Methods of modulating expression of a target nucleic acid in a cell are provided including introducing into the cell a first foreign nucleic acid encoding one or more RNAs complementary to DNA, wherein the DNA includes the target nucleic acid, introducing into the cell a second foreign nucleic acid encoding a nuclease-null Cas9 protein that binds to the DNA and is guided by the one or more RNAs, introducing into the cell a third foreign nucleic acid encoding a transcriptional regulator protein or domain, wherein the one or more RNAs, the nuclease-null Cas9 protein, and the transcriptional regulator protein or domain are expressed, wherein the one or more RNAs, the nuclease-null Cas9 protein and the transcriptional regulator protein or domain co-localize to the DNA and wherein the transcriptional regulator protein or domain regulates expression of the target nucleic acid.
FIG. 2A

Biased library of binding sites
(target nucleotide: 79%, rest: 21%)

24bp barcode
(A/C/G)

Shared Site | Target Site | Min Promoter | Barcode | dTomato

Step 1: Map barcode to corresponding target site in the library.
Step 2: Stimulate library by either a:
   1) control-TF that binds the shared site; or
   2) TALE-TF/gRNA+Cas9-TF (target-TF) that binds the target site.
Step 3: Perform RNAseq and determine expressed barcodes for each.
Step 4: Map back expressed barcodes to corresponding binding sites.
Step 5: Compute relative enrichment of target-TF vs. control-TF barcodes.
FIG. 2D

Cas9<sup>m4</sup> gRNA<sub>VP64</sub> + gRNA: two base mismatch

FIG. 2E

18mer TALE<sup>VP64</sup>

relative expression

# of mutations
FIG. 2F
18mer TALE$^{VP64}$: one base mismatch

match, mismatch=A, mismatch=C, mismatch=G, mismatch=T

relative expression

T C T G T C A C A T C C T G T C

FIG. 2G
18mer TALE$^{VP64}$: two base mismatch

match

mismatch position

mismatch position

3 5 7 9 11 13 15 17

1.125

.750

.375
<table>
<thead>
<tr>
<th>Name</th>
<th>Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas9m1</td>
<td>D10A</td>
</tr>
<tr>
<td>Cas9m2</td>
<td>D10A+H840A</td>
</tr>
<tr>
<td>Cas9m3</td>
<td>D10A+D839A+H840A</td>
</tr>
<tr>
<td>Cas9m4</td>
<td>D10A+D839A+H840A+N863A</td>
</tr>
<tr>
<td>Cas9</td>
<td>wild-type</td>
</tr>
</tbody>
</table>
FIG. 6A

FIG. 6B

Luciferase Activity (RLU)

Cas9_{nuclease-VP64} + Promoter_{OCT4}-Luciferase Reporter
FIG. 6C

Fold Change in Endogenous Gene Expression

<table>
<thead>
<tr>
<th>Sample</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>plain</td>
<td>0.00</td>
</tr>
<tr>
<td>negative</td>
<td>0.00</td>
</tr>
<tr>
<td>gRNA1-7</td>
<td>5.00</td>
</tr>
<tr>
<td>gRNA8-14</td>
<td>40.00</td>
</tr>
<tr>
<td>gRNA15-21</td>
<td>60.00</td>
</tr>
<tr>
<td>gRNA1-21</td>
<td>80.00</td>
</tr>
</tbody>
</table>

Cas9 nuclease-VP64
FIG. 8A

Construct library

+ control TALE or TALE site Cas9 site

Transfected &
induce

Sequence library

Library barcodes

Analyse sequences

Binding site transcript tag

<table>
<thead>
<tr>
<th>Binding site</th>
<th>transcript tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>Tag 1</td>
</tr>
<tr>
<td>Site 2</td>
<td>Tag 2</td>
</tr>
<tr>
<td>Site 3</td>
<td>Tag 3</td>
</tr>
</tbody>
</table>

Transcript tag count

<table>
<thead>
<tr>
<th>Transcript tag</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag 1</td>
<td>8</td>
</tr>
<tr>
<td>Tag 2</td>
<td>2</td>
</tr>
<tr>
<td>Tag 3</td>
<td>6</td>
</tr>
</tbody>
</table>

Binding site vs. transcript tag association tables

(one per construct library)

Tag counts per sample

Combine tables

<table>
<thead>
<tr>
<th>Binding site</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>94204</td>
</tr>
<tr>
<td>Site 2</td>
<td>6121</td>
</tr>
<tr>
<td>Site 3</td>
<td>3506</td>
</tr>
</tbody>
</table>

Tag counts per binding site for samples and + controls

<table>
<thead>
<tr>
<th>Binding site</th>
<th>Exp level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>5.8</td>
</tr>
<tr>
<td>Site 2</td>
<td>6.2</td>
</tr>
<tr>
<td>Site 3</td>
<td>6.4</td>
</tr>
</tbody>
</table>

Normalized expression level for sample
FIG. 9A

Cas9m4VP64+gRNA

relative expression

# of mutations

0 1 2 3 4 5 6 7 8

FIG. 9B

Cas9m4VP64+gRNA: one base mismatch

match, mismatch=A, mismatch=C, mismatch=G, mismatch=T

relative expression

G T C C C C T C C A C C A A G T G G G

G T C C C C T C C A C C A A G T G G G
FIG. 9C

Cas9m4VP64+gRNA: two base mismatch

match
FIG. 9D

...GTCCCCTCCACCCCCACAGTG CRR.... gRNA Target

repair donor

DNA break stimulates HR

donor + Cas9 + gRNA

GTCCCCTCCACCCCCACAGTG CAG
GTCCCCTCCACCCCCACAGTG CAA
GTCCCCTCCACCCCCACAGTG CGG

empty

%age rate of HR

0 0.5 1 1.5 2 2.5 3 3.5 4
FIG. 10C

10mer TALE<sup>VP64</sup>: one base mismatch

match, mismatch=A, mismatch=C, mismatch=G, mismatch=T

FIG. 10D

10mer TALE<sup>VP64</sup>: two base mismatch

match
FIG. 12A

-5.0kB

-1.0kB

SOX2

Fold Change in Endogenous Gene Expression

0 10 20 30 40 50 60 70 80 90

plain  negative  gRNA1-5  gRNA6-10  gRNA1-10

Cas9\textsubscript{N-VP64}
FIG. 13A

Cas9<sup>N</sup>-VP<sup>64</sup> + gRNA2

N.S. *** *** *** *** *** *** ***

relative expression

# of mutations

0 1 2 3 4 5 6 7 8

FIG. 13B

Cas9<sup>N</sup>-VP<sup>64</sup> + gRNA2: one base mismatch

match, mismatch=A, mismatch=C, mismatch=G, mismatch=T

relative expression

G T C C C T C A C C A C A G T G G G
FIG. 13E

**Cas9<sub>N-VP64</sub> + gRNA3: one base mismatch**

match, mismatch=A, mismatch=C, mismatch=G, mismatch=T

Relative expression

FIG. 13F

**Cas9<sub>N-VP64</sub> + gRNA3: two base mismatch**

Match positions

Mismatch position
FIG. 15A

gRNA target

...GTCCCTCCACCCACAGTG GGG....

target locus

repair donor

DNA break stimulates HR
FIG. 15B-1

AGCCACGGUGAAAAAGUUUC
UCGGUGC

matches
target sequence

GUCCCUCCACCCACAGUG

relative rate of targeting

0 1 2 3
FIG. 15B-2

GUCCCUCCACCCCCACAGUG
GUCCCUCCACCCCCACAGAG
GUCCCUCCACCCCCACACUG
GUCCCUCCACCCCCACUGUG
GUCCCUCCACCCCCAGAGUG
GUCCCUCCACCCCCUCAGUG
GUCCCUCCACCCCCUCAGUG
GUCCCUCCACCCGACAGUG
GUCCCUCCACCGGCACAGUG
GUCCCUCCACGCCCCACAGUG
GUCCCUCCACCGCCCCACAGUG
GUCCCUCCGACCCCCACAGTG

GUCCCUCCACCCCCACAGAC
GUCCCUCCACCCCCACUCUG
GUCCCUCCACCCCCUGAGUG
GUCCCUCCACCGGGACAGUG
GUCCCUCCAGGCCACAGUG
GUCCCUCCUGCCCCACAGUG
FIG. 16A

gRNA target

GGGGCCACTAGGGGACAGGAT GGG...

target locus

DNA break stimulates HR

repair donor
FIG. 16B-1

AGCCACGGUGGAAAAGAGUUC

UCGGUGCC

GAA

UAAUUU CAU

CAG

matches
target sequence

GGGCCCACUAGGACAGGAU
FIG. 16B-2

GGGCCCACUUGGACAGGAU
GGCCACUUGGACAGGAU
GGCCACUUGGACAGGAU
GCCACUUGGACAGGAU

1bp  2bp  3bp
FIG. 16C

gRNA target

GAGATGATCGCCCCCTTTTCGG...

DNA break stimulates HR

repair donor

target locus
FIG. 16D-1

matches
target sequence

5' - GAGAUGAUCGCCUCUCUUUC

AAGCCUAGGCCGUUUAUCAA
FIG. 17B

donor + Cas9 + gRNA

GTCCCTCCACCCACAGTG CAG
GTCCCTCCACCCACAGTG CAA
GTCCCTCCACCCACAGTG CGG
empty

% age rate of HR

0 0.5 1 1.5 2 2.5 3 3.5 4
FIG. 18B

![Graph showing percentage rate of HR for different conditions: donor, donor+GFP-ZFNs, donor+TALENs, with wild-type, 1 mutation, 4 mutations, and 6 mutations indicated.]
FIG. 19A

Random hexamer library of binding sites (TGTCCNNNNACCC)

24 bp barcode (A/C/G)

Target Site

Min Promoter

Shared Site

Barcode

δTomato
FIG. 19B-1

Target TALE-TF: NG NN NG HD NI NI NI NI NI NI NI HD HD HD

TGTCAAAAAAAAACCC

TGTCCCCCCCCCACCC

candidate match

mis-match

normalized expression

sequence mismatches
FIG. 21A
FIG. 21C

as3 + s4

CCCACAGTGGGCCACAC AGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
CCCACAGTGGGCCACAGACGGATTTGAGAC-AGAAAGCCCATCCCCACCTTTGAGGCTCTCTCTCTCTTAGTCTCTCTGTATATTGGGCTTACCC
RNA-GUIDED TRANSCRIPTIONAL REGULATION

RELATED APPLICATION DATA

This application is a continuation of PCT application no. PCT/US2014/040868, designating the United States and filed Jun. 4, 2014; which claims the benefit U.S. Provisional Patent Application No. 61/830,787 filed on Jun. 4, 2013; each of which are hereby incorporated by reference in their entireties.

STATEMENT OF GOVERNMENT INTERESTS

This invention was made with government support under Grant No. P50 HG005550 from the National Institutes of health and DE-FG02-02ER63445 from the Department of Energy. The government has certain rights in the invention.

BACKGROUND


SUMMARY

Aspects of the present disclosure are directed to a complex of a guide RNA, a DNA binding protein and a double stranded DNA target sequence. According to certain aspects, DNA binding proteins within the scope of the present disclosure include a protein that forms a complex with the guide RNA and with the guide RNA guiding the complex to a double stranded DNA sequence wherein the complex binds to the DNA sequence. This aspect of the present disclosure may be referred to as co-localization of the RNA and DNA binding protein to a double stranded DNA. In this manner, a DNA binding protein-guide RNA complex may be used to localize a transcriptional regulator protein or domain at target DNA so as to regulate expression of target DNA.

According to certain aspects, a method of modulating expression of a target nucleic acid in a cell is provided including introducing into the cell a first foreign nucleic acid encoding one or more RNAs (ribonucleic acids) complementary to DNA (deoxyribonucleic acid), wherein the DNA includes the target nucleic acid, introducing into the cell a second foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein that binds to the DNA and is guided by one or more RNAs, introducing into the cell a third foreign nucleic acid encoding a transcriptional regulator protein or domain, wherein the one or more RNAs, the RNA guided nuclease-null DNA binding protein, and the transcriptional regulator protein or domain are expressed, wherein the one or more RNAs, the RNA guided nuclease-null DNA binding protein and the transcriptional regulator protein or domain co-localize to the DNA and wherein the transcriptional regulator protein or domain regulates expression of the target nucleic acid.

According to one aspect, the foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein further encodes the transcriptional regulator protein or domain fused to the RNA guided nuclease-null DNA binding protein. According to one aspect, the foreign nucleic acid encoding one or more RNAs further encodes a site of an RNA-binding domain and the foreign nucleic acid encoding the transcriptional regulator protein or domain further encodes an RNA-binding domain. The RNA-guided domain fused to the transcriptional regulator protein or domain.

According to one aspect, the cell is a eukaryotic cell. According to one aspect, the cell is a yeast cell, a plant cell or an animal cell. According to one aspect, the cell is a mammalian cell.

According to one aspect, the RNA is between about 10 to about 500 nucleotides. According to one aspect, the RNA is between about 20 to about 100 nucleotides.

According to one aspect, the transcriptional regulator protein or domain is a transcriptional activator. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid to treat a disease or detrimental condition. According to one aspect, the target nucleic acid is associated with a disease or detrimental condition.

According to one aspect, the one or more RNAs is a guide RNA. According to one aspect, the one or more RNAs is a tracrRNA-crRNA fusion. According to one aspect, the guide RNA includes a spacer sequence and a tracrRNA sequence. The guide RNA may also include a tracrRNA sequence, a portion of which hybridizes to the tracrRNA sequence. The guide RNA may also include a linker nucleic acid sequence which links the tracrRNA sequence and the tracrRNA sequence to produce the tracrRNA-crRNA fusion. The spacer sequence binds to target DNA, such as by hybridization.

According to one aspect, the guide RNA includes a truncated spacer sequence. According to one aspect, the guide RNA includes a truncated spacer sequence having a base truncation at the 5' end of the spacer sequence. According to one aspect, the guide RNA includes a truncated spacer sequence having a base truncation at the 5' end of the spacer sequence. According to one aspect, the guide RNA includes a truncated spacer sequence having a base truncation at the 5' end of the spacer sequence. According to one aspect, the guide RNA includes a truncated spacer sequence having a base truncation at the 5' end of the spacer sequence. According to one aspect, the guide RNA includes a truncated spacer sequence having a base truncation at the 5' end of the spacer sequence. According to one aspect, the guide RNA includes a truncated spacer sequence having a 4
base truncation at the 5' end of the spacer sequence. Accordingly, the spacer sequence may have a 1 to 4 base truncation at the 5' end of the spacer sequence.

[0012] According to certain embodiments, the spacer sequence may include between about 16 to about 20 nucleotides which hybridize to the target nucleic acid sequence. According to certain embodiments, the spacer sequence may include about 20 nucleotides which hybridize to the target nucleic acid sequence.

[0013] According to certain aspects, the linker nucleic acid sequence may include between about 4 and about 6 nucleotides.

[0014] According to certain aspects, the spacer sequence may include between about 60 to about 500 nucleic acids. According to certain aspects, the spacer sequence may include between about 64 to about 500 nucleic acids. According to certain aspects, the spacer sequence may include between about 65 to about 500 nucleic acids. According to certain aspects, the spacer sequence may include between about 66 to about 500 nucleic acids. According to certain aspects, the spacer sequence may include between about 67 to about 500 nucleic acids. According to certain aspects, the spacer sequence may include between about 68 to about 500 nucleic acids. According to certain aspects, the spacer sequence may include between about 69 to about 500 nucleic acids. According to certain aspects, the spacer sequence may include between about 70 to about 500 nucleic acids. According to certain aspects, the spacer sequence may include between about 80 to about 500 nucleic acids. According to certain aspects, the spacer sequence may include between about 90 to about 500 nucleic acids. According to certain aspects, the spacer sequence may include between about 100 to about 500 nucleic acids.

[0015] According to certain aspects, the spacer sequence may include between about 60 to about 200 nucleic acids. According to certain aspects, the spacer sequence may include between about 64 to about 200 nucleic acids. According to certain aspects, the spacer sequence may include between about 65 to about 200 nucleic acids. According to certain aspects, the spacer sequence may include between about 66 to about 200 nucleic acids. According to certain aspects, the spacer sequence may include between about 67 to about 200 nucleic acids. According to certain aspects, the spacer sequence may include between about 68 to about 200 nucleic acids. According to certain aspects, the spacer sequence may include between about 69 to about 200 nucleic acids. According to certain aspects, the spacer sequence may include between about 70 to about 200 nucleic acids. According to certain aspects, the spacer sequence may include between about 80 to about 200 nucleic acids. According to certain aspects, the spacer sequence may include between about 90 to about 200 nucleic acids. According to certain aspects, the spacer sequence may include between about 100 to about 200 nucleic acids. According to certain aspects, the spacer sequence may include between about 100 to about 200 nucleic acids.

[0016] An exemplary guide RNA is depicted in FIG. 5B.

[0017] According to one aspect, the DNA is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

[0018] According to certain aspects, a method of modulating expression of a target nucleic acid in a cell is provided including introducing into the cell a first foreign nucleic acid encoding one or more RNAs (ribonucleic acids) complementary to DNA (deoxyribonucleic acid), wherein the DNA includes the target nucleic acid, introducing into the cell a second foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein of a Type II CRISPR System that binds to the DNA and is guided by the one or more RNAs, introducing into the cell a third foreign nucleic acid encoding a transcriptional regulator protein or domain, wherein the one or more RNAs, the RNA guided nuclease-null DNA binding protein of a Type II CRISPR System, and the transcriptional regulator protein or domain are expressed, wherein the one or more RNAs, the RNA guided nuclease-null DNA binding protein of a Type II CRISPR System and the transcriptional regulator protein or domain co-localize to the DNA and wherein the transcriptional regulator protein or domain regulates expression of the target nucleic acid.

[0019] According to one aspect, the foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein of a Type II CRISPR System further encodes the transcriptional regulator protein or domain fused to the RNA guided nuclease-null DNA binding protein of a Type II CRISPR System. According to one aspect, the foreign nucleic acid encoding one or more RNAs further encodes a target of an RNA-binding domain and the foreign nucleic acid encoding the transcriptional regulator protein or domain further encodes an RNA-binding domain fused to the transcriptional regulator protein or domain.

[0020] According to one aspect, the cell is a eukaryotic cell. According to one aspect, the cell is a yeast cell, a plant cell or an animal cell. According to one aspect, the cell is a mammalian cell.

[0021] According to one aspect, the RNA is between about 10 to about 500 nucleotides. According to one aspect, the RNA is between about 20 to about 100 nucleotides.

[0022] According to one aspect, the transcriptional regulator protein or domain is a transcriptional activator. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid to treat a disease or detrimental condition. According to one aspect, the target nucleic acid is associated with a disease or detrimental condition.

[0023] According to one aspect, the one or more RNAs is a guide RNA. According to one aspect, the one or more RNAs is a tracrRNA-crRNA fusion.

[0024] According to one aspect, the DNA is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

[0025] According to certain aspects, a method of modulating expression of a target nucleic acid in a cell is provided including introducing into the cell a first foreign nucleic acid encoding one or more RNAs (ribonucleic acids) complementary to DNA (deoxyribonucleic acid), wherein the DNA includes the target nucleic acid, introducing into the cell a second foreign nucleic acid encoding a nuclease-null Cas9 protein that binds to the DNA and is guided by the one or more RNAs, introducing into the cell a third foreign nucleic acid encoding a transcriptional regulator protein or domain, wherein the one or more RNAs, the nuclease-null Cas9 protein, and the transcriptional regulator protein or domain are expressed, wherein the one or more RNAs, the nuclease-null Cas9 protein and the transcriptional regulator protein or domain co-localize to the DNA and wherein the transcriptional regulator protein or domain regulates expression of the target nucleic acid.

[0026] According to one aspect, the foreign nucleic acid encoding a nuclease-null Cas9 protein further encodes the transcriptional regulator protein or domain fused to the nuclease-null Cas9 protein. According to one aspect, the foreign nucleic acid encoding one or more RNAs further
encodes a target of an RNA-binding domain and the foreign nucleic acid encoding the transcriptional regulator protein or domain further encodes an RNA-binding domain fused to the transcriptional regulator protein or domain.

[0027] According to one aspect, the cell is a eukaryotic cell. According to one aspect, the cell is a yeast cell, a plant cell or an animal cell. According to one aspect, the cell is a mammalian cell.

[0028] According to one aspect, the RNA is between about 10 to about 500 nucleotides. According to one aspect, the RNA is between about 20 to about 100 nucleotides.

[0029] According to one aspect, the transcriptional regulator protein or domain is a transcriptional activator. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid to treat a disease or detrimental condition. According to one aspect, the target nucleic acid is associated with a disease or detrimental condition.

[0030] According to one aspect, the one or more RNAs is a guide RNA. According to one aspect, the one or more RNAs is a tracrRNA–crRNA fusion.

[0031] According to one aspect, the DNA is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

[0032] According to one aspect a cell is provided that includes a first foreign nucleic acid encoding one or more RNAs complementary to DNA, wherein the DNA includes a target nucleic acid, a second foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein, and a third foreign nucleic acid encoding a transcriptional regulator protein or domain wherein the one or more RNAs, the RNA guided nuclease-null DNA binding protein and the transcriptional regulator protein or domain are members of a co-localization complex for the target nucleic acid.

[0033] According to one aspect, the foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein further encodes the transcriptional regulator protein or domain fused to an RNA guided nuclease-null DNA binding protein. According to one aspect, the foreign nucleic acid encoding one or more RNAs further encodes a target of an RNA-binding domain and the foreign nucleic acid encoding the transcriptional regulator protein or domain further encodes an RNA-binding domain fused to the transcriptional regulator protein or domain.

[0034] According to one aspect, the cell is a eukaryotic cell. According to one aspect, the cell is a yeast cell, a plant cell or an animal cell. According to one aspect, the cell is a mammalian cell.

[0035] According to one aspect, the RNA is between about 10 to about 500 nucleotides. According to one aspect, the RNA is between about 20 to about 100 nucleotides.

[0036] According to one aspect, the transcriptional regulator protein or domain is a transcriptional activator. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid to treat a disease or detrimental condition. According to one aspect, the target nucleic acid is associated with a disease or detrimental condition.

[0037] According to one aspect, the one or more RNAs is a guide RNA. According to one aspect, the one or more RNAs is a tracrRNA–crRNA fusion.

[0038] According to one aspect, the DNA is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

[0039] According to certain aspects, the RNA guided nuclease-null DNA binding protein is an RNA guided nuclease-null DNA binding protein of a Type II CRISPR System. According to certain aspects, the RNA guided nuclease-null DNA binding protein is a nuclease-null Cas9 protein.

[0040] According to one aspect, a method of altering a DNA target nucleic acid in a cell is provided that includes introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid, introducing into the cell a second foreign nucleic acid encoding at least one RNA guided DNA binding protein nickase and being guided by the two or more RNAs, wherein the two or more RNAs and the at least one RNA guided DNA binding protein nickase are expressed and wherein the at least one RNA guided DNA binding protein nickase co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks.

[0041] According to one aspect, a method of altering a DNA target nucleic acid in a cell is provided that includes introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid, introducing into the cell a second foreign nucleic acid encoding at least one RNA guided DNA binding protein nickase of a Type II CRISPR System and being guided by the two or more RNAs, wherein the two or more RNAs and the at least one RNA guided DNA binding protein nickase of a Type II CRISPR System are expressed and wherein the at least one RNA guided DNA binding protein nickase of a Type II CRISPR System co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks.

[0042] According to one aspect, a method of altering a DNA target nucleic acid in a cell is provided that includes introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid, introducing into the cell a second foreign nucleic acid encoding at least one Cas9 protein nickase having one inactive nuclease domain and being guided by the two or more RNAs, wherein the two or more RNAs and the at least one Cas9 protein nickase are expressed and wherein the at least one Cas9 protein nickase co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks.

[0043] According to the methods of altering a DNA target nucleic acid, the two or more adjacent nicks are on the same strand of the double stranded DNA. According to one aspect, the two or more adjacent nicks are on the same strand of the double stranded DNA and result in homologous recombination. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA and result in homologous recombination. According to one aspect, the two or more adjacent nicks on different strands of the double stranded DNA and create double stranded breaks. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks resulting in nonhomologous end joining. According to one aspect, the two or more adjacent nicks on different strands of the double stranded DNA and create double stranded breaks resulting in nonhomologous end joining.
double stranded DNA and are offset with respect to one another. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA and are offset with respect to one another and create double stranded breaks. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA and are offset with respect to one another and create double stranded breaks resulting in nonhomologous end joining. According to one aspect, the method further includes introducing into the cell a third foreign nucleic acid encoding a donor nucleic acid sequence wherein the two or more nicks results in homologous recombination of the target nucleic acid with the donor nucleic acid sequence.

[0044] According to one aspect, a method of altering a DNA target nucleic acid in a cell is provided including introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid, introducing into the cell a second foreign nucleic acid encoding at least one RNA guided DNA binding protein nickase and being guided by the two or more RNAs, and wherein the two or more RNAs and the at least one RNA guided DNA binding protein nickase are expressed and wherein the at least one RNA guided DNA binding protein nickase co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks, and wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks resulting in fragmentation of the target nucleic acid thereby preventing expression of the target nucleic acid.

[0045] According to one aspect, a method of altering a DNA target nucleic acid in a cell is provided including introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid, introducing into the cell a second foreign nucleic acid encoding at least one RNA guided DNA binding protein nickase of a Type II CRISPR system and being guided by the two or more RNAs, and wherein the two or more RNAs and the at least one RNA guided DNA binding protein nickase of a Type II CRISPR system are expressed and wherein the at least one RNA guided DNA binding protein nickase of a Type II CRISPR System co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks, and wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks resulting in fragmentation of the target nucleic acid thereby preventing expression of the target nucleic acid.

[0046] According to one aspect, a method of altering a DNA target nucleic acid in a cell is provided including introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid, introducing into the cell a second foreign nucleic acid encoding at least one Cas9 protein nickase having one inactive nuclease domain and being guided by the two or more RNAs, and wherein the two or more RNAs and the at least one Cas9 protein nickase are expressed and wherein the at least one Cas9 protein nickase co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks, and wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks resulting in fragmentation of the target nucleic acid thereby preventing expression of the target nucleic acid.

[0047] According to one aspect, a cell is provided including a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in a DNA target nucleic acid, and a second foreign nucleic acid encoding at least one RNA guided DNA binding protein nickase and wherein the two or more RNAs and the at least one RNA guided DNA binding protein nickase are members of a co-localization complex for the DNA target nucleic acid.

[0048] According to one aspect, the RNA guided DNA binding protein nickase is an RNA guided DNA binding protein nickase of a Type II CRISPR System. According to one aspect, the RNA guided DNA binding protein nickase is a Cas9 protein nickase having one inactive nuclease domain.

[0049] According to one aspect, the cell is a eukaryotic cell. According to one aspect, the cell is a yeast cell, a plant cell or an animal cell. According to one aspect, the cell is a mammalian cell.

[0050] According to one aspect, the RNA includes between about 10 to about 500 nucleotides. According to one aspect, the RNA includes between about 20 to about 100 nucleotides.

[0051] According to one aspect, the target nucleic acid is associated with a disease or detrimental condition.

[0052] According to one aspect, the two or more RNAs are guide RNAs. According to one aspect, the two or more RNAs are tRNA-crRNA fusions.

[0053] According to one aspect, the DNA target nucleic acid is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

[0054] Further features and advantages of certain embodiments of the present invention will become more fully apparent in the following description of embodiments and drawings thereof, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0055] The patent or application file contains drawings executed in color. Copies of this patent or patent application publication with the color drawings will be provided by the Office upon request and payment of the necessary fee. The foregoing and other features and advantages of the present embodiments will be more fully understood from the following detailed description of illustrative embodiments taken in conjunction with the accompanying drawings in which:

[0056] FIG. 1A and FIG. 1B are schematics of RNA-guided transcriptional activation. FIG. 1C is a design of a reporter construct (SEQ ID NOs 62 and 63). FIG. 1D shows data demonstrating that Cas9-VP64 fusions display RNA-guided transcriptional activation as assayed by both fluorescence-activated cell sorting (FACS) and immunofluorescence assays (IF). FIG. 1E shows assay data by FACS and IF demonstrating gRNA sequence-specific transcriptional activation from reporter constructs in the presence of Cas9, MS2-VP64 and gRNA bearing the appropriate MS2 aptamer binding sites. FIG. 1F depicts data demonstrating transcriptional induction by individual gRNAs and multiple gRNAs.

[0057] FIG. 2A depicts a methodology for evaluating the landscape of targeting by Cas9-gRNA complexes and TALEs. FIG. 2B depicts data demonstrating that a Cas9-gRNA complex is on average tolerant to 1-3 mutations in its target sequences. FIG. 2C depicts data demonstrating that the Cas9-gRNA complex is largely insensitive to point mutations, except those localized to the PAM sequence.
depicts heat plot data demonstrating that introduction of 2 base mismatches significantly impairs the Cas9-gRNA complex activity. FIG. 2E depicts data demonstrating that an 18-mer TALE reveals an average tolerant to 1-2 mutations in its target sequence. FIG. 2F depicts data demonstrating the 18-mer TALE is, similar to the Cas9-gRNA complexes, largely insensitive to single base mismatched in its target. FIG. 2G depicts heat plot data demonstrating that introduction of 2 base mismatches significantly impairs the 18-mer TALE activity.

[0058] FIG. 3A depicts a schematic of a guide RNA design. FIG. 3B depicts data showing percentage rate of non-homologous end joining for off-set nicks leading to 5' overhangs and off-set nicks leading to 3' overhangs. FIG. 3C depicts data showing percentage rate of targeting for off-set nicks leading to 5' overhangs and off-set nicks leading to 3' overhangs.

[0059] FIG. 4A is a schematic of a metal coordinating residue in RuVc PDB ID: 4EP4 (blue) position D7 (left), a schematic of HIN1 endonuclease domains from PDB IDs: 3M7K (orange) and 4H9D (cyan) including a coordinated Mg-ion (gray sphere) and DNA from 3M7K (purple) (middle) and a list of mutants analyzed (right). FIG. 4B depicts data showing undetectable nuclease activity for Cas9 mutants m3 and m4, and also their respective fusions with VP64. FIG. 4C is a higher-resolution examination of the data in FIG. 4B.

[0060] FIG. 5A is a schematic of a homologous recombination assay to determine Cas9-gRNA activity (SEQ ID NO:64). FIG. 5B depicts guide RNAs with random sequence insertions and percentage rate of homologous recombination (SEQ ID NO:65 and 66).

[0061] FIG. 6A is a schematic of guide RNAs for the OCT4 gene. FIG. 6B depicts transcriptional activation for a promoter-luciferase reporter construct. FIG. 6C depicts transcriptional activation via qPCR of endogenous genes.

[0062] FIG. 7A is a schematic of guide RNAs for the REX1 gene. FIG. 7B depicts transcriptional activation for a promoter-luciferase reporter construct. FIG. 7C depicts transcriptional activation via qPCR of endogenous genes.

[0063] FIG. 8A depicts in schematic a high level specificity analysis processing flow for calculation of normalized expression levels. FIG. 8B depicts data of distributions of percentages of binding sites by numbers of mismatches generated within a biased construct library. Left: Theoretical distribution. Right: Distribution observed from an actual TALE construct library. FIG. 8C depicts data of distributions of percentages of tag counts aggregated to binding sites by numbers of mismatches. Left: Distribution observed from the positive control sample. Right: Distribution observed from a sample in which a non-control TALE was induced.

[0064] FIG. 9A depicts data for analysis of the targeting landscape of a Cas9-gRNA complex showing tolerance to 1-3 mutations in its target sequence. FIG. 9B depicts data for analysis of the targeting landscape of a Cas9-gRNA complex showing insensitivity to point mutations, except those localized to the PAM sequence. FIG. 9C depicts heat plot data for analysis of the targeting landscape of a Cas9-gRNA complex showing that introduction of 2 base mismatches significantly impairs activity. FIG. 9D depicts data from a nuclease mediated HR assay confirming that the predicted PAM for the S. pyogenes Cas9 is NGG and also NAG (SEQ ID NOs:67-69).

[0065] FIG. 10A depicts data from a nuclease mediated HR assay confirming that 18-mer TALEs tolerate multiple mutations in their target sequences (SEQ ID NOs:70-73). FIG. 10B depicts data from analysis of the targeting landscape of TALEs of 3 different sizes (18-mer, 14-mer and 10-mer). FIG. 10C depicts data for 10-mer TALEs show near single-base mismatch resolution. FIG. 10D depicts heat plot data for 10-mer TALEs show near single-base mismatch resolution.

[0066] FIG. 11A depicts designed guide RNAs. FIG. 11B depicts percentage rate of non-homologous end joining for various guide RNAs (SEQ ID NOs:74-87).

[0067] FIG. 10A depicts data from a nuclease mediated HR assay confirming that 18-mer TALEs tolerate multiple mutations in their target sequences. FIG. 10B depicts data from analysis of the targeting landscape of TALEs of 3 different sizes (18-mer, 14-mer and 10-mer). FIG. 10C depicts data for 10-mer TALEs show near single-base mismatch resolution. FIG. 10D depicts heat plot data for 10-mer TALEs show near single-base mismatch resolution.

[0068] FIG. 11A depicts designed guide RNAs. FIG. 11B depicts percentage rate of non-homologous end joining for various guide RNAs.

[0069] FIG. 12A depicts the Sox2 gene. FIG. 12B depicts the Nanog gene.

[0070] FIGS. 13A-13F depict the targeting landscape of two additional Cas9-gRNA complexes.

[0071] FIG. 14A depicts the specificity profile of two gRNAs (wild-type (SEQ ID NO:88) and mutants (SEQ ID NO:89-90)). Sequence differences are highlighted in red. FIGS. 14B and 14C depict that this assay was specific for the gRNA being evaluated (data re-plotted from FIG. 13D).

[0072] FIGS. 15A-15D depict gRNA2 (FIGS. 15A-B) and gRNA3 (FIGS. 15C-D) bearing single or double-base mismatches (highlighted in red) in the spacer sequence versus the target. Sequences are set forth as SEQ ID NOs:91-131.

[0073] FIGS. 16A-16D depict a nuclease assay of two independent gRNA that were tested: gRNA1 (FIGS. 16A-B) and gRNA3 (FIGS. 16C-D) bearing truncations at the 5' end of their spacer. Sequences are set forth as SEQ ID NOs:66, 185-186 and 133-140.

[0074] FIGS. 17A-17B depict a nuclease mediated HR assay that shows the PAM for the S. pyogenes Cas9 is NGG and also NAG. Sequences are set forth as SEQ ID NOs:67-69 and 141.

[0075] FIGS. 18A-18B depict a nuclease mediated HR assay that confirmed that 18-mer TALEs tolerate multiple mutations in their target sequences. Sequences are set forth as SEQ ID NOs:70-73.

[0076] FIGS. 19A-19C depict a comparison of TALE monomer specificity versus TALE protein specificity. Sequences are set forth as SEQ ID NOs:142-150.

[0077] FIGS. 20A-20B depict data related to off-set nicking. Sequences are set forth as SEQ ID NOs:151-158.

[0078] FIGS. 21A-21C depict off-set nicking and NHEJ profiles. Sequences are set forth as SEQ ID NOs:159-184.

DETAILED DESCRIPTION

[0079] Embodiments of the present disclosure are based on the use of DNA binding proteins to co-localize transcriptional regulator proteins or domains to DNA in a manner to regulate a target nucleic acid. Such DNA binding proteins are readily known to those of skill in the art to bind to DNA for various purposes. Such DNA binding proteins may be naturally occurring. DNA binding proteins included within the scope of the present disclosure include those which may be guided by RNA, referred to herein as guide RNA. According to this aspect, the guide RNA and the RNA guided DNA binding
protein form a co-localization complex at the DNA. According to certain aspects, the DNA binding protein may be a nucleosome-null DNA binding protein. According to this aspect, the nucleosome-null DNA binding protein may result from the alteration or modification of a DNA binding protein having nucleosome activity. Such DNA binding proteins having nucleosome activity are known to those of skill in the art, and include naturally occurring DNA binding proteins having nucleosome activity, such as Cas9 proteins present, for example, in Type II CRISPR systems. Such Cas9 proteins and Type II CRISPR systems are well documented in the art. See Makarova et al., Nature Reviews, Microbiology, Vol. 9, June 2011, pp. 467-477 including all supplementary information hereby incorporated by reference in its entirety.  

[0080] Exemplary DNA binding proteins having nucleosome activity function to nick or cut double stranded DNA. Such nucleosome activity may result from the DNA binding protein having one or more polypeptide sequences exhibiting nucleosome activity. Such exemplary DNA binding proteins may have two separate nucleosome domains with each domain responsible for cutting or nicking a particular strand of the double stranded DNA. Exemplary polypeptide sequences having nucleosome activity known to those of skill in the art include the MerA-HIN1 nucleosome related domain and the RuvC-like nucleosome domain. Accordingly, exemplary DNA binding proteins are those that in nature contain one or more of the MerA-HIN1 nucleosome related domain and the RuvC-like nucleosome domain. According to certain aspects, the DNA binding protein is altered or otherwise modified to inactivate the nucleosome activity. Such alteration or modification includes removing one or more amino acids to inactivate the nucleosome activity or the nucleosome domain. Such modification includes removing the polypeptide sequence or polypeptide sequences exhibiting nucleosome activity, i.e. the nucleosome domain, from the DNA binding protein. Other modifications to inactivate nucleosome activity will be readily apparent to one of skill in the art based on the present disclosure. Accordingly, a nucleosome-null DNA binding protein includes polypeptide sequences modified to inactivate nucleosome activity or removal of a polypeptide sequence or sequences to inactivate nucleosome activity. A nucleosome-null DNA binding protein retains the ability to bind to DNA even though the nucleosome activity has been inactivated. Accordingly, the DNA binding protein includes the polypeptide sequence or sequences required for DNA binding but may lack the one or more or all of the nucleosome sequences exhibiting nucleosome activity. Accordingly, the DNA binding protein includes the polypeptide sequence or sequences required for DNA binding but may have one or more or all of the nucleosome sequences exhibiting nucleosome activity inactivated.

[0081] According to one aspect, a DNA binding protein having two or more nucleosome domains may be modified or altered to inactivate all but one of the nucleosome domains. Such a modified or altered DNA binding protein is referred to as a DNA binding protein nickase, to the extent that the DNA binding protein cuts or nicks only one strand of double stranded DNA. When guided by RNA to DNA, the DNA binding protein nickase is referred to as a RNA guided DNA binding protein nickase.  

[0082] An exemplary DNA binding protein is an RNA guided DNA binding protein of a Type II CRISPR System which lacks nucleosome activity. An exemplary DNA binding protein is a Cas9 protein nickase.  

[0083] In S. pyogenes, Cas9 generates a blunt-ended double-stranded break 3 bp upstream of the protospacer-adjacent motif (PAM) via a process mediated by two catalytic domains in the protein: an HNH domain that cleaves the complementary strand of the DNA and a RuvC-like domain that cleaves the non-complementary strand. See Jinake et al., Science 337, 816-821 (2012) hereby incorporated by reference in its entirety. Cas9 proteins are known to exist in many Type II CRISPR systems including the following as identified in the supplementary information to Makarova et al., Nature Reviews, Microbiology, Vol. 9, June 2011, pp. 467-477: Methanococcus maripaludis C7; Corynebacterium diphtheriae; Corynebacterium efficiens YS-314; Corynebacterium glutamicum ATCC 13032 Kitasato; Corynebacterium glutamicum ATCC 13032 Bielefeld; Corynebacterium glutamicum R; Corynebacterium kroppenstedtii DSM 44385; Mycobacterium ahensis DSM 19977; Nocardia farcinica IFM10152; Rhodococcus erythropolis PR4; Rhodococcus jostii RHA1; Rhodococcus opacus B4 uid36573; Acido- thermydus cellulosilyticus 113B; Arthrobacter chlorophenolica A6; Kribbella flaviga DSM 17836 uid43465; Thermomonospora curvata DSM 43183; Bifidobacterium dentium Bd1; Bifidobacterium longum DJO10A; Slackia heliotritiireducens DSM 20476; Persephonella marina EX I1; Bacteroides fragilis NCTC 9434; Capnocytophaga ochracea DSM 7271; Flavobacterium psychrophilum JIP02 86; Akkermansia muciniphila ATCC BAA 835; Roseifexus castenholzii DSM 13941; Roseifexus RS1; Synchoctis PCC6803; Elusimicrobium minutum Pei191; uncultured Ter- mite group 1 bacterium phytophyle Rs D17; Fibrobacter succi- cingenes SR8; Bacillus cereus ATCC 10847; Listeria innocua; Lactobacillus casei; Lactobacillus rhamnosus GG; Lactobacillus salivarius UCC118; Streptococcus galgae A909; Streptococcus agalactiae NEM316; Streptococcus galalactiae 2603; Streptococcus dysgalactiae equisimilis GGS 124; Streptococcus equi zoonpidemicus MGCS10565; Streptococcus galalactylus UCN34 uid46061; Streptococcus gordoni Challis subst CH1; Streptococcus mutans NN2025 uid46353; Streptococcus mutans; Streptococcus pyogenes M1 GAS; Streptococcus pyogenes MGAS5005; Streptococcus pyogenes MGAS2096; Streptococcus pyogenes MGAS9429; Streptococcus pyogenes MGAS10270; Streptococcus pyogenes MGAS6180; Streptococcus pyogenes MGAS315; Streptococcus pyogenes SS1-1; Streptococcus pyogenes MGAS10750; Streptococcus pyogenes NZ131; Streptococcus thermophilus CN1; Streptococcus thermophilus LMD-9; Streptococcus thermophilus LMG 18321; Clostridium botulinum A3 Loch Maree; Clostridium botuli- num B Ekland 17B; Clostridium botulinum Ba4 657; Clostridium botulinum F Langeland; Clostridium cellulolyti- hic H10; Finegoldia magna ATCC 29328; Enbacterium rectal ATEC 33656; Mycoplasma gallisepticum; Mycoplasma mobile 163K; Mycoplasma penetrans; Mycoplasma synoviae 53; Streptococcus moniliformis DSM 12112; Bradyrhizo- bium BTA1; Nitrobacter hamburgensis X14; Rhodopseudomonas palustris BiB 18; Rhodopseudomonas palustris B is B5; Parvibaculum lavamentovorans DS-1; Dinoroseobacter shibae DFL 12; Gluconacetobacter diazo- trophiicus Pal 5 FAPERJ; Gluconacetobacter diazotrophi-icus Pal 5 GJ1; Azospirillum B510 uid46085; Rhodospirillum rubrum ATCC 11170; Diaphorobacter TPSY uid29975; Ter- minephrobacter eisenii EF01-2; Neisseria meningitides
053442; Neisseria meningitidis alpha14; Neisseria meningitides Z2491; Desulfovibrio salteniens DSM 2638; Campylobacter jejuni doylei 269 97; Campylobacter jejuni 81116; Campylobacter jejuni; Campylobacter lari RM2100; Helicobacter hepaticus; Wolinella succinogenes; Tolumonas auensis DSM 9187; Pseudotetraodon atlantica Tc; Shewanella pealeana ATCC 700345; Legionella pneumophila Paris; Actinobacillus succinogenes 130Z; Pasteurella multocida; Francisella tularensis novicida U112; Francisella tularensis holarctica; Francisella tularensis TSC 198; Francisella tularensis tularensis; Francisella tularensis WY96-3418; and Treponema denticola ATCC 35405. Accordingly, aspects of the present disclosure are directed to a Cas9 protein present in a Type II CRISPR system, which has been rendered nuclease null or which has been rendered a nickase as described herein.

[0084] The Cas9 protein may be referred to by one of skill in the art in the literature as Cas1. The S. pyogenes Cas9 protein sequence that is the subject of experiments described herein is shown below. See Delcheva et al., Nature 471, 602-607 (2011) hereby incorporated by reference in its entirety.

(QS ID No.: 1)

MDKXGSLGQDGTSHVNAVITDEYKYPSONKLVGQHTSHSSTKRLNIGA
LLPGQSTEAATLRLRLEVELRMTERRIRKIRGRLQIEPBHFSMVEK
LEESFLVKEEKHRHP1FGQIVQDAIYKQYTPYHRLFEKLVSTDKD
LRVSYALAMLIEPQGHPLRLGSMPNSDGVDLQFLQGQYQFQYELQFEDIP
INASQVDAKALSLSXESSRLKELNIAQLPGEKXNLPGNHLALISLQIGT
NQNSRNPFLADAEALQRSLGQKTDDDLQMLNLQGQAYPQDLAAPALNLDAI
LSSDL1RVNTET1YVAPLSAMKRYDEEQGQLTLLKLARVQQPEKXEEI
FFQSOQEGHXYQIGGSLAQSGQEQRYMKIPLMQGTEELVXLRNEDLRLLR
KORSTFDQNS1YPQ1LGHSA1QLRLAQDQFQPPKLDHREIKQILKTFRP
YVQGQLRSNSRFVTREKTSSETITIPVPFEEOVQSKAGSFIETMTHFDK
NLPNEKULPNHLIELYTVNHELTKVYTEGMKAPLPLGQKKA1VD
LLFQKNRSQVQGKEDYFFKIKCPSDVGSEIGVREGPRASLQGKYDLYI
I1VXEKQOEEHEDILEQIVLTLQFEDERKIEELKTVINHFLDDVHQK
LTSSKRYTEKPNLQIKKRGJSQGETLLPYLDKSIDQXHNFHRMPQMLHID
SLTFKEQ1KQVQSUQEDSLHEHIANAGPAIKNGILQGVTRVQELVKV
MGGSHKPVNFLNMMRQENGQETQKQGKXSERMKERRIGKIQSGLQCELHLP
VETQOLCNPKLTVLYQNGMIDVQGEI1NLSDYDYDHDIVQPQKLDK
SIDEHVLRSDESDKQ1QVQAEVSKKMRQYQMLNAPIQRPFQNLH
TKAERGLGSDKFAG1KQVTQGITEQAOGDTSRIONTYENDLKL
REVKVLTLXVLSDPFRQDIYKREWVQIYHNNHKNHDANLYAVQTALI
IK
YPKLESEPYGQYTKVYDREAIKSEQEQMKATAYKQFYSNNHSFFTEI
TLANGKHDREPLDQTEGQIVNKQGQPAVTRIKLKVQLVQNIQVKEITV
QTQGQPSKESITPKSNDKLIARKEWDPQKYPGQDTSVYAVYVAKVE
KGGKHLKSVKELQGTTIERS4FENFIPDFLANKYQKVEKL1LIPLK
YSLFELENGKREKLSAGELQKXNLALSKYMNPLYLSXKYLXGSPE

[0085] According to certain aspects of methods of RNA-guided genome regulation described herein, Cas9 is altered to reduce, substantially reduce or eliminate nuclease activity. According to one aspect, Cas9 nuclease activity is reduced, substantially reduced or eliminated by altering the RuvC nuclease domain or the HNH nuclease domain. According to one aspect, the RuvC nuclease domain is inactivated. According to one aspect, the HNH1 nuclease domain is inactivated. According to one aspect, the RuvC nuclease domain and the HNH1 nuclease domain are inactivated. According to one additional aspect, Cas9 proteins are provided where the RuvC nuclease domain and the HNH nucleic acid domain are inactivated. According to an additional aspect, nuclease-null Cas9 proteins are provided insofar as the RuvC nuclease domain and the HNH nucleic acid domain are inactivated. According to an additional aspect, a Cas9 nickase is provided where either the RuvC nuclease domain or the HNH nucleic acid domain is inactivated, thereby leaving the remaining nucleic acid domain active for nuclease activity. In this manner, only one strand of the double stranded DNA is cut or nicked.

[0086] According to an additional aspect, nucleic acid Cas9 proteins are provided where one or more amino acids in Cas9 are altered or otherwise removed to provide nuclease-null Cas9 proteins. According to one aspect, the amino acids include D10 and H840. See Jinke et al., Science 357, 816-821 (2012). According to an additional aspect, the amino acids include D839 and N863. According to one aspect, one or more of all of D10, H840, D839 and H863 are substituted with an amino acid which reduces, substantially eliminates or eliminates nuclease activity. According to one aspect, one or more of all of D10, H840, D839 and H863 are substituted with alanine. According to one aspect, a Cas9 protein having one or more or all of D10, H840, D839 and H863 substituted with an amino acid which reduces, substantially eliminates or eliminates nuclease activity, such as alanine, is referred to as a nuclease-null Cas9 or Cas9N and exhibits reduced or eliminated nuclease activity, or nuclease activity is absent or substantially absent within levels of detection. Accordingly to this aspect, nuclease activity for a Cas9N may be undetectable using known assays, i.e. below the level of detection of known assays.

[0087] According to one aspect, the nuclease null Cas9 protein includes homologs and orthologs thereof which retain the ability of the protein to bind to the DNA and is altered. According to one aspect, the nuclease null Cas9 protein includes the sequence as set forth for naturally occurring Cas9 from S. pyogenes and having one or more, or all of D10, H840, D839 and H863 substituted with alanine and protein sequences having at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98% or 99% homology thereto and being a DNA binding protein, such as a RNA guided DNA binding protein.

[0088] According to one aspect, the nuclease null Cas9 protein includes the sequence as set forth for naturally occurring Cas9 from S. pyogenes excepting the protein sequence of the Ruvc nuclease domain and the HNH nucleic acid domain and also protein sequences having at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98% or 99% homology thereto.
and being a DNA binding protein, such as an RNA guided DNA binding protein. In this manner, aspects of the present disclosure include the protein sequence responsible for DNA binding, for example, for co-localizing with guide RNA and binding to DNA and protein sequences homologous thereto, and need not include the protein sequences for the RuvC nuclease domain and the HNH nuclease domain (to the extent not needed for DNA binding), as these domains may be either inactivated or removed from the protein sequence of the naturally occurring Cas9 protein to produce a nuclease null Cas9 protein.

[0099] For purposes of the present disclosure, FIG. 4A depicts metal coordinating residues in known protein structures with homology to Cas9. Residues are labeled based on position in Cas9 sequence. Left: RuvC structure, PDB ID: 4EP4 (blue) position D7, which corresponds to D10 in the Cas9 sequence, is highlighted in a Mg-ion coordinating position. Middle: Structures of HNH endonuclease domains from PDB IDs: 3M7K (orange) and 4H9D (cyan) including a coordinated Mg-ion (gray sphere) and DNA from 3M7K (purple). Residues D92 and N113 in 3M7K and 4H9D positions D53 and N77, which have sequence homology to Cas9 amino acids D839 and N863, are shown as sticks. Right: List of mutants made and analyzed for nuclease activity: Cas9 wild-type; Cas9 Y51 which substitutes alanine for D10; Cas9 N52 which substitutes alanine for D10 and alanine for H840; Cas9 N52 which substitutes alanine for D10, alanine for H840, and alanine for D839; and Cas9 N52 which substitutes alanine for D10, alanine for H840, alanine for D839, and alanine for N863.

[0100] As shown in FIG. 4B, the Cas9 mutants: m3 and m4, and also their respective fusions with VP64 showed no detectable nuclease activity upon deep sequencing at targeted loci. The plots show the mutational frequency versus genomic position, with the red lines demarcating the gRNA target. FIG. 4C is a higher-resolution examination of the data in FIG. 4B and confirms that the mutation landscape shows comparable profile as unmodified loci.

[0101] According to one aspect, an engineered Cas9-gRNA system is provided which enables RNA-guided genome regulation in human cells by tethering transcriptional activation domains to either a nuclease-null Cas9 or to guide RNAs. According to one aspect of the present disclosure, one or more transcriptional regulatory proteins or domains (such terms are used interchangeably) are joined or otherwise connected to a nuclease-deficient Cas9 or one or more guide RNA (gRNA). The transcriptional regulatory domains correspond to targeted loci. Accordingly, aspects of the present disclosure include methods and materials for localizing transcriptional regulatory domains to targeted loci by fusing, connecting or joining such domains to either Cas9 or to the gRNA.

[0102] According to one aspect, a Cas9N-fusion protein capable of transcriptional activation is provided. According to one aspect, a VP64 activation domain (see Zhang et al., Nature Biotechnology 29, 149-153 (2011) hereby incorporated by reference in its entirety) is joined, fused, connected or otherwise tethered to the C terminus of Cas9N. According to one method, the transcriptional regulatory domain is provided to the site of target genomic DNA by the Cas9N protein. According to one method, a Cas9N fused to a transcriptional regulatory domain is provided within a cell along with one or more guide RNAs. The Cas9N with the transcriptional regulatory domain fused thereto bind at or near target genomic DNA. The one or more guide RNAs bind at or near target genomic DNA. The transcriptional regulatory domain regulates expression of the target gene. According to a specific aspect, a Cas9N-VP64 fusion activated transcription of reporter constructs when combined with gRNAs targeting sequences near the promoter, thereby displaying RNA-guided transcriptional activation.

[0103] According to one aspect, a gRNA-fusion protein capable of transcriptional activation is provided. According to one aspect, a VP64 activation domain is joined, fused, connected or otherwise tethered to the gRNA. According to one method, the transcriptional regulatory domain is provided to the site of target genomic DNA by the gRNA. According to one method, a gRNA fused to a transcriptional regulatory domain is provided within a cell along with a Cas9N protein. The Cas9N binds at or near target genomic DNA. The one or more guide RNAs with the transcriptional regulatory protein or domain fused thereto bind at or near target genomic DNA. The transcriptional regulatory domain regulates expression of the target gene. According to a specific aspect, a Cas9N protein and a gRNA fused with a transcriptional regulatory domain activated transcription of reporter constructs, thereby displaying RNA-guided transcriptional activation.

[0104] The gRNA tethers capable of transcriptional regulation were constructed by identifying which regions of the gRNA will tolerate modifications by inserting random sequences into the gRNA and assaying for Cas9 function. gRNAs bearing random sequence insertions at either the 5’ end of the crRNA portion or the 3’ end of the tracrRNA portion of a chimeric gRNA retain functionality, while insertions into the tracrRNA scaffold portion of the chimeric gRNA result in loss of function. See FIGS. 5A-B summarizing gRNA flexibility to random base insertions. FIG. 5A is a schematic of a homologous recombination (HR) assay to determine Cas9-gRNA activity. As shown in FIG. 5B, gRNAs bearing random sequence insertions at either the 5’ end of the crRNA portion or the 3’ end of the tracrRNA portion of a chimeric gRNA retain functionality, while insertions into the tracrRNA scaffold portion of the chimeric gRNA result in loss of function. The points of insertion in the gRNA sequence are indicated by red nucleotides. Without wishing to be bound by scientific theory, the increased activity upon random base insertions at the 5’ end may be due to increased half-life of the longer gRNA.

[0105] To attach VP64 to the gRNA, two copies of the MS2 bacteriophage coat-protein binding RNA stem-loop were appended to the 3’ end of the gRNA. See Fusco et al., Current Biology: CB13, 161-167 (2003) hereby incorporated by reference in its entirety. These chimeric gRNAs were expressed together with Cas9N and MS2-VP64 fusion protein. Sequence-specific transcriptional activation from reporter constructs was observed in the presence of all 3 components.

[0106] FIG. 1A is a schematic of RNA-guided transcriptional activation. As shown in FIG. 1A, to generate a Cas9N-fusion protein capable of transcriptional activation, the VP64 activation domain was directly tethered to the C terminus of Cas9N. As shown in FIG. 1B, to generate gRNA tethers capable of transcriptional activation, two copies of the MS2 bacteriophage coat-protein binding RNA stem-loop were appended to the 3’ end of the gRNA. These chimeric gRNAs were expressed together with Cas9N and MS2-VP64 fusion protein. FIG. 1C shows design of reporter constructs used to assay transcriptional activation. The two reporters bear dis-
distinct gRNA target sites, and share a control TALE-TF target site. As shown in FIG. 1D, Cas9N-VP64 fusions display RNA-guided transcriptional activation as assayed by both fluorescence-activated cell sorting (FACS) and immunofluorescence assays (IF). Specifically, while the control TALE-TF activated both reporters, the Cas9N-VP64 fusion activates reporters in a gRNA sequence specific manner. As shown in FIG. 1E, gRNA sequence-specific transcriptional activation from reporter constructs only in the presence of all 3 components: Cas9N, MS2-VP64 and gRNA bearing the appropriate MS2 aptamer binding sites was observed by both FACS and IF.

[0097] According to certain aspects, methods are provided for regulating endogenous genes using Cas9N, one or more gRNAs and a transcriptional regulatory protein or domain. According to one aspect, an endogenous gene can be any desired gene, referred to herein as a target gene. According to one exemplary aspect, genes target for regulation included ZFP42 (REX1) and POUSF1 (OCT4), which are both tightly regulated genes involved in maintenance of pluripotency. As shown in FIG. 1F, 10 gRNAs targeting a ~5 kb stretch of DNA upstream of the transcription start site (DNase hypersensitive sites are highlighted in green) were designed for the REX1 gene. Transcriptional activation was assayed using either a promoter-luciferase reporter construct (see Takashashi et al., Cell 131:861-872 (2007) hereby incorporated by reference in its entirety) or directly via qPCR of the endogenous genes. As shown in FIG. 6A, Cas9N-VP64 fusions as shown in FIG. 6A. 21 gRNAs targeting a ~5 kb stretch of DNA upstream of the transcription start site were designed for the OCT4 gene. The DNase hypersensitive sites are highlighted in green. FIG. 6B shows transcriptional activation using a promoter-luciferase reporter construct. FIG. 6C shows transcriptional activation directly via qPCR of the endogenous genes. While introduction of individual gRNAs modestly stimulated transcription, multiple gRNAs acted synergistically to stimulate robust multi-fold transcriptional activation.

[0099] FIGS. 7A-C is directed to RNA-guided REX1 regulation using Cas9N, MS2-VP64 and gRNA×2×-MS2 aptamers. As shown in FIG. 7A, 10 gRNAs targeting a ~5 kb stretch of DNA upstream of the transcription start site were designed for the REX1 gene. The DNase hypersensitive sites are highlighted in green. FIG. 7B shows transcriptional activation using a promoter-luciferase reporter construct. FIG. 7C shows transcriptional activation directly via qPCR of the endogenous genes. While introduction of individual gRNAs modestly stimulated transcription, multiple gRNAs acted synergistically to stimulate robust multi-fold transcriptional activation. In one aspect, the absence of the 2×-MS2 aptamers on the gRNA does not result in transcriptional activation. See Maeder et al., Nature Methods 10, 243-245 (2013) and Perez-Pinera et al., Nature Methods 10, 239-242 (2013) each of which are hereby incorporated by reference in its entirety.

[0100] Accordingly, methods are directed to the use of multiple guide RNAs with a Cas9N protein and a transcriptional regulatory protein or domain to regulate expression of a target gene.

[0101] Both the Cas9 and gRNA tethering approaches were effective, with the former displaying ~1.5-2 fold higher potency. This difference is likely due to the requirement for 2-component as opposed to 3-component complex assembly. However, the gRNA tethering approach in principle enables different effector domains to be recruited by distinct gRNAs so long as each gRNA uses a different RNA-protein interaction pair. See Karyer-Bibens et al., Biology of the Cell/Under the Auspices of the European Cell Biology Organization 100, 125-138 (2008) hereby incorporated by reference in its entirety. According to one aspect of the present disclosure, different target genes may be regulated using specific guide RNA and a generic Cas9N protein, i.e. the same or a similar Cas9N protein for different target genes. According to one aspect, methods of multiplex gene regulation are provided using the same or similar Cas9N.

[0102] Methods of the present disclosure are also directed to editing target genes using the Cas9N proteins and guide RNAs described herein to provide multiplex genetic and epigenetic engineering of human cells. With Cas9-gRNA targeting being an issue (see Jiang et al., Nature Biotechnology 31, 233-239 (2013) hereby incorporated by reference in its entirety), methods are provided for in-depth interrogation of Cas9 affinity for a very large space of target sequence variations. Accordingly, aspects of the present disclosure provide direct high-throughput readout of Cas9 targeting in human cells, while avoiding complications introduced by dsDNA cut toxicity and mutagenic repair incurred by specificity testing with native nuclease-active Cas9.

[0103] Further aspects of the present disclosure are directed to the use of DNA binding proteins or systems in general for the transcriptional regulation of a target gene. One of skill in the art will readily identify exemplary DNA binding systems based on the present disclosure. Such DNA binding systems need not have any nucleosome activity, as with the naturally occurring Cas9 protein. Accordingly, such DNA binding systems need not have nucleosome activity inactivated. One exemplary DNA binding system is TALE. As a genome editing tool, usually TALE-Fok1 dimers are used, and for genome regulation TALE-VNP64 fusions have been shown to be highly effective. According to one aspect, TALE specificity was evaluated using the methodology shown in FIG. 2A. A construct library in which each element of the library comprises a minimal promoter driving a dTomato fluorescent protein is designed. Downstream of the transcription start site m, a 24 bp (A/C/G) random transcript tag is inserted, while two TF binding sites are placed upstream of the promoter: one is a constant DNA sequence shared by all library elements, and the second is a variable feature that bears a “biased” library of binding sites which are engineered to span a large collection of sequences that present many combinations of mutations away from the target sequence the programmable DNA targeting complex was designed to bind. This is achieved using degenerate oligonucleotides engineered to bear nucleotide frequencies at each position such that the target sequence nucleotide appears at a 79% frequency and each other nucleotide occurs at 7% frequency. See Patwardhan et al., Nature Biotechnology 30, 265-270 (2012) hereby incorporated by reference in its entirety. The reporter library is then sequenced to reveal the associations between the 24 bp dTomato transcript tags and their corresponding “biased” target site in the library element. The large diversity of the transcript tags assures that sharing of tags between different targets will be extremely rare, while the biased construction of the target sequences means that sites with few mutations will be associated with more tags than sites with more mutations. Next, transcription of the dTomato reporter genes is stimulated with either a control-TF engineered to bind the shared DNA site, or the target-TF that was engineered to bind the target site. The abundance of each expressed transcript tag is measured in...
each sample by conducting RNAseq on the stimulated cells, which is then mapped back to their corresponding binding sites using the association table established earlier. The control-TF is expected to excite all library members equally since its binding site is shared across all library elements, while the target-TF is expected to skew the distribution of the expressed members to those that are preferentially targeted by it. This assumption is used in step 5 to compute a normalized expression level for each binding site by dividing the tag counts obtained for the target-TF by those obtained for the control-TF.

[0104] As shown in FIG. 2B, the targeting landscape of a Cas9-gRNA complex reveals that it is on average tolerant to 1-3 mutations in its target sequences. As shown in FIG. 2C, the Cas9-gRNA complex is also largely insensitive to point mutations, except those localized to the PAM sequence. Notably this data reveals that the predicted PAM for the S. pyogenes Cas9 is not just NGG but also NAG. As shown in FIG. 2D, introduction of 2 base mismatches significantly impairs the Cas9-gRNA complex activity, however only when these are localized to the 8-10 bases nearer the 3’ end of the gRNA target sequence (in the heat plot the target sequence positions are labeled from 1-23 starting from the 5’ end).

[0105] The mutational tolerance of another widely used genome editing tool, TALE domains, was determined using the transcriptional specificity assay described herein. As shown in FIG. 2E, the TALE off-targeting data for an 18-mer TALE reveals that it can tolerate on average 1-2 mutations in its target sequence, and fails to activate a large majority of 3 base mismatch variants in its targets. As shown in FIG. 2F, the 18-mer TALE is, similar to the Cas9-gRNA complexes, largely insensitive to single base mismatched in its target. As shown in FIG. 2G, introduction of 2 base mismatches significantly impairs the 18-mer TALE activity. TALE activity is more sensitive to mismatches nearer the 5’ end of its target sequence (in the heat plot the target sequence positions are labeled from 1-18 starting from the 5’ end).

[0106] Results were confirmed using targeted experiments in a nuclease assay which is the subject of FIGS. 10A-C directed to evaluating the landscape of targeting by TALEs of different sizes. As shown in FIG. 10A, using a nuclease mediated HR assay, it was confirmed that 18-mer TALEs tolerate multiple mutations in their target sequences. As shown in FIG. 10B, using the approach described in FIG. 2, the targeting landscape of TALEs of 3 different sizes (18-mer, 14-mer and 10-mer) was analyzed. Shorter TALEs (14-mer and 10-mer) are progressively more specific in their targeting but also reduced in activity by nearly an order of magnitude. As shown in FIGS. 10C and 10D, 10-mer TALEs show near single base mismatch resolution, losing almost all activity against targets bearing 2 mismatches (in the heat plot the target sequence positions are labeled from 1-10 starting from the 5’ end). Taken together, these data imply that engineering shorter TALEs can yield higher specificity in genome engineering applications, while the requirement for FokI dimerization in TALE nuclease applications is essential to avoid off-target effect. See Kim et al., *Proceedings of the National Academy of Sciences of the United States of America* 93, 1156-1160 (1996) and Pattanayak et al., *Nature Methods* 8, 765-770 (2011) each of which are hereby incorporated by reference in its entirety.

[0107] FIGS. 8A-C is directed to high level specificity analysis processing for calculation of normalized expression levels illustrated with examples from experimental data. As shown in FIG. 8A, construct libraries are generated with a biased distribution of binding site sequences and random sequence 24 bp tags that will be incorporated into reporter gene transcripts (top). The transcribed tags are highly degenerate so that they should map many-to-one to Cas9 or TALE binding sequences. The construct libraries are sequenced (3rd level, left) to establish which tags co-occur with binding sites, resulting in an association table of binding sites vs. transcribed tags (4th level, left). Multiple construct libraries built for different binding sites may be sequenced at once using library barcodes (indicated here by the light blue and light yellow colors; levels 1-4, left). A construct library is then transferred into a cell population and a set of different Cas9/gRNA or TALE transcription factors are induced in samples of the populations (2nd level, right). One sample is always induced with a fixed TALE activator targeted to a fixed binding site sequence within the construct (top level, green box); this sample serves as a positive control (green sample, also indicated by a + sign). cDNAs generated from the reporter mRNA molecules in the induced samples are then sequenced and analyzed to obtain tag counts for each tag in each sample (3rd and 4th level, right). As with the construct library sequencing, multiple samples, including the positive control, are sequenced and analyzed together by appending sample barcodes. Here the light red color indicates one non-control sample that has been sequenced and analyzed with the positive control (green). Because only the transcribed tags and not the construct binding sites appear in each read, the binding site vs. tag association table obtained from construct library sequencing is then used to tally up total counts of tags expressed from each binding site in each sample (5th level). The tallies for each non-positive control sample are then converted to normalized expression levels for each binding site by dividing them by the tallies obtained in the positive control sample. Examples of plots of normalized expression levels by numbers of mismatches are provided in FIGS. 2B and 2E, and in FIG. 9A and FIG. 10B. Not covered in this overall process flow are several levels of filtering for erroneous tags, for tags not associated with a construct library, and for tags apparently shared with multiple binding sites. FIG. 8B depicts example distributions of percentages of binding sites by numbers of mismatches generated within a biased construct library. Left: Theoretical distribution. Right: Distribution observed from an actual TALE construct library. FIG. 8C depicts example distributions of percentages of tag counts aggregated to binding sites by numbers of mismatches. Left: Distribution observed from the positive control sample. Right: Distribution observed from a sample in which a non-control TALE was induced. As the positive control TALE binds to a fixed site in the construct, the distribution of aggregated tag counts closely reflects the distribution of binding sites in FIG. 8B, while the distribution is skewed to the left for the non-control TALE sample because sites with fewer mismatches induce higher expression levels. Below: Computing the relative enrichment between these by dividing the tag counts obtained for the target-TF by those obtained for the control-TF reveals the average expression level versus the number of mutations in the target site.

[0108] These results are further reaffirmed by specificity data generated using a different Cas9-gRNA complex. As shown in FIG. 9A, a different Cas9-gRNA complex is tolerant to 1-3 mutations in its target sequence. As shown in FIG. 9B, the Cas9-gRNA complex is also largely insensitive to point mutations, except those localized to the PAM sequence. As
shown in FIG. 9C, introduction of 2 base mismatches however significantly impairs activity (in the heat plot the target sequence positions are labeled from 1-23 starting from the 5' end). As shown in FIG. 9D, it was confirmed using a nuclelease mediated HR assay that the predicted PAM for the S. pyogenes Cas9 is NGG and also NAG.

[0109] According to certain aspects, binding specificity is increased according to methods described herein. Because synergy between multiple complexes is a factor in target gene activation by Cas9-VP64, transcriptional regulation applications of Cas9N is naturally quite specific as individual off-target binding events should have minimal effect. According to one aspect, off-set nicks are used in methods of genome editing. A large majority of nicks seldom result in NHEJ events, (see Certo et al., Nature Methods 8, 671-676 (2011) hereby incorporated by reference in its entirety) thus minimizing the effects of off-target nicking. In contrast, inducing off-set nicks to generate double stranded breaks (DSBs) is highly effective at inducing gene disruption. According to certain aspects, 5' overhangs generate more significant NHEJ events as opposed to 3' overhangs. Similarly, 3' overhangs favor HR over NHEJ events, although the total number of HR events is significantly lower than when a 5' overhang is generated. Accordingly, methods are provided for using nicks for homologous recombination and off-set nicks for generating double stranded breaks to minimize the effects of off-target Cas9-gRNA activity.

[0110] FIGS. 3A-C is directed to multiplex off-set nicking and methods for reducing the off-target binding with the guide RNAs. As shown in FIG. 3A, the traffic light reporter was used to simultaneously assay for HR and NHEJ events upon introduction of targeted nicks or breaks. DNA cleavage events resolved through the HDR pathway restore the GFP sequence, whereas mutagenic NHEJ causes frameshift rendering the GFP out of frame and the downstream mCherry sequence in frame. For the assay, 14 gRNAs covering a 200 bp stretch of DNA; 7 targeting the sense strand (U1-7) and 7 the antisense strand (D1-7) were designed. Using the Cas9D10A mutant, which lacks the complementary strand, different two-way combinations of the gRNAs were used to induce a range of programmed 5' or 3' overhangs (the nicking sites for the 14 gRNAs are indicated). As shown in FIG. 3B, inducing off-set nicks to generate double stranded breaks (DSBs) is highly effective at inducing gene disruption. Notably off-set nicks leading to 5' overhangs result in more NHEJ events as opposed to 3' overhangs. As shown in FIG. 3C, generating 3' overhangs also favors the ratio of HR over NHEJ events, but the total number of HR events is significantly lower than when a 5' overhang is generated.

[0111] FIGS. 11A-B is directed to Cas9D10A nickase mediated NHEJ. As shown in FIG. 11A, the traffic light reporter was used to assay NHEJ events upon introduction of targeted nicks or double-stranded breaks. Briefly, upon introduction of DNA cleavage events, if the break goes through mutagenic NHEJ, the GFP is translated out of frame and the downstream mCherry sequences are rendered in frame resulting in red fluorescence. 14 gRNAs covering a 200 bp stretch of DNA: 7 targeting the sense strand (U1-7) and 7 the antisense strand (D1-7) were designed. As shown in FIG. 11B, it was observed that unlike the wild-type Cas9 which results in DSBs and robust NHEJ across all targets, most nicks (using the Cas9D10A mutant) seldom result in NHEJ events. All 14 sites are located within a contiguous 200 bp stretch of DNA and over 10-fold differences in targeting efficiencies were observed.

[0112] According to certain aspects, methods are described herein of modulating expression of a target nucleic acid in a cell that include introducing one or more, two or more or a plurality of foreign nucleic acids into the cell. The foreign nucleic acids introduced into the cell encode for a guide RNA or guide RNAs, a nuclelease-null Cas9 protein or proteins and a transcriptional regulator protein or domain. Together, a guide RNA, a nuclelease-null Cas9 protein and a transcriptional regulator protein or domain are referred to as a co-localization complex as that term is understood by one of skill in the art to the extent that the guide RNA, the nucleases-null Cas9 protein and the transcriptional regulator protein or domain bind to DNA and regulate expression of a target nucleic acid. According to certain additional aspects, the foreign nucleic acids introduced into the cell encode for a guide RNA or guide RNAs and a Cas9 protein nickase. Together, a guide RNA and a Cas9 protein nickase are referred to as a co-localization complex as that term is understood by one of skill in the art to the extent that the guide RNA and the Cas9 protein nickase bind to DNA and nick a target nucleic acid.

[0113] Cells according to the present disclosure include any cell into which foreign nucleic acids can be introduced and expressed as described herein. It is to be understood that the basic concepts of the present disclosure described herein are not limited by cell types. Cells according to the present disclosure include eukaryotic cells, prokaryotic cells, animal cells, plant cells, fungal cells, archaeal cells, eubacterial cells and the like. Cells include eukaryotic cells such as yeast cells, plant cells, and animal cells. Particular cells include mammalian cells. Further, cells include any in which it would be beneficial or desirable to regulate a target nucleic acid. Such cells may include those which are deficient in expression of a particular protein leading to a disease or detrimental condition. Such diseases or detrimental conditions are readily known to those of skill in the art. According to the present disclosure, the nucleic acid responsible for expressing the particular protein may be targeted by the methods described herein and a transcriptional activator resulting in upregulation of the target nucleic acid and corresponding expression of the particular protein. In this manner, the methods described herein provide therapeutic treatment.

[0114] Target nucleic acids include any nucleic acid sequence to which a co-localization complex as described herein can be useful to either regulate or nick. Target nucleic acids include genes. For purposes of the present disclosure, DNA, such as double stranded DNA, can include the target nucleic acid and a co-localization complex can bind to or otherwise co-localize with the DNA at or adjacent or near the target nucleic acid and in a manner in which the co-localization complex may have a desired effect on the target nucleic acid. Such target nucleic acids can include endogenous (or naturally occurring) nucleic acids and exogenous (or foreign) nucleic acids. One of skill based on the present disclosure will readily be able to identify or design guide RNAs and Cas9 proteins which co-localize to a DNA including a target nucleic acid. One of skill will further be able to identify transcriptional regulator proteins or domains which likewise co-localize to a DNA including a target nucleic acid. DNA includes genomic DNA, mitochondrial DNA, viral DNA or exogenous DNA.
Foreign nucleic acids (i.e. those which are not part of a cell’s natural nucleic acid composition) may be introduced into a cell using any method known to those skilled in the art for such introduction. Such methods include transfection, transduction, viral transduction, microinjection, lipofection, nucleofection, nanoparticle bombardment, transformation, conjugation and the like. One of skill in the art will readily understand and adapt such methods using readily identifiable literature sources.

Transcriptional regulator proteins or domains which are transcriptional activators include VP16 and VP64 and others readily identifiable by those skilled in the art based on the present disclosure.

Diseases and detrimental conditions are those characterized by abnormal loss of expression of a particular protein. Such diseases or detrimental conditions can be treated by upregulation of the particular protein. Accordingly, methods of treating a disease or detrimental condition are provided where the co-localization complex as described herein associates or otherwise binds to DNA including a target nucleic acid, and the transcriptional activator of the co-localization complex upregulates expression of the target nucleic acid. For example upregulating PRDM16 and other genes promoting brown fat differentiation and increased metabolic uptake can be used to treat metabolic syndrome or obesity. Activating anti-inflammatory genes are useful in autoimmunity and cardiovascular disease. Activating tumor suppressor genes is useful in treating cancer. One of skill in the art will readily identify such diseases and detrimental conditions based on the present disclosure.

The following examples are set forth as being representative of the present disclosure. These examples are not to be construed as limiting the scope of the present disclosure as these and other equivalent embodiments will be apparent in view of the present disclosure, figures and accompanying claims.

Example I

Cas9 Mutants

Sequences homologous to Cas9 with known structure were searched to identify candidate mutations in Cas9 that could ablate the natural activity of its RuvC and HNH domains. Using HHpred (world wide website tool kit.tuebingen.mp.lhepred), the full sequence of Cas9 was queried against the full Protein Data Bank (January 2013). This search returned two different HNH endonucleases that had significant sequence homology to the HNH domain of Cas9; Pad and a putative endonuclease (PDB ID: 3M7K and 4H9D respectively). These proteins were examined to find residues involved in magnesium ion coordination. The corresponding residues were then identified in the sequence alignment to Cas9. Two Mg-coordinating side-chains in each structure were identified that aligned to the same amino acid type in Cas9. They are 3M7K D92 and N113, and 4H9D D53 and N77. These residues corresponded to Cas9 D839 and N863. It was also reported that mutations of Pad residues D92 and N113 to alanine rendered the nuclease catalytically deficient. The Cas9 mutations D839A and N863A were made based on this analysis. Additionally, HHpred also predicts homology between Cas9 and the N-terminus of a Thermus thermophilus RuvC (PDB ID: 4EP4).

Example II

Plasmid Construction

The Cas9 mutants were generated using the Quickchange kit (Agilen technologies). The target gRNA expression constructs were either (1) directly ordered as individual gBlocks from IDT and cloned into the pCR-BluntII-TOPO vector (Invitrogen); or (2) custom synthesized by Genewiz; or (3) assembled using Gibson assembly of oligonucleotides into the gRNA cloning vector (plasmid #41824). The vectors for the HR reporter assay involving a broken GFP were constructed by fusion PCR assembly of the GFP sequence bearing the stop codon and appropriate fragment assembled into the EGFP lentivector from Addgene (plasmid #26777). These lentivectors were then used to establish the GFP reporter stable lines. TALENs used in this study were constructed using standard protocols. See Sanjana et al., Nature Protocols 7, 171-192 (2012) hereby incorporated by reference in its entirety. Cas9N and MS2 VP64 fusions were performed using standard PCR fusion protocol procedures. The promoter luciferase constructs for OCT4 and REX1 were obtained from Addgene (plasmid #17221 and plasmid #17222).

Example III

Cell Culture and Transfections

HEK 293T cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM, Invitrogen) high glucose supplemented with 10% fetal bovine serum (FBS, Invitrogen), penicillin/streptomycin (pen/strep, Invitrogen), and non-essential amino acids (NEAA, Invitrogen). Cells were maintained at 37°C and 5% CO2 in a humidified incubator.

Transfections involving nuclease assays were as follows: 0.4 x 10⁶ cells were transfected with 2 µg Cas9 plasmid, 2 µg gRNA and/or 2 µg DNA donor plasmid using Lipofectamine 2000 as per the manufacturer’s protocols. Cells were harvested 3 days after transfection and either analyzed by FACs, or for direct assay of genomic cuts the genomic DNA of ~1 x 10⁶ cells was extracted using DNAeasy kit (Qiagen). For these PCR was conducted to amplify the target region with genomic DNA derived from the cells and amplicons were deep sequenced by MiSeq Personal Sequencer (Illumina) with coverage >200,000 reads. The sequencing data was analyzed to estimate NHEJ efficiencies.

For transfections involving transcriptional activation assays: 0.4 x 10⁶ cells were transfected with (1) 2 µg Cas9N-VP64 plasmid, 2 µg gRNA and/or 0.25 µg of reporter construct; or (2) 2 µg Cas9N plasmid, 2 µg MS2-VP64, 2 µg gRNA-2XMS2apamter and/or 0.25 µg of reporter construct. Cells were harvested 24-48 hrs post transfection and assayed using FACs or immunofluorescence methods, or their total RNA was extracted and these were subsequently analyzed by RT-PCR. Here standard taqman probes from Invitrogen for OCT4 and REX1 were used, with normalization for each sample performed against GAPDH.
For transfections involving transcriptional activation assays for specificity profile of Cas9-gRNA complexes and TALEs: 0.4x10^6 cells were transfected with (1) 2 μg Cas9-N-VP64 plasmid, 2 μg gRNA and 0.25 μg of reporter library; or (2) 2 μg TALE-TF plasmid and 0.25 μg of reporter library; or (3) 2 μg control-TF plasmid and 0.25 μg of reporter library. Cells were harvested 24 hrs post transfection (to avoid the stimulation of reporters being in saturation mode). Total RNA extraction was performed using RNAeasy-plus kit (Qiagen), and standard RT-PCR performed using Superscript-III (Invitrogen). Libraries for next-generation sequencing were generated by targeted re-amplification of the transcript tags.

Example IV
Computational and Sequence Analysis for Calculation of Cas9-TF and TALE-TF Reporter Expression Levels

The high-level logic flow for this process is depicted in FIG. 8A, and additional details are given here. For details on construct library composition, see FIGS. 8A (level 1) and 8B.

Sequencing:

For Cas9 experiments, construct library (FIG. 8A, level 3, left) and reporter gene cDNA sequences (FIG. 8A, level 3, right) were obtained as 150 bp overlapping paired end reads on an Illumina MiSeq, while for TALE experiments, corresponding sequences were obtained as 51 bp non-overlapping paired end reads on an Illumina HiSeq.

Construct Library Sequence Processing:

Alignment: For Cas9 experiments, novoalign V2.07.17 (world wide website novocraft.com/main/index.php) was used to align paired reads to a set of 250 bp reference sequences that corresponded to 234 bp of the constructs flanked by the pairs of 8 bp library barcodes (see FIG. 8A, 3rd level, left). In the reference sequences supplied to novoalign, the 23 bp degenerate Cas9 binding site regions and the 24 bp degenerate transcript tag regions (FIG. 8A, first level) were specified as Ns, while the construct library barcodes were explicitly provided. For TALE experiments, the same procedures were used except that the reference sequences were 203 bp in length and the degenerate binding site regions were 18 bp vs. 23 bp in length. Validity checking: Novoalign output for comprised files in which left and right reads for each read pair were individually aligned to the reference sequences. Only read pairs that were both uniquely aligned to the reference sequence were subjected to additional validity conditions, and only read pairs that passed all of these conditions were retained. The validity conditions included: (i) Each of the two construct library barcodes must align in at least 4 positions to a reference sequence barcode, and the two barcodes must to the barcode pair for the same construct library. (ii) All bases aligning to the N regions of the reference sequence must be called by novoalign as As, Cs, Gs or Ts. Note that for neither Cas9 nor TALE experiments did left and right reads overlap in a reference N region, so that the possibility of ambiguous novoalign calls of these N bases did not arise. (iii) Likewise, no novoalign-called inserts or deletions must appear in these regions. (iv) No Ts must appear in the transcript tag region (as these random sequences were generated from As, Cs, and Gs only). Read pairs for which any one of these conditions were violated were collected in a rejected read pair file. These validity checks were implemented using custom perl scripts.

Induced Sample Reporter Gene cDNA Sequence Processing:

Alignment: SeqPrep (downloaded from world wide website github.com/jstjohn/SeqPrep) was first used to merge the overlapping read pairs to the 79 bp common segment, after which novoalign (version above) was used to align these 79 bp common segments as unpaired single reads to a set of reference sequences (see FIG. 8A, 3rd level, right) in which (as for the construct library sequencing) the 24 bp degenerate transcript tag was specified as Ns while the sample barcodes were explicitly provided. Both TALE and Cas9 cDNA sequence regions corresponded to the same 63 bp regions of cDNA flanked by pairs of 8 bp sample barcode sequences. Validity checking: The same conditions were applied as for construct library sequencing (see above) except that: (a) Here, due prior SeqPrep merging of read pairs, validity processing did not have to filter for unique alignments of both reads in a read pair but only for unique alignments of the merged reads. (b) Only transcript tags appeared in the cDNA sequence reads, so that validity processing only applied these tag regions of the reference sequences and not also to a separate binding site region.

Assembly of Table of Binding Sites Vs. Transcript Tag Associations:

Custom perl was used to generate these tables from the validated construct library sequences (FIG. 8A, 4th level, left). Although the 24 bp tag sequences composed of A, C, and G bases should be essentially unique across a construct library (probability of sharing≈2.8e-11), early analysis of binding site vs. tag associations revealed that a non-negligible fraction of tag sequences were in fact shared by multiple binding sequences, likely mainly caused by a combination of sequence errors in the binding sequences, or oligo synthesis errors in the oligos used to generate the construct libraries. In addition to tag sharing, tags found associated with binding sites in validated read pairs might also be found in the construct library read pair reject file if it was not clear, due to barcode mismatches, which construct library they might be from. Finally, the tag sequences themselves might contain sequence errors. To deal with these sources of error, tags were categorized with three attributes: (i) safe vs. unsafe, where unsafe meant the tag could be in the construct library rejected read pair file; shared vs. nonshared, where shared meant the tag was found associated with multiple binding site sequences, and 2+ vs. 1-only, where 2+ meant that the tag appeared at least twice among the validated construct library sequences and so presumed to be less likely to contain sequence errors. Combining these three criteria yielded 8 classes of tags associated with each binding site, the most secure (but least abundant) class comprising only safe, nonshared, 2+ tags; and the least secure (but most abundant) class comprising all tags regardless of safety, sharing, or number of occurrences.

Computation of Normalized Expression Levels:

Custom perl code was used to implement the steps indicated in FIG. 8A, levels 5-6. First, tag counts obtained for each induced sample were aggregated for each binding site, using the binding site vs. transcript tag table previously computed for the construct library (see FIG. 8C). For each sample, the aggregated tag counts for each binding site were then
divided by the aggregated tag counts for the positive control sample to generate normalized expression levels. Additional considerations relevant to these calculations included:

1. For each sample, a subset of "novel" tags were found among the validity-checked cDNA gene sequences that could not be found in the binding site vs. transcript tag association table. These tags were ignored in the subsequent calculations.

2. The aggregations of tag counts described above were performed for each of the eight classes of tags described above in binding site vs. transcript tag association table. Because the binding sites in the construct libraries were biased to generate sequences similar to a central sequence frequently, but sequences with increasing numbers of mismatches increasingly rarely, binding sites with few mismatches generally aggregated to large numbers of tags, while binding sites with more mismatches aggregated to smaller numbers. Thus, although use of the most secure tag class was generally desirable, evaluation of binding sites with two or more mismatches might be based on small numbers of tags per binding site, making the secure counts and ratios less statistically reliable even if the tags themselves were more reliable. In such cases, all tags were used. Some compensation for this consideration is obtained from the fact that the number of separate aggregated tag counts for n mismatching positions grows with the number of combinations of mismatching positions (equal to \( \binom{L}{n} \)).

3\(^\circ\), and so dramatically increases with n; thus the averages of aggregated tag counts for different numbers n of mismatches (shown in FIGS. 2b, 2c, and in FIGS. 9A and 10B) are based on a statistically very large set of aggregated tag counts for n=2.

3. Finally, the binding site built into the TALE construct libraries was by bp and tag associations were assigned based on these 18 bp sequences, but some experiments were conducted with TALEs programmed to bind central 14 bp or 10 bp regions within the 18 bp construct binding site regions. In computing expression levels for these TALEs, tags were aggregated to binding sites based on the corresponding regions of the 18 bp binding sites in the association table, so that binding site mismatches outside of this region were ignored.

**Example V**

RNA-Guided SOX2 and NANOG Regulation Using Cas9 vs. VP64

[0131] The sgRNA (aptamer-modified single guide RNA) tethering approach described herein allows different effector domains to be recruited by distinct sgRNAs so long as each sgRNA uses a different RNA-protein interaction pair, enabling multiplex gene regulation using the same Cas9-Np protein. For the FIG. 12A SOX2 and FIG. 12B NANO genes, 10 gRNAs were designed targeting a 1 kb stretch of DNA upstream of the transcription start site. The DNase hypersensitive sites are highlighted in green. Transcriptional activation via qPCR of the endogenous genes was assayed. In both instances, while introduction of individual gRNAs modestly stimulated transcription, multiple gRNAs acted synergistically to stimulate robust multi-fold transcriptional activation. Data are means+/-SEM (N=3). As shown in FIGS. 12A-B, two additional genes, SOX2 and NANOG, were regulated via sgRNAs targeting within an upstream -1 kb stretch of promoter DNA. The sgRNAs proximal to the transcriptional start site resulted in robust gene activation.

**Example VI**

Evaluating the Landscape of Targeting by Cas9-gRNA Complexes

[0132] Using the approach described in FIG. 2, the targeting landscape of two additional Cas9-gRNA complexes (FIGS. 13A-C) and (FIGS. 13D-F) was analyzed. The two gRNAs have vastly different specificity profiles with gRNA2 tolerating up to 2-3 mismatches and gRNA3 only up to 1. These aspects are reflected in both the one base mismatch (FIGS. 13B, 13E) and two base mismatch plots (FIGS. 13C, 13F). In FIGS. 13C and 13F, base mismatch pairs for which insufficient data were available to calculate a normalized expression level are indicated as gray boxes containing an 'x', while, to improve data display, mismatch pairs whose normalized expression levels are outliers that exceed the top of the color scale are indicated as yellow boxes containing an asterisk '*'. Statistical significance symbols are: *** for P<0.0005/n, ** for P<0.005/n, * for P<0.05/n, and N.S. (Non-Significant) for P>0.05/n, where n is the number of comparisons (refer Table 2).

**Example VII**

Validations, Specificity of Reporter Assay

[0133] As shown in FIGS. 14A-C, specificity data was generated using two different sgRNA:Cas9 complexes. It was confirmed that the assay was specific for the sgRNA being evaluated, as a corresponding mutant sgRNA was unable to stimulate the reporter library. FIG. 14A: The specificity profile of two gRNAs (wild-type and mutant; sequence differences are highlighted in red) were evaluated using a reporter library designed against the wild-type gRNA target sequence. FIG. 14B: It was confirmed that this assay was specific for the gRNA being evaluated (data re-plotted from FIG. 13D), as the corresponding mutant gRNA is unable to stimulate the reporter library. Statistical significance symbols are: *** for P<0.0005/n, ** for P<0.005/n, * for P<0.05/n, and N.S. (Non-Significant) for P>0.05/n, where n is the number of comparisons (refer Table 2). Different sgRNAs can have different specificity profiles (FIGS. 13A, 13D), specifically, sgRNA2 tolerates up to 3 mismatches and sgRNA3 only up to 1. The greatest sensitivity to mismatches was localized to the 3′ end of the spacer, albeit mismatches at other positions were also observed to affect activity.

**Example VIII**

Validations, Single and Double-Base gRNA Mismatches

[0134] As shown in FIGS. 15A-D, it was confirmed by targeted experiments that single-base mismatches within 12 bp of the 3′ end of the spacer in the assayed sgRNAs resulted in detectable targeting. However, 2 bp mismatches in this region resulted in significant loss of activity. Using a nuclease assay, 2 independent gRNAs were tested: gRNA2 (FIGS. 15A-B) and gRNA3 (FIGS. 15C-D) bearing single or double-
base mismatches (highlighted in red) in the spacer sequence versus the target. It was confirmed that single-base mismatches within 12 bp of the 3' end of the spacer in the assayed gRNAs result in detectable targeting, however 2 bp mismatches in this region result in rapid loss of activity. These results further highlight the differences in specificity profiles between different gRNAs consistent with the results in FIG. 13. Data are means+/−SEM (N=3).

Example IX

Validations, 5' gRNA Truncations

[0135] As shown in FIGS. 16A-D, truncations in the 5' portion of the spacer resulted in retention of sgRNA activity. Using a nuclease assay, 2 independent gRNA were tested: gRNA1 (FIGS. 16A-B) and gRNA3 (FIGS. 16C-D) bearing truncations at the 5' end of their spacer. It was observed that 1-3 bp 5' truncations are well tolerated, but larger deletions lead to loss of activity. Data are means+/−SEM (N=3).

Example X

Validations, S. pyogenes PAM

[0136] As shown in FIGS. 17A-B, it was confirmed using a nuclease mediated HR assay that the PAM for the S. pyogenes Cas9 is NGG and also NAG. Data are means+/−SEM (N=3). According to an additional investigation, a generated set of about 190K Cas9 targets in human exons that had no alternate NGG targets starting the last 13 nt of the targeting sequence was scanned for the presence of alternate NAG sites or for NGG sites with a mismatch in the prior 13 nt. Only 0.4% were found to have no such alternate targets.

Example XI

Validations, TALE Mutations

[0137] Using a nuclease mediated HR assay (FIGS. 18A-B) it was confirmed that 18-mer TALEs tolerate multiple mutations in their target sequences. As shown in FIGS. 18A-B certain mutations in the middle of the target lead to higher TALE activity, as determined via targeted experiments in a nuclease assay.

Example XII

TALE Monomer Specificity Versus TALE Protein Specificity

[0138] To decouple the role of individual repeat-variable diresidues (RVDs), it was confirmed that choice of RVDs did contribute to base specificity but TALE specificity is also a function of the binding energy of the protein as a whole. FIGS. 19A-C shows a comparison of TALE monomer specificity versus TALE protein specificity. FIG. 19A: Using a modification of approach described in FIG. 2, the targeting landscape of 2 14-mer TALE-TFs bearing a contiguous set of 6 NI or 6 NH repeats was analyzed. In this approach, a reduced library of reporters bearing a degenerate 6-mer sequence in the middle was created and used to assay the TALE-TF specificity. FIGS. 19B-C: In both instances, it was noted that the expected target sequence is enriched (i.e. one bearing 6 As for NI repeats, and 6 Gs for NH repeats). Each of these TALEs still tolerate 1-2 mismatches in the central 6-mer target sequence. While choice of monomers does contribute to base specificity, TALE specificity is also a function of the binding energy of the protein as a whole. According to one aspect, shorter engineered TALEs or TALEs bearing a composition of high and low affinity monomers result in higher specificity in genome engineering applications and FokI dimerization in nuclease applications allows for further reduction in off-target effects when using shorter TALEs.

Example XIII

Off-Set Nicking, Native Locus

[0139] FIGS. 20A-B shows data related to off-set nicking. In the context of genome-editing, off-set nicked were created to generate DSBs. A large majority of nicks do not result in non-homologous end joining (NHEJ) mediated indels and thus when inducing off-set nick, off-target single nick events will likely result in very low indel rates. Inducing off-set nick to generate DSBs is effective at inducing gene disruption at both integrated reporter loci and at the native AAVS1 genomic locus.

[0140] FIG. 20A: The native AAVS1 locus with 8 gRNAs covering a 200 bp stretch of DNA was targeted; 4 targeting the sense strand (s1-4) and 4 the antisense strand (as1-4). Using the Cas9D10A mutant, which nicked the complementary strand, different two-way combinations of the gRNAs was used to induce a range of programmed 5' or 3' overhangs. FIG. 20B: Using a Sanger sequencing based assay, it was observed that while single gRNAs did not induce detectable NHEJ events, inducing off-set nick to generate DSBs is highly effective at inducing gene disruption. Notably off-set nick leading to 5' overhangs result in more NHEJ events as opposed to 3' overhangs. The number of Sanger sequencing clones is highlighted above the bars, and the predicted overhang lengths are indicated below the corresponding y-axis legends.

Example XIV

Off-Set Nicking, NHEJ Profiles

[0141] FIGS. 21A-C is directed to off-set nicking and NHEJ profiles. Representative Sanger sequencing results of three different off-set nicking combinations is shown with positions of the targeting gRNAs highlighted by boxes. Furthermore, consistent with the standard model for homologous recombination (HR) mediated repair, engineering of 5' overhangs via off-set nick generated more robust NHEJ events than 3' overhangs (FIG. 3B). In addition to a stimulation of NHEJ, robust induction of HR was observed when the 5' overhangs were created. Generation of 3' overhangs did not result in improvement of HR rates (FIG. 3C).
**Example XV**

**TABLE 1**

<table>
<thead>
<tr>
<th>gRNA Name</th>
<th>gRNA Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>REXL 1</td>
<td>ctcgcgcgaagtctctggtgtaa ggg</td>
</tr>
<tr>
<td>REXL 2</td>
<td>cctcgagcccccataagtact tgg</td>
</tr>
<tr>
<td>REXL 3</td>
<td>acgcctcctctctaagggatg tgg</td>
</tr>
<tr>
<td>REXL 4</td>
<td>ccaagtaaagccagcagc tgg</td>
</tr>
<tr>
<td>REXL 5</td>
<td>gcccaccacccacctctac tgg</td>
</tr>
<tr>
<td>REXL 6</td>
<td>aataaattctaatttctgtaa ggg</td>
</tr>
<tr>
<td>REXL 7</td>
<td>gcctgctgaggttaagctct ggg</td>
</tr>
<tr>
<td>REXL 8</td>
<td>ttagaataaagctcctat tgg</td>
</tr>
<tr>
<td>REXL 9</td>
<td>ccggcgagcaagggcttgca ggg</td>
</tr>
<tr>
<td>REXL 10</td>
<td>tcttcctcctccttccttggtg tgg</td>
</tr>
<tr>
<td>OCT4 1</td>
<td>tcttcctcctccttccttgtg tgg</td>
</tr>
<tr>
<td>OCT4 2</td>
<td>tcttcctcctccttccttggtg tgg</td>
</tr>
<tr>
<td>OCT4 3</td>
<td>aatcgtgctgctgctgtgtaa ggg</td>
</tr>
<tr>
<td>OCT4 4</td>
<td>ctcggccctcctccttggc ggg</td>
</tr>
<tr>
<td>OCT4 5</td>
<td>ggttttgttggaaggggaag tgg</td>
</tr>
<tr>
<td>OCT4 6</td>
<td>ggttttgttggaaggggaag tgg</td>
</tr>
<tr>
<td>OCT4 7</td>
<td>cgctcctctccttccttggtg tgg</td>
</tr>
<tr>
<td>OCT4 8</td>
<td>tcttcctcctccttccttggtg tgg</td>
</tr>
<tr>
<td>OCT4 9</td>
<td>cccagccagccagccagc tgg</td>
</tr>
<tr>
<td>OCT4 10</td>
<td>gatcagctcctcctctccttg tgg</td>
</tr>
<tr>
<td>OCT4 11</td>
<td>tgaagtaaagccagcagc tgg</td>
</tr>
<tr>
<td>OCT4 12</td>
<td>ccagagcagccagcagc tgg</td>
</tr>
<tr>
<td>OCT4 13</td>
<td>ccagagcagccagcagc tgg</td>
</tr>
<tr>
<td>OCT4 14</td>
<td>aaaggagtctcctctccttg tgg</td>
</tr>
<tr>
<td>OCT4 15</td>
<td>atcgtcctcttgggtgtaa ggg</td>
</tr>
<tr>
<td>OCT4 16</td>
<td>gttgttcctcttgggtgtaa ggg</td>
</tr>
<tr>
<td>OCT4 17</td>
<td>tgggttgttggaaggggaag tgg</td>
</tr>
<tr>
<td>OCT4 18</td>
<td>ggttttgttggaaggggaag tgg</td>
</tr>
<tr>
<td>OCT4 19</td>
<td>tgggttgttggaaggggaag tgg</td>
</tr>
<tr>
<td>OCT4 20</td>
<td>tcttcctcctccttccttggtg tgg</td>
</tr>
<tr>
<td>OCT4 21</td>
<td>gcctcctctccttccttggtg tgg</td>
</tr>
<tr>
<td>S0X2 1</td>
<td>cacacccacccacccacccac ccagag</td>
</tr>
<tr>
<td>S0X2 2</td>
<td>gtaagccagctcctctcctccttg tgg</td>
</tr>
<tr>
<td>S0X2 3</td>
<td>taagacccacccacccacccacccagag</td>
</tr>
<tr>
<td>S0X2 4</td>
<td>ggcagagcagcagcagcagcag</td>
</tr>
<tr>
<td>S0X2 5</td>
<td>cccacccacccacccacccacccagag</td>
</tr>
<tr>
<td>S0X2 6</td>
<td>ctccctcctctccttcctccttg tgg</td>
</tr>
<tr>
<td>S0X2 7</td>
<td>gagcagagcagcagcagcagcag</td>
</tr>
<tr>
<td>S0X2 8</td>
<td>aaaaccgcagcagcagcagcag</td>
</tr>
<tr>
<td>S0X2 9</td>
<td>ccttccttcctccttccttccttg tgg</td>
</tr>
<tr>
<td>S0X2 10</td>
<td>aggagcagcagcagcagcag</td>
</tr>
</tbody>
</table>

**Example XVI**

**TABLE 2**

<table>
<thead>
<tr>
<th>FIG.</th>
<th>Expression level comparison: mutations vs. mutations</th>
<th>t-test</th>
<th>P-value</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>2b</td>
<td>1-samp 7.8E-05 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2-samp 1.4E-06 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3-samp 4.0E-01 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4-samp 0 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5-samp 0 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6-samp 1.0E-21 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7-samp 1.7E-43 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8-samp 3.7E-02 N.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2e</td>
<td>1-samp 8.9E-01 N.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2-samp 1.9E-06 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3-samp 5.0E-147 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4-samp 0 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5-samp 0 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6-samp 4.2E-62 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7-samp 1.6E-03 *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8-samp 4.7E-01 N.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S7a</td>
<td>1-samp 5.2E-02 N.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2-samp 2.8E-05 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3-samp 3.5E-21 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4-samp 1.4E-58 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5-samp 8.3E-101 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6-samp 6.8E-04 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7-samp 1.6E-01 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8-samp 8.1E-24 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S7d and S8d</td>
<td>1-samp 2.3E-18 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2-samp 2.4E-08 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3-samp 6.2E-54 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4-samp 4.0E-141 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5-samp 1.9E-20 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6-samp 1.2E-03 *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7-samp 3.8E-05 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8-samp 9.4E-01 N.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S8c</td>
<td>1-samp 7.2E-03 N.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2-samp 5.0E-01 N.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3-samp 3.9E-84 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4-samp 8.5E-133 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5-samp 8.6E-76 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6-samp 1.6E-05 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7-samp 7.1E-01 N.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8-samp 7.8E-02 N.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S13a (left)</td>
<td>1-samp 7.3E-01 N.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2-samp 2.4E-06 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3-samp 7.2E-140 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4-samp 0 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5-samp 0 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6-samp 1.0E-72 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7-samp 4.9E-03 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S13a (middle)</td>
<td>1-samp 9.4E-02 N.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1-samp 5.2E-09 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2-samp 7.9E-86 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3-samp 2.9E-53 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4-samp 3.5E-10 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S13a (right)</td>
<td>1-samp 1.3E-13 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1-samp 1.1E-04 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2-samp 3.7E-08 **</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2—continued

**Summary of Statistical Analysis of Cas9-gRNA and TALE Specificity Data**

<table>
<thead>
<tr>
<th>seed start position</th>
<th>Number position pairs</th>
<th>both in seed</th>
<th>not both in seed</th>
<th>-log10(P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>171</td>
<td>19</td>
<td>3.11</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>153</td>
<td>37</td>
<td>1.46</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>136</td>
<td>54</td>
<td>2.01</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>120</td>
<td>70</td>
<td>3.34</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>105</td>
<td>85</td>
<td>5.65</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>91</td>
<td>99</td>
<td>7.34</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>78</td>
<td>112</td>
<td>6.61</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>66</td>
<td>124</td>
<td>7.10</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>55</td>
<td>135</td>
<td>9.72</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>45</td>
<td>145</td>
<td>9.83</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>36</td>
<td>154</td>
<td>10.44</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>28</td>
<td>152</td>
<td>10.72</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>21</td>
<td>169</td>
<td>8.97</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>15</td>
<td>175</td>
<td>5.61</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>10</td>
<td>180</td>
<td>3.34</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
<td>7</td>
<td>184</td>
<td>2.26</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>3</td>
<td>187</td>
<td></td>
</tr>
</tbody>
</table>

Table 2(a) P-values for comparisons of normalized expression levels of TALE or Cas9-VP64 activators binding to target sequences with particular numbers of target site mutations. Normalized expression values have been indicated by boxplots in the Figs., indicated in the FEG column, where the boxes represent the distribution of these levels by numbers of mismatches from the target site. P-values were calculated using t-tests for each consecutive pair of numbers of mismatches in each boxplot, where the t-tests were either one sample or two sample t-tests (see Methods). Statistical significance was assessed using Bonferroni-corrected P-value thresholds, where the correction was based on the number of comparisons within each boxplot.

Statistical significance symbols are: **** P < 0.005, *P < 0.05, n.s. (Non-Significant) for P > 0.05, where n is the number of comparisons.

Table 2(b) Statistical characterization of seed region in FEG, 3D, log10(P-values) indicating the degree of separation between expression values for Cas9-VP64 and gRNA binding to target sequences with two mutations for those position pairs mutated within candidate seed regions at the 3’ and 3’ substitution sites; all other position pairs. The greatest separation, indicated by the largest log10(P-values) (highlighted above), is found in the last 8-9 bp of the target site. These positions may be interpreted as indicating the start of the “seed” region of this target site. See the sections “Statistical characterization of seed region” in Methods for information on how the P-values were computed.

---

**Example XVII**

Sequences of Proteins and RNAs in the Examples

[0144] A. Sequences of the Cas9<sub>VP64</sub> activator constructs based on the m4 mutant are displayed below. Three versions were constructed with the Cas9<sub>VP64</sub> and Cas9<sub>VP64</sub>N fusion protein formats showing highest activity. Corresponding vectors for the m3 and m2 mutants (Fig. 4A) were also constructed (NLS and VP64 domains are highlighted).

**Cas9<sub>VP64</sub>**

[0145]
Cas9\textsubscript{RNP} Sequences

(SEQ ID NO. 3)

gccaccATGCCAAGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA

[0146]
B. Sequences of the MS2-activator constructs and corresponding gRNA backbone vector with 2xMS2 aptamer domains is provided below (NLS, VP64, gRNA spacer, and MS2-binding RNA stem loop domains are highlighted). Two versions of the former were constructed with the MS2<sub>rpoA</sub>-N fusion protein format showing highest activity.
MS2/VP6aN
[0148]

-continued

TCTTGCTAGTTTGGCAGTTTTAAAATTAAAAGATGACTACATCATA
TGCTTACCCTAATTCAAGATTTTTGGCATTTTTGGCTTAATATCTTG
TGAAAGAGGAGAAGACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
AAAGAAAAAGCTATGACCGGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAG
AGCAGCGAGGGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
CTAGTGACGTAATCTCCGGTTCTACATTAGCTCACTCTGAATCGGAG
TAGAAAAGAGCATGAGATCTTAATCTGCTAGGGGAGGAGGAGGAGGAGGAG

C. dTomato fluorescence based transcriptional activation reporter sequences are listed below (ISce1 control-1F target, gRNA targets, minC-MV promoter and FLAG tag-dTomato sequences are highlighted).

TF Reporter 1
[0150]

-continued

TAGAGGATACAGGCGATATGACCTCTCCCTCCACCCCCCCCGGAGCAGCT
AGCCGATTCGCAGCTGGAATCCGCGAGCTAGCGCTGAGCTGACAGCT
GAGCCGCGCGCCGCGCCCGCCCGGCAATGAGGACCCGCGAGCTCTCG
TTCAATTGGAGGCGCTGATGCACTTGCGAGGAGGAACGAGGAGGAGGAG
TTCGCGCGACAGACTTGCGCGCGCGGCGCGCGCGCGCGCGCGCGGCGCGGCG
GAGCGCGAGCGAGCGAGCGAGACGAGACGAGACGAGACGAGACGAGACGAG
GAGCGCGAGCGAGCGAGCGAGACGAGACGAGACGAGACGAGACGAGACGAG
GAGCGCGAGCGAGCGAGCGAGACGAGACGAGACGAGACGAGACGAGACGAG

TF Reporter 2
[0151]
D. General format of the reporter libraries used for TALE and Cas9-gRNA specificity assays is provided below (iSceI control target, gRNA/TALE target site (23 bp for gRNAs and 18 bp for TALEs), minCMV promoter, RNA barcode, and dTomato sequences are highlighted).

```plaintext
SEQ LEANING LISTING

<160> NUMBER OF SEQ ID NOS: 186
<210> SEQ ID NO 1
<211> LENGTH: 1368
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes
<400> SEQUENCE: 1

Met Asp Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
5
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Phe
10
Gly Lys Val Gly Asn Thr Asp Arg His Ser Ile Lys Asn Leu Ile
15
Gly Ala Leu Leu Phe Asp Ser Gly Thr Ala Glu Ala Thr Arg Leu
20
Gly Arg Thr Ala Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
25
Gly Lys Val Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Ser
30
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
35
```
<table>
<thead>
<tr>
<th>Residue</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>His</td>
<td>115</td>
</tr>
<tr>
<td>Glu</td>
<td>120</td>
</tr>
<tr>
<td>Arg</td>
<td>125</td>
</tr>
<tr>
<td>His</td>
<td>130</td>
</tr>
<tr>
<td>Glu</td>
<td>135</td>
</tr>
<tr>
<td>Arg</td>
<td>140</td>
</tr>
<tr>
<td>Pro</td>
<td>145</td>
</tr>
<tr>
<td>Ile</td>
<td>150</td>
</tr>
<tr>
<td>Pro</td>
<td>155</td>
</tr>
<tr>
<td>Thr</td>
<td>160</td>
</tr>
<tr>
<td>Tyr</td>
<td>165</td>
</tr>
<tr>
<td>Tyr</td>
<td>170</td>
</tr>
<tr>
<td>His</td>
<td>175</td>
</tr>
<tr>
<td>Residue</td>
<td>180</td>
</tr>
<tr>
<td>Asp</td>
<td>185</td>
</tr>
<tr>
<td>Ser</td>
<td>190</td>
</tr>
<tr>
<td>Asp</td>
<td>195</td>
</tr>
<tr>
<td>Val</td>
<td>200</td>
</tr>
<tr>
<td>Asp</td>
<td>205</td>
</tr>
<tr>
<td>Lys</td>
<td>210</td>
</tr>
<tr>
<td>Ala</td>
<td>215</td>
</tr>
<tr>
<td>Ser</td>
<td>220</td>
</tr>
<tr>
<td>Ala</td>
<td>225</td>
</tr>
<tr>
<td>Leu</td>
<td>230</td>
</tr>
<tr>
<td>Leu</td>
<td>235</td>
</tr>
<tr>
<td>Ser</td>
<td>240</td>
</tr>
<tr>
<td>Asp</td>
<td>245</td>
</tr>
<tr>
<td>Ala</td>
<td>250</td>
</tr>
<tr>
<td>Lys</td>
<td>255</td>
</tr>
<tr>
<td>Asp</td>
<td>260</td>
</tr>
<tr>
<td>Ala</td>
<td>265</td>
</tr>
<tr>
<td>Glu</td>
<td>270</td>
</tr>
<tr>
<td>Asp</td>
<td>275</td>
</tr>
<tr>
<td>Leu</td>
<td>280</td>
</tr>
<tr>
<td>Asp</td>
<td>285</td>
</tr>
<tr>
<td>Ala</td>
<td>290</td>
</tr>
<tr>
<td>Ala</td>
<td>295</td>
</tr>
<tr>
<td>Ser</td>
<td>300</td>
</tr>
<tr>
<td>Ala</td>
<td>305</td>
</tr>
<tr>
<td>Leu</td>
<td>310</td>
</tr>
<tr>
<td>Ser</td>
<td>315</td>
</tr>
<tr>
<td>Ala</td>
<td>320</td>
</tr>
<tr>
<td>Met</td>
<td>325</td>
</tr>
<tr>
<td>Ile</td>
<td>330</td>
</tr>
<tr>
<td>Leu</td>
<td>335</td>
</tr>
<tr>
<td>Leu</td>
<td>340</td>
</tr>
<tr>
<td>Ser</td>
<td>345</td>
</tr>
<tr>
<td>Ala</td>
<td>350</td>
</tr>
<tr>
<td>Gly</td>
<td>355</td>
</tr>
<tr>
<td>Ser</td>
<td>360</td>
</tr>
<tr>
<td>Gly</td>
<td>365</td>
</tr>
<tr>
<td>Lys</td>
<td>370</td>
</tr>
<tr>
<td>Gln</td>
<td>375</td>
</tr>
<tr>
<td>Gly</td>
<td>390</td>
</tr>
<tr>
<td>Thr</td>
<td>395</td>
</tr>
<tr>
<td>Glu</td>
<td>400</td>
</tr>
<tr>
<td>Arg</td>
<td>405</td>
</tr>
<tr>
<td>Ser</td>
<td>410</td>
</tr>
<tr>
<td>Gly</td>
<td>415</td>
</tr>
<tr>
<td>Asp</td>
<td>420</td>
</tr>
<tr>
<td>Ser</td>
<td>425</td>
</tr>
<tr>
<td>Gly</td>
<td>430</td>
</tr>
<tr>
<td>Arg</td>
<td>435</td>
</tr>
<tr>
<td>Thr</td>
<td>440</td>
</tr>
<tr>
<td>Phe</td>
<td>445</td>
</tr>
<tr>
<td>Asp</td>
<td>450</td>
</tr>
<tr>
<td>Thr</td>
<td>455</td>
</tr>
<tr>
<td>Asp</td>
<td>460</td>
</tr>
<tr>
<td>Lys</td>
<td>465</td>
</tr>
<tr>
<td>Ser</td>
<td>470</td>
</tr>
<tr>
<td>Glu</td>
<td>475</td>
</tr>
<tr>
<td>Pro</td>
<td>480</td>
</tr>
<tr>
<td>Ser</td>
<td>485</td>
</tr>
<tr>
<td>Ala</td>
<td>490</td>
</tr>
<tr>
<td>Ser</td>
<td>495</td>
</tr>
<tr>
<td>Asp</td>
<td>500</td>
</tr>
<tr>
<td>Ala</td>
<td>505</td>
</tr>
<tr>
<td>Glu</td>
<td>510</td>
</tr>
<tr>
<td>Leu</td>
<td>515</td>
</tr>
</tbody>
</table>
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gin
515 520 525
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Glu 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
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 Gin 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 Gin Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Sex Asp Gly Phe
675 680 685
Ala Asn Arg Asp Leu Asn Phe Met Gin Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gin Lys Ala Gin Val Ser Gly Gin 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 Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Gln Met Ala Arg Glu Asn Gin
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gin Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gin Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gin Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val His Ile Val Pro Gin Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Val Pro Ser Glu Glu Val Val Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gin Leu Leu Asn Ala Lys Leu Ile Thr Gin 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 Gin Leu Val Glu Thr Arg Gin 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 Val Lys Val Ile Thr Leu Lys Ser 945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Glu Phe Tyr Lys Val Arg 965 970 975
Glu Ile Asn Asn Tyr His His Ala Asp Ala Tyr Leu Asn Ala Val 980 985 990
Val Gly Thr Ala Leu Ile Lys Tyr Pro Lys Leu Glu Ser Glu Phe 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
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 Thr 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 Glu Asp Gly Arg Lys Arg Met Leu Ala Ser Ala Gly 1205 1210 1215
Glu Leu Glu 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 Glu 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 Arg Leu Gly Ala Pro Ala Ala 1310 1315 1320
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> SEQ ID NO 2
<211> LENGTH: 4332
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE: 
<223> OTHER INFORMATION: VP64-activator construct
<400> SEQUENCE: 2

gccaccatgg acgaaagta ctctcatggg ctgcctctcg gcacaacag cgtagcttg 60
gcgtgctta gcggaggtga caagggtcgg agcaaaasat tcaagtctt ggcacaatcc 120
gatgcgccac gcataagaa gaaccttcat ggcgccttcc tggctgactc cgggagagc 180
gcgcagaga gcgcgctcaa aagaacagca cggcgcgat atacccgca aagaatacg 240
atctgcatac cgcggagagat ctttagtaat gagatgtgta agtggtgagct cttctcttc 300
catgagcttg agagctgcttt tttggtggag gaggataaa agacgcagcc ccaaccaatc 360
tttggacata ctggtggagca ggtggctgtac cttgaaagat acaccaacct atactcatct 420
aggaagagcc tgtgtaagac tactgtaag ggtgcgtgca ggtgtgactta tctggtgctg 480
ggctatagta tcaatatttcg gggacaccttc cttctgaggg ggcacccgaa ccaacgaac 540
agcgatctgc acaaacctctt tatacaacgt tctggacactt acaacgctt tttggagag 600
aaccgcatac agcagcagcc agtggaagcc aaacataac ccagccgtgag cttgctcaga 660
tccgggctg ccgaagacct ctcatgcaag ctcctgaggg aggaagagaa ccgctgtt 720
gtatttacca tggcctgtgc acctggtcgtg acacccaaact tttaactcata ctggactcgt 780
ggcggagct ccaacggcag acgcagcttc acgtgctttc agggtggaga aatggcgagc 840
tgcgccgacgg gcgcgcacgg ggcagcggaa ttgcttcttg cggccgaaag caagccgagc 900
gcctctctcg tgtgtgtgtc tctggctggc agaccgcaac tcaaccaagttc acacggcgg 960
gctagtata tcaagagctc gcggagactc cagccggcagc taaggtgtc cagggctc 1020
gtgcacagc aacgtgcctg gaagtcacgg aataattttc tcggcgccat ttaaatgcgc 1080
tagccgaggt aatccggagc ggggagcgc cggagggatt tttcaacatt tattaggcc 1140
aattccgcct gctcaggtgga caagagggag ccattggact tctcaacca aaacgaagaa 1200
ctgcggcctg cagctcaccgg gccaatgactt gcggccaggg gaaagggga 1260
ctgcggcctg cagctcaccgg gccaatgactt gcggccaggg gaaagggga 1260

tggatcagc aacctcctca atttgggata cccatctt ctcagcaggta cagcagccatg 1320
aagaggaga aatcctccct cggccgctctt cccgtggtct caggtggtct 1380
aattcctag ctcggtggtgg cactgtgcaaa cctagttc ttcgaggttc agttcgctttt 1440
gacgcgtgct gcgtgctgctgc ttcgaggttc agttcgctttt 1500
gactctgattctgcggctgc cagctcgtcct cggccgctctt cctacgtgctgct 1560
acacagtta cagccccctc cagctcgtcct cggccgctctt cctacgtgctgct 1620
-continued

ttcctgcctg gagagcagaa gaaagctatc gttgacctcc tctccaagac gacccggaaa 1680
gttacgcga aagacgcta aaagaactat ttcnaanaag aaggatagtt cagactgcttt 1740
gaatacgag gtagtgaggta cgctgctca cgctgctcag gcacagtata cagactgcttt 1800
aaaatcataa aaccgaagaa ctccctggac aatgagagaa acaggagcaat tctggagac 1860
attgtcttca cccctctagt gttggaagat agggagatga ttgagagac ccattgaaact 1920
tacgtcactct ttcctcagaa caaagctatc aaccacgca gaggagcctc atatacgga 1980
tgaggggccc tgcgcaagaa actgatcact gggatcggag acaagcagac tggacagaca 2040
atctgctgaatt ttcttacgct cgatggtatt gcacacccgga actctagctta ctggctatc 2100
gagatcctc tcatcctttaa gggacagcc cagaaacgc aacgttctcg ccagagggggc 2160
agctcctgag acgcctgcag taacactggc aagtcgcccc gcctcaaaa gagcaatact 2220
cagacagtta aggtgctgga tgaacctgcg aagtaataag gaaggcattaa gccgagcact 2280
atcgtatcg agatggcctg aggacacaca ctaacctcgcag ggagacgagaa aacactgag 2340
gaaaggtaga agaggtgatga agaggttaaa aagaaactgg ggctccaaat ccttaagga 2400
cacccagtta aaaaacccac gatcctcaga gagaagctct acctgtacta cctgagcaga 2460
gggaggacat cgtcactggct ccagagactc gacatcactc ggtgctcaga 2520
gcgtctgtcg tcgccccagtc tttttcctaa gatgatcctta tgtagataata aagtggcaca 2580
agatcogata aagctacgag ggaagctgat aagctcctgt cagaaagctg tggataaaaa 2640
atgaaaatt attcgtcagcag gcggtctgaaac gcacacactg tcatcacaacgg gaagttcact 2700
aatctgactga aagttagcagc agztgtgctgg tgtgatctgg ataaagcccgc ttctcataaa 2760
aggacagtgg tggagacacg cctgcgcgcc gcacagcttt gcattcaccg 2820
atgacacccc aagctatgagc aattgacaa ctggagcagcag agttagagac 2880
atgtctcaag cgatgcgtcag ttcgaaagaggac cagttttactt ttttaagtct ggcgagagatc 2940
aacatcagcc accatcgcag cgtcgctcagc cttgactgcag tggtaggccg tggagctttc 3000
aaaaatcata ccagctgctga atcctgtattt gttgagagag acataaaggt tggatgatgt 3060
agggatagga tgcgcaaggtc tgcggccgag aagctgaggc cccgccttaag ttccttttct 3120	
tacgcaataa tctgacatat ttcggaagaca gatttgaccg gctgcagccgg 3180
aagcggaccc ttcgcaaaac aacggaagaa acaggagaaaa cgtgtcgggg caaggttagg 3240
gatgtccgga ctaggcggaa ggctcgtgcc atggcagggc tggacatgtg taaaagccac 3300
gaatgacagca cggagaggct cttgccagga aagttcctcc cgcagaaggg aacgacagaa 3360
cattgcgcag gcacacagaggg ctcggagccc aacaaatcag acggttccga tttcctcaca 3420
gctcgtcata cgtatctggt tgtggcacaag tggagagagc ggaagttcattt gaaactcaca 3480
agcgtcagag aaccttctgag ccattcagagc atggagccagc cagttgtggg caaaaacc 3540
atcgtacctgg caggggaag aaggatgaa cagctcagat cattacggt 3600
cagcaagctcc tctctgccag gttgaaaaac gcgcggaaag gactgtcgcg tgggcccggc 3660
agcgtgcaag caggttatcga gcggccagct gctcccaaat aagttttcaatt ctgctatttg 3720
gccagcact acgaagagct caaaggggtc cccacgagta atgcgcagaa gcagctgcttcc 3780

agagttgatt ctccgacagc cattccctctt ggtgtttgct ccattcagac cagttgtcttctg 3840
agagttgatt ctccgacagc cattccctctt ggtgtttgct ccattcagac cagttgtcttctg 3900
gataagccca tcggggagca ggcaagaaac attattccct tgtttactct gcaccaacttgg 3960
gggogcgctg cagccttcaaa acaactcg gcaacctcag acaggaaagcg gtaacactct 4020
aacaagggag tcctgagcc acaacttct gatcagctaa tcgcgggctc ctatggaaca 4080
agaatcgacg ccctcttgctc agtgcggc ggcgggaaga acaacaaagaa gaaagagggag 4140
gttggagccag cggcgttgcc aacgggtgca gcgattgagcc atttgagact gatatatctgct 4200
ggaagctgaac ccctcgagtc tttttgcc tccacggtga gctgacgtagc ccctgggtac 4260
ttttgacgtc acagctgtagc cagttgcgcc cttgatgcatt cctagcggaa cttgatgtcg 4320
aactcctag qa 4332

<210> SEQ ID NO: 3
<211> LENGTH: 4365
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE: 
<222> OTHER INFORMATION: VP64-activator construct

<400> SEQUENCE: 3

ggcacacatgt caagagaagaa ggaggggtgc ggaaggggggc tggacaaagaa gttactcattcat 60
gggctgctata ggacaaacaa caacgtgctg acggcgtctc ttaacggagct caacaggttct 120
cggacaaaaa aatccaaagt cccggcgaatt caacagtgcac acagcatgaa gaaagacttc 180
attggcggcag ccctcttgctc agtgcggcagcccgacagcacgcaagagcgctt ccacagggag 240
gcaggggtcag tataacccag caagaaagcat cctgtctggtc acggctcgag cattttctagg 300
aattgcgtgg ccctcttgctc agtgcggcagcccgacagcacgcaagagcgctt ccacagggag 360
ggcagagtaaa aacgcaacgc ggcagacccag atctttggttctcctgactgagac gtaggtaggg 420
tacttcatgaa agtaaccaac catatatcatg ctaggagaga aagcttagtca tagctctgat 480
aaggtggtcatgtg ccctcttgctc agtgcggcagcccgacagcacgcaagagcgctt ccacagggag 540
ttcctctcatgc agggggacat gaaocagcag acaagcagtc gctaaactgt cttatttccgacaa 600
ctcggttgaag ccctcttgctc agtgcggcagcccgacagcacgcaagagcgctt ccacagggag 660
ggcacacagc tcggaggcgc tagctctgcc aacctccggg gcggcggggc aacagtccagca 720
cggctccgttc ggagggcttg tttggatgaac ttattcgcocctt gtaactcgagcgg 780
tgacacccagc actttacata caacttcgcag ttcggccaag atgcaagctg taacagtgac 840
aagacagctc aagatgataga ttctgcaaat ctggagccc acagagggcc caagtaacgca 900
gaccccttttt tcggggagaag caacgctgca gcggcagctc ggcgggtcgg tcgctagtga 960
gtcgacagc acgttcacgctc aacgctcccg aatgctcctc tgaatgattc gcctcagttcgag 1020
ccacacacag ctctgtcactg cttgagccgc ccctctctgacg aacacacggc tgaagaacgt 1080
aaggaacttt tcctctgcatc ccctctaata cggactgcgg gttactattgc ccgggcggaggaco 1140
agccggagag attttacaa attttaaag cccatcgttgg aaaaaatgga ccggagccag 1200
aggtcgctgc taagctgttaa cagagaaatag ttcgctgcag aacagacgcac tcctcactaat 1260
ggagacatcgc cccacaccag tcacgtgggg cggacgagct catccccagc ggcggcaagag 1320
gattttctgc ccccttttga ccataacaggg gaaaagattg agaaaaactc ccccttgcag 1380
attacctct atgtgggccc cttcggccgg gaaaccacca gttcggctgt gatgactgcgct 1440
aaatcagggc agccacactcag tccccggaac cttgagggag tggcttttagg cgggctcct 1500
-continued

gcctcaacct tcatcgaagaag atgacactaacttttgataaacttcgcttaaaccgaaggttaa
1560
cctctataaac aactctctgtgtgagattaccttttaaattcagagcttcacaagggacttc
1620
aatcaggtcaagagaagggacaggaaacgaccttctgtgctggaggaacgaaagagctt
1680
atggtagaccctctcttctgagcaggaacccagaaagttccgctgtaaaaccagctcaaaagaagagc
1740
atttttacaaagaagttcagtttatgatctgtgaatatactgctgacggaagactttgtgtgaa
1800
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
1860
gcagagcggg cagagctctgagacagttctgagagagctttggtggaacgcagatgc
1920
atagagctagctaattagaccatcaaatgtcctagtcaagcagtgcagacttc
1980
atagagctacctgctggagacagttcagactgcagtgcctttggtggaacgcagatgc
2040
aactcctgcttcgtgctggagacagttcagactgcagtgcctttggtggaacgcagatgc
2100
cttggcccaaccggaaggctgagttcagactgcagtgcctttggtggaacgcagatgc
2160
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
2220
gcagagcggg cagagctctgagagagctttggtggaacgcagatgc
2280
gtccaaattgagacagttcagactgcagtgcctttggtggaacgcagatgc
2340
ccaacttaaccggaaggctgagttcagactgcagtgcctttggtggaacgcagatgc
2400
atatatttcagcagagctctgagagagctttggtggaacgcagatgc
2460
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
2520
ttttgccaagcttggagagagctttggtggaacgcagatgc
2580
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
2640
gtaactccttccacagagagagctttggtggaacgcagatgc
2700
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
2760
ttgggtttgagcagagctctgagagagctttggtggaacgcagatgc
2820
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
2880
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
2940
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3000
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3060
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3120
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3180
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3240
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3300
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3360
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3420
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3480
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3540
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3600
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3660
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3720
aaccattacacctggagattgtaactgatcttcacctttataaatgagacagtgc
3780
tcctccgagtcataatgaccaagaagctgctggtcattgacacaacctatacctgat
3840
gatctacagcaagcataaagcagcataactaaacactactacgcaagaactac
3900
gataagttccttgtcagcttgtctgctagtgaagcagaaaaagccaccggt
3960
acgatatactctgtgacctaatctggtgctgggtgctttgtctgttcctgtc
4020
gaccccaccagaagagagcttcgtttgtctttgtttgtctttgtttttctgg
4080
attcactgtctttaagcttcggctcttttcttgtttttctgctttttctgtc
4140
gagagacagctgtgtgtgtgtgtttgtgtgtgtgtgtgtgtgtgtgtgtgt
4200
gacgctttgctacacccgatgtgtgtgctgtgtgtgtgtgtgtgtgtgtgtg
4260
cttggcagcattggtttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtg
4320
gacccagctatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtg
4380

<210> SEQ ID NO: 4
<211> LENGTH: 4425
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: VP64-activator construct

<400> SEQUENCE: 4

gccacacattg ctcacattg ccacctgctt ccacattg gcaccacag cgtggtgtgg
60
gacgtgtttta ccacgcgtcag ccccacaccag cccacacccag cccacaccag cccacacccag cccacacccac cccacacaccag cccacacacccag cccacacacccag
120
gatcgcacca gtaataac gatcgcacca gatcgcacca gatcgcacca gatcgcacca gatcgcacca gatcgcacca gatcgcacca gatcgcacca
180
gcgcacccac cgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtg
<p>| aatccagat tcgccgtgat gacctgcaca tcgaagaga acactcacttc gtggaacttc | 1440 |
| ggggaatctgg tgatgaaggg gccttggcgg aagctcctca tcgaagagag gactaacttt | 1500 |
| gatgaaatct gcctcacaag aagaagtgtt ccataaaacct ctctgtctga cgagtcacctc | 1560 |
| aagggtttata cgagatocca caaggtcaca ttcgtcgcag aagggatgag aagccagca | 1620 |
| ttctgtctgg gagaagcagaa gaagctatcc gttgaccttc ttcttaagac gaaacggaaa | 1680 |
| gttacgcgta aacaagcctgaa gaaagacatt ttcataaaga ttgaatgtttt cgacgcctg | 1740 |
| gaatccagac gtaggaggga tcggcttcaac gcactcctgg gaacgatcca cgactccctg | 1800 |
| aacaatctta aagcacagga cttcccgac aattgagagaa acagggacat tcctgagggc | 1860 |
| attgctctc cccctcaaggt gtgtgagagag ttggaagcac ccgacagtca gatcagctcag | 1920 |
| tcagctccat cctcagcaga caaagcagc gaaagatgca agagggcggc ctatacaggc | 1980 |
| ttggtggggggtgtcagaa aactgctcaag ggattcggag acaagcagag tcgaagagca | 2040 |
| atctctgatt ttcttattgc gcattgtgctt gcacaccgga cttccctgca gttgacttc | 2100 |
| gttgagctcg ctcctcggga cggagcactc cagaaagcc agaagttctgg ccagggggca | 2160 |
| agtctcttga aagggagctgg caatcgtttc gataaatcgg ggagctcatc gaagaatctg | 2220 |
| cagagaagtta gtaactcgcg ttcagagcgc aatcagcccg gggacaagaa gaaagattgg | 2280 |
| atgccttctcc agtgctgccg cagaagcccc aactccccgaa aggagcaagaa gaaacattgg | 2340 |
| gaaagatgag aaggtggtaga aagaggttata aagagtag caaagcagga cttcggagaa | 2400 |
| ccacagcttg aacaacctca gctccgagaa gagaagctct atoctgacta cctgcagaca | 2460 |
| ggaggaggca tcgcttgctg tcggaaactg gcatacattc ggtctctaga ccagaagctg | 2520 |
| gttgtcagct tcggccagatgt tttctctcaga gatgattctc ttgatatataagtt ctgttgaga | 2580 |
| agatcagata aagctcaggg gaaagtgatg acgtctccct cagaaagat gtcaagaaaaa | 2640 |
| atgaaaaatt atcgtcgggca gcctctggaac gccaatctga tccaaacagc gaaggctgat | 2700 |
| aacctgacta aggtgctcaag aggttgctctg ctctggaggg ttataaggcct ttcttactaa | 2760 |
| aggccaggtgg tggcacaag ccagccagcc aacagtgcttg ccccaatct ccgaggtgctg | 2820 |
| atgaaaaaca agtctcggta aatccgacaa ctagttccag aggtaagtct tattactctg | 2880 |
| agtctcagag tgcgcagttc tttcagaaag gcctttcact gtattataagtt gtagagacag | 2940 |
| aacaaaaacc aacctgctgca ttagtgttctc tcgaagtccg ttgcagggcgc gcagactctc | 3000 |
| aaaaaaatct acctgctgag agtctatatt ttgatagagat gactataagtt gtgtggatgt | 3060 |
| aagaaaaata tcggcagagtc gctggggagaa atacgcatgg ccaagcttca gatacctttt | 3120 |
| ttgcaggtat cctcagccgg cattctcaggg gatggcaggt cagagttcag | 3180 |
| aaggggacc acatatgaa aagggagaa acagggagaat ctggtgctgg caaggtctgg | 3240 |
| gatctctgga ctgctgcagc atgcgctctgg gttgaccgag tcaaacatcg taaaaagacc | 3300 |
| gaagaatcaga cggaggtgctc tctcaagggaa agttatctcc caagaaaggaa cagagcaag | 3360 |
| cgtatcgccg gcagaaaaa gttggagccc aagaaaaatgc gcgatttgcg tttctctaca | 3420 |
| gttcgtctcata gttgtctctt ttggcgccaa gtgcgagagag ggagatttca aaaaactcaaa | 3480 |
| agagctcaggg aactcgtcggg cattcaacac atggagagcgt cagagttcag aaaaaacc | 3540 |
| atcgttccac tcgaggggca aagatatata cgctgcagaa gaaacaccct catttaagct | 3600 |
| cccagagtct cttcttttga ctgtgaaaga ggctgaggac gaattgcctg cttgacggyg | 3660 |</p>
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaggtgctga aagggtaacga gcggcactg ccctttssaat acgtaatttt cttgtatcctg 3720</td>
<td></td>
</tr>
<tr>
<td>gcccagcact atgaaagagt caaagggtct ccgaaagata atgacgcaag gaagotgtgc 3780</td>
<td></td>
</tr>
<tr>
<td>gtggacacac aaaccaacaa cttgtagtag atacatgagg aataaagcga atttccccaa 3840</td>
<td></td>
</tr>
<tr>
<td>agagtgttcc tcggcgcgac taacactgtat agggtgtttt cttgatcaca aaacgccccg 3900</td>
<td></td>
</tr>
<tr>
<td>gataaagccac tcaaggggaca gcgcagaaac attaaccact tgtttacctct gcaccaacctg 3960</td>
<td></td>
</tr>
<tr>
<td>gcgcgggctcg cacgcttccac gtaacggtgac accaccatac acagaaaggg gcacacttcct 4020</td>
<td></td>
</tr>
<tr>
<td>acaacagagg ctcggtgaagc cacactgtatt ccacagtctaa ttacggaagg ctaacgaaca 4080</td>
<td></td>
</tr>
<tr>
<td>agaatgacac cttcctcgact cggtgacagac agcagaggcc attcccaagaa gaacaggagag 4140</td>
<td></td>
</tr>
<tr>
<td>gtggaggacca ggctggtagc ggcgctgac gcagttggcag atttggtagc cggatatgctg 4200</td>
<td></td>
</tr>
<tr>
<td>ggagagtgaag ccctcagatga ttggtgaccct gacatgtgctg gtctgggtgct cctgtgac 4260</td>
<td></td>
</tr>
<tr>
<td>ttgctgacgt cagctgtagc cagttgacgcc cttgaggtatt ctagcctggg ctagcctgtt 4320</td>
<td></td>
</tr>
<tr>
<td>aaacctcgcgg gcgcgcagac tcaaaaaag aagacaaaaag ctagatcaca aaaaagaga 4380</td>
<td></td>
</tr>
<tr>
<td>ggcattgacac ccaaaaaag aacagaagtga gtaacgctgag ctag 4425</td>
<td></td>
</tr>
</tbody>
</table>

**<210> SEQ ID NO 5**
**<211> LENGTH: 587**
**<212> TYPE: DNA**
**<213> ORGANISM: Artificial**
**<220> FEATURE:**
**<223> OTHER INFORMATION: MS2-activator construct**

**<400> SEQUENCE:**

cacaatggg acatgaagaa aagaagggcct gcggtgcctg tttcataagg gttcttaact 60

ttacctcggt cttgctgccg gaaatccgagg gaacggcggc gctgactgtc gecacgaagca 120

**<210> SEQ ID NO 6**
**<211> LENGTH: 661**
**<212> TYPE: DNA**
**<213> ORGANISM: Artificial**
**<220> FEATURE:**
**<223> OTHER INFORMATION: MS2-activator construct**

**<400> SEQUENCE:**

gccacacatgg gacccagaa aagagagggc gttggggtgag ctttctagaat ggccttcaca 60

tttacctcggt cttgctgccg gaaatccgagg caagctgaagt ctgacgcaagca 120

**<210> SEQ ID NO 7**
**<211> LENGTH: 696**
**<212> TYPE: DNA**
**<213> ORGANISM: Artificial**
**<220> FEATURE:**
**<223> OTHER INFORMATION: MS2-activator construct**

**<400> SEQUENCE:**

gggctgctgg gcgcgcagac tcaaaaaag aagacaaaaag ctagatcaca aaaaagaga 60

tttacctcggt cttgctgccg gaaatccgagg caagctgaagt ctgacgcaagca 120
acgaattccg acctgcatc tattgctaag gcaatgcaag gctctctaaa agatgggaac 360
ccgattaecg cagcaatgac agcaaatcag ggcactaagcg aggccacagtgg ccggtggag 420
gtcgacgcc ctgacgcgttc atgcgtgctg aatgtcgagcg acgactaagcg ccgcttgcttg 480
gcagctgctgg acggctgctg aatgctggagc tcaagaacgc cagagatccaa ggggctag 540
aaaagagca ggaagctgctg tcacaaagag aagagaggaag tagatccaa aaagaagaga 600
aaggtgaata ggggccgcata g 681

<210> SEQ ID NO 7
<211> LENGTH: 557
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: N52-activator construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (320) .. (399)
<223> OTHER INFORMATION: wherein N is G, A, T or C
<400> SEQUENCE: 7
tgtacaaaaa acgacaggtt taaagaacc acgtcataca cctggatcgg tagcaaggtc 60
ccgaggagc agggcctatt ttcctgtatt cttctatatt tgcataatacg tataaaggtc 120
gttacagaga taattgatag taatgctacc gtaaaccacc aagcattagc aacacaagcc 180
tgatgtagga atgataatatt ttctgggttag tttgcagttt taaaattgct tttaaaaatg 240
gtactataga tcggccagtc aacagctgctg tttcgggttt atatatctttg 300
ccggaaagc gaaacagcgn nnnnnnnnn nnnnnnnnn nnnnnnnttt gcgatacgca aagaaaagc 360
agtttataa atggctagct gcgtacatgc ttcacaaacgc gacccagcgcc gtcgctgcgc 420
agttgccagc tagaaacagc cattgctgca gtaacccggg gctttatttga cgaattgcagc 480
tcaagagta ccaattaaggc tagaaacagc gagttacacc cattgctgca gtcgctgcgc 540
agaaaatttt tttgcag 557

<210> SEQ ID NO 8
<211> LENGTH: 882
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Activation reporter construct
<400> SEQUENCE: 8
tattgataac atggataag tgcctccttc accccacagt ggaggccaggt aggctggttc 60
gctacggc ctcataaagc agagctccgt ttgtagaacc gcgggcatg gtagaatttc 120
ccacccagcc atcctacagc ttggagcagt gcaaccccgg gtaggagggag cctgctcacc 180
gttacaccc gggagaggtt cttcaaaaaa ttcggtcgtc ttaaagtctg gataggagggc 240
tccatgaaac gcaccaggtt ctagactcag gcgggaggcc gcgggagggc etcggagggc 300
accagccac ccgacgagtc gtagacccac gcgggccccg ccgcctctgctctggtgaccctc 360
gcctcccccc agtcaatgta cggcttcaag gtagactgtga agcaacgccg ccatcctccc 420
ccggactagc atgggtgcct cccccgggcc tcaagttgag agcggctatg gaaacttctgag 480
gaggccgctc tggtatccgtg gaccggcage tctctctgctg agagggccac gctgtctcag 540
-continued

aagtgtaag agcgcgcgac caaatctccc cccgcagcgc ccgtaatgac gaaagagacc 600
atggctgcttg agggctcaag ccagagcctg taccctcgcc aagcgtggtct gaaagccgag 660
atcacaacagg cctctgaagct gaagagcgcc ggcactacc tgggtagatg caggaaccatc 720
tacactgcca agagcctcggt gactacctgc ggtgtaaatc agttgcacac caagacggtac 780
atcactcccc caacagcagga ctacacccct gtggacagtc agcagcgcct cggagccgac 840
cacacccgtgt cctgtgaacc gcattgcagct ctgatc 882

<210> SEQ ID NO 9
<211> LENGTH: 882
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Activation reporter construct

<400> SEQUENCE: 9

taggataaac attggggtct tgggcgccct agggacagaa tgggaggtc agggctgtac 60
ggtggaggg cttataacg agcagtctgt tctgtaaccg tcaagtcgcc tggagatctc 120
gccacagtt gactcaagga tgaacacgat aaaaatcccg tgggccggct ggggccacc 180
tgagcaggg gggagaggtg catcaaaagag tcctagtgcg tcaaggtgcg catagagggc 240
tccatgaacgc gcacagaggt cgaagctgag ggcagagggg aggggcccgc ctagagggcc 300
eccagcagcc ccaagctggaa ggtgaccaag ggccgcgcct cggccctggc ctagcc 360
cgtcccccc caaggtccatgta cggctcccaag cgcagctgtga agccacccgc gcacatcccc 420
gatcacaagct agctgtcctt ccctggaggg cttcaagttgg agcgcgtgtg aaacttcgag 480
gaccgcctgc tggctgacgct gacccaggac tctccctgcg agacccggcg ctagctctac 540
aagtcgtaagg agcgcgcgttg aagagcgtcc gcagttcccgc cccccccttc gaaagagacc 600
atgggtgctgg agggctcaag ccagagctgg taccctcgcc aagcgtggtct gaaagccgag 660
atcacaacagg cctctgaagct gaagagcgcc ggcactacc tgggtagatg caggaaccatc 720
tacactgcca agagcctcggt gactacctgc ggtgtaaatc agttgcacac caagacggtac 780
atcactcccc caacagcagga ctacacccct gtggacagtc agcagcgcct cggagccgac 840
cacacccgtgt cctgtgaacc gcattgcagct ctgatc 882

<210> SEQ ID NO 10
<211> LENGTH: 912
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Specificity reporter library
<220> FEATURE:
<223> OTHER INFORMATION: wherein N is G, A, T or C
<220> FEATURE:
<223> OTHER INFORMATION: wherein N is G, A, T or C

<400> SEQUENCE: 10

taggataaac attggggtct tgggcgccct agggacagaa tgggaggtc agggctgtac 60
ggtggaggg cttataacg agcagtctgt tctgtaaccg tcaagtcgcc tggagatctc 120
gccacagtt gactcaagga tgaacacgat aaaaatcccg tgggccggct ggggccacc 180
<table>
<thead>
<tr>
<th>Sequence ID No.</th>
<th>Length</th>
<th>Type</th>
<th>Organism</th>
<th>Feature</th>
<th>Other Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ ID NO 11</td>
<td>23</td>
<td>DNA</td>
<td>Artificial</td>
<td>Feature</td>
<td>Target probe</td>
</tr>
<tr>
<td>SEQ ID NO 12</td>
<td>23</td>
<td>DNA</td>
<td>Artificial</td>
<td>Feature</td>
<td>Target probe</td>
</tr>
<tr>
<td>SEQ ID NO 13</td>
<td>23</td>
<td>DNA</td>
<td>Artificial</td>
<td>Feature</td>
<td>Target probe</td>
</tr>
<tr>
<td>SEQ ID NO 14</td>
<td>23</td>
<td>DNA</td>
<td>Artificial</td>
<td>Feature</td>
<td>Target probe</td>
</tr>
<tr>
<td>SEQ ID NO 15</td>
<td>23</td>
<td>DNA</td>
<td>Artificial</td>
<td>Feature</td>
<td>Target probe</td>
</tr>
</tbody>
</table>
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 15

gccacacca agcgatcaaa tgg

<210> SEQ ID NO 16
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 16

aaataataca ttctaaggta agg

<210> SEQ ID NO 17
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 17

gctactgggg aggctgaggc agg

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

<400> SEQUENCE: 18

tagcaatatca gtcacatattaa tgg

<210> SEQ ID NO 19
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 19

cctcatgtgat ccccccgtcct cgg

<210> SEQ ID NO 20
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 20

cgggagacg agtgaacgcg cgg

<210> SEQ ID NO 21
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 21

ttccttcct ctcctgtctg tgg

<210> SEQ ID NO 22
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 22

ttcctgcagaac gcctcttcgtg agg

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

<400> SEQUENCE: 23

aatgcaggtg cgcgcgcgag tgg

<210> SEQ ID NO 24
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 24

cctccgctc tacaatctg cgg

<210> SEQ ID NO 25
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 25

gagtoaagt cctctttact agg

<210> SEQ ID NO 26
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 26

agtgtgctgg atttggagata agg

<210> SEQ ID NO 27
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 27
cagcactca tcctccagtg agg 23

<210> Seq ID: NO 28
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 28
tcttacaacc agggagcatcag agg 23

<210> Seq ID: NO 29
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 29
cacaggctcg ccaaggatcc agg 23

<210> Seq ID: NO 30
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 30
gatggaagc tgaagaacac tgg 23

<210> Seq ID: NO 31
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 31
tgaaattgacg gcatacaatc agg 23

<210> Seq ID: NO 32
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 32
ccaagtc ccctggccttc tgg 23

<210> Seq ID: NO 33
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 33
cccagaasaa cagasccctga agg 23

<210> Seq ID: NO 34
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 34

aaggttga gacttttta ggg

<210> SEQ ID NO 35
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 35

atgtctgagt tttgggttag ggg

<210> SEQ ID NO 36
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 36

ggtccottga aaggggagta ggg

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

<400> SEQUENCE: 37

tggcaagtcta cctttgaaga tgg

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

<400> SEQUENCE: 38

ggcacagtgc cagagctctg tgg

<210> SEQ ID NO 39
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 39

taaaaataaa asactaaca ggg

<210> SEQ ID NO 40
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 40

tctgtggggg acctgcactg agg  23

<210> SEQ ID NO 41
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 41
ggcccagagt caaggctagt ggg  23

<210> SEQ ID NO 42
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 42
cacgacgaaaacctttota cgg  23

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

<400> SEQUENCE: 43
gttgaatgsa gcagcgtctag tgg  23

<210> SEQ ID NO 44
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 44
taagaacaga gcagtcgtcag tgg  23

<210> SEQ ID NO 45
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 45
tgtaaggtaa gcagagtag cgg  23

<210> SEQ ID NO 46
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 46
tgcacacca actcctgcac tgg

<210> SEQ ID NO 47
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 47

ttaccacact tcttcaagaa agg

<210> SEQ ID NO 48
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 48

gtggtgtgca ggtggtgctct ggg

<210> SEQ ID NO 49
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 49

cctccoccg gcctccoccg cgg

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

<400> SEQUENCE: 50

cacacacgc ccgagcag ggg

<210> SEQ ID NO 51
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 51

aggagcgccc ggctgctgat tgg

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

<400> SEQUENCE: 52

cacacacacc cacacagagat ggg

<210> SEQ ID NO 53
-continued

```
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 53

gaggaagcta aagagcaga ggg
```

```
<211> SEQ ID NO 54
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 54

atgagaattt caataaacgc agg
```

```
<211> SEQ ID NO 55
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 55

toccogcttg tgcccaaggg tgg
```

```
<211> SEQ ID NO 56
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 56

cagacaccca cccacatgca cgg
```

```
<211> SEQ ID NO 57
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 57

toccaatgta ctgggatatcg agg
```

```
<211> SEQ ID NO 58
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 58

tgatttaaaa gttggaaacg tgg
```

```
<211> SEQ ID NO 59
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
```
<223> OTHER INFORMATION: Target probe
<400> SEQUENCE: 59

ttcgttccc cacctagtct ggg 23

<210> SEQ ID NO 60
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 60

gattaactga gaatccaa ggg 23

<210> SEQ ID NO 61
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 61

cgcagaggg ggtggtctta agg 23

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

<400> SEQUENCE: 62

gtcacctcca cccacagtg ggg 23

<210> SEQ ID NO 63
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Reporter construct

<400> SEQUENCE: 63

ggcacatca gggacagat tgg 23

<210> SEQ ID NO 64
<211> LENGTH: 71
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target Oligonucleotide sequence

<400> SEQUENCE: 64

ataaatcttt atcgtcccc tccacccccg aatggggcca ctaggacag gattgggtcac 60
agaaaaagccc c 71

<210> SEQ ID NO 65
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target Oligonucleotide sequence
-continued

<400> SEQUENCE: 65

ggggccacta ggagacaggt

<210> SEQ ID NO 66
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Guide RNA

<400> SEQUENCE: 66

guuuauagc uagaaauagc aaguuuaaau aagcuagcu uguuuaaagc uugaaaaagu
ggacacagu cggugcuuuu

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

<400> SEQUENCE: 67
gtccotcca cccccactg cag

<210> SEQ ID NO 68
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 68
gtccotcca cccccacgt cca

<210> SEQ ID NO 69
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 69
gtccotcca cccccacgt cgg

<210> SEQ ID NO 70
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 70
tgtccotcc aacccacagc ggggocacta ggagacaggt tgtgacaga aa

<210> SEQ ID NO 71
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 71
tgtccccc ccacctagt ggggccacta gggacaggt tgtgtacaga aa  52

<210> SEQ ID NO 72
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 72

aaacccctcc accccccagt ggggccacta gggacaggt tgtgtacaga aa  52

<210> SEQ ID NO 73
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 73

tgtccocctcc ttttttcagt ggggccacta gggacaggt tgtgtacaga aa  52

<210> SEQ ID NO 74
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 74

caccgggttg tgtccocatcc tgtg  23

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

<400> SEQUENCE: 75

ggtgccccatcc tgtgtcaggtgtc tgtg  23

<210> SEQ ID NO 76
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 76

cccatcctgg tgcagcgtgaa cgg  23

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

<400> SEQUENCE: 77

gggccacagtt tacggtgtgtc cgg  23

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

<400> SEQUENCE: 78

cgcaataag agtcaccta cgg

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

<400> SEQUENCE: 79

cggaagttca ttcgacacc cgg

<210> SEQ ID NO: 80
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 80

cgcgcaagot ggcggtgcoc tgg

<210> SEQ ID NO: 81
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 81

gaccaggtag gcacacacc cgg

<210> SEQ ID NO: 82
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 82

gcgccgcacg tcagaccgaa tgg

<210> SEQ ID NO: 83
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 83

ggcggccacg cgtgcaccttg tgg

<210> SEQ ID NO: 84
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 94

```
taacaggtta atgtcagggc cgg
```

SEQUENCE: 85

```
aggtgacgct tcattgcatg agg
```

SEQUENCE: 86

```
cctcagggcg cgcttgccct cgg
```

SEQUENCE: 87

```
ggcagccgc agcgtgctgg tgg
```

SEQUENCE: 88

```
gatgatgacc cccctctcgg tgg
```

SEQUENCE: 89

```
gagatgatcg cccctctcgg
```

SEQUENCE: 90

```
```
gtgatgaccc gcccagttttc
<210> SEQ ID NO 91
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 91
gtccctcca cccacagtg ggg
<210> SEQ ID NO 92
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 92
gagatgatcg ccgccctttc tgg
<210> SEQ ID NO 93
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 93
gucccucua cccacagug
<210> SEQ ID NO 94
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 94
gucccucua cccacaguc
<210> SEQ ID NO 95
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 95
gucccucua cccacacag
<210> SEQ ID NO 96
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 96
gucccucua cccacacagc
<210> SEQ ID NO 97
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 97

guccccuca cccacagug 20

<210> SEQ ID NO 98
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 98

guccccuca cccacagug 20

<210> SEQ ID NO 99
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 99

guccccuca cccucacug 20

<210> SEQ ID NO 100
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 100

guccccuca cccacagug 20

<210> SEQ ID NO 101
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 101

guccccuca cgcacagug 20

<210> SEQ ID NO 102
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 102

guccccuca cgcacagug 20

<210> SEQ ID NO 103
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
OTHER INFORMATION: RNA target sequence

SEQUENCE: 103

guuccucca gcccacagug

SEQ ID NO 104
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: RNA target sequence

SEQUENCE: 104

guuccuccu cccacagug

SEQ ID NO 105
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: RNA target sequence

SEQUENCE: 105

guuccucgca cccacagug

SEQ ID NO 106
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: RNA target sequence

SEQUENCE: 106

guuccuccuc ccacacacac

SEQ ID NO 107
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: RNA target sequence

SEQUENCE: 107

guuccuccuc ccacacagug

SEQ ID NO 108
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: RNA target sequence

SEQUENCE: 108

guuccuccuc ccacugagug

SEQ ID NO 109
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: RNA target sequence

SEQUENCE: 109

guuccuccuc ccacucagug
GUCCUCUGCGGACAGUG

SEQ ID NO: 110
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial
OTHER INFORMATION: RNA target sequence

GUCCUCUGGGGACACAGUG

SEQ ID NO: 111
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial
OTHER INFORMATION: RNA target sequence

GUCCUCUGCUGCGGACAGUG

SEQ ID NO: 112
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial
OTHER INFORMATION: Target oligonucleotide sequence

GGGGCACTAGGAGGGGAGGG

SEQ ID NO: 113
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial
OTHER INFORMATION: RNA target sequence

GAGAUGACUGCCUCCUCUC

SEQ ID NO: 114
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial
OTHER INFORMATION: RNA target sequence

GAGAUGACUGCCUCCUCUUG

SEQ ID NO: 115
LENGTH: 20
TYPE: RNA
ORGANISM: Artificial
OTHER INFORMATION: RNA target sequence

GAGAUGACUGCCUCCUCUAC

SEQ ID NO: 116
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: DNA target sequence

<400> SEQUENCE: 116

gagauaucg ccuccucauc

20

<210> SEQ ID NO 117
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: DNA target sequence

<400> SEQUENCE: 117

gagauaucg ccuccugnucc

20

<210> SEQ ID NO 118
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: DNA target sequence

<400> SEQUENCE: 118

gagauaucg ccuccuacuc

20

<210> SEQ ID NO 119
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: DNA target sequence

<400> SEQUENCE: 119

gagauaucg ccuccauucuc

20

<210> SEQ ID NO 120
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: DNA target sequence

<400> SEQUENCE: 120

gagauaucg ccgcuucucuc

20

<210> SEQ ID NO 121
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: DNA target sequence

<400> SEQUENCE: 121

gagauaucg ccgcuucucuc

20

<210> SEQ ID NO 122
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 122

gagaugauc gcuccuucuc  20

<210> SEQ ID NO 123
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 123

gagaugauc gcuccuucuc  20

<210> SEQ ID NO 124
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 124

gagaugauc gcuccuucuc  20

<210> SEQ ID NO 125
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 125

gagaugau ggccuucuc  20

<210> SEQ ID NO 126
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 126

gagaugauc gcuccuucag  20

<210> SEQ ID NO 127
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 127

gagaugauc gcccuugau  20

<210> SEQ ID NO 128
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 128
-continued

gagaugac gccaacuc
<210> SEQ ID NO 129
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 129
gagaugac ggguuucuuc

<210> SEQ ID NO 130
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 130
gagaugac gcuccuucuuc

<210> SEQ ID NO 131
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 131
gagaugac ccccuucuuc

<210> SEQ ID NO 132
<211> LENGTH: 23
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 132
gagatgatcg cccctcttc tgg

<210> SEQ ID NO 133
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 133
ggggacacua ggcacagau

<210> SEQ ID NO 134
<211> LENGTH: 19
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 134
ggccacacug ggcacagau

<210> SEQ ID NO 135
<211> LENGTH: 18
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 135

ggccacuagg gacaggau

<210> SEQ ID NO 136
<211> LENGTH: 17
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 136

gccacuagg acaggau

<210> SEQ ID NO 137
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 137

gagaugacg cccucuccu

<210> SEQ ID NO 138
<211> LENGTH: 16
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 138
gaugaugcc ccuuccu

<210> SEQ ID NO 139
<211> LENGTH: 15
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 139
gaugcgcccu ucucu

<210> SEQ ID NO 140
<211> LENGTH: 11
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 140
gccccuuucu c

<210> SEQ ID NO 141
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 141

gtccctccta ccccacagtg c
  21

<210> SEQ ID NO 142
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: (5)...(10)
  <223> OTHER INFORMATION: wherein N is G, A, T or C

<400> SEQUENCE: 142
tgtnnnnnnn accc
  14

<210> SEQ ID NO 143
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
  <223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 143
tgtcaaaaaa accc
  14

<210> SEQ ID NO 144
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
  <223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 144
tgtnccgggg accc
  14

<210> SEQ ID NO 145
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
  <223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 145
tgtcaaaaaa accc
  14

<210> SEQ ID NO 146
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
  <223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 146
tgtnccgggg accc
  14

<210> SEQ ID NO 147
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 147

tgcccccccc accc

14

<210> SEQ ID NO 148
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 148

tgctttttttt accc

14

<210> SEQ ID NO 149
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 149

tgcccccccc accc

14

<210> SEQ ID NO 150
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 150

tgctttttttt accc

14

<210> SEQ ID NO 151
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 151

ggatcgctgtg tcccgagct ggg

23

<210> SEQ ID NO 152
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 152

gttasgcttg cttggttct ggg

23

<210> SEQ ID NO 153
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 153
gaggccacata gggcagggat tgg

<210> SEQ ID NO 154
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 154
cctccctgcttcctgatatt ggg

<210> SEQ ID NO 155
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 155
tgctgccagc tgggggacac agg

<210> SEQ ID NO 156
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 156	aggaccagag ccacattacg cgg

<210> SEQ ID NO 157
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 157
gtcaacatc ctgctcctag tgg

<210> SEQ ID NO 158
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 158
agacccacat tccagagact agg

<210> SEQ ID NO 159
<211> LENGTH: 75
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 159
gggatctcgt gtcccctgac tggaggaccct ttatatctcc aaggccggttt aatgtggtctc

tggttcctgg tactt
<210> SEQ ID NO 160
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE: 
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 160

ggatccctgt gtccegcagc ttgacacc acctcatcttc acggcctgtt aatgtgttc
60
tgggt actt
69

<210> SEQ ID NO 161
<211> LENGTH: 113
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE: 
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 161

ggatccctgt gtccegcagc ttgacacc acctcatcttc acggcctgtt cctggtgacc
60
acccatat tccagggcag ggcgc gttgtttaaa tgtgctcttg gttctgggta ctt
113

<210> SEQ ID NO 162
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE: 
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 162

ggatccctgt gtccegcagc ttgacacc actt
34

<210> SEQ ID NO 163
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE: 
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 163

ggatccctgt gtccegcagc ttgacacc tttacttg gatct
47

<210> SEQ ID NO 164
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE: 
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 164

ggatccctgt gatct
17

<210> SEQ ID NO 165
<211> LENGTH: 93
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE: 
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 165

acggcccggtt aatgtgtgtc tgggtgttgg ttcttttatc tgtcccctcc acccccacagtt 60
<210> SEQ ID NO 166
<211> LENGTH: 83
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 166
agggccccgtt aatgaatgt gctctggttc tgggtacttt tattctgcccc tctccacccc
  60
cagttgggccc actagacaga aaaa
  83

<210> SEQ ID NO 167
<211> LENGTH: 76
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 167
agggccccgtt aatgtggtct tgtttcggg tacttttct ctcocccccag tggggccact
  60
gattgtgac agaaaa
  76

<210> SEQ ID NO 168
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 168
agggccccgtt caggttggt gacagaaaa
  29

<210> SEQ ID NO 169
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 169
agggccccgtt aatgtggcga tgtgtgacag aaaa
  34

<210> SEQ ID NO 170
<211> LENGTH: 63
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 170
agggccccgtt aatgtgccgtc tgtttctggg tactttttct ctcocccgat tgtgtgacaga
  60
aaa
  63

<210> SEQ ID NO 171
<211> LENGTH: 84
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 171
agggcggtt aatgtgggtc tggttctggg tactttttac tgccoccctc aoccacagct 60
ggggacagga ttggtgacag aaaa 84

<210> SEQ ID NO 172
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 172
agggcggtt aatgtgggtg cagaaaa 27

<210> SEQ ID NO 173
<211> LENGTH: 165
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 173
agggcggtt aatgtgggtc tggttctggc tacattttac tgccoccctc aoccacagggg 60
acagctctgc ccctcacaac caggacaggg attggtgaca gaaaa 105

<210> SEQ ID NO 174
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 174
agggcggtt aatgtgggtc tggttctggg tacatttttac tgccoccctc aaccataggg 60
acaggatgg tgacagaaaa 80

<210> SEQ ID NO 175
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 175
cccacagtgg ggcacactgg gacaggatgg gtcagcagaa acggcctata ccc 53

<210> SEQ ID NO 176
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 176
cccacagtgg ggcacactcc c 22

<210> SEQ ID NO 177
<211> LENGTH: 96
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 177
cccacagtgg ggcacagt agaaaagccc cattcctaggg cctcccccct ccttaggctt 60
tctccttcct agtctctctga tttaggtct aacccc 96

<210> SEQ ID NO 178
<211> LENGTH: 94
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 178
cccacagtgg ggcacagt gacaggattg gtgacagaaaa gcoccccatcc ttaggcctg 60
tctccttcct gtcctgtata ttaggctt aaacc 94

<210> SEQ ID NO 179
<211> LENGTH: 62
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 179
cccacagtgg ggcacacott ggcocctcct cttcctagtt tctgatatg gggttaacc 60
c 62

<210> SEQ ID NO 180
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 180
cccacagtgg ggcacagt gatattgggt cttaaacc 38

<210> SEQ ID NO 181
<211> LENGTH: 94
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: target oligonucleotide sequence

<400> SEQUENCE: 181
cccacagtgg ggcacagtgg gacagggattg gtgacagaaaa gcoccccatcc ttaggcctg 60
tctccttcct gtcctgtata ttaggctt aaacc 94

<210> SEQ ID NO 182
<211> LENGTH: 65
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 182
cccacagtgg ggcacagtgg gacagggctcc ttcctctctta gtcctgtata ttagggtcta 60
aacccc 65
1. A method of modulating expression of a target nucleic acid in a cell comprising
providing to the cell a guide RNA complementary to the target nucleic acid sequence including a transcriptional activator or repressor domain as a fusion protein for modulating target nucleic acid expression in vivo, providing to the cell a nuclease null Cas9 protein that interacts with the guide RNA and binds to the target nucleic acid sequence in a site specific manner, wherein the guide RNA including the transcriptional activator or repressor domain as a fusion protein and the Cas9 protein co-localize to the target nucleic acid sequence and wherein the transcriptional activator or repressor domain modulates expression of the target nucleic acid.

2. The method of claim 1 wherein the guide RNA including the transcriptional activator or repressor domain as a fusion protein is provided to the cell by introducing to the cell a nucleic acid encoding the guide RNA including the transcriptional activator or repressor domain as a fusion protein, wherein the Cas9 protein is provided to the cell by introducing to the cell a nucleic acid encoding the Cas9 protein, and wherein the cell expresses the guide RNA including the transcriptional activator or repressor domain as a fusion protein and the Cas9 protein.

3. The method of claim 1 wherein the cell is a eukaryotic cell.

4. The method of claim 1 wherein the cell is a yeast cell, a plant cell or a mammalian cell.

5. The method or claim 1 wherein the cell is a human cell.

6. The method of claim 1 wherein the guide RNA is between about 10 to about 250 nucleotides.

7. The method of claim 1 wherein the guide RNA is between about 20 to about 100 nucleotides.

8. The method of claim 1 wherein the guide RNA is between about 100 to about 250 nucleotides.
9. The method of claim 1 wherein the target nucleic acid is genomic DNA, mitochondrial DNA, viral DNA or exogenous DNA.

10. A method of modulating expression of a target nucleic acid in a cell comprising providing to the cell a guide RNA complementary to viral DNA including the target nucleic acid sequence, wherein the guide RNA includes a transcriptional activator or repressor domain as a fusion protein for modulating target nucleic acid expression in vivo, providing to the cell a nuclease null Cas9 protein that interacts with the guide RNA and binds to the target nucleic acid sequence in a site specific manner, wherein the guide RNA including the transcriptional activator or repressor domain as a fusion protein and the Cas9 protein co-localize to the target nucleic acid sequence and wherein the transcriptional activator or repressor domain modulates expression of the target nucleic acid.

11. The method of claim 10 wherein the guide RNA including the transcriptional activator or repressor domain as a fusion protein is provided to the cell by introducing to the cell a nucleic acid encoding the guide RNA including the transcriptional activator or repressor domain as a fusion protein, wherein the Cas9 protein is provided to the cell by introducing to the cell a nucleic acid encoding the Cas9 protein, and wherein the cell expresses the guide RNA including the transcriptional activator or repressor domain as a fusion protein and the Cas9 protein.

12. The method of claim 10 wherein the cell is a eukaryotic cell.

13. The method of claim 10 wherein the cell is a yeast cell, a plant cell or a mammalian cell.

14. The method or claim 10 wherein the cell is a human cell.

15. The method of claim 10 wherein the guide RNA is between about 10 to about 250 nucleotides.

16. The method of claim 10 wherein the guide RNA is between about 20 to about 100 nucleotides.

17. The method of claim 10 wherein the guide RNA is between about 100 to about 250 nucleotides.

* * * * *