactggagcg	gctgaagaaa	480
gacggagaag	tgcgcggtc	gatcaaccgc	ttcaagacct	cggactacgt	gaaggaggcc	540
aagcagctcc	tgaaagtgca	aaaggcctat	caccaacttg	accagtcctt	tatcgatacc	600
tacatcgatc	tgctcgagac	tcggcggact	tactacgagg	gtccagggga	gggctcccca	660
tttggttga	aggatattaa	ggagtggtag	gaaatgctga	tgggacactg	cacatacttc	720
cctgaggagc	tgcggagcgt	gaaatacgc	tacaacgcag	acctgtacaa	cgcgctgaac	780
gacctgaaca	atctcgtgat	cacccgggac	gagaacgaaa	agctcgagta	ttacgaaaag	840
ttccagatta	ttgagaacgt	gttcaaacag	aagaagaagc	cgacactgaa	gcagattgcc	900
aaggaaatcc	tcgtgaacga	agaggacatc	aagggtatc	gagtgacctc	aacgggaaag	960
ccggagttca	ccaatctgaa	ggtctaccac	gacatcaaag	acattaccgc	ccggaaggag	1020
atcattgaga	acgcggagct	gttggaccag	attgcgaaga	ttctgaccat	ctaccaatcc	1080

8009W000_SequenceListing.txt

tccgaggata	ttcaggaaga	actcaccaac	ctcaacagcg	aactgaccca	ggaggagata	1140
gagcaaatct	ccaacctgaa	gggctacacc	ggaactcata	acctgagcct	gaaggccatc	1200
aacttgatcc	tggacgagct	gtggcacacc	aacgataacc	agatcgctat	tttcaatcgg	1260
ctgaagctgg	tccccaagaa	agtggacctc	tcacaacaaa	aggagatccc	tactaccctt	1320
gtggacgatt	tcattctgtc	ccccgtggtc	aagagaagct	tcatacagtc	aatcaaagtg	1380
atcaatgcca	ttatcaagaa	atacggctctg	cccaacgaca	ttatcattga	gctcgcccg	1440
gagaagaact	cgaaggacgc	ccagaagatg	attaacgaaa	tgcagaagag	gaaccgacag	1500
actaacgaac	ggatcgaaga	aatcatccgg	accaccggga	aggaaaacgc	gaagtacctg	1560
atcgaaaaga	tcaagctcca	tgacatgcag	gaaggaaaagt	gtctgtactc	gctggaggcc	1620
attccgctgg	aggacttgct	gaacaaccct	tttaactacg	aagtggatca	tatcattccg	1680
aggagcgtgt	cattcgacaa	ttccttcaac	aacaagggtcc	tcgtgaagca	ggaggaaaac	1740
tcgaagaagg	gaaaccgcac	gccgttccag	tacctgagca	gcagcgactc	caagatttcc	1800
tacgaaacct	tcaagaagca	catcctcaac	ctggcaaagg	ggaagggtcg	catctccaag	1860
accaagaagg	aatatctgct	ggaagaaaga	gacatcaaca	gattctccgt	gcaaaaaggac	1920
ttcatcaacc	gcaacctcgt	ggatactaga	tacgctactc	ggggtctgat	gaacctcctg	1980
agaagctact	ttagagtga	caatctggac	gtgaagggtca	agtcgattaa	cggaggtttc	2040
acctccttcc	tgcggcgcaa	gtggaagttc	aagaaggaac	ggaacaaggg	ctacaagcac	2100
cacgccgagg	acgccctgat	cattgccaac	gccgacttca	tcttcaaaga	atggaagaaa	2160
cttgacaagg	ctaagaaggt	catggaaaac	cagatgttcg	aagaaaagca	ggccgagtct	2220
atgcctgaaa	tcgagactga	acaggagtac	aaggaaatct	ttattacgcc	acaccagatc	2280
aaacacatca	aggatttcaa	ggattacaag	tactcacatc	gcgtggacaa	aaagccgaac	2340
agggaactga	tcaacgacac	cctctactcc	acccggaagg	atgacaaaagg	gaataccctc	2400
atcgtcaaca	accttaacgg	cctgtacgac	aaggacaacg	ataagctgaa	gaagctcatt	2460
aacaagtcgc	ccgaaaagtt	gctgatgtac	caccacgacc	ctcagactta	ccagaagctc	2520
aagctgatca	tggagcagta	tggggacgag	aaaaacccgt	tgtacaagta	ctacgaagaa	2580
actgggaatt	atctgactaa	gtactccaag	aaagataacg	gccccgtgat	taagaagatt	2640
aagtactacg	gcaacaagct	gaacgcccac	ctggacatca	ccgatgacta	ccctaattcc	2700
cgcaacaagg	tcgtcaagct	gagcctcaag	ccctaccggt	ttgatgtgta	ccttgacaat	2760
ggagtgtaca	agttcgtgac	tgtgaagaac	cttgacgtga	tcaagaagga	gaactactac	2820
gaagtcaact	ccaagtgtca	cgaggaagca	aagaagttga	agaagatctc	gaaccaggcc	2880
gagttcattg	cctccttcta	taacaacgac	ctgattaaga	tcaacggcga	actgtaccgc	2940
gtcattggcg	tgaacaacga	tctcctgaac	cgcatcgaag	tgaacatgat	cgacatcact	3000

8009W000_SequenceListing.txt

taccgggaat acctggagaa tatgaacgac aagcgcccg cccggatcat taagactatc	3060
gcctcaaaga cccagtcgat caagaagtac agcaccgaca tcctgggcaa cctgtacgag	3120
gtcaaatacga agaagcaccc ccagatcatc aagaaggga	3159

<210> 10
 <211> 3159
 <212> DNA
 <213> Staphylococcus aureus

<400> 10	
atgaaaagga actacattct ggggctggcc atcgggatta caagcgtggg gtatgggatt	60
attgactatg aaacaaggga cgtgatcgac gcaggcgtca gactgttcaa ggaggccaac	120
gtggaaaaca atgagggacg gagaagcaag aggggagcca ggcgcctgaa acgacggaga	180
aggcacagaa tccagagggt gaagaaactg ctgttcgatt acaacctgct gaccgaccat	240
tctgagctga gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg	300
tcagaggaag agttttccgc agctctgctg cacctggcta agcgccgagg agtgcataac	360
gtcaatgagg tggaagagga caccggcaac gagctgtcta caaaggaaca gatctcacgc	420
aatagcaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa	480
gatggcgagg tgagagggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc	540
aagcagctgc tgaaagtgca gaaggcttac caccagctgg atcagagctt catcgatact	600
tatatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga agggagcccc	660
ttcggatgga aagacatcaa ggaatggtac gagatgctga tgggacattg cacctatfff	720
ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgccctgaat	780
gacctgaaca acctgggcat caccagggat gaaaacgaga aactggaata ctatgagaag	840
ttccagatca tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct	900
aaggagatcc tggatcaacga agaggacatc aagggtacc gggtgacaag cactggaaaa	960
ccagagttca ccaatctgaa agtgtatcac gatattaagg acatcacagc acggaaagaa	1020
atcattgaga acgccgaact gctggatcag attgctaaga tcctgactat ctaccagagc	1080
tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgaccca ggaagagatc	1140
gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc	1200
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg	1260
ctgaagctgg tccccaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg	1320
gtggacgatt tcattctgtc acccggtggc aagcggagct tcatccagag catcaaagtg	1380
atcaacgcca tcatcaagaa gtacggcctg cccaatgata tcattatcga gctggctagg	1440
gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag	1500

8009W000_SequenceListing.txt

accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg	1560
attgaaaaaa tcaagctgca cgatatgcag gagggaaaagt gtctgtattc tctggaggcc	1620
atccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgatca tattatcccc	1680
agaagcgtgt ccttcgacaa ttcctttaac aacaagggtc tggtaagca ggaagagaac	1740
tctaaaaagg gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct	1800
tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaagggccg catcagcaag	1860
acaaaaaagg agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat	1920
tttattaacc ggaatctggt ggacacaaga tacgctactc gcggcctgat gaatctgctg	1980
cgatcctatt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cggcgggttc	2040
acatcttttc tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac	2100
catgccgaag atgctctgat tatcgcaaat gccgacttca tctttaagga gtggaaaaag	2160
ctggacaaaag ccaagaaaagt gatggagaac cagatgttcg aagagaagca ggccgaatct	2220
atgcccgaag tcgagacaga acaggagtac aaggagattt tcatcactcc tcaccagatc	2280
aagcatatca aggtattcaa ggactacaag tactctcacc ggggtggataa aaagcccaac	2340
agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg	2400
attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgac	2460
aacaaaagtc ccgagaagct gctgatgtac caccatgac ctcagacata tcagaaactg	2520
aagctgatta tggagcagta cggcgacgag aagaaccac tgtataagta ctatgaagag	2580
actgggaact acctgaccaa gtatagcaaa aaggataatg gccccgtgat caagaagac	2640
aagtactatg ggaacaagct gaatgcccat ctggacatca cagacgatta ccctaacagt	2700
cgcaacaagg tggtaagct gtcactgaag ccatacagat tcgatgtcta tctggacaac	2760
ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat	2820
gaagtgaata gcaagtgcta cgaagaggct aaaaagctga aaaagattag caaccaggca	2880
gagttcatcg cctcctttta caacaacgac ctgattaaga tcaatggcga actgtatagg	2940
gtcatcgggg tgaacaatga tctgctgaac cgattgaag tgaatatgat tgacatcact	3000
taccgagagt atctggaaaa catgaatgat aagcgcccc ctcgaattat caaaacaatt	3060
gcctctaaga ctcagagtat caaaaagtac tcaaccgaca ttctgggaaa cctgtatgag	3120
gtgaagagca aaaagcacc tcagattatc aaaaagggc	3159

<210> 11
 <211> 3159
 <212> DNA
 <213> Staphylococcus aureus

8009W000_SequenceListing.txt

```

<400> 11
atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg gtatgggatt      60
attgactatg aaacaaggga cgtgatcgac gcaggcgtca gactgttcaa ggaggccaac      120
gtggaaaaca atgaggggacg gagaagcaag aggggagcca ggcgcctgaa acgacggaga      180
aggcacagaa tccagagggg gaagaaactg ctgttcgatt acaacctgct gaccgaccat      240
tctgagctga gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg      300
tcagaggaag agttttccgc agctctgctg cacctggcta agcgccgagg agtgcataac      360
gtcaatgagg tggaagagga caccggcaac gagctgtcta caaaggaaca gatctcacgc      420
aatagcaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa      480
gatggcgagg tgagaggggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc      540
aagcagctgc tgaaagtgca gaaggcttac caccagctgg atcagagctt catcgatact      600
tatatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga agggagcccc      660
ttcggatgga aagacatcaa ggaatggtag gagatgctga tgggacattg cacctatttt      720
ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgccctgaat      780
gacctgaaca acctggtcat caccagggat gaaaacgaga aactggaata ctatgagaag      840
ttccagatca tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct      900
aaggagatcc tgggtcaacga agaggacatc aagggtacc gggtgacaag cactggaaaa      960
ccagagttca ccaatctgaa agtgtatcac gatattaagg acatcacagc acggaaagaa     1020
atcattgaga acgccgaact gctggatcag attgctaaga tcctgactat ctaccagagc     1080
tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgacca ggaagagatc     1140
gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc     1200
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg     1260
ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg     1320
gtggacgatt tcattctgtc acccggtgtc aagcggagct tcatccagag catcaaagtg     1380
atcaacgcca tcatcaagaa gtacggcctg cccaatgata tcattatcga gctggctagg     1440
gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag     1500
accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg     1560
attgaaaaaa tcaagctgca cgatatgcag gagggaaaagt gtctgtattc tctggaggcc     1620
atccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgatca tattatcccc     1680
agaagcgtgt ccttcgacaa ttctttaaac aacaagggtc tggtaagca ggaagaggcc     1740
tctaaaaagg gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct     1800
tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaagggccg catcagcaag     1860

```

8009W000_SequenceListing.txt

```

accaaaaaagg agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat 1920
tttattaacc ggaatctggt ggacacaaga tacgctactc gcggcctgat gaatctgctg 1980
cgatcctatt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cggcgggttc 2040
acatcttttc tgaggcgcaa atggaagttt aaaaaggagc gcaacaaaagg gtacaagcac 2100
catgccgaag atgctctgat tatcgcaa at gccgacttca tctttaagga gtggaaaaag 2160
ctggacaaaag ccaagaaaagt gatggagaac cagatgttcg aagagaagca ggccgaatct 2220
atgcccga aa tcgagacaga acaggagtac aaggagattt tcatcactcc tcaccagatc 2280
aagcatatca aggatttcaa ggactacaag tactctcacc ggggtggataa aaagcccaac 2340
agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataacctg 2400
attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgac 2460
aacaaaagtc ccgagaagct gctgatgtac caccatgac ctcagacata tcagaaactg 2520
aagctgatta tggagcagta cggcgacgag aagaaccac tgtataagta ctatgaagag 2580
actgggaact acctgaccaa gtatagcaaa aaggataatg gccccgtgat caagaagatc 2640
aagtactatg ggaacaagct gaatgccc at ctggacatca cagacgatta ccctaacagt 2700
cgacaacaagg tggtaagct gtcactgaag ccatacagat tcgatgtcta tctggacaac 2760
ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat 2820
gaagtgaata gcaagtgtc cgaagaggct aaaaagctga aaaagattag caaccaggca 2880
gagttcatcg cctcctttta caacaacgac ctgattaaga tcaatggcga actgtatagg 2940
gtcatcgggg tgaacaatga tctgctgaac cgcattgaag tgaatatgat tgacatcact 3000
taccgagagt atctggaaaa catgaatgat aagcgcccc ctcgaattat caaaacaatt 3060
gcctctaaga ctcagagtat caaaaagtac tcaaccgaca ttctgggaaa cctgtatgag 3120
gtgaagagca aaaagcacc tcagattatc aaaaagggc 3159

```

```

<210> 12
<211> 1082
<212> PRT
<213> Neisseria meningitidis
<400> 12

```

```

Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Asp
1           5           10          15

```

```

Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Asp
          20          25          30

```

```

Glu Asn Pro Ile Cys Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg
          35          40          45

```

8009W000_SequenceListing.txt

Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu
50 55 60

Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu
65 70 75 80

Arg Ala Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp
85 90 95

Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln
100 105 110

Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser
115 120 125

Ala Val Leu Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg
130 135 140

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

Gly Val Ala Asp Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr
165 170 175

Pro Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile
180 185 190

Arg Asn Gln Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu
195 200 205

Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn
210 215 220

Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
225 230 235 240

Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly
245 250 255

His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr
260 265 270

Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
275 280 285

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr
290 295 300

8009W000_SequenceListing.txt

Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala
 305 310 315 320
 Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg
 325 330 335
 Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala
 340 345 350
 Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys
 355 360 365
 Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile Gly Thr
 370 375 380
 Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
 385 390 395 400
 Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser
 405 410 415
 Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
 420 425 430
 Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
 435 440 445
 Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
 450 455 460
 Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
 465 470 475 480
 Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
 485 490 495
 Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
 500 505 510
 Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
 515 520 525
 Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
 530 535 540
 Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu

8009W000_SequenceListing.txt

```

545                               550                               555                               560

Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly
      565      570      575

Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
      580      585      590

Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
      595      600      605

Ser Glu Asn Gln Asn Lys Gly Asn Gln Thr Pro Tyr Glu Tyr Phe Asn
      610      615      620

Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
      625      630      635      640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
      645      650      655

Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr
      660      665      670

Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp Arg Met Arg Leu Thr
      675      680      685

Gly Lys Gly Lys Lys Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
      690      695      700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
      705      710      715      720

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
      725      730      735

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
      740      745      750

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln
      755      760      765

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met
      770      775      780

Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
      785      790      795      800

```

8009W000_SequenceListing.txt

Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
 805 810 815
 Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
 820 825 830
 Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys
 835 840 845
 Ser Ala Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu
 850 855 860
 Thr Gln Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg
 865 870 875 880
 Glu Pro Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys
 885 890 895
 Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys
 900 905 910
 Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val
 915 920 925
 Gln Lys Thr Gly Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn
 930 935 940
 Ala Thr Met Val Arg Val Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr
 945 950 955 960
 Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp
 965 970 975
 Arg Ala Val Val Gln Gly Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp
 980 985 990
 Asp Ser Phe Asn Phe Lys Phe Ser Leu His Pro Asn Asp Leu Val Glu
 995 1000 1005
 Val Ile Thr Lys Lys Ala Arg Met Phe Gly Tyr Phe Ala Ser Cys
 1010 1015 1020
 His Arg Gly Thr Gly Asn Ile Asn Ile Arg Ile His Asp Leu Asp
 1025 1030 1035
 His Lys Ile Gly Lys Asn Gly Ile Leu Glu Gly Ile Gly Val Lys
 1040 1045 1050

8009W000_SequenceListing.txt

Thr Ala Leu Ser Phe Gln Lys Tyr Gln Ile Asp Glu Leu Gly Lys
 1055 1060 1065

Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
 1070 1075 1080

<210> 13
 <211> 3249
 <212> DNA
 <213> Neisseria meningitidis

<220>
 <221> misc_feature
 <222> (1)..(3249)
 <223> Exemplary codon optimized Cas9

<400> 13
 atggccgcct tcaagcccaa ccccatcaac tacatcctgg gcctggacat cggcatcgcc 60
 agcgtgggct gggccatggt ggagatcgac gaggacgaga accccatctg cctgatcgac 120
 ctgggtgtgc gcgtgttcga gcgcgctgag gtgcccaaga ctggtgacag tctggctatg 180
 gctcgccggc ttgctcgctc tgttcggcgc cttactcgcc ggcgcgctca ccgccttctg 240
 cgcgctcgcc gcctgctgaa gcgcgagggg gtgctgcagg ctgccgactt cgacgagaac 300
 ggctgatca agagcctgcc caacactcct tggcagctgc gcgctgccgc tctggaccgc 360
 aagctgactc ctctggagtg gagcgccgtg ctgctgcacc tgatcaagca ccgcggctac 420
 ctgagccagc gcaagaacga gggcgagacc gccgacaagg agctgggtgc tctgctgaag 480
 ggcgtggccg acaacgcca cgccctgcag actggtgact tccgcactcc tgctgagctg 540
 gccctgaaca agttcgagaa ggagagcggc cacatccgca accagcgcgg cgactacagc 600
 cacaccttca gccgcaagga cctgcaggcc gagctgatcc tgctgttcga gaagcagaag 660
 gagttcggca acccccacgt gagcggcggc ctgaaggagg gcatcgagac cctgctgatg 720
 acccagcgcc ccgccctgag cggcgacgcc gtgcagaaga tgctggggcca ctgcaccttc 780
 gagccagccg agcccaaggc cgccaagaac acctacaccg ccgagcgctt catctggctg 840
 accaagctga acaacctgcg catcctggag cagggcagcg agcgccccct gaccgacacc 900
 gagcgcgcca ccctgatgga cgagccctac cgcaagagca agctgacctt cgcccaggcc 960
 cgcaagctgc tgggtctgga ggacaccgcc ttcttcaagg gcctgcgcta cggcaaggac 1020
 aacgccgagg ccagcaccct gatggagatg aaggcctacc acgccatcag ccgcgccctg 1080
 gagaaggagg gcctgaagga caagaagagt cctctgaacc tgagccccga gctgcaggac 1140
 gagatcggca ccgccttcag cctgttcaag accgacgagg acatcaccgg ccgcctgaag 1200
 gaccgcatcc agcccagat cctggaggcc ctgctgaagc acatcagctt cgacaagttc 1260

8009W000_SequenceListing.txt

gtgcagatca	gcctgaaggc	cctgcgccgc	atcgtgcccc	tgatggagca	gggcaagcgc	1320
tacgacgagg	cctgcgccga	gatctacggc	gaccactacg	gcaagaagaa	caccgaggag	1380
aagatctacc	tgcttcctat	ccccgccgac	gagatccgca	accccgtggt	gctgcgcgcc	1440
ctgagccagg	cccgcaaggt	gatcaacggc	gtggtgcgcc	gctacggcag	ccccgcccgc	1500
atccacatcg	agaccgcccg	cgagggtggc	aagagcttca	aggaccgcaa	ggagatcgag	1560
aagcgccagg	aggagaaccg	caaggaccgc	gagaaggccg	ccgccaagtt	ccgcgagtac	1620
ttccccaact	tcgtgggcga	gccaagagc	aaggacatcc	tgaagctgcg	cctgtacgag	1680
cagcagcacg	gcaagtgcct	gtacagcggc	aaggagatca	acctgggccg	cctgaacgag	1740
aagggctacg	tggagatcga	ccacgccctg	cccttcagcc	gcacctggga	cgacagcttc	1800
aacaacaagg	tgctggtgct	gggcagcgag	aaccagaaca	agggaacca	gaccccctac	1860
gagtacttca	acggcaagga	caacagccgc	gagtggcagg	agttcaaggc	ccgcgtggag	1920
accagccgct	tccccgcag	caagaagcag	cgcctcctgc	tgcagaagtt	cgacgaggac	1980
ggcttcaagg	agcgcaacct	gaacgacacc	cgctacgtga	accgcttcct	gtgccagttc	2040
gtggccgacc	gcatgcgcct	gaccggcaag	ggcaagaagc	gcgtgttcgc	cagcaacggc	2100
cagatcacca	acctgctgcg	cggcttctgg	ggcctgcgca	aggcgcgcgc	cgagaacgac	2160
cgccaccacg	ccctggacgc	cgtggtggtg	gcctgcagca	ccgtggccat	gcagcagaag	2220
atcacccgct	tcgtgcgcta	caaggagatg	aacgccttcg	acggtaaaac	catcgacaag	2280
gagaccggcg	agggtgctga	ccagaagacc	cacttcccc	agccctggga	gttcttcgcc	2340
caggaggtga	tgatccgcgt	gttcggcaag	cccgcgcca	agcccgagtt	cgaggaggcc	2400
gacacccccg	agaagctgcg	caccctgctg	gccgagaagc	tgagcagccg	ccctgaggcc	2460
gtgcacgagt	acgtgactcc	tctgttcgtg	agccgcgccc	ccaaccgcaa	gatgagcggc	2520
cagggtcaca	tggagaccgt	gaagagcgcc	aagcgcctgg	acgagggcgt	gagcgtgctg	2580
cgcgcgcccc	tgacccagct	gaagctgaag	gacctggaga	agatggtgaa	ccgcgagcgc	2640
gagcccaagc	tgtacgaggc	cctgaaggcc	cgctggagg	cccacaagga	cgaccccgcc	2700
aaggccttcg	ccgagccctt	ctacaagtac	gacaaggccg	gcaaccgcac	ccagcaggcg	2760
aaggccgtgc	gcgtggagca	ggtgcagaag	accggcgtgt	gggtgcgcaa	ccacaacggc	2820
atcgccgaca	acgccaccat	ggtgcgcgtg	gacgtgttcg	agaagggcga	caagtactac	2880
ctggtgcca	tctacagctg	gcagggtggc	aagggcatcc	tggccgaccg	cgccgtggtg	2940
cagggcaagg	acgaggagga	ctggcagctg	atcgacgaca	gcttcaactt	caagttcagc	3000
ctgcacccca	acgacctggc	ggagggtgatc	accaagaagg	cccgcagtgt	cggctacttc	3060
gccagctgcc	accgcggcac	cggcaacatc	aacatccgca	tccacgacct	ggaccacaag	3120
atcggaaga	acggcatcct	ggagggcatc	ggcgtgaaga	ccgccctgag	cttcagagaag	3180

8009W000_SequenceListing.txt

taccagatcg acgagctggg caaggagatc cgcccctgcc gcctgaagaa gcgccctcct 3240
gtgcgctaa 3249

<210> 14
<211> 859
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Cas9 consensus sequence derived from Sm, Sp, St, and Li
Cas9 sequences

<220>
<221> misc_feature
<222> (4)..(4)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (8)..(18)
<223> N-terminal RuvC-like domain

<220>
<221> misc_feature
<222> (21)..(21)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (23)..(23)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (24)..(24)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (26)..(26)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (29)..(31)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (33)..(33)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (36)..(36)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature

8009W000_SequenceListing.txt

<222> (45)..(45)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (54)..(54)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (63)..(63)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (71)..(71)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (75)..(75)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (76)..(76)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (78)..(80)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (82)..(82)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (84)..(84)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (85)..(85)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (89)..(89)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (90)..(90)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (98)..(98)
<223> Xaa can be any naturally occurring amino acid

8009W000_SequenceListing.txt

```

<220>
<221> misc_feature
<222> (100)..(100)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (106)..(106)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (113)..(113)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (116)..(116)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (125)..(125)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (126)..(126)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (128)..(133)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (135)..(135)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (137)..(137)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (139)..(147)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (153)..(155)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (157)..(157)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (159)..(159)

```

<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (161)..(161)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (163)..(163)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (166)..(168)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (170)..(170)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (171)..(171)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (173)..(175)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (177)..(177)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (183)..(183)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (185)..(187)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (189)..(189)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (192)..(195)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (198)..(198)
<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature
<222> (199)..(199)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (202)..(202)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (206)..(206)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (207)..(207)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (210)..(210)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (212)..(212)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (213)..(213)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (219)..(219)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (220)..(220)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (222)..(222)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (224)..(224)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (226)..(226)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (236)..(236)
<223> Xaa can be any naturally occurring amino acid

8009W000_SequenceListing.txt

```

<220>
<221> misc_feature
<222> (240)..(240)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (241)..(241)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (244)..(246)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (248)..(248)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (249)..(249)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (250)..(250)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (252)..(254)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (256)..(256)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (257)..(257)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (268)..(268)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (271)..(271)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (273)..(273)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature

```

<222> (277)..(277)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (280)..(280)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (281)..(281)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (283)..(283)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (289)..(289)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (290)..(290)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (292)..(294)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (301)..(301)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (308)..(308)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (319)..(322)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (328)..(328)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (329)..(329)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (335)..(337)
<223> Each Xaa can independently be any naturally occurring amino acid

8009W000_SequenceListing.txt

```

<220>
<221> misc_feature
<222> (346)..(346)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (347)..(347)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (356)..(361)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (363)..(363)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (370)..(373)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (375)..(375)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (376)..(376)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (379)..(379)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (386)..(390)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (393)..(393)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (395)..(395)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (396)..(396)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (398)..(398)

```

<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (400)..(400)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (403)..(403)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (407)..(407)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (410)..(410)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (411)..(411)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (413)..(416)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (418)..(418)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (422)..(422)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (428)..(428)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (431)..(431)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (433)..(433)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (437)..(439)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>

<221> misc_feature
<222> (445)..(445)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (451)..(451)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (456)..(456)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (459)..(459)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (465)..(469)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (481)..(481)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (482)..(482)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (484)..(484)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (490)..(490)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (494)..(494)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (495)..(502)
<223> RuvC-like domain

<220>
<221> misc_feature
<222> (497)..(497)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (506)..(506)
<223> Xaa can be any naturally occurring amino acid

8009W000_SequenceListing.txt

```

<220>
<221> misc_feature
<222> (510)..(510)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (513)..(513)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (514)..(514)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (517)..(517)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (520)..(520)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (525)..(525)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (526)..(526)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (529)..(529)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (531)..(531)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (532)..(532)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (534)..(534)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (542)..(542)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature

```

8009W000_SequenceListing.txt

<222> (546)..(546)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (547)..(547)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (553)..(553)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (555)..(575)
<223> HNH-like domain

<220>
<221> misc_feature
<222> (556)..(556)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (560)..(560)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (563)..(563)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (565)..(565)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (567)..(567)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (579)..(579)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (582)..(582)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (583)..(583)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (585)..(585)
<223> Xaa can be any naturally occurring amino acid

8009W000_SequenceListing.txt

```

<220>
<221> misc_feature
<222> (588)..(588)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (590)..(590)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (592)..(592)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (594)..(596)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (610)..(610)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (616)..(616)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (628)..(628)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (631)..(631)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (633)..(633)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (634)..(634)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (636)..(636)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (638)..(641)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (643)..(645)

```

<223> Each Xaa can independently be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (653)..(653)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (657)..(657)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (659)..(659)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (660)..(660)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (665)..(665)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (666)..(666)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (668)..(668)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (669)..(669)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (670)..(677)

<223> RuvC-like domain

<220>

<221> misc_feature

<222> (680)..(680)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (681)..(681)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (683)..(683)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature
<222> (686)..(686)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (696)..(696)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (697)..(697)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (704)..(704)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (708)..(708)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (710)..(710)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (711)..(711)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (714)..(714)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (717)..(720)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (722)..(722)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (725)..(725)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (727)..(727)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (733)..(735)
<223> Each Xaa can independently be any naturally occurring amino acid

8009W000_SequenceListing.txt

```

<220>
<221> misc_feature
<222> (740)..(740)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (742)..(742)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (749)..(754)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (758)..(761)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (763)..(768)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (771)..(771)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (774)..(777)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (782)..(782)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (784)..(786)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (788)..(788)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (790)..(790)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (795)..(795)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature

```

8009W000_SequenceListing.txt

```

<222> (799)..(799)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (801)..(801)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (802)..(802)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (804)..(804)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (806)..(813)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (815)..(818)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (820)..(820)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (823)..(827)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (829)..(829)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (830)..(830)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (832)..(832)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (835)..(837)
<223> Each Xaa can independently be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (842)..(844)
<223> Each Xaa can independently be any naturally occurring amino acid

```

8009W000_SequenceListing.txt

```

<220>
<221> misc_feature
<222> (846)..(846)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (848)..(848)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (851)..(851)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (857)..(857)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (857)..(857)
<223> Xaa can be any naturally occurring amino acid

<400> 14

Met Lys Tyr Xaa Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp
1          5          10          15

Ala Val Thr Asp Xaa Tyr Xaa Xaa Lys Xaa Lys Gly Xaa Xaa Xaa Ile
20          25          30

Xaa Lys Asn Xaa Gly Leu Phe Asp Gly Thr Ala Arg Xaa Arg Thr Ala
35          40          45

Arg Arg Arg Arg Arg Xaa Asn Arg Ile Tyr Leu Gln Ile Phe Xaa Glu
50          55          60

Met Asp Phe Phe Arg Leu Xaa Ser Phe Val Xaa Xaa Lys Xaa Xaa Xaa
65          70          75          80

Pro Xaa Phe Xaa Xaa Glu Tyr His Xaa Xaa Pro Thr Ile Tyr His Leu
85          90          95

Arg Xaa Leu Xaa Lys Asp Leu Arg Leu Xaa Tyr Leu Ala Leu Ala His
100         105         110

Xaa Ile Lys Xaa Arg Gly Asn Phe Leu Ile Glu Gly Xaa Xaa Asn Xaa
115         120         125

Xaa Xaa Xaa Xaa Xaa Tyr Xaa Phe Xaa Ile Xaa Xaa Xaa Xaa Xaa
130         135         140

```

8009W000_SequenceListing.txt

Xaa Xaa Xaa Pro Glu Lys Gly Phe Xaa Xaa Xaa Leu Xaa Gly Xaa Phe
 145 150 155 160
 Xaa Phe Xaa Leu Glu Xaa Xaa Xaa Lys Xaa Xaa Tyr Xaa Xaa Xaa Leu
 165 170 175
 Xaa Leu Leu Ile Gly Asp Xaa Tyr Xaa Xaa Xaa Phe Xaa Ala Lys Xaa
 180 185 190
 Xaa Xaa Xaa Leu Ser Xaa Xaa Val Thr Xaa Ala Leu Ser Xaa Xaa Met
 195 200 205
 Ile Xaa Arg Xaa Xaa His Asp Leu Leu Lys Xaa Xaa Tyr Xaa Glu Xaa
 210 215 220
 Phe Xaa Lys Gly Tyr Ala Gly Tyr Ile Asp Gly Xaa Gln Phe Tyr Xaa
 225 230 235 240
 Xaa Lys Leu Xaa Xaa Xaa Gly Xaa Xaa Xaa Lys Xaa Xaa Xaa Glu Xaa
 245 250 255
 Xaa Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Xaa Ile Pro Xaa Gln
 260 265 270
 Xaa His Leu Glu Xaa Ala Ile Xaa Xaa Gln Xaa Tyr Pro Phe Leu Asn
 275 280 285
 Xaa Xaa Ile Xaa Xaa Xaa Thr Phe Arg Ile Pro Tyr Xaa Val Gly Pro
 290 295 300
 Leu Ala Gly Xaa Ser Phe Ala Trp Arg Lys Ile Pro Trp Asn Xaa Xaa
 305 310 315 320
 Xaa Xaa Asp Ser Ala Phe Ile Xaa Xaa Met Thr Asp Leu Pro Xaa Xaa
 325 330 335
 Xaa Val Leu Pro Lys His Ser Leu Tyr Xaa Xaa Val Tyr Asn Glu Leu
 340 345 350
 Thr Lys Val Xaa Xaa Xaa Xaa Xaa Xaa Lys Xaa Ile Phe Lys Arg Lys
 355 360 365
 Val Xaa Xaa Xaa Xaa Gly Xaa Xaa Phe Asn Xaa Ser Thr Tyr His Asp
 370 375 380
 Leu Xaa Xaa Xaa Xaa Xaa Leu Asp Xaa Asn Xaa Xaa Glu Xaa Ile Xaa
 385 390 395 400

8009W000_SequenceListing.txt

Leu Thr Xaa Phe Glu Asp Xaa Met Ile Xaa Xaa Leu Xaa Xaa Xaa Xaa
 405 410 415
 Lys Xaa Leu Arg Arg Xaa Tyr Thr Gly Trp Gly Xaa Leu Ser Xaa Leu
 420 425 430
 Xaa Gly Ile Arg Xaa Xaa Xaa Ser Thr Ile Leu Asp Xaa Leu Asp Asn
 435 440 445
 Arg Asn Xaa Met Gln Leu Ile Xaa Asp Leu Xaa Phe Lys Ile Lys Gln
 450 455 460
 Xaa Xaa Xaa Xaa Xaa Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln
 465 470 475 480
 Xaa Xaa Lys Xaa Val Asp Glu Leu Val Xaa Met Gly Pro Xaa Ile Val
 485 490 495
 Xaa Glu Met Ala Arg Glu Asn Gln Thr Xaa Gly Asn Ser Xaa Arg Lys
 500 505 510
 Xaa Xaa Lys Glu Xaa Gly Ser Xaa Ile Leu Lys Glu Xaa Xaa Asn Leu
 515 520 525
 Xaa Asn Xaa Xaa Leu Xaa Leu Tyr Tyr Leu Gln Asn Gly Xaa Asp Met
 530 535 540
 Tyr Xaa Xaa Leu Asp Ile Leu Ser Xaa Tyr Asp Xaa Asp His Ile Xaa
 545 550 555 560
 Pro Gln Xaa Phe Xaa Asp Xaa Ser Ile Asp Asn Val Leu Ser Asn Arg
 565 570 575
 Lys Asp Xaa Val Pro Xaa Xaa Val Xaa Lys Lys Xaa Trp Xaa Leu Xaa
 580 585 590
 Leu Xaa Xaa Xaa Arg Lys Phe Asp Leu Thr Lys Ala Glu Arg Gly Gly
 595 600 605
 Leu Xaa Asp Lys Ala Phe Ile Xaa Arg Gln Leu Val Glu Thr Arg Gln
 610 615 620
 Ile Thr Lys Xaa Val Ala Xaa Leu Xaa Xaa Asn Xaa Asp Xaa Xaa Xaa
 625 630 635 640
 Xaa Val Xaa Xaa Xaa Thr Leu Lys Ser Leu Val Ser Xaa Phe Arg Lys
 645 650 655

8009W000_SequenceListing.txt

Xaa Phe Xaa Xaa Leu Tyr Lys Val Xaa Xaa Asn Xaa Xaa His His Ala
660 665 670

His Asp Ala Tyr Leu Asn Val Xaa Xaa Leu Xaa Tyr Pro Xaa Leu Glu
675 680 685

Glu Phe Val Tyr Gly Asp Tyr Xaa Xaa Lys Ala Thr Lys Phe Tyr Xaa
690 695 700

Asn Ile Met Xaa Phe Xaa Xaa Gly Glu Xaa Trp Lys Xaa Xaa Xaa Xaa
705 710 715 720

Val Xaa Met Gln Xaa Asn Xaa Val Lys Lys Glu Gln Xaa Xaa Xaa Pro
725 730 735

Lys Asn Ser Xaa Leu Xaa Lys Asp Lys Tyr Gly Gly Xaa Xaa Xaa Xaa
740 745 750

Xaa Xaa Lys Gly Lys Xaa Xaa Xaa Xaa Ile Xaa Xaa Xaa Xaa Xaa
755 760 765

Phe Leu Xaa Gly Tyr Xaa Xaa Xaa Xaa Leu Pro Lys Tyr Xaa Leu Xaa
770 775 780

Xaa Xaa Gly Xaa Arg Xaa Leu Ala Ser Glu Xaa Lys Gly Asn Xaa Leu
785 790 795 800

Xaa Xaa Leu Xaa Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Glu Xaa Xaa
805 810 815

Xaa Xaa Phe Xaa Ala Asn Xaa Xaa Xaa Xaa Xaa Leu Xaa Xaa Gly Xaa
820 825 830

Ala Phe Xaa Xaa Xaa Ile Arg Arg Tyr Xaa Xaa Xaa Thr Xaa Ile Xaa
835 840 845

Gln Ser Xaa Thr Gly Leu Tyr Glu Xaa Arg Leu
850 855

<210> 15
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> RuvC-like domain

8009W000_SequenceListing.txt

<220>
 <221> VARIANT
 <222> (2)..(2)
 <223> Xaa is Val or His

<220>
 <221> VARIANT
 <222> (3)..(3)
 <223> Xaa is Ile, Leu, or Val

<220>
 <221> VARIANT
 <222> (5)..(5)
 <223> Xaa is Met or Thr

<400> 15

Ile Xaa Xaa Glu Xaa Ala Arg Glu
 1 5

<210> 16
 <211> 8
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> RuvC-like domain

<220>
 <221> VARIANT
 <222> (3)..(3)
 <223> Xaa is Ile, Leu, or Val

<400> 16

Ile Val Xaa Glu Met Ala Arg Glu
 1 5

<210> 17
 <211> 8
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> RuvC-like domain

<220>
 <221> VARIANT
 <222> (4)..(4)
 <223> Xaa is His or Leu

<220>
 <221> VARIANT
 <222> (7)..(7)
 <223> Xaa is Arg or Val

<220>
 <221> VARIANT
 <222> (8)..(8)

<223> Xaa is Glu or Val

<400> 17

His His Ala Xaa Asp Ala Xaa Xaa
1 5

<210> 18

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> RuvC-like domain

<400> 18

His His Ala His Asp Ala Tyr Leu
1 5

<210> 19

<211> 30

<212> PRT

<213> Artificial Sequence

<220>

<223> N-terminal RuvC-like domain

<220>

<221> VARIANT

<222> (2)..(2)

<223> Xaa is Lys or Pro

<220>

<221> VARIANT

<222> (4)..(4)

<223> Xaa is Val, Leu, Ile, or Phe

<220>

<221> VARIANT

<222> (5)..(5)

<223> Xaa is Gly, Ala, or Ser

<220>

<221> VARIANT

<222> (6)..(6)

<223> Xaa is Leu, Ile, Val, or Phe

<220>

<221> MOD_RES

<222> (7)..(26)

<223> N-terminal RuvC-like domain, each Xaa can be any amino acid or absent, region may encompass 5-20 residues

<220>

<221> VARIANT

<222> (29)..(29)

<223> Xaa is Asp, Glu, Asn, or Gln

<400> 19

8009W000_SequenceListing.txt

Lys Xaa Tyr Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Thr Asp Xaa Tyr
20 25 30

<210> 20
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> N-terminal RuvC-like domain

<220>
<221> VARIANT
<222> (2)..(2)
<223> Xaa is Ile, Val, Met, Leu, or Thr

<220>
<221> VARIANT
<222> (4)..(4)
<223> Xaa is Thr, Ile, Val, Ser, Asn, Tyr, Glu, or Leu

<220>
<221> VARIANT
<222> (5)..(5)
<223> Xaa is Asn, Ser, Gly, Ala, Asp, Thr, Arg, Met, or Phe

<220>
<221> VARIANT
<222> (6)..(6)
<223> Xaa is Ser, Tyr, Asn, or Phe

<220>
<221> VARIANT
<222> (7)..(7)
<223> Xaa is Val, Ile, Leu, Cys, Thr, or Phe

<220>
<221> VARIANT
<222> (9)..(9)
<223> Xaa is Trp, Phe, Val, Tyr, Ser, or Leu

<220>
<221> VARIANT
<222> (10)..(10)
<223> Xaa is Ala, Ser, Cys, Val, or Gly

<220>
<221> VARIANT
<222> (11)..(11)
<223> Xaa is Val, Ile, Leu, Ala, Met, or His

<220>
<221> VARIANT
<222> (12)..(12)
<223> Any amino acid or absent

8009W000_SequenceListing.txt

<400> 20

Asp Xaa Gly Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa
 1 5 10

<210> 21

<211> 12

<212> PRT

<213> Artificial Sequence

<220>

<223> N-terminal RuvC-like domain

<220>

<221> VARIANT

<222> (2)..(2)

<223> Xaa is Ile, Val, Met, Leu, or Thr

<220>

<221> VARIANT

<222> (4)..(4)

<223> Xaa is Thr, Ile, Val, Ser, Asn, Tyr, Glu, or Leu

<220>

<221> VARIANT

<222> (5)..(5)

<223> Xaa is Asn, Ser, Gly, Ala, Asp, Thr, Arg, Met, or Phe

<220>

<221> VARIANT

<222> (7)..(7)

<223> Xaa is Val, Ile, Leu, Cys, Thr, or Phe

<220>

<221> VARIANT

<222> (9)..(9)

<223> Xaa is Trp, Phe, Val, Tyr, Ser, or Leu

<220>

<221> VARIANT

<222> (10)..(10)

<223> Xaa is Ala, Ser, Cys, Val, or Gly

<220>

<221> VARIANT

<222> (11)..(11)

<223> Xaa is Val, Ile, Leu, Ala, Met, or His

<220>

<221> VARIANT

<222> (12)..(12)

<223> Any amino acid or absent

<400> 21

Asp Xaa Gly Xaa Xaa Ser Xaa Gly Xaa Xaa Xaa Xaa
 1 5 10

<210> 22

<211> 12

<212> PRT
 <213> Artificial Sequence

<220>
 <223> N-terminal RuvC-like domain

<220>
 <221> VARIANT
 <222> (4)..(4)
 <223> Xaa is Thr, Ile, Val, Ser, Asn, Tyr, Glu, or Leu

<220>
 <221> VARIANT
 <222> (5)..(5)
 <223> Xaa is Asn, Ser, Gly, Ala, Asp, Thr, Arg, Met, or Phe

<220>
 <221> VARIANT
 <222> (11)..(11)
 <223> Xaa is Val, Ile, Leu, Ala, Met, or His

<220>
 <221> VARIANT
 <222> (12)..(12)
 <223> Any amino acid or absent

<400> 22

Asp Ile Gly Xaa Xaa Ser Val Gly Trp Ala Xaa Xaa
 1 5 10

<210> 23
 <211> 12
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> N-terminal RuvC-like domain

<220>
 <221> MOD_RES
 <222> (12)..(12)
 <223> Any non-polar alkyl amino acid or a hydroxyl amino acid

<400> 23

Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Xaa
 1 5 10

<210> 24
 <211> 73
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> HNH-like domain

<220>

```

<221> VARIANT
<222> (8)..(8)
<223> Xaa is Lys or Arg

<220>
<221> VARIANT
<222> (12)..(12)
<223> Xaa is Val or Thr

<220>
<221> VARIANT
<222> (13)..(13)
<223> Xaa is Gly or Asp

<220>
<221> VARIANT
<222> (14)..(14)
<223> Xaa is Glu, Gln, or Asp

<220>
<221> VARIANT
<222> (15)..(15)
<223> Xaa is Glu or Asp

<220>
<221> VARIANT
<222> (19)..(19)
<223> Xaa is Asp, Asn, or His

<220>
<221> VARIANT
<222> (20)..(20)
<223> Xaa is Tyr, Arg, or Asn

<220>
<221> VARIANT
<222> (23)..(23)
<223> Xaa is Gln, Asp, or Asn

<220>
<221> MOD_RES
<222> (25)..(64)
<223> HNH-like domain, each Xaa can be any amino acid or absent, region
may encompass 15-40 residues

<220>
<221> VARIANT
<222> (67)..(67)
<223> Xaa is Gly or Glu

<220>
<221> VARIANT
<222> (69)..(69)
<223> Xaa is Ser or Gly

<220>
<221> VARIANT
<222> (71)..(71)
<223> Xaa is Asp or Asn

<400> 24

```

Leu Tyr Tyr Leu Gln Asn Gly Xaa Asp Met Tyr Xaa Xaa Xaa Xaa Leu

8009W000_SequenceListing.txt

```

1              5              10              15
Asp Ile Xaa Xaa Leu Ser Xaa Tyr Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
      20              25              30
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
      35              40              45
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
      50              55              60
Asn Arg Xaa Lys Xaa Asp Xaa Val Pro
65              70

```

```

<210> 25
<211> 27
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> HNH-like domain

```

```

<220>
<221> VARIANT
<222> (1)..(1)
<223> Xaa is Asp, Glu, Gln, or Asn

```

```

<220>
<221> VARIANT
<222> (2)..(2)
<223> Xaa is Leu, Ile, Arg, Gln, Val, Met, or Lys

```

```

<220>
<221> VARIANT
<222> (3)..(3)
<223> Xaa is Asp or Glu

```

```

<220>
<221> VARIANT
<222> (5)..(5)
<223> Xaa is Ile, Val, Thr, Ala, or Leu

```

```

<220>
<221> VARIANT
<222> (6)..(6)
<223> Xaa is Val, Tyr, Ile, Leu, Phe, or Trp

```

```

<220>
<221> VARIANT
<222> (8)..(8)
<223> Xaa is Gln, His, Arg, Lys, Tyr, Ile, Leu, Phe, or Trp

```

```

<220>
<221> VARIANT
<222> (9)..(9)
<223> Xaa is Ser, Ala, Asp, Thr, or Lys

```

8009W000_SequenceListing.txt

<220>
<221> VARIANT
<222> (10)..(10)
<223> Xaa is Phe, Leu, Val, Lys, Tyr, Met, Ile, Arg, Ala, Glu, Asp, or
Gln

<220>
<221> VARIANT
<222> (11)..(11)
<223> Xaa is Leu, Arg, Thr, Ile, Val, Ser, Cys, Tyr, Lys, Phe, or Gly

<220>
<221> VARIANT
<222> (12)..(12)
<223> Xaa is Lys, Gln, Tyr, Thr, Phe, Leu, Trp, Met, Ala, Glu, Gly, or
Ser

<220>
<221> VARIANT
<222> (13)..(13)
<223> Xaa is Asp, Ser, Asn, Arg, Leu, or Thr

<220>
<221> VARIANT
<222> (14)..(14)
<223> Xaa is Asp, Asn, or Ser

<220>
<221> VARIANT
<222> (15)..(15)
<223> Xaa is Ser, Ala, Thr, Gly, or Arg

<220>
<221> VARIANT
<222> (16)..(16)
<223> Xaa is Ile, Leu, Phe, Ser, Arg, Tyr, Gln, Trp, Asp, Lys, or His

<220>
<221> VARIANT
<222> (17)..(17)
<223> Xaa is Asp, Ser, Ile, Asn, Glu, Ala, His, Phe, Leu, Gln, Met,
Gly, Tyr, or Val

<220>
<221> VARIANT
<222> (19)..(19)
<223> Xaa is Lys, Leu, Arg, Met, Thr, or Phe

<220>
<221> VARIANT
<222> (20)..(20)
<223> Xaa is Val, Leu, Ile, Ala, or Thr

<220>
<221> VARIANT
<222> (21)..(21)
<223> Xaa is Leu, Ile, Val, or Ala

<220>
<221> VARIANT
<222> (22)..(22)
<223> Xaa is Thr, Val, Cys, Glu, Ser, or Ala

8009W000_SequenceListing.txt

<220>
 <221> VARIANT
 <222> (23)..(23)
 <223> Xaa is Arg, Phe, Thr, Trp, Glu, Leu, Asn, Cys, Lys, Val, Ser,
 Gln, Ile, Tyr, His, or Ala

<220>
 <221> VARIANT
 <222> (24)..(24)
 <223> Xaa is Ser, Pro, Arg, Lys, Asn, Ala, His, Gln, Gly, or Leu

<220>
 <221> VARIANT
 <222> (25)..(25)
 <223> Xaa is Asp, Gly, Thr, Asn, Ser, Lys, Ala, Ile, Glu, Leu, Gln,
 Arg, or Tyr

<220>
 <221> VARIANT
 <222> (26)..(26)
 <223> Xaa is Lys, Val, Ala, Glu, Tyr, Ile, Cys, Leu, Ser, Thr, Gly,
 Lys, Met, Asp, or Phe

<400> 25

Xaa Xaa Xaa His Xaa Xaa Pro Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 1 5 10 15

Xaa Asn Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asn
 20 25

<210> 26
 <211> 27
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> HNH-like domain

<220>
 <221> VARIANT
 <222> (1)..(1)
 <223> Xaa is Asp or Glu

<220>
 <221> VARIANT
 <222> (2)..(2)
 <223> Xaa is Leu, Ile, Arg, Gln, Val, Met, or Lys

<220>
 <221> VARIANT
 <222> (3)..(3)
 <223> Xaa is Asp or Glu

<220>
 <221> VARIANT
 <222> (5)..(5)
 <223> Xaa is Ile, Val, Thr, Ala, or Leu

<220>

8009W000_SequenceListing.txt

<221> VARIANT
<222> (6)..(6)
<223> Xaa is Val, Tyr, Ile, Leu, Phe, or Trp

<220>
<221> VARIANT
<222> (8)..(8)
<223> Xaa is Gln, His, Arg, Lys, Tyr, Ile, Leu, Phe, or Trp

<220>
<221> VARIANT
<222> (10)..(10)
<223> Xaa is Phe, Leu, Val, Lys, Tyr, Met, Ile, Arg, Ala, Glu, Asp, or Gln

<220>
<221> VARIANT
<222> (11)..(11)
<223> Xaa is Leu, Arg, Thr, Ile, Val, Ser, Cys, Tyr, Lys, Phe, or Gly

<220>
<221> VARIANT
<222> (12)..(12)
<223> Xaa is Lys, Gln, Tyr, Thr, Phe, Leu, Trp, Met, Ala, Glu, Gly, or Ser

<220>
<221> VARIANT
<222> (16)..(16)
<223> Xaa is Ile, Leu, Phe, Ser, Arg, Tyr, Gln, Trp, Asp, Lys, or His

<220>
<221> VARIANT
<222> (17)..(17)
<223> Xaa is Asp, Ser, Ile, Asn, Glu, Ala, His, Phe, Leu, Gln, Met, Gly, Tyr, or Val

<220>
<221> VARIANT
<222> (22)..(22)
<223> Xaa is Thr, Val, Cys, Glu, Ser, or Ala

<220>
<221> VARIANT
<222> (23)..(23)
<223> Xaa is Arg, Phe, Thr, Trp, Glu, Leu, Asn, Cys, Lys, Val, Ser, Gln, Ile, Tyr, His, or Ala

<220>
<221> VARIANT
<222> (24)..(24)
<223> Xaa is Ser, Pro, Arg, Lys, Asn, Ala, His, Gln, Gly, or Leu

<220>
<221> VARIANT
<222> (25)..(25)
<223> Xaa is Asp, Gly, Thr, Asn, Ser, Lys, Ala, Ile, Glu, Leu, Gln, Arg, or Tyr

<220>
<221> VARIANT
<222> (26)..(26)
<223> Xaa is Lys, Val, Ala, Glu, Tyr, Ile, Cys, Leu, Ser, Thr, Gly,

Lys, Met, Asp, or Phe

<400> 26

Xaa Xaa Xaa His Xaa Xaa Pro Xaa Ser Xaa Xaa Xaa Asp Asp Ser Xaa
 1 5 10 15

Xaa Asn Lys Val Leu Xaa Xaa Xaa Xaa Xaa Asn
 20 25

<210> 27

<211> 27

<212> PRT

<213> Artificial Sequence

<220>

<223> HNH-like domain

<220>

<221> VARIANT

<222> (1)..(1)

<223> Xaa is Asp or Glu

<220>

<221> VARIANT

<222> (3)..(3)

<223> Xaa is Asp or Glu

<220>

<221> VARIANT

<222> (8)..(8)

<223> Xaa is Gln, His, Arg, Lys, Tyr, Ile, Leu, or Trp

<220>

<221> VARIANT

<222> (10)..(10)

<223> Xaa is Phe, Leu, Val, Lys, Tyr, Met, Ile, Arg, Ala, Glu, Asp, or Gln

<220>

<221> VARIANT

<222> (11)..(11)

<223> Xaa is Leu, Arg, Thr, Ile, Val, Ser, Cys, Tyr, Lys, Phe, or Gly

<220>

<221> VARIANT

<222> (12)..(12)

<223> Xaa is Lys, Gln, Tyr, Thr, Phe, Leu, Trp, Met, Ala, Glu, Gly, or Ser

<220>

<221> VARIANT

<222> (16)..(16)

<223> Xaa is Ile, Leu, Phe, Ser, Arg, Tyr, Gln, Trp, Asp, Lys, or His

<220>

<221> VARIANT

<222> (17)..(17)

<223> Xaa is Asp, Ser, Ile, Asn, Glu, Ala, His, Phe, Leu, Gln, Met, Gly, Tyr, or Val

8009W000_SequenceListing.txt

<220>
 <221> VARIANT
 <222> (23)..(23)
 <223> Xaa is Arg, Phe, Thr, Trp, Glu, Leu, Asn, Cys, Lys, Val, Ser,
 Gln, Ile, Tyr, His, or Ala

<220>
 <221> VARIANT
 <222> (24)..(24)
 <223> Xaa is Ser, Pro, Arg, Lys, Asn, Ala, His, Gln, Gly, or Leu

<220>
 <221> VARIANT
 <222> (25)..(25)
 <223> Xaa is Asp, Gly, Thr, Asn, Ser, Lys, Ala, Ile, Glu, Leu, Gln,
 Arg, or Tyr

<220>
 <221> VARIANT
 <222> (26)..(26)
 <223> Xaa is Lys, Val, Ala, Glu, Tyr, Ile, Cys, Leu, Ser, Thr, Gly,
 Lys, Met, Asp, or Phe

<400> 27

Xaa	Val	Xaa	His	Ile	Val	Pro	Xaa	Ser	Xaa	Xaa	Xaa	Asp	Asp	Ser	Xaa
1				5					10					15	

Xaa	Asn	Lys	Val	Leu	Thr	Xaa	Xaa	Xaa	Xaa	Asn
			20					25		

<210> 28
 <211> 27
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> HNH-like domain

<220>
 <221> VARIANT
 <222> (2)..(2)
 <223> Xaa is Ile or Val

<220>
 <221> VARIANT
 <222> (6)..(6)
 <223> Xaa is Ile or Val

<220>
 <221> VARIANT
 <222> (9)..(9)
 <223> Xaa is Ala or Ser

<220>
 <221> VARIANT
 <222> (11)..(11)
 <223> Xaa is Ile or Leu

8009W000_SequenceListing.txt

```

<220>
<221> VARIANT
<222> (12)..(12)
<223> Xaa is Lys or Thr

<220>
<221> VARIANT
<222> (14)..(14)
<223> Xaa is Asp or Asn

<220>
<221> VARIANT
<222> (19)..(19)
<223> Xaa is Arg, Lys, or Leu

<220>
<221> VARIANT
<222> (22)..(22)
<223> Xaa is Thr or Val

<220>
<221> VARIANT
<222> (23)..(23)
<223> Xaa is Ser or Arg

<220>
<221> VARIANT
<222> (25)..(25)
<223> Xaa is Lys, Asp, or Ala

<220>
<221> VARIANT
<222> (26)..(26)
<223> Xaa is Glu, Lys, Gly, or Asn

<400> 28

Asp Xaa Asp His Ile Xaa Pro Gln Xaa Phe Xaa Xaa Asp Xaa Ser Ile
1          5          10          15

Asp Asn Xaa Val Leu Xaa Xaa Ser Xaa Xaa Asn
          20          25

<210> 29
<211> 116
<212> RNA
<213> Artificial Sequence

<220>
<223> gRNA

<220>
<221> misc_feature
<222> (1)..(20)
<223> a, c, u, g, unknown or other

<220>
<221> misc_feature
<222> (1)..(20)
<223> targeting region

```

8009W000_SequenceListing.txt

```

<220>
<221> misc_feature
<222> (21)..(42)
<223> first complementarity domain

<220>
<221> misc_feature
<222> (43)..(46)
<223> linking domain

<220>
<221> misc_feature
<222> (47)..(70)
<223> second complementarity domain

<400> 29
nnnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguuu uggaaacaaa acagcauagc      60
aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugc          116

<210> 30
<211> 116
<212> RNA
<213> Artificial Sequence

<220>
<223> gRNA

<220>
<221> misc_feature
<222> (1)..(20)
<223> a, c, u, g, unknown or other

<220>
<221> misc_feature
<222> (1)..(20)
<223> targeting region

<220>
<221> misc_feature
<222> (21)..(42)
<223> first complementarity domain

<220>
<221> misc_feature
<222> (43)..(46)
<223> linking domain

<220>
<221> misc_feature
<222> (47)..(70)
<223> second complementarity domain

<400> 30
nnnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguau uggaaacaau acagcauagc      60
aaguuaauau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugc          116

<210> 31

```

```

<211> 96
<212> RNA
<213> Artificial Sequence

<220>
<223> gRNA

<220>
<221> misc_feature
<222> (1)..(20)
<223> a, c, u, g, unknown or other

<220>
<221> misc_feature
<222> (1)..(20)
<223> targeting domain

<220>
<221> misc_feature
<222> (21)..(32)
<223> first complementarity domain

<220>
<221> misc_feature
<222> (33)..(36)
<223> linking domain

<220>
<221> misc_feature
<222> (37)..(50)
<223> second complementarity domain

<220>
<221> misc_feature
<222> (51)..(62)
<223> proximal domain

<400> 31
nnnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuuuuuu aaggcuaguc      60
cguuaucaac uugaaaaaagu ggcaccgagu cggugc                                96

<210> 32
<211> 47
<212> RNA
<213> Artificial Sequence

<220>
<223> gRNA proximal and tail domains derived from S. pyogenes

<400> 32
aaggcuaguc cguuaucaac uugaaaaaagu ggcaccgagu cggugcu                      47

<210> 33
<211> 49
<212> RNA
<213> Artificial Sequence

<220>
<223> gRNA proximal and tail domains

```

8009W000_SequenceListing.txt

<400> 33
aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cgguggugc 49

<210> 34
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> gRNA proximal and tail domains

<400> 34
aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcggau c 51

<210> 35
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> gRNA proximal and tail domains

<400> 35
aaggcuaguc cguuaucaac uugaaaaagu g 31

<210> 36
<211> 18
<212> RNA
<213> Artificial Sequence

<220>
<223> gRNA proximal and tail domains

<400> 36
aaggcuaguc cguuauca 18

<210> 37
<211> 12
<212> RNA
<213> Artificial Sequence

<220>
<223> gRNA proximal and tail domains

<400> 37
aaggcuaguc cg 12

<210> 38
<211> 102
<212> RNA
<213> Artificial Sequence

<220>
<223> Unimolecular gRNA derived from S. aureus

<220>

8009W000_SequenceListing.txt

```

<221> misc_feature
<222> (1)..(20)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (1)..(20)
<223> Targeting domain

<400> 38
nnnnnnnnnnn nnnnnnnnnn guuuuaguac ucuggaaaca gaaucuacua aaacaaggca      60
aaaugccgug uuuaucucgu caacuuguug gcgagauuuu uu                        102

<210> 39
<211> 42
<212> RNA
<213> Artificial

<220>
<223> Modular gRNA derived from S. pyogenes

<220>
<221> misc_feature
<222> (1)..(20)
<223> a, c, u, g, unknown or other

<220>
<221> misc_feature
<222> (1)..(20)
<223> Targeting domain

<220>
<221> misc_feature
<222> (21)..(42)
<223> First complementarity domain

<400> 39
nnnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguuu ug                        42

<210> 40
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Modular gRNA derived from S. pyogenes

<220>
<221> misc_feature
<222> (1)..(9)
<223> 5' extension domain

<220>
<221> misc_feature
<222> (10)..(33)
<223> Second complementarity domain

<220>

```

8009W000_SequenceListing.txt

<221> misc_feature
 <222> (34)..(45)
 <223> Proximal domain

 <220>
 <221> misc_feature
 <222> (46)..(85)
 <223> Tail domain

 <400> 40
 ggaaccauuc aaaacagcau agcaaguuaa aauaaaggcua guccguuauc aacuugaaaa 60
 aguggcaccg agucggugcu uuuuuu 85

 <210> 41
 <211> 62
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Unimolecular gRNA derived from *S. pyogenes*

 <220>
 <221> misc_feature
 <222> (1)..(20)
 <223> a, c, u, g, unknown or other

 <220>
 <221> misc_feature
 <222> (1)..(20)
 <223> Targeting domain

 <220>
 <221> misc_feature
 <222> (21)..(32)
 <223> First complementarity domain

 <220>
 <221> misc_feature
 <222> (33)..(36)
 <223> Linking domain

 <220>
 <221> misc_feature
 <222> (37)..(50)
 <223> Second complementarity domain

 <220>
 <221> misc_feature
 <222> (51)..(62)
 <223> Proximal domain

 <400> 41
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
 cg 62

 <210> 42
 <211> 102
 <212> RNA

<213> Artificial Sequence

<220>

<223> Unimolecular gRNA derived from *S. pyogenes*

<220>

<221> misc_feature

<222> (1)..(20)

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

<220>

<221> misc_feature

<222> (1)..(20)

<223> Targeting domain

<220>

<221> misc_feature

<222> (21)..(32)

<223> First complementarity domain

<220>

<221> misc_feature

<222> (33)..(36)

<223> Linking domain

<220>

<221> misc_feature

<222> (37)..(50)

<223> Second complementarity domain

<220>

<221> misc_feature

<222> (51)..(62)

<223> Proximal domain

<220>

<221> misc_feature

<222> (63)..(102)

<223> Tail domain

<400> 42

nnnnnnnnnnn nnnnnnnnnnn guuuuagagc uagaaaauagc aaguuaaaaau aaggcuaguc 60

cguuaucaac uugaaaaaagu ggcaccgagu cggugcuuuu uu 102

<210> 43

<211> 75

<212> RNA

<213> Artificial Sequence

<220>

<223> Unimolecular gRNA derived from *S. pyogenes*

<220>

<221> misc_feature

<222> (1)..(20)

<223> Targeting domain

<220>

<221> misc_feature

```

<222> (1)..(20)
<223> a, c, u, g, unknown or other

<220>
<221> misc_feature
<222> (21)..(36)
<223> First complementarity domain

<220>
<221> misc_feature
<222> (37)..(40)
<223> Linking domain

<220>
<221> misc_feature
<222> (41)..(58)
<223> Second complementarity domain

<220>
<221> misc_feature
<222> (59)..(70)
<223> Proximal domain

<220>
<221> misc_feature
<222> (71)..(75)
<223> Tail domain

<400> 43
nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcugaaa agcauagcaa guuaaaauaa      60
ggcuaguccg uuauuc                                                         75

<210> 44
<211> 87
<212> RNA
<213> Artificial Sequence

<220>
<223> Unimolecular gRNA derived from S. pyogenes

<220>
<221> misc_feature
<222> (1)..(20)
<223> a, c, u, g, unknown or other

<220>
<221> misc_feature
<222> (1)..(20)
<223> Targeting domain

<220>
<221> misc_feature
<222> (21)..(32)
<223> First complementarity domain

<220>
<221> misc_feature
<222> (43)..(46)
<223> Linking domain

```

8009W000_SequenceListing.txt

```

<220>
<221> misc_feature
<222> (57)..(70)
<223> Second complementarity domain

<220>
<221> misc_feature
<222> (71)..(82)
<223> Proximal domain

<220>
<221> misc_feature
<222> (83)..(87)
<223> Tail domain

<400> 44
nnnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguuu uggaacaaa acagcauagc      60
aaguuaaaaau aaggcuaguc cguauc                                           87

<210> 45
<211> 42
<212> RNA
<213> Artificial Sequence

<220>
<223> Modular gRNA derived from S. thermophilus

<220>
<221> misc_feature
<222> (1)..(20)
<223> a, c, u, g, unknown or other

<220>
<221> misc_feature
<222> (1)..(20)
<223> Targeting domain

<220>
<221> misc_feature
<222> (21)..(42)
<223> First complementarity domain

<400> 45
nnnnnnnnnnn nnnnnnnnnn guuuuagagc uguguuguuu cg                        42

<210> 46
<211> 78
<212> RNA
<213> Artificial Sequence

<220>
<223> Modular gRNA derived from S. thermophilus

<220>
<221> misc_feature
<222> (1)..(3)
<223> 5' extension domain

```

8009W000_SequenceListing.txt

```

<220>
<221> misc_feature
<222> (4)..(27)
<223> Second complementarity domain

<220>
<221> misc_feature
<222> (28)..(40)
<223> Proximal domain

<220>
<221> misc_feature
<222> (41)..(78)
<223> Tail domain

<400> 46
gggcgaaaca acacagcgag uaaaaauaag gcuuaguccg uacucaacuu gaaaaggugg      60
caccgauucg guguuuuu                                                    78

<210> 47
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Modular gRNA derived from S. pyogenes

<400> 47
gaaccauuca aaacagcaua gcaaguuaaa auaaggcuag uccguuauca acuugaaaaa      60
guggcaccga gucggugcuu uuuuuu                                           85

<210> 48
<211> 96
<212> RNA
<213> Artificial Sequence

<220>
<223> gRNA from S. pyogenes

<220>
<221> misc_feature
<222> (1)..(20)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (1)..(20)
<223> Targeting domain

<400> 48
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc      60
cguaaucaac uugaaaaagu ggcaccgagu cggugc                                96

<210> 49
<211> 96
<212> RNA

```

<213> Artificial Sequence

<220>

<223> gRNA

<220>

<221> misc_feature

<222> (1)..(20)

<223> n is a, c, g, or u

<220>

<221> misc_feature

<222> (1)..(20)

<223> Targeting domain

<220>

<221> misc_feature

<222> (21)..(32)

<223> First complementarity domain

<220>

<221> misc_feature

<222> (33)..(36)

<223> Linking domain

<220>

<221> misc_feature

<222> (37)..(50)

<223> Second complementarity domain

<220>

<221> misc_feature

<222> (51)..(62)

<223> Proximal domain

<400> 49

nnnnnnnnnnn nnnnnnnnnn guauuagagc uagaaauagc aaguuaauau aaggcuaguc 60

cguuaucaac uugaaaaaagu ggcaccgagu cggugc 96

<210> 50

<211> 104

<212> RNA

<213> Artificial Sequence

<220>

<223> gRNA

<220>

<221> misc_feature

<222> (1)..(20)

<223> n is a, c, g, or u

<220>

<221> misc_feature

<222> (1)..(20)

<223> Targeting domain

<220>

<221> misc_feature

8009W000_SequenceListing.txt

```

<222> (21)..(36)
<223> First complementarity domain

<220>
<221> misc_feature
<222> (37)..(40)
<223> Linking domain

<220>
<221> misc_feature
<222> (41)..(58)
<223> Second complementarity domain

<400> 50
nnnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcugaaa agcauagcaa guuaaaaaua      60
ggcuaguccg uuaucaacuu gaaaaagugg caccgagucg gugc                          104

<210> 51
<211> 106
<212> RNA
<213> Artificial Sequence

<220>
<223> gRNA

<220>
<221> misc_feature
<222> (1)..(20)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (1)..(20)
<223> Targeting domain

<220>
<221> misc_feature
<222> (21)..(37)
<223> First complementarity domain

<220>
<221> misc_feature
<222> (38)..(41)
<223> Linking domain

<220>
<221> misc_feature
<222> (42)..(60)
<223> Second complementarity domain

<400> 51
nnnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuggaa acagcauagc aaguuaaaau      60
aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cgugc                          106

<210> 52
<211> 12
<212> PRT
<213> Peptoniphilus duerdenii

```

8009W000_SequenceListing.txt

<400> 52

Asp Ile Gly Thr Ala Ser Val Gly Trp Ala Val Thr
1 5 10

<210> 53

<211> 12

<212> PRT

<213> Treponema denticola

<400> 53

Asp Val Gly Thr Gly Ser Val Gly Trp Ala Val Thr
1 5 10

<210> 54

<211> 12

<212> PRT

<213> S. mutans

<400> 54

Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Val
1 5 10

<210> 55

<211> 12

<212> PRT

<213> S. pyogenes

<400> 55

Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile
1 5 10

<210> 56

<211> 12

<212> PRT

<213> L. innocua

<400> 56

Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Leu
1 5 10

<210> 57

<211> 12

<212> PRT

<213> Flavobacterium branchiophilum FL-15

<400> 57

Asp Leu Gly Thr Asn Ser Ile Gly Trp Ala Val Val
1 5 10

<210> 58

8009W000_SequenceListing.txt

```

<211> 11
<212> PRT
<213> Pedobacter glucosidilyticus

<400> 58
Asp Leu Gly Thr Asn Ser Ile Gly Trp Ala Ile
1          5          10

<210> 59
<211> 12
<212> PRT
<213> Bacteroides fragilis, NCTC 9343

<400> 59
Asp Leu Gly Thr Asn Ser Ile Gly Trp Ala Leu Val
1          5          10

<210> 60
<211> 12
<212> PRT
<213> Fusobacterium nucleatum

<400> 60
Asp Ile Gly Thr Asn Ser Val Gly Trp Cys Val Thr
1          5          10

<210> 61
<211> 12
<212> PRT
<213> Acidaminococcus sp. D21

<400> 61
Asp Ile Gly Thr Asn Ser Val Gly Tyr Ala Val Thr
1          5          10

<210> 62
<211> 12
<212> PRT
<213> Coprococcus catus GD-7

<400> 62
Asp Met Gly Thr Gly Ser Leu Gly Trp Ala Val Thr
1          5          10

<210> 63
<211> 12
<212> PRT
<213> Oenococcus kitaharae DSM 17330

<400> 63
Asp Ile Gly Thr Ser Ser Val Gly Trp Ala Ala Ile
1          5          10

```


8009W000_SequenceListing.txt

<210> 64
 <211> 12
 <212> PRT
 <213> Catenibacterium mitsuokai DSM 15897

<400> 64

Asp Leu Gly Thr Gly Ser Val Gly Trp Ala Val Val
 1 5 10

<210> 65
 <211> 12
 <212> PRT
 <213> Mycoplasma gallisepticum str. F

<400> 65

Asp Leu Gly Val Gly Ser Val Gly Trp Ala Ile Val
 1 5 10

<210> 66
 <211> 12
 <212> PRT
 <213> Mycoplasma ovipneumoniae SC01

<400> 66

Asp Leu Gly Ile Ala Ser Ile Gly Trp Ala Ile Ile
 1 5 10

<210> 67
 <211> 12
 <212> PRT
 <213> Mycoplasma canis PG 14

<400> 67

Asp Leu Gly Ile Ala Ser Val Gly Trp Ala Ile Val
 1 5 10

<210> 68
 <211> 12
 <212> PRT
 <213> Mycoplasma synoviae 53

<400> 68

Asp Leu Gly Val Ala Ser Val Gly Trp Ser Ile Val
 1 5 10

<210> 69
 <211> 12
 <212> PRT
 <213> Eubacterium rectale

<400> 69

8009W000_SequenceListing.txt

Asp Ile Gly Ile Ala Ser Val Gly Trp Ala Ile Leu
1 5 10

<210> 70
<211> 12
<212> PRT
<213> Enterococcus faecalis TX0012

<400> 70

Asp Leu Gly Ile Ser Ser Val Gly Trp Ser Val Ile
1 5 10

<210> 71
<211> 12
<212> PRT
<213> Ilyobacter polytropus DSM 2926

<400> 71

Asp Ile Gly Ile Ala Ser Val Gly Trp Ser Val Ile
1 5 10

<210> 72
<211> 12
<212> PRT
<213> Ruminococcus albus 8

<400> 72

Asp Val Gly Ile Gly Ser Ile Gly Trp Ala Val Ile
1 5 10

<210> 73
<211> 12
<212> PRT
<213> Elusimicrobium minutum Pei191

<400> 73

Asp Leu Gly Val Gly Ser Ile Gly Phe Ala Ile Val
1 5 10

<210> 74
<211> 12
<212> PRT
<213> Akkermansia muciniphila

<400> 74

Asp Ile Gly Tyr Ala Ser Ile Gly Trp Ala Val Ile
1 5 10

<210> 75
<211> 12
<212> PRT

<213> Prevotella ruminicola

<400> 75

Asp Thr Gly Thr Asn Ser Leu Gly Trp Ala Ile Val
1 5 10

<210> 76

<211> 12

<212> PRT

<213> Cand. Puniceispirillum marinum

<400> 76

Asp Leu Gly Thr Asn Ser Ile Gly Trp Cys Leu Leu
1 5 10

<210> 77

<211> 12

<212> PRT

<213> Rhodospirillum rubrum

<400> 77

Asp Ile Gly Thr Asp Ser Leu Gly Trp Ala Val Phe
1 5 10

<210> 78

<211> 12

<212> PRT

<213> Lactobacillus rhamnosus GG

<400> 78

Asp Ile Gly Ser Asn Ser Ile Gly Phe Ala Val Val
1 5 10

<210> 79

<211> 12

<212> PRT

<213> Sphaerochaeta globus str. Buddy

<400> 79

Asp Leu Gly Val Gly Ser Ile Gly Val Ala Val Ala
1 5 10

<210> 80

<211> 12

<212> PRT

<213> Rhodopseudomonas palustris

<400> 80

Asp Leu Gly Ile Ala Ser Cys Gly Trp Gly Val Val
1 5 10

8009W000_SequenceListing.txt

<210> 81
 <211> 12
 <212> PRT
 <213> Mycoplasma mobile 163K
 <400> 81
 Asp Leu Gly Ile Ala Ser Val Gly Trp Cys Leu Thr
 1 5 10

<210> 82
 <211> 12
 <212> PRT
 <213> Streptococcus thermophilus LMD-9
 <400> 82

Asp Ile Gly Ile Gly Ser Val Gly Val Gly Ile Leu
 1 5 10

<210> 83
 <211> 12
 <212> PRT
 <213> Staphylococcus lugdunensis M23590
 <400> 83

Asp Ile Gly Ile Thr Ser Val Gly Tyr Gly Leu Ile
 1 5 10

<210> 84
 <211> 12
 <212> PRT
 <213> Eubacterium dolichum DSM 3991
 <400> 84

Asp Ile Gly Ile Thr Ser Val Gly Phe Gly Ile Ile
 1 5 10

<210> 85
 <211> 12
 <212> PRT
 <213> Lactobacillus coryniformis KCTC 3535
 <400> 85

Asp Val Gly Ile Thr Ser Thr Gly Tyr Ala Val Leu
 1 5 10

<210> 86
 <211> 12
 <212> PRT
 <213> Nitratifractor salsuginis DSM 16511
 <400> 86

Asp Leu Gly Ile Thr Ser Phe Gly Tyr Ala Ile Leu
 Page 97

```

1                5                10

<210> 87
<211> 12
<212> PRT
<213> Bifidobacterium bifidum S17

<400> 87
Asp Ile Gly Asn Ala Ser Val Gly Trp Ser Ala Phe
1                5                10

<210> 88
<211> 12
<212> PRT
<213> Lactobacillus gasseri

<400> 88
Asp Val Gly Thr Asn Ser Cys Gly Trp Val Ala Met
1                5                10

<210> 89
<211> 12
<212> PRT
<213> Acidothermus cellulolyticus 11B

<400> 89
Asp Val Gly Glu Arg Ser Ile Gly Leu Ala Ala Val
1                5                10

<210> 90
<211> 12
<212> PRT
<213> Bifidobacterium longum DJ010A

<400> 90
Asp Val Gly Leu Asn Ser Val Gly Leu Ala Ala Val
1                5                10

<210> 91
<211> 12
<212> PRT
<213> Bifidobacterium dentium

<400> 91
Asp Val Gly Leu Met Ser Val Gly Leu Ala Ala Ile
1                5                10

<210> 92
<211> 12
<212> PRT
<213> Corynebacterium diphtheriae

```

<400> 92

Asp Val Gly Thr Phe Ser Val Gly Leu Ala Ala Ile
 1 5 10

<210> 93

<211> 12

<212> PRT

<213> Staphylococcus pseudintermedius ED99

<400> 93

Asp Ile Gly Thr Gly Ser Val Gly Tyr Ala Cys Met
 1 5 10

<210> 94

<211> 12

<212> PRT

<213> Capnocytophaga ochracea

<400> 94

Asp Leu Gly Thr Thr Ser Ile Gly Phe Ala His Ile
 1 5 10

<210> 95

<211> 12

<212> PRT

<213> Prevotella denticola

<400> 95

Asp Leu Gly Thr Asn Ser Ile Gly Ser Ser Val Arg
 1 5 10

<210> 96

<211> 12

<212> PRT

<213> Ralstonia solanacearum

<400> 96

Asp Ile Gly Thr Asn Ser Ile Gly Trp Ala Val Ile
 1 5 10

<210> 97

<211> 12

<212> PRT

<213> Pasteurella multocida str. Pm70

<400> 97

Asp Leu Gly Ile Ala Ser Val Gly Trp Ala Val Val
 1 5 10

<210> 98

<211> 12

8009W000_SequenceListing.txt

<212> PRT
<213> Comamonas granuli
<400> 98
Asp Ile Gly Ile Ala Ser Val Gly Trp Ala Val Leu
1 5 10

<210> 99
<211> 12
<212> PRT
<213> Helicobacter mustelae 12198
<400> 99
Asp Ile Gly Ile Ala Ser Ile Gly Trp Ala Val Ile
1 5 10

<210> 100
<211> 12
<212> PRT
<213> Agathobacter rectalis
<400> 100
Asp Ile Gly Ile Ala Ser Val Gly Trp Ala Ile Ile
1 5 10

<210> 101
<211> 12
<212> PRT
<213> Clostridium cellulolyticum H10
<400> 101
Asp Val Gly Ile Ala Ser Val Gly Trp Ala Val Ile
1 5 10

<210> 102
<211> 11
<212> PRT
<213> Methylophilus sp. OH31
<400> 102
Asp Ile Gly Ile Ala Ser Val Gly Trp Ala Leu
1 5 10

<210> 103
<211> 12
<212> PRT
<213> Neisseria meningitidis
<400> 103
Asp Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val
1 5 10

8009W000_SequenceListing.txt

<210> 104
 <211> 12
 <212> PRT
 <213> Clostridium perfringens
 <400> 104
 Asp Ile Gly Ile Thr Ser Val Gly Trp Ala Val Ile
 1 5 10

<210> 105
 <211> 12
 <212> PRT
 <213> Wolinella succinogenes DSM 1740
 <400> 105

Asp Leu Gly Ile Ser Ser Leu Gly Trp Ala Ile Val
 1 5 10

<210> 106
 <211> 12
 <212> PRT
 <213> Azospirillum sp. B510
 <400> 106

Asp Leu Gly Thr Asn Ser Ile Gly Trp Gly Leu Leu
 1 5 10

<210> 107
 <211> 12
 <212> PRT
 <213> Verminephrobacter eiseniae
 <400> 107

Asp Leu Gly Ser Thr Ser Leu Gly Trp Ala Ile Phe
 1 5 10

<210> 108
 <211> 12
 <212> PRT
 <213> Campylobacter jejuni, NCTC 11168
 <400> 108

Asp Ile Gly Ile Ser Ser Ile Gly Trp Ala Phe Ser
 1 5 10

<210> 109
 <211> 12
 <212> PRT
 <213> Parvibaculum lavamentivorans DS-1
 <400> 109

8009W000_SequenceListing.txt

Asp Ile Gly Thr Thr Ser Ile Gly Phe Ser Val Ile
 1 5 10

<210> 110
 <211> 12
 <212> PRT
 <213> Dinoroseobacter shibae DFL 12

<400> 110

Asp Ile Gly Thr Ser Ser Ile Gly Trp Trp Leu Tyr
 1 5 10

<210> 111
 <211> 12
 <212> PRT
 <213> Nitrobacter hamburgensis X14

<400> 111

Asp Leu Gly Ser Asn Ser Leu Gly Trp Phe Val Thr
 1 5 10

<210> 112
 <211> 12
 <212> PRT
 <213> Bradyrhizobium sp. BTAi1

<400> 112

Asp Leu Gly Ala Asn Ser Leu Gly Trp Phe Val Val
 1 5 10

<210> 113
 <211> 15
 <212> PRT
 <213> Bacillus cereus

<400> 113

Asp Ile Gly Leu Arg Ile Gly Ile Thr Ser Cys Gly Trp Ser Ile
 1 5 10 15

<210> 114
 <211> 12
 <212> PRT
 <213> Sutterella wadsworthensis

<400> 114

Asp Met Gly Ala Lys Tyr Thr Gly Val Phe Tyr Ala
 1 5 10

<210> 115
 <211> 12
 <212> PRT
 <213> Wolinella succinogenes DSM 1740

8009W000_SequenceListing.txt

<400> 115

Asp Leu Gly Gly Lys Asn Thr Gly Phe Phe Ser Phe
1 5 10

<210> 116

<211> 12

<212> PRT

<213> Francisella tularensis

<400> 116

Asp Leu Gly Val Lys Asn Thr Gly Val Phe Ser Ala
1 5 10

<210> 117

<211> 12

<212> PRT

<213> Gamma proteobacterium HTCC5015

<400> 117

Asp Leu Gly Ala Lys Phe Thr Gly Val Ala Leu Tyr
1 5 10

<210> 118

<211> 12

<212> PRT

<213> Legionella pneumophila str. Paris

<400> 118

Asp Leu Gly Gly Lys Phe Thr Gly Val Cys Leu Ser
1 5 10

<210> 119

<211> 12

<212> PRT

<213> Parasutterella excrementihominis

<400> 119

Asp Leu Gly Gly Thr Tyr Thr Gly Thr Phe Ile Thr
1 5 10

<210> 120

<211> 12

<212> PRT

<213> S. thermophilus

<400> 120

Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Thr
1 5 10

<210> 121

8009W000_SequenceListing.txt

<211> 12
<212> PRT
<213> Eubacterium yurii

<400> 121

Asp Val Gly Thr Asn Ser Val Gly Trp Ala Val Thr
1 5 10

<210> 122
<211> 12
<212> PRT
<213> Butyrivibrio hungatei

<400> 122

Asp Met Gly Thr Asn Ser Val Gly Trp Ala Val Thr
1 5 10

<210> 123
<211> 12
<212> PRT
<213> Solobacterium moorei F0204

<400> 123

Asp Val Gly Thr Ser Ser Val Gly Trp Ala Val Thr
1 5 10

<210> 124
<211> 27
<212> PRT
<213> Treponema denticola

<400> 124

Asp Ile Asp His Ile Tyr Pro Gln Ser Lys Ile Lys Asp Asp Ser Ile
1 5 10 15

Ser Asn Arg Val Leu Val Cys Ser Ser Cys Asn
20 25

<210> 125
<211> 27
<212> PRT
<213> Coprococcus catus GD-7

<400> 125

Asp Ile Asp His Ile Tyr Pro Gln Ser Lys Thr Met Asp Asp Ser Leu
1 5 10 15

Asn Asn Arg Val Leu Val Lys Lys Asn Tyr Asn
20 25

<210> 126

8009W000_SequenceListing.txt

<211> 27
<212> PRT
<213> Peptoniphilus duerdenii

<400> 126

Asp Gln Asp His Ile Tyr Pro Lys Ser Lys Ile Tyr Asp Asp Ser Leu
1 5 10 15

Glu Asn Arg Val Leu Val Lys Lys Asn Leu Asn
20 25

<210> 127
<211> 27
<212> PRT
<213> Catenibacterium mitsuokai DSM 15897

<400> 127

Gln Ile Asp His Ile Val Pro Gln Ser Leu Val Lys Asp Asp Ser Phe
1 5 10 15

Asp Asn Arg Val Leu Val Val Pro Ser Glu Asn
20 25

<210> 128
<211> 27
<212> PRT
<213> S. mutans

<400> 128

Asp Ile Asp His Ile Ile Pro Gln Ala Phe Ile Lys Asp Asn Ser Ile
1 5 10 15

Asp Asn Arg Val Leu Thr Ser Ser Lys Glu Asn
20 25

<210> 129
<211> 27
<212> PRT
<213> S. thermophilus

<400> 129

Asp Ile Asp His Ile Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile
1 5 10 15

Asp Asn Lys Val Leu Val Ser Ser Ala Ser Asn
20 25

<210> 130
<211> 27
<212> PRT
<213> Oenococcus kitaharae DSM 17330

8009W000_SequenceListing.txt

<400> 130

Asp Ile Asp His Ile Ile Pro Gln Ala Tyr Thr Lys Asp Asn Ser Leu
1 5 10 15

Asp Asn Arg Val Leu Val Ser Asn Ile Thr Asn
20 25

<210> 131

<211> 27

<212> PRT

<213> L. inocua

<400> 131

Asp Ile Asp His Ile Val Pro Gln Ser Phe Ile Thr Asp Asn Ser Ile
1 5 10 15

Asp Asn Leu Val Leu Thr Ser Ser Ala Gly Asn
20 25

<210> 132

<211> 27

<212> PRT

<213> S. pyogenes

<400> 132

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile
1 5 10 15

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn
20 25

<210> 133

<211> 27

<212> PRT

<213> Acidaminococcus sp. D21

<400> 133

Asn Ile Asp His Ile Tyr Pro Gln Ser Met Val Lys Asp Asp Ser Leu
1 5 10 15

Asp Asn Lys Val Leu Val Gln Ser Glu Ile Asn
20 25

<210> 134

<211> 27

<212> PRT

<213> Lactobacillus rhamnosus GG

<400> 134

8009W000_SequenceListing.txt

Asp Ile Asp His Ile Leu Pro Gln Ser Leu Ile Lys Asp Asp Ser Leu
 1 5 10 15

Asp Asn Arg Val Leu Val Asn Ala Thr Ile Asn
 20 25

<210> 135
 <211> 27
 <212> PRT
 <213> Lactobacillus gasseri

<400> 135

Asp Ile Asp His Ile Leu Pro Gln Ser Phe Ile Lys Asp Asp Ser Leu
 1 5 10 15

Glu Asn Arg Val Leu Val Lys Lys Ala Val Asn
 20 25

<210> 136
 <211> 27
 <212> PRT
 <213> Staphylococcus pseudintermedius ED99

<400> 136

Glu Val Asp His Ile Phe Pro Arg Ser Phe Ile Lys Asp Asp Ser Ile
 1 5 10 15

Asp Asn Lys Val Leu Val Ile Lys Lys Met Asn
 20 25

<210> 137
 <211> 27
 <212> PRT
 <213> Olsenella uli

<400> 137

Glu Val Asp His Ile Ile Pro Arg Ser Tyr Ile Lys Asp Asp Ser Phe
 1 5 10 15

Glu Asn Lys Val Leu Val Tyr Arg Glu Glu Asn
 20 25

<210> 138
 <211> 27
 <212> PRT
 <213> Bifidobacterium bifidum S17

<400> 138

Asp Ile Asp His Ile Ile Pro Gln Ala Val Thr Gln Asn Asp Ser Ile
 1 5 10 15

8009W000_SequenceListing.txt

Asp Asn Arg Val Leu Val Ala Arg Ala Glu Asn
20 25

<210> 139
<211> 27
<212> PRT
<213> Mycoplasma gallisepticum str. F

<400> 139

Glu Ile Asp His Ile Ile Pro Tyr Ser Ile Ser Phe Asp Asp Ser Ser
1 5 10 15

Ser Asn Lys Leu Leu Val Leu Ala Glu Ser Asn
20 25

<210> 140
<211> 27
<212> PRT
<213> Mycoplasma canis PG 14

<400> 140

Glu Ile Asp His Ile Ile Pro Tyr Ser Leu Cys Phe Asp Asp Ser Ser
1 5 10 15

Ala Asn Lys Val Leu Val His Lys Gln Ser Asn
20 25

<210> 141
<211> 27
<212> PRT
<213> Ilyobacter polytropus DSM 2926

<400> 141

Asp Ile Asp His Ile Ile Pro Tyr Ser Arg Ser Met Asp Asp Ser Tyr
1 5 10 15

Ser Asn Lys Val Leu Val Leu Ser Gly Glu Asn
20 25

<210> 142
<211> 27
<212> PRT
<213> Uncultured Termite group 1 bacterium

<400> 142

Asp Ile Asp His Ile Ile Pro Tyr Ser Lys Ser Met Asp Asp Ser Phe
1 5 10 15

Asn Asn Lys Val Leu Cys Leu Ala Glu Glu Asn
20 25

8009W000_SequenceListing.txt

<210> 143
 <211> 27
 <212> PRT
 <213> Campylobacter jejuni

<400> 143

Glu Ile Asp His Ile Tyr Pro Tyr Ser Arg Ser Phe Asp Asp Ser Tyr
 1 5 10 15

Met Asn Lys Val Leu Val Phe Thr Lys Gln Asn
 20 25

<210> 144
 <211> 27
 <212> PRT
 <213> Clostridium cellulolyticum H10

<400> 144

Gln Ile Asp His Ile Tyr Pro Tyr Ser Arg Ser Met Asp Asp Ser Tyr
 1 5 10 15

Met Asn Lys Val Leu Val Leu Thr Asp Glu Asn
 20 25

<210> 145
 <211> 27
 <212> PRT
 <213> Clostridium perfringens

<400> 145

Glu Ile Asp His Ile Ile Pro Phe Ser Arg Ser Phe Asp Asp Ser Leu
 1 5 10 15

Ser Asn Lys Ile Leu Val Leu Gly Ser Glu Asn
 20 25

<210> 146
 <211> 27
 <212> PRT
 <213> N. meningitides

<400> 146

Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp Ser Phe
 1 5 10 15

Asn Asn Lys Val Leu Val Leu Gly Ser Glu Asn
 20 25

<210> 147

8009W000_SequenceListing.txt

<211> 27
<212> PRT

<213> Pasteurella multocida str. Pm70

<400> 147

Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp Ser Phe
1 5 10 15

Asn Asn Lys Val Leu Val Leu Ala Ser Glu Asn
20 25

<210> 148
<211> 27
<212> PRT

<213> Enterococcus faecalis TX0012

<400> 148

Glu Ile Asp His Ile Ile Pro Ile Ser Ile Ser Leu Asp Asp Ser Ile
1 5 10 15

Asn Asn Lys Val Leu Val Leu Ser Lys Ala Asn
20 25

<210> 149
<211> 27
<212> PRT

<213> Eubacterium dolichum DSM 3991

<400> 149

Glu Val Asp His Ile Ile Pro Ile Ser Ile Ser Leu Asp Asp Ser Ile
1 5 10 15

Thr Asn Lys Val Leu Val Thr His Arg Glu Asn
20 25

<210> 150
<211> 27
<212> PRT

<213> Acidovorax ebreus

<400> 150

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

Asn Asn Lys Val Leu Val Leu Thr His Glu Asn
20 25

<210> 151
<211> 27
<212> PRT

<213> Streptococcus thermophilus LMD-9

8009W000_SequenceListing.txt

<400> 151

Glu Val Asp His Ile Leu Pro Leu Ser Ile Thr Phe Asp Asp Ser Leu
1 5 10 15

Ala Asn Lys Val Leu Val Tyr Ala Thr Ala Asn
20 25

<210> 152

<211> 27

<212> PRT

<213> Eubacterium rectale

<400> 152

Glu Ile Asp His Ile Ile Pro Arg Ser Ile Ser Phe Asp Asp Ala Arg
1 5 10 15

Ser Asn Lys Val Leu Val Tyr Arg Ser Glu Asn
20 25

<210> 153

<211> 27

<212> PRT

<213> Staphylococcus lugdunensis M23590

<400> 153

Glu Val Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Tyr
1 5 10 15

His Asn Lys Val Leu Val Lys Gln Ser Glu Asn
20 25

<210> 154

<211> 27

<212> PRT

<213> Roseburia intestinalis

<400> 154

Asp Ile Asp His Ile Leu Pro Tyr Ser Ile Thr Phe Asp Asp Ser Phe
1 5 10 15

Arg Asn Lys Val Leu Val Thr Ser Gln Glu Asn
20 25

<210> 155

<211> 27

<212> PRT

<213> Wolinella succinogenes DSM 1740

<400> 155

8009W000_SequenceListing.txt

Glu Ile Asp His Ile Leu Pro Arg Ser Arg Ser Ala Asp Asp Ser Phe
 1 5 10 15

Ala Asn Lys Val Leu Cys Leu Ala Arg Ala Asn
 20 25

<210> 156
 <211> 27
 <212> PRT
 <213> Cand. Puniceispirillum marinum

<400> 156

Glu Ile Glu His Leu Leu Pro Phe Ser Leu Thr Leu Asp Asp Ser Met
 1 5 10 15

Ala Asn Lys Thr Val Cys Phe Arg Gln Ala Asn
 20 25

<210> 157
 <211> 27
 <212> PRT
 <213> Azospirillum sp. B510

<400> 157

Asp Ile Asp His Ile Leu Pro Phe Ser Val Ser Leu Asp Asp Ser Ala
 1 5 10 15

Ala Asn Lys Val Val Cys Leu Arg Glu Ala Asn
 20 25

<210> 158
 <211> 27
 <212> PRT
 <213> Bradyrhizobium sp. BTAi1

<400> 158

Asp Ile Asp His Leu Ile Pro Phe Ser Ile Ser Trp Asp Asp Ser Ala
 1 5 10 15

Ala Asn Lys Val Val Cys Met Arg Tyr Ala Asn
 20 25

<210> 159
 <211> 27
 <212> PRT
 <213> Nitrobacter hamburgensis X14

<400> 159

Asp Ile Asp His Ile Leu Pro Val Ala Met Thr Leu Asp Asp Ser Pro
 1 5 10 15

8009W000_SequenceListing.txt

Ala Asn Lys Ile Ile Cys Met Arg Tyr Ala Asn
20 25

<210> 160
<211> 27
<212> PRT
<213> Dinoroseobacter shibae

<400> 160

Asp Val Asp His Ile Leu Pro Tyr Ser Arg Thr Leu Asp Asp Ser Phe
1 5 10 15

Pro Asn Arg Thr Leu Cys Leu Arg Glu Ala Asn
20 25

<210> 161
<211> 27
<212> PRT
<213> Verminephrobacter eiseniae

<400> 161

Glu Ile Glu His Ile Leu Pro Phe Ser Arg Thr Leu Asp Asp Ser Leu
1 5 10 15

Asn Asn Arg Thr Val Ala Met Arg Arg Ala Asn
20 25

<210> 162
<211> 27
<212> PRT
<213> Lactobacillus coryniformis KCTC 3535

<400> 162

Glu Val Asp His Ile Ile Pro Tyr Ser Ile Ser Trp Asp Asp Ser Tyr
1 5 10 15

Thr Asn Lys Val Leu Thr Ser Ala Lys Cys Asn
20 25

<210> 163
<211> 27
<212> PRT
<213> Rhodopseudomonas palustris

<400> 163

Gln Val Asp His Ile Leu Pro Trp Ser Arg Phe Gly Asp Asp Ser Tyr
1 5 10 15

Leu Asn Lys Thr Leu Cys Thr Ala Arg Ser Asn
20 25

8009W000_SequenceListing.txt

<210> 164
 <211> 27
 <212> PRT
 <213> *Ralstonia syzygii* R24

<400> 164

Gln Val Asp His Ile Leu Pro Phe Ser Lys Thr Leu Asp Asp Ser Phe
 1 5 10 15

Ala Asn Lys Val Leu Ala Gln His Asp Ala Asn
 20 25

<210> 165
 <211> 27
 <212> PRT
 <213> *Helicobacter mustelae* 12198

<400> 165

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

Ser Asn Lys Val Leu Cys Leu Thr Ser Ser Asn
 20 25

<210> 166
 <211> 27
 <212> PRT
 <213> *Mycoplasma mobile* 163K

<400> 166

Asp Ile Asp His Ile Val Pro Arg Ser Ile Ser Phe Asp Asp Ser Phe
 1 5 10 15

Ser Asn Leu Val Ile Val Asn Lys Leu Asp Asn
 20 25

<210> 167
 <211> 27
 <212> PRT
 <213> *Mycoplasma ovipneumoniae* SC01

<400> 167

Glu Ile Glu His Ile Ile Pro Tyr Ser Met Ser Tyr Asp Asn Ser Gln
 1 5 10 15

Ala Asn Lys Ile Leu Thr Glu Lys Ala Glu Asn
 20 25

<210> 168

8009W000_SequenceListing.txt

<211> 27

<212> PRT

<213> Mycoplasma synoviae 53

<400> 168

Glu Ile Asp His Val Ile Pro Tyr Ser Lys Ser Ala Asp Asp Ser Trp
1 5 10 15

Phe Asn Lys Leu Leu Val Lys Lys Ser Thr Asn
20 25

<210> 169

<211> 27

<212> PRT

<213> Aminomonas paucivorans DSM 12260

<400> 169

Glu Met Asp His Ile Leu Pro Tyr Ser Arg Ser Leu Asp Asn Gly Trp
1 5 10 15

His Asn Arg Val Leu Val His Gly Lys Asp Asn
20 25

<210> 170

<211> 27

<212> PRT

<213> Ruminococcus albus 8

<400> 170

Glu Val Asp His Ile Val Pro Tyr Ser Leu Ile Leu Asp Asn Thr Ile
1 5 10 15

Asn Asn Lys Ala Leu Val Tyr Ala Glu Glu Asn
20 25

<210> 171

<211> 27

<212> PRT

<213> Fibrobacter succinogenes

<400> 171

Glu Ile Glu His Val Ile Pro Gln Ser Leu Tyr Phe Asp Asp Ser Phe
1 5 10 15

Ser Asn Lys Val Ile Cys Glu Ala Glu Val Asn
20 25

<210> 172

<211> 27

<212> PRT

<213> Bacteroides fragilis, NCTC 9343

8009W000_SequenceListing.txt

<400> 172

Asp Ile Glu His Ile Ile Pro Gln Ala Arg Leu Phe Asp Asp Ser Phe
1 5 10 15

Ser Asn Lys Thr Leu Glu Ala Arg Ser Val Asn
20 25

<210> 173

<211> 27

<212> PRT

<213> Capnocytophaga sputigena

<400> 173

Glu Ile Glu His Ile Val Pro Lys Ala Arg Val Phe Asp Asp Ser Phe
1 5 10 15

Ser Asn Lys Thr Leu Thr Phe His Arg Ile Asn
20 25

<210> 174

<211> 28

<212> PRT

<213> Finegoldia magna

<400> 174

Asp Lys Asp His Ile Ile Pro Gln Ser Met Lys Lys Asp Asp Ser Ile
1 5 10 15

Ile Asn Asn Leu Val Leu Val Asn Lys Asn Ala Asn
20 25

<210> 175

<211> 27

<212> PRT

<213> Parvibaculum lavamentivorans DS-1

<400> 175

Glu Val Glu His Ile Trp Pro Arg Ser Arg Ser Phe Asp Asn Ser Pro
1 5 10 15

Arg Asn Lys Thr Leu Cys Arg Lys Asp Val Asn
20 25

<210> 176

<211> 27

<212> PRT

<213> Bacillus cereus

<400> 176

8009W000_SequenceListing.txt

Ile Val Asn His Ile Ile Pro Tyr Asn Arg Ser Phe Asp Asp Thr Tyr
 1 5 10 15

His Asn Arg Val Leu Thr Leu Thr Glu Thr Lys
 20 25

<210> 177
 <211> 27
 <212> PRT
 <213> Prevotella micans

<400> 177

Asp Met Glu His Thr Ile Pro Lys Ser Ile Ser Phe Asp Asn Ser Asp
 1 5 10 15

Gln Asn Leu Thr Leu Cys Glu Ser Tyr Tyr Asn
 20 25

<210> 178
 <211> 27
 <212> PRT
 <213> Prevotella ruminicola

<400> 178

Asp Ile Glu His Thr Ile Pro Arg Ser Ala Gly Gly Asp Ser Thr Lys
 1 5 10 15

Met Asn Leu Thr Leu Cys Ser Ser Arg Phe Asn
 20 25

<210> 179
 <211> 27
 <212> PRT
 <213> Flavobacterium columnare

<400> 179

Asp Ile Glu His Thr Ile Pro Arg Ser Ile Ser Gln Asp Asn Ser Gln
 1 5 10 15

Met Asn Lys Thr Leu Cys Ser Leu Lys Phe Asn
 20 25

<210> 180
 <211> 27
 <212> PRT
 <213> Rhodospirillum rubrum

<400> 180

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

8009W000_SequenceListing.txt

Asp Asn Met Val Leu Cys Gln Ser Asp Ala Asn
20 25

<210> 181
<211> 27
<212> PRT
<213> Elusimicrobium minutum Pei191

<400> 181

Asp Ile Glu His Leu Phe Pro Ile Ala Glu Ser Glu Asp Asn Gly Arg
1 5 10 15

Asn Asn Leu Val Ile Ser His Ser Ala Cys Asn
20 25

<210> 182
<211> 27
<212> PRT
<213> Sphaerochaeta globus str. Buddy

<400> 182

Asp Val Asp His Ile Phe Pro Arg Asp Asp Thr Ala Asp Asn Ser Tyr
1 5 10 15

Gly Asn Lys Val Val Ala His Arg Gln Cys Asn
20 25

<210> 183
<211> 27
<212> PRT
<213> Nitratifractor salsuginis DSM 16511

<400> 183

Asp Ile Glu His Ile Val Pro Gln Ser Leu Gly Gly Leu Ser Thr Asp
1 5 10 15

Tyr Asn Thr Ile Val Thr Leu Lys Ser Val Asn
20 25

<210> 184
<211> 27
<212> PRT
<213> Acidothermus cellulolyticus 11B

<400> 184

Glu Leu Asp His Ile Val Pro Arg Thr Asp Gly Gly Ser Asn Arg His
1 5 10 15

Glu Asn Leu Ala Ile Thr Cys Gly Ala Cys Asn
20 25

8009W000_SequenceListing.txt

<210> 185
 <211> 28
 <212> PRT
 <213> Bifidobacterium longum DJ010A

<400> 185

Glu Met Asp His Ile Val Pro Arg Lys Gly Val Gly Ser Thr Asn Thr
 1 5 10 15

Arg Thr Asn Phe Ala Ala Val Cys Ala Glu Cys Asn
 20 25

<210> 186
 <211> 28
 <212> PRT
 <213> Bifidobacterium dentium

<400> 186

Glu Met Asp His Ile Val Pro Arg Lys Gly Val Gly Ser Thr Asn Thr
 1 5 10 15

Arg Val Asn Leu Ala Ala Ala Cys Ala Ala Cys Asn
 20 25

<210> 187
 <211> 28
 <212> PRT
 <213> Corynebacterium diphtheriae

<400> 187

Glu Met Asp His Ile Val Pro Arg Ala Gly Gln Gly Ser Thr Asn Thr
 1 5 10 15

Arg Glu Asn Leu Val Ala Val Cys His Arg Cys Asn
 20 25

<210> 188
 <211> 33
 <212> PRT
 <213> Sutterella wadsworthensis

<400> 188

Glu Ile Asp His Ile Leu Pro Arg Ser Leu Ile Lys Asp Ala Arg Gly
 1 5 10 15

Ile Val Phe Asn Ala Glu Pro Asn Leu Ile Tyr Ala Ser Ser Arg Gly
 20 25 30

Asn

8009W000_SequenceListing.txt

<210> 189
 <211> 33
 <212> PRT
 <213> Gamma proteobacterium HTCC5015

<400> 189

Glu Ile Asp His Ile Ile Pro Arg Ser Leu Thr Gly Arg Thr Lys Lys
 1 5 10 15

Thr Val Phe Asn Ser Glu Ala Asn Leu Ile Tyr Cys Ser Ser Lys Gly
 20 25 30

Asn

<210> 190
 <211> 33
 <212> PRT
 <213> Parasutterella excrementihominis

<400> 190

Glu Ile Asp His Ile Ile Pro Arg Ser Leu Thr Leu Lys Lys Ser Glu
 1 5 10 15

Ser Ile Tyr Asn Ser Glu Val Asn Leu Ile Phe Val Ser Ala Gln Gly
 20 25 30

Asn

<210> 191
 <211> 33
 <212> PRT
 <213> Legionella pneumophila str. Paris

<400> 191

Glu Ile Asp His Ile Tyr Pro Arg Ser Leu Ser Lys Lys His Phe Gly
 1 5 10 15

Val Ile Phe Asn Ser Glu Val Asn Leu Ile Tyr Cys Ser Ser Gln Gly
 20 25 30

Asn

<210> 192
 <211> 33
 <212> PRT

<213> Wolinella succinogenes DSM 1740

<400> 192

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

Thr Val Phe Asn Pro Glu Gly Asn Leu Ile Tyr Val His Gln Lys Cys
20 25 30

Asn

<210> 193

<211> 30

<212> PRT

<213> Francisella tularensis

<400> 193

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

Asn Asp Glu Ala Asn Leu Ile Cys Val Thr Arg Gly Asp Asn
20 25 30

<210> 194

<211> 27

<212> PRT

<213> Akkermansia muciniphila

<400> 194

Glu Leu Glu His Ile Val Pro His Ser Phe Arg Gln Ser Asn Ala Leu
1 5 10 15

Ser Ser Leu Val Leu Thr Trp Pro Gly Val Asn
20 25

<210> 195

<211> 27

<212> PRT

<213> Solobacterium moorei F0204

<400> 195

Asp Ile Asp His Ile Tyr Pro Arg Ser Lys Ile Lys Asp Asp Ser Ile
1 5 10 15

Thr Asn Arg Val Leu Val Glu Lys Asp Ile Asn
20 25

<210> 196

<211> 28

8009W000_SequenceListing.txt

```

<212> PRT
<213> Veillonella atypica ACS-134-V-Col7a

<400> 196

Tyr Asp Ile Asp His Ile Tyr Pro Arg Ser Leu Thr Lys Asp Asp Ser
1          5          10          15

Phe Asp Asn Leu Val Leu Cys Glu Arg Thr Ala Asn
          20          25

<210> 197
<211> 28
<212> PRT
<213> Fusobacterium nucleatum

<400> 197

Asp Ile Asp His Ile Tyr Pro Arg Ser Lys Val Ile Lys Asp Asp Ser
1          5          10          15

Phe Asp Asn Leu Val Leu Val Leu Lys Asn Glu Asn
          20          25

<210> 198
<211> 27
<212> PRT
<213> Filifactor alocis

<400> 198

Asp Arg Asp His Ile Tyr Pro Gln Ser Lys Ile Lys Asp Asp Ser Ile
1          5          10          15

Asp Asn Leu Val Leu Val Asn Lys Thr Tyr Asn
          20          25

<210> 199
<211> 5
<212> DNA
<213> Streptococcus thermophilus

<220>
<221> misc_feature
<222> (1)..(1)
<223> Any nucleotide (e.g., A, G, C, or T)

<220>
<221> misc_feature
<222> (4)..(4)
<223> Any nucleotide (e.g., A, G, C, or T)

<400> 199
nggng

```

<210> 200
 <211> 7
 <212> DNA
 <213> Streptococcus thermophilus

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> Any nucleotide (e.g., A, G, C, or T)

<220>
 <221> misc_feature
 <222> (2)..(2)
 <223> Any nucleotide (e.g., A, G, C, or T)

<220>
 <221> misc_feature
 <222> (7)..(7)
 <223> A or T

<400> 200
 nnagaaw

7

<210> 201
 <211> 4
 <212> DNA
 <213> Streptococcus mutans

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> Any nucleotide (e.g., A, G, C, or T)

<220>
 <221> misc_feature
 <222> (4)..(4)
 <223> A or G

<400> 201
 naar

4

<210> 202
 <211> 5
 <212> DNA
 <213> Staphylococcus aureus

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> Any nucleotide (e.g., A, G, C, or T)

<220>
 <221> misc_feature
 <222> (2)..(2)
 <223> Any nucleotide (e.g., A, G, C, or T)

<220>
 <221> misc_feature

<222> (4)..(4)
 <223> A or G

<220>
 <221> misc_feature
 <222> (5)..(5)
 <223> A or G

<400> 202
 nngrr

5

<210> 203
 <211> 6
 <212> DNA
 <213> Staphylococcus aureus

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> Any nucleotide (e.g., A, G, C, or T)

<220>
 <221> misc_feature
 <222> (2)..(2)
 <223> Any nucleotide (e.g., A, G, C, or T)

<220>
 <221> misc_feature
 <222> (4)..(4)
 <223> A or G

<220>
 <221> misc_feature
 <222> (5)..(5)
 <223> A or G

<220>
 <221> misc_feature
 <222> (6)..(6)
 <223> Any nucleotide (e.g., A, G, C, or T)

<400> 203
 nngrrn

6

<210> 204
 <211> 6
 <212> DNA
 <213> Staphylococcus aureus

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> Any nucleotide (e.g., A, G, C, or T)

<220>
 <221> misc_feature
 <222> (2)..(2)
 <223> Any nucleotide (e.g., A, G, C, or T)

8009W000_SequenceListing.txt

<220>
 <221> misc_feature
 <222> (4)..(4)
 <223> A or G

<220>
 <221> misc_feature
 <222> (5)..(5)
 <223> A or G

<400> 204
 nngrrt

6

<210> 205
 <211> 6
 <212> DNA
 <213> Staphylococcus aureus

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> Any nucleotide (e.g., A, G, C, or T)

<220>
 <221> misc_feature
 <222> (2)..(2)
 <223> Any nucleotide (e.g., A, G, C, or T)

<220>
 <221> misc_feature
 <222> (4)..(4)
 <223> A or G

<220>
 <221> misc_feature
 <222> (5)..(5)
 <223> A or G

<220>
 <221> misc_feature
 <222> (6)..(6)
 <223> A, G, or C

<400> 205
 nngrrv

6

<210> 206

<400> 206
 000

<210> 207

<400> 207
 000

<210> 208

<400> 208
 000

8009W000_SequenceListing.txt

<210> 209

<400> 209
000

<210> 210

<400> 210
000

<210> 211

<400> 211
000

<210> 212

<400> 212
000

<210> 213

<400> 213
000

<210> 214

<400> 214
000

<210> 215

<400> 215
000

<210> 216

<400> 216
000

<210> 217

<400> 217
000

<210> 218

<400> 218
000

<210> 219

<400> 219
000

<210> 220

<400> 220
000

<210> 221

8009W000_SequenceListing.txt

<400> 221
000

<210> 222

<400> 222
000

<210> 223

<400> 223
000

<210> 224

<400> 224
000

<210> 225

<400> 225
000

<210> 226

<400> 226
000

<210> 227

<400> 227
000

<210> 228

<400> 228
000

<210> 229

<400> 229
000

<210> 230

<400> 230
000

<210> 231

<400> 231
000

<210> 232

<400> 232
000

<210> 233

<400> 233
000

8009W000_SequenceListing.txt

<210> 234

<400> 234
000

<210> 235

<400> 235
000

<210> 236

<400> 236
000

<210> 237

<400> 237
000

<210> 238

<400> 238
000

<210> 239

<400> 239
000

<210> 240

<400> 240
000

<210> 241

<400> 241
000

<210> 242

<400> 242
000

<210> 243

<400> 243
000

<210> 244

<400> 244
000

<210> 245

<400> 245
000

<210> 246

<400> 246

8009W000_SequenceListing.txt

000

<210> 247

<400> 247

000

<210> 248

<400> 248

000

<210> 249

<400> 249

000

<210> 250

<400> 250

000

<210> 251

<211> 17

<212> RNA

<213> Artificial Sequence

<220>

<223> Targeting domain

<400> 251

gaaggaaacu agcuaaa

17

<210> 252

<211> 20

<212> RNA

<213> Artificial Sequence

<220>

<223> Targeting domain

<400> 252

ggagaaggaa acuagcuaaa

20

<210> 253

<211> 20

<212> RNA

<213> Artificial Sequence

<220>

<223> Targeting domain

<400> 253

gggagaaggaa aacuagcuaa

20

<210> 254

<211> 20

<212> RNA

<213> Artificial Sequence

8009W000_SequenceListing.txt

<220>
 <223> Targeting domain
 <400> 254
 guauccucua ugaugggaga 20

<210> 255
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 255
 guuuccuucu ccacaucauag 20

<210> 256
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 256
 guccugguau ccucuaugau 20

<210> 257
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 257
 agaaggaaac uagcuaa 17

<210> 258
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 258
 uccucuauga ugggaga 17

<210> 259
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 259
ccugguaucc ucuauga 17

<210> 260
<211> 17
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 260
ccaucauaga ggauacc 17

<210> 261
<211> 17
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 261
uccuucuccc aucauag 17

<210> 262
<211> 17
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 262
uagcaguaucc cucuugg 17

<210> 263
<211> 17
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 263
uuagcaguau ccucuug 17

<210> 264
<211> 17
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 264
aacuggaaug acugaau 17

8009W000_SequenceListing.txt

<210>	265	
<211>	17	
<212>	RNA	
<213>	Artificial Sequence	
<220>		
<223>	Targeting domain	
<400>	265	
	cugguauccu cuaugau	17
<210>	266	
<211>	17	
<212>	RNA	
<213>	Artificial Sequence	
<220>		
<223>	Targeting domain	
<400>	266	
	aaauagcagu auccucu	17
<210>	267	
<211>	17	
<212>	RNA	
<213>	Artificial Sequence	
<220>		
<223>	Targeting domain	
<400>	267	
	auuagcagua uccucuu	17
<210>	268	
<211>	20	
<212>	RNA	
<213>	Artificial Sequence	
<220>		
<223>	Targeting domain	
<400>	268	
	aguccuggua uccucuauga	20
<210>	269	
<211>	20	
<212>	RNA	
<213>	Artificial Sequence	
<220>		
<223>	Targeting domain	
<400>	269	
	cucccaucau agaggauacc	20
<210>	270	
<211>	20	

```

<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 270
aauuagcagu auccucuugg                20

<210> 271
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 271
aaauuagcag uauccucuug                20

<210> 272
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 272
aaaaacugga augacugaau                20

<210> 273
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 273
aaaaauuagc aguauccucu                20

<210> 274
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 274
aaaauuagca guauccucu                20

<210> 275
<211> 17
<212> RNA
<213> Artificial Sequence

```


<220>
 <223> Targeting domain
 <400> 275
 gaaucggaac aaggcaa 17

<210> 276
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 276
 gaccaauagc cuugaca 17

<210> 277
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 277
 ggcuaauuggu caaggca 17

<210> 278
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 278
 gucaaggcua uugguca 17

<210> 279
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 279
 guguguggaa cugcuga 17

<210> 280
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 280
 gggccggcgg cuggcua 17

<210> 281
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 281
 gaguauccag ugaggcc 17

<210> 282
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 282
 gcugacaaaa gaagucc 17

<210> 283
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 283
 ggccaggggc cggcggc 17

<210> 284
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 284
 gggaaggggc cccaag 17

<210> 285
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 285
 gagauagugu ggggaag 17

8009W000_SequenceListing.txt

<210> 286
 <211> 17
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 286
 guauccagug aggccag 17

 <210> 287
 <211> 17
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 287
 gugaggccag gggccgg 17

 <210> 288
 <211> 17
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 288
 gcuggccaac ccauggg 17

 <210> 289
 <211> 17
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 289
 ggcuaaacuc cacccau 17

 <210> 290
 <211> 17
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 290
 ggauacucua agacuau 17

 <210> 291
 <211> 17

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 291
 ggggccggcg gcuggcu 17

 <210> 292
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 292
 ggcuagggau gaagaauaaa 20

 <210> 293
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 293
 gagugugugg aacugcugaa 20

 <210> 294
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 294
 ggaaugacug aaucggaaca 20

 <210> 295
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 295
 gcauugagau agugugggga 20

 <210> 296
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 296
 gcuaauugguc aaggcaaggc 20

<210> 297
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 297
 guggggaagg ggcccccaag 20

<210> 298
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 298
 ggcaaggcug gccaaccgau 20

<210> 299
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 299
 guuugccuug ucaaggcuau 20

<210> 300
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 300
 gcuaaacucc acccaugggu 20

<210> 301
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 301		
caaauaucug ucugaaa		17
<210> 302		
<211> 17		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 302		
uagggaugaa gaauaaa		17
<210> 303		
<211> 17		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 303		
ugagauagug uggggaa		17
<210> 304		
<211> 17		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 304		
uguguggaac ugcugaa		17
<210> 305		
<211> 17		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 305		
augacugaau cggaaca		17
<210> 306		
<211> 17		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 306		
caaggcuggc caaccaca		17

8009W000_SequenceListing.txt

<210> 307
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 307
 uggcuaaacu ccaccca 17

<210> 308
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 308
 uggguggagu uuagcca 17

<210> 309
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 309
 aguauccagu gaggcca 17

<210> 310
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 310
 ucaaguuugc cuuguca 17

<210> 311
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 311
 uugagauagu gugggga 17

<210> 312
 <211> 17

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 312
 auaaaauuaga gaaaaaac 17

<210> 313
 <211> 17
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 313
 ccggccccug gccucac 17

<210> 314
 <211> 17
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 314
 agccagccgc cggcccc 17

<210> 315
 <211> 17
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 315
 cugucugaaa cgguccc 17

<210> 316
 <211> 17
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 316
 auggguggag uuuagcc 17

<210> 317
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 317
 caucccuagc cagccgc 17

<210> 318
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 318
 auuggucaag gcaaggc 17

<210> 319
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 319
 ccagugaggc caggggc 17

<210> 320
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 320
 uuccacacac ucgcuuc 17

<210> 321
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 321
 cgcuucugga acgucug 17

<210> 322
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 322
 ucuuagagua uccagug 17

<210> 323
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 323
 uuugcauuga gauagug 17

<210> 324
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 324
 uuccagaagc gagugug 17

<210> 325
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 325
 ugcauugaga uagugug 17

<210> 326
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 326
 aaggcuggcc aaccgau 17

<210> 327
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 327
 ugccuuguca aggcua 17

8009W000_SequenceListing.txt

<210> 328
<211> 17
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 328
aaacuccacc caugggu 17

<210> 329
<211> 17
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 329
uugcauugag auagugu 17

<210> 330
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 330
augcaauau cugucugaaa 20

<210> 331
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 331
acugaaucgg aacaaggcaa 20

<210> 332
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 332
cauugagaua guguggggaa 20

<210> 333
<211> 20

```

<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 333
cuugaccaau agccuugaca 20

<210> 334
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 334
aggcaaggcu ggccaaccca 20

<210> 335
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 335
cccuggcuaa acuccaccca 20

<210> 336
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 336
ccaugggugg aguuuagcca 20

<210> 337
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 337
uagaguaucc agugaggcca 20

<210> 338
<211> 20
<212> RNA
<213> Artificial Sequence

```

<220>
 <223> Targeting domain
 <400> 338
 caaggcuauu ggucaaggca 20
 <210> 339
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 339
 cuugucaagg cuauugguca 20
 <210> 340
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 340
 uggucaaguu ugccuuguca 20
 <210> 341
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 341
 cgagugugug gaacugcuga 20
 <210> 342
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 342
 caggggccgg cggcuggcua 20
 <210> 343
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 343
agaauaaaauu agagaaaaaac 20

<210> 344
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 344
ccgccggccc cuggccucac 20

<210> 345
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 345
ccuagccagc cgccggcccc 20

<210> 346
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 346
uaucugucug aaacgguccc 20

<210> 347
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 347
cccaugggug gaguuuagcc 20

<210> 348
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 348
uuagaguauc cagugaggcc 20

8009W000_SequenceListing.txt

<210> 349
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 349
 acggcugaca aaagaagucc 20

<210> 350
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 350
 cuucaucccu agccagccgc 20

<210> 351
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 351
 ugaggccagg ggccggcggc 20

<210> 352
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 352
 uauccaguga ggccaggggc 20

<210> 353
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 353
 caguuccaca cacucgcuuc 20

<210> 354
 <211> 20

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 354
 auugagauag uguggggaag 20

 <210> 355
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 355
 agaguaucca gugaggccag 20

 <210> 356
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 356
 ccagugaggc caggggccgg 20

 <210> 357
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 357
 aaggcuggcc aaccauggg 20

 <210> 358
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 358
 acucgcuucu ggaacgucug 20

 <210> 359
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

 <400> 359
 uagucuuaga guauccagug 20

 <210> 360
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 360
 auauuuugcau ugagauagug 20

 <210> 361
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 361
 acguuccaga agcgagugug 20

 <210> 362
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 362
 auuugcauug agauagugug 20

 <210> 363
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 363
 ccuggcuaaa cuccacccau 20

 <210> 364
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 364
 acuggauacu cuaagacuau 20

<210> 365
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 365
 ccaggggccg gcggcuggcu 20

<210> 366
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 366
 uauuuucauu gagauagugu 20

<210> 367
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 367
 guuuccuucu cccaucua 19

<210> 368
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 368
 gcuaguuucc uucucccauc aua 23

<210> 369
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 369
 gaauaaaaua gagaaaaac 19

8009W000_SequenceListing.txt

<210> 370
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 370
 gaagaauaaa uuagagaaaa ac 22

 <210> 371
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 371
 ggaagaauaa auuagagaaa aac 23

 <210> 372
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 372
 gggaagaaua aaauagagaa aaac 24

 <210> 373
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 373
 gaaggaaacu agcuaaaggg 20

 <210> 374
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 374
 gagaaggaaa cuagcuaaag gg 22

 <210> 375
 <211> 23

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 375
 ggagaaggaa acuagcuaaa ggg 23

<210> 376
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 376
 gggagaagga aacuagcuaa aggg 24

<210> 377
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 377
 uuuccuucuc ccaucaua 18

<210> 378
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 378
 aguuuccuuc ucccacaua 20

<210> 379
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 379
 uaguuccuu cuccacau a 21

<210> 380
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 380
 cuaguuuccu ucucccauca ua 22

<210> 381
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 381
 agcuaguuuuc cuucucccau caua 24

<210> 382
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 382
 agagaaaaaac uggaauga 18

<210> 383
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 383
 uagagaaaaa cuggaauga 19

<210> 384
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 384
 uuagagaaaa acuggaauga 20

<210> 385
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 385		
auuagagaaa aacuggaug a		21
<210> 386		
<211> 22		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 386		
aaauagagaa aaacuggaau ga		22
<210> 387		
<211> 23		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 387		
aaaauagaga aaaacuggaa uga		23
<210> 388		
<211> 24		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 388		
uaaaauagag aaaaacugga auga		24
<210> 389		
<211> 18		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 389		
aauaaaauag agaaaaaac		18
<210> 390		
<211> 21		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 390		
aagaauaaa uagagaaaaa c		21

8009W000_SequenceListing.txt

<210> 391
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 391
 aggaaacuag cuaaaggg 18

 <210> 392
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 392
 aaggaaacua gcuaaaggg 19

 <210> 393
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 393
 agaaggaaac uagcuaaagg g 21

 <210> 394
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 394
 uggggaaggg gcccccaa 18

 <210> 395
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 395
 guggggaagg ggcccccaa 19

 <210> 396
 <211> 20

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 396
 uguggggaag gggcccccaa 20

<210> 397
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 397
 guguggggaa gggccccca a 21

<210> 398
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 398
 agugugggga aggggcccc aa 22

<210> 399
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 399
 uagugugggg aaggggcccc caa 23

<210> 400
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 400
 auaguguggg gaaggggcccc ccaa 24

<210> 401
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 401
 accucagacg uuccagaa 18

<210> 402
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 402
 aaccucagac guuccagaa 19

<210> 403
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 403
 uaaccucaga cguuccagaa 20

<210> 404
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 404
 auaaccucag acguuccaga a 21

<210> 405
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 405
 gauaaccuca gacguuccag aa 22

<210> 406
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 406
 ugauaaccuc agacguucca gaa 23

<210> 407
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 407
 uugauaaccu cagacguucc agaa 24

<210> 408
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 408
 cgccggcccc uggccuca 18

<210> 409
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 409
 ccgccggccc cuggccuca 19

<210> 410
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 410
 gccgccggcc ccuggccuca 20

<210> 411
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 411
 agccgccggc cccuggccuc a 21

8009W000_SequenceListing.txt

<210> 412
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 412
 cagccgccgg ccccuggccu ca 22

 <210> 413
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 413
 ccagccgccg gccccuggcc uca 23

 <210> 414
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 414
 gccagccgcc ggccccuggc cuca 24

 <210> 415
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 415
 ggcaaggcug gccaaccc 18

 <210> 416
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 416
 aggcaaggcu ggccaaccc 19

 <210> 417
 <211> 20

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 417
 aaggcaaggc uggccaaccc 20

<210> 418
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 418
 caaggcaagg cuggccaacc c 21

<210> 419
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 419
 ucaaggcaag gcuggccaac cc 22

<210> 420
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 420
 gucaaggcaa ggcuggccaa ccc 23

<210> 421
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 421
 ggucaaggca aggcuggcca accc 24

<210> 422
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 422
 ggcuggccaa cccauggg 18

<210> 423
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 423
 aggcuggcca acccauggg 19

<210> 424
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 424
 caaggcuggc caaccgaug g 21

<210> 425
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 425
 gcaaggcugg ccaaccgaug gg 22

<210> 426
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 426
 ggcaaggcug gccaaccgau ggg 23

<210> 427
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 427	
aggcaaggcu ggccaaccca uggg	24
<210> 428	
<211> 18	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 428	
gagugugugg aacugcug	18
<210> 429	
<211> 19	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 429	
cgagugugug gaacugcug	19
<210> 430	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 430	
gcgagugugu ggaacugcug	20
<210> 431	
<211> 21	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 431	
agcgagugug uggaacugcu g	21
<210> 432	
<211> 22	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 432	
aagcgagugu guggaacugc ug	22

8009W000_SequenceListing.txt

<210> 433
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 433
 gaagcgagug uguggaacug cug 23

 <210> 434
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 434
 agaagcgagu guguggaacu gcug 24

 <210> 435
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 435
 ccuggcuaaa cuccaccc 18

 <210> 436
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 436
 cccuggcuaa acuccaccc 19

 <210> 437
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 437
 ucccuggcua aacuccaccc 20

 <210> 438
 <211> 21

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 438
 gucccuggcu aaacuccacc c 21

 <210> 439
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 439
 ggucccuggc uaaacuccac cc 22

 <210> 440
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 440
 cggucccugg cuaaacucca ccc 23

 <210> 441
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 441
 acggucccug gcuaaacucc accc 24

 <210> 442
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 442
 ggcggcuggc uagggau 18

 <210> 443
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 443
 cggcggcugg cuagggau 19

<210> 444
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 444
 ccggcggcug gcuagggau 20

<210> 445
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 445
 gccggcggcu ggcuaaggau g 21

<210> 446
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 446
 ggccggcggc uggcuaggga ug 22

<210> 447
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 447
 gggccggcgg cuggcuaggg aug 23

<210> 448
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 448
 ggggccggcg gcuggcuagg gaug 24

 <210> 449
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 449
 aggggccggc ggcuggcu 18

 <210> 450
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 450
 caggggccgg cggcuggcu 19

 <210> 451
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 451
 gccaggggcc ggcggcuggc u 21

 <210> 452
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 452
 ggccaggggc cggcggcugg cu 22

 <210> 453
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 453
 aggccagggg ccggcggcug gcu 23

8009W000_SequenceListing.txt

<210> 454
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 454
 gaggccagg ggcggcggu ggcu 24

 <210> 455
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 455
 aaacuugacc aaauagucu 18

 <210> 456
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 456
 caacuugac caauagucu 19

 <210> 457
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 457
 gcaacuuga ccaauagucu 20

 <210> 458
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 458
 ggcaacuug accaauaguc u 21

 <210> 459
 <211> 22

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 459
 aggcaaacuu gaccaauagu cu 22

 <210> 460
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 460
 aaggcaaacu ugaccaauag ucu 23

 <210> 461
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 461
 caaggcaaac uugaccaaua gucu 24

 <210> 462
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 462
 uucagacaga uauuugca 18

 <210> 463
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 463
 uuucagacag auauuugca 19

 <210> 464
 <211> 20
 <212> RNA
 <213> Artificial Sequence

8009W000_SequenceListing.txt

<220>
 <223> Targeting domain
 <400> 464
 guuucagaca gauauuuugca 20
 <210> 465
 <211> 21
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 465
 cguuucagac agauauuuugc a 21
 <210> 466
 <211> 22
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 466
 ccguuucaga cagauauuuug ca 22
 <210> 467
 <211> 23
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 467
 accguuucag acagauauuu gca 23
 <210> 468
 <211> 24
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 468
 gaccguuuca gacagauauu ugca 24
 <210> 469
 <211> 18
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 469		
aguuuccuuc ucccauca		18
<210> 470		
<211> 19		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 470		
uaguuuccuu cucccauca		19
<210> 471		
<211> 20		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 471		
cuaguuuccu ucucccauca		20
<210> 472		
<211> 21		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 472		
gcuaguuucc uucucccauc a		21
<210> 473		
<211> 22		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 473		
agcuaguuuc cuucucccau ca		22
<210> 474		
<211> 23		
<212> RNA		
<213> Artificial Sequence		
<220>		
<223> Targeting domain		
<400> 474		
uagcuaguuu ccuucuccca uca		23

8009W000_SequenceListing.txt

<210> 475
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 475
 uuagcuaguu uccuucuccc auca 24

 <210> 476
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 476
 auugagauag ugugggga 18

 <210> 477
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 477
 cauugagaua gugugggga 19

 <210> 478
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 478
 ugcauugaga uagugugggg a 21

 <210> 479
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 479
 uugcauugag auaguguggg ga 22

 <210> 480
 <211> 23

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 480
 uuugcauuga gauagugugg gga 23

<210> 481
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 481
 auuugcauug agauagugug ggga 24

<210> 482
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 482
 ucccaucaua gaggauac 18

<210> 483
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 483
 cucccaucau agaggauac 19

<210> 484
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 484
 ucucccauca uaggauac 20

<210> 485
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 485
 uucucccauc auagaggaua c 21

<210> 486
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 486
 cuucucccau cauagaggau ac 22

<210> 487
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 487
 ccuucuccca ucauagagga uac 23

<210> 488
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 488
 uccuucuccc aucauagagg auac 24

<210> 489
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 489
 uguggggaag gggccccc 18

<210> 490
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 490
 guguggggaa ggggcccc 19

<210> 491
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 491
 agugugggga aggggcccc 20

<210> 492
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 492
 uagugugggg aaggggcccc c 21

<210> 493
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 493
 auaguguggg gaaggggccc cc 22

<210> 494
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 494
 gauagugugg ggaaggggcc ccc 23

<210> 495
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 495
 agauagugug gggaaggggc cccc 24

8009W000_SequenceListing.txt

```

<210> 496
<211> 18
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 496
caugggugga guuuagcc 18

<210> 497
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 497
ccaugggugg aguuuagcc 19

<210> 498
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 498
acccaugggu ggaguuuagc c 21

<210> 499
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 499
aaccauggg uggaguuuag cc 22

<210> 500
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 500
caaccaugg guggaguuua gcc 23

<210> 501
<211> 24

```

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 501
 ccaacccaug gguggaguuu agcc 24

<210> 502
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 502
 ccaugggugg aguuuagc 18

<210> 503
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 503
 cccaugggug gaguuuagc 19

<210> 504
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 504
 acccaugggu ggaguuuagc 20

<210> 505
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 505
 aaccauggg uggaguuuag c 21

<210> 506
 <211> 22
 <212> RNA
 <213> Artificial Sequence

8009W000_SequenceListing.txt

<220>
 <223> Targeting domain

 <400> 506
 caacccaugg guggaguua gc 22

 <210> 507
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 507
 ccaacccaug gguggaguuu agc 23

 <210> 508
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 508
 gccaacccau gguggaguu uagc 24

 <210> 509
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 509
 ugauaaccuc agacguuc 18

 <210> 510
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 510
 uugauaaccu cagacguuc 19

 <210> 511
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

<400> 511 auugauaacc ucagacguuc	20
<210> 512 <211> 21 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 512 uauugauaac cucagacguu c	21
<210> 513 <211> 22 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 513 uuauugauaa ccucagacgu uc	22
<210> 514 <211> 23 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 514 cuuauugaua accucagacg uuc	23
<210> 515 <211> 24 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 515 gcuuauugau aaccucagac guuc	24
<210> 516 <211> 18 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 516 cauugagaua gugugggg	18

8009W000_SequenceListing.txt

<210> 517
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 517
 gcauugagau agugugggg 19

 <210> 518
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 518
 ugcaugaga uagugugggg 20

 <210> 519
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 519
 uugcauagag auaguguggg g 21

 <210> 520
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 520
 uuugcauga gauagugugg gg 22

 <210> 521
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 521
 auuugcauug agauagugug ggg 23

 <210> 522
 <211> 24

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 522
 uauuuugcauu gagauagugu gggg 24

 <210> 523
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 523
 aggcuggcca acccaugg 18

 <210> 524
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 524
 aaggcuggcc aaccaugg 19

 <210> 525
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 525
 caaggcuggc caaccaugg 20

 <210> 526
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 526
 gcaaggcugg ccaaccaug g 21

 <210> 527
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 527
 ggcaaggcug gccaacccau gg 22

<210> 528
 <211> 23
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 528
 aggcaaggcu ggccaacca ugg 23

<210> 529
 <211> 24
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 529
 aaggcaaggc uggccaaccc augg 24

<210> 530
 <211> 18
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 530
 agcgagugug uggaacug 18

<210> 531
 <211> 19
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 531
 aagcgagugu guggaacug 19

<210> 532
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain

<400> 532	
gaagcgagug uguggaacug	20
<210> 533	
<211> 21	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 533	
agaagcgagu guguggaacu g	21
<210> 534	
<211> 22	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 534	
cagaagcgag uguguggaac ug	22
<210> 535	
<211> 23	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 535	
ccagaagcga guguguggaa cug	23
<210> 536	
<211> 24	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 536	
uccagaagcg agugugugga acug	24
<210> 537	
<211> 18	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 537	
auuugcauug agauagug	18

8009W000_SequenceListing.txt

<210>	538	
<211>	19	
<212>	RNA	
<213>	Artificial Sequence	
<220>		
<223>	Targeting domain	
<400>	538	
	uauuuugcauu gagauagug	19
<210>	539	
<211>	21	
<212>	RNA	
<213>	Artificial Sequence	
<220>		
<223>	Targeting domain	
<400>	539	
	gauauuuugca uugagauagu g	21
<210>	540	
<211>	22	
<212>	RNA	
<213>	Artificial Sequence	
<220>		
<223>	Targeting domain	
<400>	540	
	agauauuuugc auugagauag ug	22
<210>	541	
<211>	23	
<212>	RNA	
<213>	Artificial Sequence	
<220>		
<223>	Targeting domain	
<400>	541	
	cagauauuuug cauugagaua gug	23
<210>	542	
<211>	24	
<212>	RNA	
<213>	Artificial Sequence	
<220>		
<223>	Targeting domain	
<400>	542	
	acagauauuu gcauugagau agug	24
<210>	543	
<211>	18	

<212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 543
 guuccagaag cgagugug 18

<210> 544
 <211> 19
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 544
 cguuccagaa gcgagugug 19

<210> 545
 <211> 21
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 545
 gacguuccag aagcgagugu g 21

<210> 546
 <211> 22
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 546
 agacguucca gaagcgagug ug 22

<210> 547
 <211> 23
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 547
 cagacguucc agaagcgagu gug 23

<210> 548
 <211> 24
 <212> RNA
 <213> Artificial Sequence

8009W000_SequenceListing.txt

<220>
 <223> Targeting domain
 <400> 548
 ucagacguuc cagaagcgag ugug 24
 <210> 549
 <211> 18
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 549
 uugcauugag auagugug 18
 <210> 550
 <211> 19
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 550
 uuugcauuga gauagugug 19
 <210> 551
 <211> 21
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 551
 uauuuugcau gagauagugu g 21
 <210> 552
 <211> 22
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 552
 auauuuugcau ugagauagug ug 22
 <210> 553
 <211> 23
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain

<400> 553	
gauauuuugca uugagauagu gug	23
<210> 554	
<211> 24	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 554	
agauauuuugc auugagauag ugug	24
<210> 555	
<211> 18	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 555	
uauuuugcauu gagauagu	18
<210> 556	
<211> 19	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 556	
auauuuugcau ugagauagu	19
<210> 557	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 557	
gauauuuugca uugagauagu	20
<210> 558	
<211> 21	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 558	
agauauuuugc auugagauag u	21

8009W000_SequenceListing.txt

<210> 559
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 559
 cagauauuug cauugagaua gu 22

 <210> 560
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 560
 acagauauuu gc auugagau agu 23

 <210> 561
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 561
 gacagauauu ugcauugaga uagu 24

 <210> 562
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 562
 cguuccagaa gcgagugu 18

 <210> 563
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 563
 acguuccaga agcgagugu 19

 <210> 564
 <211> 20

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 564
 gacguuccag aagcgagugu 20

<210> 565
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 565
 agacguucca gaagcgagug u 21

<210> 566
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 566
 cagacguucc agaagcgagu gu 22

<210> 567
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 567
 ucagacguuc cagaagcgag ugu 23

<210> 568
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 568
 cucagacguu ccagaagcga gugu 24

<210> 569
 <211> 18
 <212> RNA
 <213> Artificial Sequence

8009W000_SequenceListing.txt

<220>
 <223> Targeting domain

<400> 569
 uuugcauuga gauagugu 18

<210> 570
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 570
 auuugcauug agauagugu 19

<210> 571
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 571
 auauuugcau ugagauagug u 21

<210> 572
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 572
 gauauuugca uugagauagu gu 22

<210> 573
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 573
 agauauuugc auugagauag ugu 23

<210> 574
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 574
 cagauauuug cauugagaua gugu 24

<210> 575
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 575
 gaauaaauua gagaaaaa 18

<210> 576
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 576
 agaauaaauu agagaaaaa 19

<210> 577
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 577
 aagaauaaau uagagaaaaa 20

<210> 578
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 578
 gaagaauaaa uuagagaaaa a 21

<210> 579
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 579
 ggaagaauaa auuagagaaa aa 22

8009W000_SequenceListing.txt

```

<210> 580
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 580
gggaagaaua aaauagagaa aaa                23

<210> 581
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 581
aggaagaau aaauagaga aaaa                24

<210> 582
<211> 18
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 582
cuagggauga agaauaaa                18

<210> 583
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 583
gcuagggaug aagaauaaa                19

<210> 584
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 584
uggcuaggga ugaagaauaa a                21

<210> 585
<211> 22

```

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 585
 cuggcuaggg augaagaaua aa 22

 <210> 586
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 586
 gcuggcuagg gaugaagaau aaa 23

 <210> 587
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 587
 ggcuggcuag ggaugaagaa uaaa 24

 <210> 588
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 588
 agaaggaaac uagcuaaa 18

 <210> 589
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 589
 gagaaggaaa cuagcuaaa 19

 <210> 590
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 590
 gggagaagga aacuagcuaa a 21

<210> 591
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 591
 ugggagaagg aaacuagcua aa 22

<210> 592
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 592
 augggagaag gaaacuagcu aaa 23

<210> 593
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 593
 gaugggagaa ggaaacuagc uaaa 24

<210> 594
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 594
 aaaacuggaa ugacugaa 18

<210> 595
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 595
 aaaaacugga augacugaa 19

<210> 596
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 596
 gaaaaacugg aaugacugaa 20

<210> 597
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 597
 agaaaaacug gaaugacuga a 21

<210> 598
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 598
 gagaaaaacu ggaaugacug aa 22

<210> 599
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 599
 agagaaaaac uggaugacu gaa 23

<210> 600
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 600
 uagagaaaaa cuggaaugac ugaa 24

8009W000_SequenceListing.txt

<210> 601
<211> 18
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 601
gcuagggau aagaauaa 18

<210> 602
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 602
ggcuaggau gaagaauaa 19

<210> 603
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 603
uggcuaggga ugaagaauaa 20

<210> 604
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 604
cuggcuagg augaagaau a 21

<210> 605
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 605
gcuggcuagg gaugaagaau aa 22

<210> 606
<211> 23

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 606
 ggcuggcuag ggaugaagaa uaa 23

<210> 607
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 607
 cggcuggcua gggaugaaga auaa 24

<210> 608
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 608
 gagaaggaaa cuagcuaa 18

<210> 609
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 609
 ggagaaggaa acuagcuaa 19

<210> 610
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 610
 ugggagaagg aaacuagcua a 21

<210> 611
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 611
 augggagaag gaaacuagcu aa 22
 <210> 612
 <211> 23
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 612
 gaugggagaa ggaaacuagc uaa 23
 <210> 613
 <211> 24
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 613
 ugaugggaga aggaaacuag cuaa 24
 <210> 614
 <211> 18
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 614
 auagucuag agaucca 18
 <210> 615
 <211> 19
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 615
 aaagucuua gaguaacca 19
 <210> 616
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain

<400> 616 caauagucuu agaguaucca	20
<210> 617 <211> 21 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 617 ccaauagucu uagaguaucc a	21
<210> 618 <211> 22 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 618 accaauaguc uuagaguauc ca	22
<210> 619 <211> 23 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 619 gaccaauagu cuuagaguau cca	23
<210> 620 <211> 24 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 620 ugaccaauag ucuuagagua ucca	24
<210> 621 <211> 18 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 621 auccucuaug augggaga	18

8009W000_SequenceListing.txt

<210> 622
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 622
 uauccucuau gaugggaga 19

 <210> 623
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 623
 gguaucucu augaugggag a 21

 <210> 624
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 624
 ugguaucuc uauguggga ga 22

 <210> 625
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 625
 cugguauc cuauguggg aga 23

 <210> 626
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 626
 ccugguauc ucuaugugg gaga 24

 <210> 627
 <211> 18

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 627
 uccugguauc cucuauga 18

<210> 628
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 628
 guccugguau ccucuauga 19

<210> 629
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 629
 aaguccuggu auccucuaug a 21

<210> 630
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 630
 gaaguccugg uauccucuaug ga 22

<210> 631
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 631
 agaaguccug guauccucua uga 23

<210> 632
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

 <400> 632
 aagaaguuccu gguauccucu auga 24

 <210> 633
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 633
 ggagaaggaa acuagcua 18

 <210> 634
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 634
 gggagaagga aacuagcua 19

 <210> 635
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 635
 ugggagaagg aaacuagcua 20

 <210> 636
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 636
 augggagaag gaaacuagcu a 21

 <210> 637
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 637
 gaugggagaa ggaaacuagc ua 22

<210> 638
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 638
 ugaugggaga aggaaacuag cua 23

<210> 639
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 639
 augaugggag aaggaaacua gcua 24

<210> 640
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 640
 aaaggaaga auaaaauua 18

<210> 641
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 641
 uaaaggaag aaauaaauua 19

<210> 642
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 642
 cuaaaggga gaauaaauua 20

8009W000_SequenceListing.txt

```

<210> 643
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 643
gcuaaaaggga agaauaaaau a                21

<210> 644
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 644
agcuaaaaggg aagaauaaaau ua            22

<210> 645
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 645
uagcuaaaagg gaagaauaaa uua            23

<210> 646
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 646
cuagcuaaag ggaagaauaa auua            24

<210> 647
<211> 18
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 647
agaguaucca gugaggcc                    18

<210> 648
<211> 19

```

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 648
 uagaguaucc agugaggcc 19

<210> 649
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 649
 cuuagaguau ccagugaggc c 21

<210> 650
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 650
 ucuuagagua uccagugagg cc 22

<210> 651
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 651
 gucuuagagu auccagugag gcc 23

<210> 652
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 652
 agucuuagag uauccaguga ggcc 24

<210> 653
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 653
 uagaguaucc agugaggc 18

<210> 654
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 654
 uuagaguau cagugaggc 19

<210> 655
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 655
 cuuagaguau ccagugaggc 20

<210> 656
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 656
 ucuuagagua uccagugagg c 21

<210> 657
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 657
 gucuuagagu auccagugag gc 22

<210> 658
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 658 agucuuagag uauccaguga ggc	23
<210> 659 <211> 24 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 659 uagucuuaga guauccagug aggc	24
<210> 660 <211> 18 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 660 caggggccgg cggcuggc	18
<210> 661 <211> 19 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 661 ccaggggccg gcggcuggc	19
<210> 662 <211> 20 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 662 gccaggggcc gccggcuggc	20
<210> 663 <211> 21 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 663 ggccaggggc cggcggcugg c	21

8009W000_SequenceListing.txt

<210> 664
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 664
 aggccagggg ccggcggcug gc 22

 <210> 665
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 665
 gaggccaggg gccggcggcu ggc 23

 <210> 666
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 666
 ugaggccagg ggccggcggc uggc 24

 <210> 667
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 667
 aaaaauagca guauccuc 18

 <210> 668
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 668
 aaaaauagc aguauccuc 19

 <210> 669
 <211> 20

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 669
 aaaaaauuag caguauccuc 20

<210> 670
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 670
 aaaaaauua gcaguauccu c 21

<210> 671
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 671
 aaaaaaaauu agcaguaucc uc 22

<210> 672
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 672
 uaaaaaaaau uagcaguauc cuc 23

<210> 673
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 673
 auaaaaaaaa uuagcaguau ccuc 24

<210> 674
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 674
 guuccacaca cucgcuuc 18

<210> 675
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 675
 aguuccacac acucgcuuc 19

<210> 676
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 676
 gcaguuccac acacucgcuu c 21

<210> 677
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 677
 agcaguucca cacacucgcu uc 22

<210> 678
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 678
 cagcaguucc acacacucgc uuc 23

<210> 679
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 679
 ucagcaguuc cacacacucg cuuc 24

<210> 680
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 680
 uauccucuau gaugggag 18

<210> 681
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 681
 guauccucua ugaugggag 19

<210> 682
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 682
 gguauccucu augaugggag 20

<210> 683
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 683
 ugguauccuc uaugauggga g 21

<210> 684
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 684
 cugguauccu cuaugauggg ag 22

8009W000_SequenceListing.txt

<210> 685
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 685
 ccugguaucc ucuaugaugg gag 23

 <210> 686
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 686
 uccugguauuc cucuaugaug ggag 24

 <210> 687
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 687
 gccggcgguu ggcuaggg 18

 <210> 688
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 688
 ggccggcggu ugguaggg 19

 <210> 689
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 689
 gggccggcgg cuggcuaggg 20

 <210> 690
 <211> 21

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 690
 ggggccggcg gcuggcuagg g 21

<210> 691
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 691
 aggggccggc ggcuggcuag gg 22

<210> 692
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 692
 caggggccgg cggcuggcua ggg 23

<210> 693
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 693
 ccaggggccg gcggcuggcu aggg 24

<210> 694
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 694
 ugguauccuc uaugagg 18

<210> 695
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 695
 cugguauccu cuaugaugg 19

<210> 696
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 696
 ccugguaucc ucuaugaugg 20

<210> 697
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 697
 uccugguaucc cucuaugaug g 21

<210> 698
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 698
 guccugguau ccucuaugau gg 22

<210> 699
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 699
 aguccuggua uccucuauga ugg 23

<210> 700
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 700 aaguccuggu auccucuaug augg	24
<210> 701 <211> 18 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 701 guccugguau ccucuaug	18
<210> 702 <211> 19 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 702 aguccuggua uccucuaug	19
<210> 703 <211> 20 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 703 aaguccuggu auccucuaug	20
<210> 704 <211> 21 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 704 gaaguccugg uauccucuaug g	21
<210> 705 <211> 22 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 705 agaaguccug guauccucua ug	22

8009W000_SequenceListing.txt

<210> 706
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 706
 aagaaguccu gguauccucu aug 23

 <210> 707
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 707
 aaagaagucc ugguauccuc uaug 24

 <210> 708
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 708
 cuaaagggaa gaauaaau 18

 <210> 709
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 709
 gcuaaagggg agaauaaau 19

 <210> 710
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 710
 agcuaaaggg aagaauaaau 20

 <210> 711
 <211> 21

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 711
 uagcuaaagg gaagaauaaa u 21

<210> 712
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 712
 cuagcuaaag ggaagaauaa au 22

<210> 713
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 713
 acuagcuaaa gggaagaaua aa 23

<210> 714
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 714
 aacuagcuaa aggaagaau aa 24

<210> 715
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 715
 aaacuggaau gacugaau 18

<210> 716
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 716
 aaaacuggaa ugacugaau 19

<210> 717
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 717
 gaaaaacugg aaugacugaa u 21

<210> 718
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 718
 agaaaaacug gaaugacuga au 22

<210> 719
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 719
 gagaaaaacu ggaaugacug aaau 23

<210> 720
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 720
 agagaaaaac uggaugacu gaau 24

<210> 721
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 721
 ccugguaucc ucuaugau 18

<210> 722
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 722
 uccugguauuc cucuaugau 19

<210> 723
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 723
 aguccuggua uccucuauga u 21

<210> 724
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 724
 aaguccuggu auccucuaug au 22

<210> 725
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 725
 gaaguccugg uauccucuaug gau 23

<210> 726
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 726
 agaaguccug guauccucua ugau 24

8009W000_SequenceListing.txt

```

<210> 727
<211> 18
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 727
aaauuagcag uauccucu 18

<210> 728
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 728
aaaauuagca guauccucu 19

<210> 729
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 729
aaaaaaauuag caguauccuc u 21

<210> 730
<211> 22
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 730
aaaaaaaaua gcaguauccu cu 22

<210> 731
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 731
aaaaaaaaauu agcaguaucc ucu 23

<210> 732
<211> 24

```

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 732
 uaaaaaaaaau uagcaguauc cucu 24

 <210> 733
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 733
 cacucgcuuc uggaacgu 18

 <210> 734
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 734
 acacucgcuu cuggaacgu 19

 <210> 735
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 735
 cacacucgcu ucuggaacgu 20

 <210> 736
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 736
 acacacucgc uucuggaacg u 21

 <210> 737
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 737
 cacacacucg cuucuggaac gu 22

<210> 738
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 738
 ccacacacuc gcuucuggaa cgu 23

<210> 739
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 739
 uccacacacu cgcucugga acgu 24

<210> 740
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 740
 cucaaugcaa auaucugu 18

<210> 741
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 741
 ucucuaugca aaauaucugu 19

<210> 742
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 742 aucucaaugc aaauaucugu	20
<210> 743 <211> 21 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 743 uaucaucaug caaaauaucug u	21
<210> 744 <211> 22 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 744 cuaucau gcaaaauaucu gu	22
<210> 745 <211> 23 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 745 acuaucuaa ugcaaaauauc ugu	23
<210> 746 <211> 24 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 746 cacuaucua augcaaaauau cugu	24
<210> 747 <211> 18 <212> RNA <213> Artificial Sequence	
<220> <223> Targeting domain	
<400> 747 aguuccacac acucgcuu	18

8009W000_SequenceListing.txt

<210> 748
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 748
 caguuccaca cacucgcuu 19

 <210> 749
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 749
 gcaguuccac acacucgcuu 20

 <210> 750
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 750
 agcaguucca cacacucgcu u 21

 <210> 751
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 751
 cagcaguucc acacacucgc uu 22

 <210> 752
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 752
 ucagcaguuc cacacacucg cuu 23

 <210> 753
 <211> 24

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 753
 uucagcaguu ccacacacuc gcuu 24

 <210> 754
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 754
 aaauagcagu auccucuu 18

 <210> 755
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 755
 aaauuagcag uauccucuu 19

 <210> 756
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 756
 aaaaauuagc aguauccucu u 21

 <210> 757
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 757
 aaaaaauuag caguauccuc uu 22

 <210> 758
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 758
 aaaaaaaauua gcaguauccu cuu 23

<210> 759
 <211> 24
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 759
 aaaaaaaauu agcaguaucc ucuu 24

<210> 760
 <211> 17
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 760
 gaagaaaacu agcuaaa 17

<210> 761
 <211> 17
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 761
 gcagcaguau ccucuug 17

<210> 762
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 762
 ggagaagaaa acuagcuaaa 20

<210> 763
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain

<400> 763
 gggagaagaa aacuagcuaa 20

<210> 764
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 764
 guccugguau cuucuauggu 20

<210> 765
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 765
 agaagaaaac uagcuaa 17

<210> 766
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 766
 aguccuggua ucuucua 17

<210> 767
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 767
 ccaccuauaga agauacc 17

<210> 768
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 768
 ccugguaucu ucuaugg 17

8009W000_SequenceListing.txt

<210> 769
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 769
 cagcaguauc cucuugg 17

<210> 770
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 770
 aaauuggaaug acugaau 17

<210> 771
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 771
 cugguaucuu cuauggu 17

<210> 772
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 772
 agcagcagua uccucuu 17

<210> 773
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 773
 agaaguccug guaucuucua 20

<210> 774
 <211> 20

<212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 774
 cuccaccacau agaagauacc 20

<210> 775
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 775
 aguccuggua ucuucuaugg 20

<210> 776
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 776
 aagcagcagu auccucuugg 20

<210> 777
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 777
 uaagcagcag uauccucuug 20

<210> 778
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 778
 agaauaaaau agagaaaaau 20

<210> 779
 <211> 20
 <212> RNA
 <213> Artificial Sequence

8009W000_SequenceListing.txt

<220>
 <223> Targeting domain
 <400> 779
 aaaaauugga augacugaau 20

<210> 780
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 780
 auuaagcagc aguauccucu 20

<210> 781
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 781
 uuaagcagca guauccucuu 20

<210> 782
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 782
 auaaaauaga gaaaaau 17

<210> 783
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 783
 aagcagcagu auccucu 17

<210> 784
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 784	
gaagaaaacu agcuaaaggg	20
<210> 785	
<211> 22	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 785	
gagaagaaaa cuagcuaaag gg	22
<210> 786	
<211> 23	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 786	
ggagaagaaa acuagcuaaa ggg	23
<210> 787	
<211> 24	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 787	
gggagaagaa aacuagcuaa aggg	24
<210> 788	
<211> 19	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 788	
gaauaaaaua gagaaaaau	19
<210> 789	
<211> 22	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Targeting domain	
<400> 789	
gaagaauaaa uuagagaaaa au	22

8009W000_SequenceListing.txt

```

<210> 790
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 790
ggaagaauaa auuagagaaa aaau                                     23

<210> 791
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 791
gggaagaaua aaauagagaa aaau                                     24

<210> 792
<211> 18
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 792
agagaaaaau uggaauga                                             18

<210> 793
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 793
uagagaaaaa uuggaauga                                             19

<210> 794
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 794
uuagagaaaa auuggaauga                                           20

<210> 795
<211> 21

```

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 795
 auuagagaaa aaauuggaaug a 21

<210> 796
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 796
 aaauagagaa aaauuggaau ga 22

<210> 797
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 797
 aaauuagaga aaaauuggaa uga 23

<210> 798
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 798
 uaaauuagag aaaaauugga auga 24

<210> 799
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 799
 agaaaacuag cuaaaggg 18

<210> 800
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 800
 aagaaaacua gcuaaaggg 19

<210> 801
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 801
 agaagaaaac uagcuaaagg g 21

<210> 802
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 802
 aauaaaauag agaaaaau 18

<210> 803
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 803
 aagaauaaaau uagagaaaaa u 21

<210> 804
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 804
 aguuuucuuc ucccacca 18

<210> 805
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 805
 uaguuuuuuu cucccacca 19

<210> 806
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 806
 cuaguuuuuu ucucccacca 20

<210> 807
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 807
 gcuaguuuuu uucucccacc a 21

<210> 808
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 808
 agcuaguuuu cuucucccac ca 22

<210> 809
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 809
 uagcuaguuu ucuucuccca cca 23

<210> 810
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 810
 uuagcuaguu uucuuuccc acca 24

8009W000_SequenceListing.txt

<210> 811
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 811
 ucccaccua gaagauac 18

<210> 812
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 812
 cucccaccu agaagauac 19

<210> 813
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 813
 ucuccacca uagaagauac 20

<210> 814
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 814
 uucucccacc auagaagaua c 21

<210> 815
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 815
 cuucucccac cauagaagau ac 22

<210> 816
 <211> 23

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 816
 ucuucuccca ccuagaaga uac 23

<210> 817
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 817
 uucuuuccc accuagaag auac 24

<210> 818
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 818
 agaagaaaac uagcuaaa 18

<210> 819
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 819
 gagaagaaa cuagcuaaa 19

<210> 820
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 820
 gggagaagaa aacuagcuaa a 21

<210> 821
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 821
 ugggagaaga aaacuagcua aa 22
 <210> 822
 <211> 23
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 822
 gugggagaag aaaacuagcu aaa 23
 <210> 823
 <211> 24
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 823
 ggugggagaa gaaaacuagc uaaa 24
 <210> 824
 <211> 18
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 824
 aaaaauuggaa ugacugaa 18
 <210> 825
 <211> 19
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 825
 aaaaauugga augacugaa 19
 <210> 826
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain

<400> 826
 gaaaaauugg aaugacugaa 20

<210> 827
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 827
 agaaaaauug gaaugacuga a 21

<210> 828
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 828
 gagaaaaauu ggaaugacug aa 22

<210> 829
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 829
 agagaaaaau uggaaugacu gaa 23

<210> 830
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 830
 uagagaaaaa uuggaaugac ugaa 24

<210> 831
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 831
 gagaagaaaa cuagcuaa 18

8009W000_SequenceListing.txt

<210> 832
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 832
 ggagaagaaa acuagcuaa 19

 <210> 833
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 833
 ugggagaaga aaacuagcua a 21

 <210> 834
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 834
 gugggagaag aaaacuagcu aa 22

 <210> 835
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 835
 ggugggagaa gaaaacuagc uaa 23

 <210> 836
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 836
 uggugggaga agaaaacuag cuaa 24

 <210> 837
 <211> 18

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 837
 ggagaagaaa acuagcua 18

<210> 838
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 838
 gggagaagaa aacuagcua 19

<210> 839
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 839
 ugggagaaga aaacuagcua 20

<210> 840
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 840
 gugggagaag aaaacuagcu a 21

<210> 841
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 841
 ggugggagaa gaaaacuagc ua 22

<210> 842
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 842
 uggugggaga agaaaacuag cua 23

<210> 843
 <211> 24
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 843
 auggugggag aagaaaacua gcua 24

<210> 844
 <211> 18
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 844
 uuaagcagca guauccuc 18

<210> 845
 <211> 19
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 845
 auuaagcagc aguaucuc 19

<210> 846
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain
 <400> 846
 aaauaagcag caguaucuc 20

<210> 847
 <211> 21
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 847
 aaauuaagca gcaguauccu c 21

<210> 848
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 848
 aaaaauaagc agcaguaucc uc 22

<210> 849
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 849
 aaaaauaag cagcaguauc cuc 23

<210> 850
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 850
 aaaaaauuaa gcagcaguau ccuc 24

<210> 851
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 851
 uaucuucuau ggugggag 18

<210> 852
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 852
 guaucuucua uggugggag 19

8009W000_SequenceListing.txt

<210> 853
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 853
 gguaucuucu auggugggag 20

<210> 854
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 854
 ugguaucuuc uaugguggga g 21

<210> 855
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 855
 cugguaucu cuauuggugg ag 22

<210> 856
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 856
 ccugguaucu ucuauggugg gag 23

<210> 857
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 857
 uccugguauc uucuauggug ggag 24

<210> 858
 <211> 18

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 858
 uccugguauc uucuaugg 18

<210> 859
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 859
 guccugguau cuucuaugg 19

<210> 860
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 860
 aaguccuggu aucuucuaug g 21

<210> 861
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 861
 gaaguccugg uaucuucuaug gg 22

<210> 862
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 862
 agaaguccug guaucuucua ugg 23

<210> 863
 <211> 24
 <212> RNA
 <213> Artificial Sequence

8009W000_SequenceListing.txt

<220>
 <223> Targeting domain
 <400> 863
 aagaaguccu gguaucuucu augg 24

<210> 864
 <211> 18
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 864
 ugguaucuuc uauggugg 18

<210> 865
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 865
 cugguaucu cuauuggugg 19

<210> 866
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 866
 ccugguaucu ucuauggugg 20

<210> 867
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 867
 uccugguauc uucuauggug g 21

<210> 868
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 868
guccugguau cuucuauggu gg 22

<210> 869
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 869
aguccuggua ucuucuaugg ugg 23

<210> 870
<211> 24
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 870
aaguccuggu aucuucuaug gugg 24

<210> 871
<211> 18
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 871
guccugguau cuucuaug 18

<210> 872
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 872
aguccuggua ucuucuaug 19

<210> 873
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Targeting domain

<400> 873
aaguccuggu aucuucuaug 20

8009W000_SequenceListing.txt

<210> 874
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 874
 gaaguccugg uaucuucua u g 21

 <210> 875
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 875
 agaaguccug guaucuucua ug 22

 <210> 876
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 876
 aagaaguccu gguaucuucu aug 23

 <210> 877
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 877
 aaagaagucc ugguaucuuc uaug 24

 <210> 878
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 878
 aaauuggaau gacugaau 18

 <210> 879
 <211> 19

<212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 879
 aaaaauuggaa ugacugaau 19

<210> 880
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 880
 gaaaaauugg aaugacugaa u 21

<210> 881
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 881
 agaaaaauug gaaugacuga au 22

<210> 882
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 882
 gagaaaaauu ggaaugacug aaau 23

<210> 883
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 883
 agagaaaaau uggaugacu gaau 24

<210> 884
 <211> 18
 <212> RNA
 <213> Artificial Sequence

8009W000_SequenceListing.txt

<220>
 <223> Targeting domain
 <400> 884
 uaagcagcag uauccucu 18

<210> 885
 <211> 19
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 885
 uuaagcagca guauccucu 19

<210> 886
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 886
 aaauaagcag caguauccuc u 21

<210> 887
 <211> 22
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 887
 aaauaagca gcaguauccu cu 22

<210> 888
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain
 <400> 888
 aaaauaagc agcaguaucc ucu 23

<210> 889
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

8009W000_SequenceListing.txt

<400> 889
 aaaaaauaag cagcaguauc cucu 24

 <210> 890
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 890
 ccugguauc ucuauggu 18

 <210> 891
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 891
 uccugguauc uucuauggu 19

 <210> 892
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 892
 aguccuggua ucuucuaugg u 21

 <210> 893
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 893
 aaguccuggu aucuucuaug gu 22

 <210> 894
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 894
 gaaguccugg uaucuucuaau ggu 23

8009W000_SequenceListing.txt

<210> 895
 <211> 24
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 895
 agaaguccug guaucuucua uggu 24

 <210> 896
 <211> 18
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 896
 aagcagcagu auccucuu 18

 <210> 897
 <211> 19
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 897
 uaagcagcag uauccucuu 19

 <210> 898
 <211> 21
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 898
 auuaagcagc aguauccucu u 21

 <210> 899
 <211> 22
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> Targeting domain

 <400> 899
 aaauaagcag caguauccuc uu 22

 <210> 900
 <211> 23

<212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 900
 aaauuaagca gcaguauccu cuu

23

<210> 901
 <211> 24
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> Targeting domain

<400> 901
 aaaaauaagc agcaguaucc ucuu

24

<210> 902
 <211> 4758
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (2716)..(2719)
 <223> HPFH deletion site (4 bp del -225 to -222)

<220>
 <221> misc_feature
 <222> (2748)..(2753)
 <223> GATA1 binding motif

<220>
 <221> misc_feature
 <222> (2762)..(2767)
 <223> GATA1 binding motif

<220>
 <221> misc_feature
 <222> (2791)..(2799)
 <223> FKLf transcription factor binding motif

<220>
 <221> misc_feature
 <222> (2823)..(2830)
 <223> CP1/Coup TFII binding motif

<220>
 <221> misc_feature
 <222> (2824)..(2836)
 <223> HPFH deletion site (13 bp del -114 to -102)

<220>
 <221> misc_feature
 <222> (2991)..(2993)
 <223> Start codon

8009W000_SequenceListing.txt

<400> 902

tttaggaagt	caaggttttag	gcagggatag	ccatttctatt	ttattagggg	caatactatt	60
tccaacggca	tctggctttt	ctcagccctt	gtgaggctct	acagggaggt	tgagggtgta	120
gagatcagag	caggaaacag	gtttttcttt	ccacggtaac	tacaatgaag	tgatccttac	180
tttactaagg	aacttttcat	tttaagtgtt	gacgcatgcc	taaagaggtg	aaattaatcc	240
cataccctta	agtctacaga	ctggtcacag	catttcaagg	aggagacctc	attgtaagct	300
tctagggagg	tggggactta	ggtgaaggaa	atgagccagc	agaagctcac	aagtcagcat	360
cagcgtgtca	tgtctcagca	gcagaacagc	acggtcagat	gaaaatatag	tgtgaagaat	420
ttgtataaca	ttaattgaga	aggcagattc	actggagttc	ttatataatt	gaaagttaat	480
gcacgttaat	aagcaagagt	ttagtttaat	gtgatgggtg	tatgaactta	acgcttgtgt	540
ctccagaaaa	ttcacatgct	gaatcccaa	ctccaattg	gctccatttg	tgggggaggc	600
tttgaaaaag	taatcaggtt	tagaggagct	catgagagca	gatccccatc	atagaattat	660
tttcctcatc	agaagcagag	agattagcca	tttctcttcc	ttctggtgag	gacacagtgg	720
gaagtcagcc	acctgcaacc	caggaagaga	gccctgacca	ggaaccagca	gaaaagttag	780
aaaaaatcct	gttgttgaag	tcaccagtc	tatgctat	tggttatagca	ccttgcacta	840
agtaaggcag	atgaagaaa	agaaaaaat	aagcttcggt	gttcagtgga	ttagaaacca	900
tgtttatctc	aggtttacaa	atctccactt	gtcctctgtg	tttcagaata	aaataccaac	960
tctactactc	tcattctgt	gatgcaaata	gtaagcctga	gcccttctgt	ctaactttga	1020
attctat	ttcttcaacg	tacttttaggc	ttgtaatgtg	tttatataca	gtgaaatgtc	1080
aagttctttc	tttatat	tttctttctt	ttttttcctc	agcctcagag	ttttccacat	1140
gcccttccta	ctttcaggaa	cttctttctc	caaacgtctt	ctgcctgggt	ccatcaaatc	1200
ataaaggacc	cacttcaaat	gccatcactc	actaccattt	cacaattcgc	actttctttc	1260
tttgtccttt	tttttttttag	taaaacaagt	ttataaaaaa	ttgaagggaat	aaatgaatgg	1320
ctacttcata	ggcagagtag	acgcaagggc	tactggttgc	cgatttttat	tggtattttt	1380
caatagtatg	ctaaacaagg	ggtagattat	ttatgctgcc	cattttttaga	ccataaaaaga	1440
taacttcctg	atgttgccat	ggcatttttt	tcctttta	tttatttcat	ttcattttta	1500
tttcgaagg	acatgtgcag	gatgtgcagg	cttgttacat	gggtaaatgt	gtgtctttct	1560
ggccttttag	ccatctgtat	caatgagcag	atataagctt	tacacaggat	catgaaggat	1620
gaaagaat	caccaatatt	ataataat	caatcaacct	gatagcttag	gggataaact	1680
aatttgaaga	tacagcttgc	ctccgataag	ccagaattcc	agagcttctg	gcattataat	1740
ctagcaagg	tagagatcat	ggatcacttt	cagagaaaaa	caaaaacaaa	ctaaccaaaa	1800
gcaaaacaga	acaaaaaac	caccataaat	acttcctacc	ctgttaatgg	tccaatatgt	1860

8009W000_SequenceListing.txt

cagaaacagc	actgtgttag	aaataaaagct	gtctaaagta	cactaatatt	cgagttataa	1920
tagtgtgtgg	actattagtc	aataaaaaca	acccttgcct	ctttagagtt	gttttccatg	1980
tacacgcaca	tcttatgtct	tagagtaaga	ttccctgaga	agtgaaccta	gcattttatac	2040
aagataatta	attctaatacc	acagtacctg	ccaaagaaca	ttctaccatc	atctttactg	2100
agcatagaag	agctacgcca	aaaccctggg	tcatcagcca	gcacacacac	ttatccagtg	2160
gtaaatacac	atcatctggt	gtatacatac	atacctgaat	atggaatcaa	atatttttct	2220
aagatgaaac	agtcattgatt	tattttcaa	aggtacggat	aagtagatat	tgaggtaagc	2280
attaggtctt	atattatgta	acactaatct	attactgcgc	tgaaactgtg	gctttataga	2340
aattgttttc	actgcactat	tgagaaatta	agagataatg	gcaaaagtca	caaagagtat	2400
attcaaaaag	aagtatagca	ctttttcctt	agaaaccact	gctaactgaa	agagactaag	2460
atttgtcccg	tcaaaaatcc	tggacctatg	cctaaaacac	atttcacaat	ccctgaactt	2520
ttcaaaaatt	ggtagatgct	ttagctttta	actacaggcc	tcactggagc	tagagacaag	2580
aaggtaaaaa	acggctgaca	aaagaagtcc	tggatccctc	tatgatggga	gaaggaaaact	2640
agctaaaggg	aagaataaat	tagagaaaaa	ctggaatgac	tgaatcggaa	caaggcaaag	2700
gctataaaaa	aaattagcag	tatcctcttg	ggggccctt	ccccacacta	tctcaatgca	2760
aatatctgtc	tgaaacggtc	cctggctaaa	ctccacccat	gggttgcca	gccttgcctt	2820
gaccaatagc	cttgacaagg	caaacttgac	caatagtctt	agagtatcca	gtgaggccag	2880
gggccggcgg	ctggctaggg	atgaagaata	aaaggaagca	cccttcagca	gttccacaca	2940
ctcgcttctg	gaacgtctga	ggttatcaat	aagctcctag	tccagacgcc	atgggtcatt	3000
tcacagagga	ggacaaggct	actatcaca	gcctgtgggg	caaggatgaat	gtggaagatg	3060
ctggaggaga	aaccctggga	aggtaggctc	tggtagaccag	gacaagggag	ggaaggaagg	3120
accctgtgcc	tggcaaaagt	ccaggctcgt	tctcaggatt	tgtggcacct	tctgactgtc	3180
aaactgttct	tgtcaatctc	acaggctcct	ggttgtctac	ccatggaccc	agaggttctt	3240
tgacagcttt	ggcaacctgt	cctctgcctc	tgccatcatg	ggcaacccca	aagtcaaggc	3300
acatggcaag	aagggtgctga	cttccttggg	agatgccaca	aagcacctgg	atgatctcaa	3360
gggcaccttt	gcccagctga	gtgaactgca	ctgtgacaag	ctgcatgtgg	atcctgagaa	3420
cttcaagggtg	agtccaggag	atgtttcagc	cctgttgcct	ttagtctcga	ggcaacttag	3480
acaacggagt	attgatctga	gcacagcagg	gtgtgagctg	tttgaagata	ctgggggttg	3540
gggtgaagaa	actgcagagg	actaactggg	ctgagaccca	gtggtaaatgt	tttagggcct	3600
aaggagtgcc	tctaaaaatc	tagatggaca	attttgactt	tgagaaaaga	gagggtggaaa	3660
tgaggaaaat	gacttttctt	tattagattc	cagtagaaaag	aactttcatc	tttccctcat	3720
ttttgttgtt	ttaaaacatc	tatctggagg	caggacaagt	atggctcgtta	aaaagatgca	3780

8009W000_SequenceListing.txt

```

ggcagaaggc atatattggc tcagtcaaag tggggaactt tggtgccaa acatacattg 3840
ctaaggctat tcctatatca gctggacaca tataaaatgc tgctaattgct tcattacaaa 3900
cttatatcct ttaattccag atgggggcaa agtatgtcca ggggtgagga acaattgaaa 3960
catttgggct ggagtagatt ttgaaagtca gctctgtgtg tgtgtgtgtg tgtgcgcgcg 4020
cgcggtgtgtg tgtgtgtgtc agcgtgtgtt tcttttaacg tcttcagcct acaacataca 4080
gggttcatgg tggcaagaag atagcaagat ttaaattatg gccagtgact agtgcttgaa 4140
ggggaacaac tacctgcatt taatgggaag gcaaaatctc aggctttgag ggaagttaac 4200
ataggcttga ttctgggtgg aagcttggtg tgtagttatc tggaggccag gctggagctc 4260
tcagctcact atgggttcat ctttattgtc tcctttcatc tcaacagctc ctgggaaatg 4320
tgctggtgac cgttttggca atccatttcg gcaaagaatt caccctgag gtgcaggctt 4380
cctggcagaa gatggtgact gcagtggcca gtgccctgtc ctccagatac cactgagctc 4440
actgcccattg attcagagct ttcaaggata ggctttattc tgcaagcaat acaaataata 4500
aatctattct gctgagagat cacacatgat tttcttcagc tctttttttt acatcttttt 4560
aaatatatga gccacaaagg gtttatattg aggggaagtgt gtatgtgtat ttctgcatgc 4620
ctgtttgtgt ttgtggtgtg tgcattgctc tcatttatat ttatatgaga tgtgcatttt 4680
gatgagcaaa taaaagcagt aaagacactt gtacacggga gttctgcaag tgggagtaaa 4740
tggtgtagga gaaatccg 4758

```

```

<210> 903
<211> 4773
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> (1233)..(1238)
<223> GATA1 binding site

```

```

<220>
<221> misc_feature
<222> (2672)..(2677)
<223> GATA1 binding site

```

```

<220>
<221> misc_feature
<222> (2686)..(2691)
<223> GATA1 binding site

```

```

<220>
<221> misc_feature
<222> (2715)..(2723)
<223> FKLf transcription factor binding motif

```

```

<220>

```

8009W000_SequenceListing.txt

```

<221> misc_feature
<222> (2747)..(2754)
<223> CP1/Coup TFII binding motif

<220>
<221> misc_feature
<222> (2748)..(2760)
<223> HPFH deletion site (13 bp del -114 to -102)

<220>
<221> misc_feature
<222> (2915)..(2917)
<223> Start codon

<400> 903
ttatgtcatt accagagtta aaattctata atggcttctc actccctacc actgaggaca      60
agtttatgtc cttaggttta tgcttccctg aaacaatacc acctgctatt ctccacttta      120
catatcaacg gcactgggtc tttatctaac tctctggcac agcaggagtt tgttttcttc      180
tgcttcagag ctttgaatth actatttcag cttctaaact ttatttggca atgccttccc      240
atggcagatt ccttctgtca ttttgcctct gttcgaatac tttctcctta atttcattct      300
tagttaataa tatctgaaat tattttgttg ttttaactta ttattaatth tatgtatgtt      360
ctacctagat tataatcttc agaggaaagt tttattctct gacttattta acttaaagtc      420
ccactactth aaaaattatg acatttattt aacagatatt tgctgaacaa atgtttgaaa      480
atacatggga aagaatgctt gaaaacactt gaaattgctt gtgtaaagaa acagttttat      540
cagttaggat ttaatcaatg tcagaagcaa tgatatagga aaaatcgagg aataagacag      600
ttatggataa ggagaaatca acaaactctt aaaagatatt gcctcaaaaag cataagagga      660
aataaggggt tatacatgac ttttagaaca ctgccttggg ttttggataa atgggggaagt      720
tgtttgaaaa caggagggat cctagatatt ccttagtctg aggaggagca attaagattc      780
acttgtttag aggctgggag tgggtggctca cgcctgtaat cccagaatth tgggaggcca      840
aggcaggcag atcacctgag gtcaagagtt caagaccaac ctggccaaca tggtgaaatc      900
ccatctctac aaaaatacaa aaattagaca ggcatgatgg caagtgcctg taatcccagc      960
tacttgggag gctgaggaag gagaattgct tgaacctgga aggcaggagt tgcagtgagc     1020
cgagatcata ccactgcact ccagcctggg tgacagaaca agactctgtc tcaaaaaaaaa     1080
aaaagagaga ttcaaaagat tcaattgtht aggccttagc gggcttagac accagtctct     1140
gacacattct taaagggtcag gctctacaaa tggaacccaa ccagactctc agatatggcc     1200
aaagatctat acacacccat ctacagatc ccctatctta aagagaccct aatttgggtht     1260
cacctcagtc tctataatct gtaccagcat accaataaaa atctttctca cccatcctta     1320
gattgagaga agtcacttat tattatgtga gtaactggaa gatactgata agttgacaaa     1380
tctttttctt tcctttctta ttcaacttht attttaactt ccaaagaaca agtgcaatat     1440

```

8009W000_SequenceListing.txt

gtgcagcttt	gttgcgagg	tcaacatgta	tctttctggt	cttttagccg	cctaacactt	1500
tgagcagata	taagccttac	acaggattat	gaagtctgaa	aggattccac	caatattatt	1560
ataattccta	tcaacctgat	aggtagggg	aaggtagagc	tctcctcaa	taagccagat	1620
ttccagagtt	tctgacgtca	taatctacca	aggatcatgga	tcgagttcag	agaaaaaaca	1680
aaagcaaaac	caaacctacc	aaaaaataaa	aatcccaaag	aaaaaataaa	gaaaaaaaca	1740
gcataaatac	ttcctgccat	gttaagtggc	caatatgtca	gaaacagcac	tgagttacag	1800
ataaagatgt	ctaaactaca	gtgacatccc	agctgtcaca	gtgtgtggac	tattagtcaa	1860
taaaacagtc	cctgcctctt	aagagttggt	ttccatgcaa	atacatgtct	tatgtcttag	1920
aataagattc	cctaagaagt	gaacctagca	tttatacaag	ataattaatt	ctaataccata	1980
gtatctggta	aagagcattc	taccatcatc	tttaccgagc	atagaagagc	tacaccaaaa	2040
ccctgggtca	tcagccagca	catacactta	tccagtata	aatacacatc	atcggtgccc	2100
tacatacata	cctgaatata	aaaaaaatac	ttttgctgag	atgaaacagg	cgtgatttat	2160
ttcaaatagg	tacggataag	tagatattga	agtaaggatt	cagtcttata	ttatattaca	2220
taacattaat	ctattcctgc	actgaaaactg	ttgctttata	ggatttttca	ctacactaat	2280
gagaacttaa	gagataatgg	cctaaaacca	cagagagtat	attcaaagat	aagtatagca	2340
cttcttattt	ggaaaccaat	gcttactaaa	tgagactaag	acgtgtccca	tcaaaaatcc	2400
tggacctatg	cctaaaacac	atttcacaat	ccctgaactt	ttcaaaaatt	ggtacatgct	2460
ttaactttta	actacaggcc	tcaactggagc	tacagacaag	aagggtgaaa	acggctgaca	2520
aaagaagtcc	tggtatcttc	tatggtggga	gaagaaaact	agctaaaggg	aagaataaat	2580
tagagaaaaa	ttggaatgac	tgaatcggaa	caaggcaaag	gctataaaaa	aaattaagca	2640
gcagtatcct	cttggggggc	ccttccccac	actatctcaa	tgcaaataatc	tgtctgaaac	2700
ggtccctggc	taaactccac	ccatgggttg	gccagccttg	ccttgaccaa	tagccttgac	2760
aaggcaaaact	tgaccaatag	tcttagagta	tccagtgagg	ccaggggccg	gcggctggct	2820
agggatgaag	aataaaagga	agcacccttc	agcagttcca	cacactcgct	tctggaacgt	2880
ctgagggttat	caataagctc	ctagtccaga	cgccatgggt	catttcacag	aggaggacaa	2940
ggctactatc	acaagcctgt	ggggcaaggt	gaatgtggaa	gatgctggag	gagaaaccct	3000
gggaaggtag	gctctggtga	ccaggacaag	ggagggaagg	aaggaccctg	tgcttggaac	3060
aagtccaggt	cgcttctcag	gatttgtggc	accttctgac	tgtcaaaactg	ttcttgtcaa	3120
tctcacaggc	tcctggttgt	ctacccatgg	acccagaggt	tctttgacag	ctttggcaac	3180
ctgtcctctg	cctctgccat	catgggcaac	cccaaagtca	aggcacatgg	caagaagggtg	3240
ctgacttcct	tgggagatgc	cataaagcac	ctggatgatc	tcaagggcac	ctttgcccag	3300
ctgagtgaac	tgactgtga	caagctgcat	gtggatcctg	agaacttcaa	ggtgagtcca	3360

8009W000_SequenceListing.txt

```

ggagatgttt cagcactgtt gccttttagtc tcgaggcaac ttagacaact gagtattgat      3420
ctgagcacag caggggtgtga gctgtttgaa gatactgggg ttgggagtga agaaactgca      3480
gaggactaac tgggctgaga cccagtggca atgttttagg gcctaaggag tgcctctgaa      3540
aatctagatg gacaactttg actttgagaa aagagagggtg gaaatgagga aaatgacttt      3600
tctttattag atttcggtag aaagaacttt cacctttccc ctatttttgt tattcgtttt      3660
aaaacatcta tctggaggca ggacaagtat ggtcattaaa aagatgcagg cagaaggcat      3720
atattggctc agtcaaagtg gggaactttg gtggccaaac atacattgct aaggctattc      3780
ctatatcagc tggacacata taaaatgctg ctaatgcttc attacaaact tatatccttt      3840
aattccagat gggggcaaaag tatgtccagg ggtgaggaac aattgaaaca tttgggctgg      3900
agtagatttt gaaagtcagc tctgtgtgtg tgtgtgtgtg tgtgcgcgcg tgtgtttgtg      3960
tgtgtgtgag agcgtgtgtt tcttttaacg ttttcagcct acagcataca gggttcatgg      4020
tggcaagaag ataacaagat ttaaattatg gccagtgact agtgctgcaa gaagaacaac      4080
tacctgcatt taatgggaaa gcaaaatctc aggctttgag ggaagttaac ataggcttga      4140
ttctgggtgg aagcttggtg tgtagttatc tggaggccag gctggagctc tcagctcact      4200
atgggttcat ctttattgtc tcctttcatc tcaacagctc ctgggaaatg tgctggtgac      4260
cgttttggca atccatttcg gcaaagaatt caccctgag gtgcaggctt cctggcagaa      4320
gatggtgact ggagtggcca gtgccctgtc ctccagatac cactgagctc actgcccatg      4380
atgcagagct ttcaaggata ggctttattc tgcaagcaat caaataataa atctattctg      4440
ctaagagatc acacatgggt gtcttcagtt ctttttttat gtctttttaa atatatgagc      4500
cacaaagggg tttatgttga gggatgtgtt tatgtgtatt tatacatggc tatgtgtgtt      4560
tgtgtcatgt gcacactcca cacttttttg ttacgtttag atgtgggttt tgatgagcaa      4620
ataaaagaac taggcaataa agaaacttgt acatgggagt tctgcaagtg ggagtaaaag      4680
gtgcaggaga aatctggttg gaagaaagac ctctatagga caggactcct cagaaacaga      4740
tgttttggaa gagatgggga aaggttcagt gaa      4773

```

```

<210> 904
<211> 87
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> ssODN1 5' homology arm

```

```

<400> 904
gggtgcttcc ttttattctt catccctagc cagccgccgg cccctggcct cactggatac      60
tctaagacta ttggtcaagt ttgcctt      87

```

8009W000_SequenceListing.txt

```

<210> 905
<211> 89
<212> DNA
<213> Artificial Sequence

<220>
<223> ssODN1 3' homology arm

<400> 905
gtcaaggcaa ggctggccaa cccatgggtg gagtttagcc agggaccgtt tcagacagat      60
atttgcat tg agatagtgtg gggaagggg                                         89

<210> 906
<211> 176
<212> DNA
<213> Artificial Sequence

<220>
<223> ssODN1

<400> 906
gggtgcttcc ttttattctt catccctagc cagccgccgg cccctggcct cactggatac      60
tctaagacta ttggtcaagt ttgccttgtc aaggcaaggc tggccaaccc atgggtggag      120
tttagccagg gaccgtttca gacagatatt tgcattgaga tagtgtgggg aagggg         176

<210> 907
<211> 87
<212> DNA
<213> Artificial Sequence

<220>
<223> PhTx ssODN1 5' homology arm

<220>
<221> misc_feature
<222> (1)..(1)
<223> Modified to contain phosphothioate

<400> 907
gggtgcttcc ttttattctt catccctagc cagccgccgg cccctggcct cactggatac      60
tctaagacta ttggtcaagt ttgcctt                                         87

<210> 908
<211> 89
<212> DNA
<213> Artificial Sequence

<220>
<223> PhTx ssODN1 3' homology arm

<220>
<221> misc_feature
<222> (89)..(89)

```

<223> Modified to contain phosphothioate
 <400> 908
 gtcaaggcaa ggctggccaa cccatgggtg gagtttagcc agggaccgtt tcagacagat 60
 atttgcatg agatagtgtg gggaagggg 89
 <210> 909
 <211> 176
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PhTx ssODN1
 <220>
 <221> misc_feature
 <222> (1)..(1)
 <223> Modified to contain phosphothioate
 <220>
 <221> misc_feature
 <222> (176)..(176)
 <223> Modified to contain phosphothioate
 <400> 909
 gggtgcttcc ttttattctt catccctagc cagccgccgg cccctggcct cactggatac 60
 tctaagacta ttgttcaagt ttgccttgtc aaggcaaggc tggccaaccc atgggtggag 120
 ttttagccagg gaccgtttca gacagatatt tgcattgaga tagtgtgggg aagggg 176
 <210> 910
 <211> 17
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> targeting domain
 <400> 910
 ggctattggt caaggca 17
 <210> 911
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> targeting domain
 <400> 911
 caaggctatt ggtcaaggca 20
 <210> 912
 <211> 17
 <212> DNA
 <213> Artificial Sequence

8009W000_SequenceListing.txt

<220>		
<223>	targeting domain	
<400>	912	
	tgccctgtca aggctat	17
<210>	913	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	targeting domain	
<400>	913	
	gtttgccttg tcaaggctat	20
<210>	914	
<211>	17	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	targeting domain	
<400>	914	
	gaccaatagc cttgaca	17
<210>	915	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	targeting domain	
<400>	915	
	cttgaccaat agccttgaca	20
<210>	916	
<211>	17	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	targeting domain	
<400>	916	
	gtcaaggcta ttggtca	17
<210>	917	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	targeting domain	

8009W000_SequenceListing.txt

<400> 917
 cttgtcaagg ctattggtca 20

<210> 918
 <211> 17
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> targeting domain

<400> 918
 tcaagtttgc cttgtca 17

<210> 919
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> targeting domain

<400> 919
 tggtaagtt tgccttgtca 20

<210> 920
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> targeting domain plus PAM (NGG)

<400> 920
 ggcuauuggu caaggcaagg 20

<210> 921
 <211> 23
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> targeting domain plus PAM (NGG)

<400> 921
 caaggcuauu ggucaggca agg 23

<210> 922
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> targeting domain plus PAM (NGG)

<400> 922
 ugccuuguca aggcuaugg 20

8009W000_SequenceListing.txt

<210> 923
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> targeting domain plus PAM (NGG)

 <400> 923
 guuugccuug ucaaggcuau ugg 23

 <210> 924
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> targeting domain plus PAM (NGG)

 <400> 924
 gaccaauagc cuugacaagg 20

 <210> 925
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> targeting domain plus PAM (NGG)

 <400> 925
 cuugaccaau agccuugaca agg 23

 <210> 926
 <211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> targeting domain plus PAM (NGG)

 <400> 926
 gucaaggcua uuggucaagg 20

 <210> 927
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> targeting domain plus PAM (NGG)

 <400> 927
 cuugucaagg cuauugguca agg 23

 <210> 928

8009W000_SequenceListing.txt

<211> 20
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> targeting domain plus PAM (NGG)

 <400> 928
 ucaaguuugc cuugucaagg 20

 <210> 929
 <211> 23
 <212> RNA
 <213> Artificial Sequence

 <220>
 <223> targeting domain plus PAM (NGG)

 <400> 929
 uggucaaguu ugccuuguca agg 23

 <210> 930
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> targeting domain plus PAM (NGG)

 <400> 930
 ggctatttgt caaggcaagg 20

 <210> 931
 <211> 23
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> targeting domain plus PAM (NGG)

 <400> 931
 caaggctatt ggtcaaggca agg 23

 <210> 932
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> targeting domain plus PAM (NGG)

 <400> 932
 tgccttgtca aggctattgg 20

 <210> 933
 <211> 23
 <212> DNA
 <213> Artificial Sequence

8009W000_SequenceListing.txt

```

<220>
<223>  targeting domain plus PAM (NGG)

<400>  933
gtttgccttg tcaaggctat tgg                                23

<210>  934
<211>  20
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  targeting domain plus PAM (NGG)

<400>  934
gaccaatagc cttgacaagg                                20

<210>  935
<211>  23
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  targeting domain plus PAM (NGG)

<400>  935
cttgaccaat agccttgaca agg                                23

<210>  936
<211>  20
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  targeting domain plus PAM (NGG)

<400>  936
gtcaaggcta ttggtcaagg                                20

<210>  937
<211>  23
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  targeting domain plus PAM (NGG)

<400>  937
cttgtcaagg ctattggtca agg                                23

<210>  938
<211>  20
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  targeting domain plus PAM (NGG)

```

8009W000_SequenceListing.txt

<400> 938
tcaagtttgc cttgtcaagg 20

<210> 939
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> targeting domain plus PAM (NGG)

<400> 939
tggtcaagtt tgccttgtca agg 23