Abstract

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. 1A
CAS9 nuclease.

FIG. 1B
MS2 VP64
gRNA + aptamer2X

FIG. 1C
Min Promoter
1d:Tomato
gRNA Target
GGGCACTAGGAGAT TGG
TALE Target
GTCCCTCCACCCACAGTG GGG
TALE Target2
GGGCACTAGGAGAT TGG
FIG. 1D-2
FIG. 1E-1

Cas9 nuclease-MS2VP64+ gRNA1-aptamer2X

Cas9 nuclease-MS2VP64+ gRNA2-aptamer2X

Cas9 nuclease-MS2VP64+ gRNA1+gRNA2

TALEVP64

Reporter

SSC-A

PE-A

(x1,000)

(102, 103, 104, 105)
FIG. 1E-2

TALE<sup>VP64</sup>^5

Cas<sub>9 nuclease+</sub> MS2<sup>VP64</sup>+gRNA2

Cas<sub>9 nuclease+</sub> MS2<sup>VP64</sup>+gRNA1-aptamer2X

Cas<sub>9 nuclease+</sub> MS2<sup>VP64</sup>+gRNA2-aptamer2X

Reporter 2

(x 1,000)

SSC-A

51.1%

10.6%

10.2%

40.5%
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. 2B

Cas9_{m4}^{VP64} + gRNA

FIG. 2C

Cas9_{m4}^{VP64} + gRNA: one base mismatch

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

relative expression

relative expression

G G G G C A C T A G G A C A G A T T G G
FIG. 2D

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

FIG. 2E

**18mer TALE**<sup>VP64</sup>
FIG. 2F
18mer TALE$^{VP64}$: one base mismatch
match, mismatch=A, mismatch=C, mismatch=G, mismatch=T

FIG. 2G
18mer TALE$^{VP64}$: two base mismatch
<table>
<thead>
<tr>
<th>Name</th>
<th>Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas9</td>
<td>wild-type</td>
</tr>
<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>
</tbody>
</table>

**FIG. 4A**
FIG. 4C

- Cas9_m4 + gRNA
- Cas9_m3 + gRNA
- Plain 293Ts

[Diagram showing data points and percentages for different samples]
FIG. 5B-1

AGCCACGGUGAAAAGUUC
UCGGUGC

3'
gRNA scaffold

matches target sequence

GAA
VAAVAAVCAA
GAAVAAVACAAA

5' - GGGCCCTAGGGACGAT
AAGCUAGGCUUACAA
FIG. 5B-2

%age rate of HR

Plain, donor, gRNA, gRNA5+1bp, gRNA5+5bp, gRNA5+10bp, gRNA5+20bp, gRNA5+40bp, gRNAmiddle+1bp, gRNAmiddle+5bp, gRNAmiddle+10bp, gRNAmiddle+20bp, gRNAmiddle+40bp, gRNA3+1bp, gRNA3+5bp, gRNA3+10bp, gRNA3+20bp, gRNA3+40bp

Donor + hCas9
FIG. 6C

**Fold Change in Endogenous Gene Expression**

- **Cas9_nuclease\_VP64**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain</td>
<td>0.00</td>
</tr>
<tr>
<td>Negative</td>
<td>0.00</td>
</tr>
<tr>
<td>gRNAs1.7</td>
<td>0.00</td>
</tr>
<tr>
<td>gRNAs8.14</td>
<td>0.00</td>
</tr>
<tr>
<td>gRNAs15.21</td>
<td>0.00</td>
</tr>
<tr>
<td>gRNAs1.21</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Error Bars** indicate standard deviation.
FIG. 7C

Fold Change in Endogenous Gene Expression

- plain
- negative
- gRNA1-5
- gRNA1-2XMS2
- gRNA6-10
- gRNA6-10_2XMS2

Cas9_{nuclease} + MS2_{VP64}
FIG. 8A

Construct library

+ control TALE or TALE site Cas9 site 

Transfect & induce

Sequence cDNAs

Sample barcodes

Library barcodes

Analyze sequences

Analyze 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>
</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)

Combine tables

Binding site count

<table>
<thead>
<tr>
<th>Binding site</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>552812</td>
</tr>
<tr>
<td>Site 2</td>
<td>38601</td>
</tr>
<tr>
<td>Site 3</td>
<td>22556</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. 8B**

Theoretical distribution and observed distribution for 18mer TALE-TF library.

**FIG. 8C**

Comparison of Control-TF and 18mer TALE-TF for the number of mutations.

18mer TALE-TF activation vs. number of target site mutations.
FIG. 9A

Cas$_{m4}^{VP64+}$gRNA

![Box plot showing relative expression vs. number of mutations](image)

FIG. 9B

Cas$_{m4}^{VP64+}$gRNA: one base mismatch

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

![Bar chart showing relative expression](image)
FIG. 9C

\[ \text{Cas}^{m4}\text{VP64}^{gRNA}: \text{two base mismatch} \]
FIG. 9D

...GTCCCCTCCACCCACAGTG CRR.... gRNA Target

repair donor

target locus

\[ \text{DNA break stimulates HR} \]

donor + Cas9 + gRNA

\[ \begin{align*}
\text{GTCCCCTCCACCCACAGTG CAG} \\
\text{GTCCCCTCCACCCACAGTG CAA} \\
\text{GTCCCCTCCACCCACAGTG CGG} \\
\text{empty}
\end{align*} \]

\%age rate of HR
FIG. 10A-2

- Percentage rate of HR
- 4 mutations
- 1 mutation
- Wild-type
FIG. 10B
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
**FIG. 13A**

**Cas9<sub>N</sub>-VP64 +gRNA2**

<table>
<thead>
<tr>
<th># of mutations</th>
<th>Relative Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>N.S.</td>
</tr>
<tr>
<td>1</td>
<td>***</td>
</tr>
<tr>
<td>2</td>
<td>***</td>
</tr>
<tr>
<td>3</td>
<td>***</td>
</tr>
<tr>
<td>4</td>
<td>***</td>
</tr>
<tr>
<td>5</td>
<td>***</td>
</tr>
<tr>
<td>6</td>
<td>***</td>
</tr>
<tr>
<td>7</td>
<td>***</td>
</tr>
<tr>
<td>8</td>
<td>***</td>
</tr>
</tbody>
</table>

**FIG. 13B**

**Cas9<sub>N</sub>-VP64 +gRNA2: one base mismatch**

match, mismatch=A, mismatch=C, mismatch=G, mismatch=T
FIG. 13C

Cas9<sub>N</sub>-VP64 +gRNA2: two base mismatch

FIG. 13D

Cas9<sub>N</sub>-VP64 +gRNA3

Relative expression

# of mutations
FIG. 13E

\( \text{Cas9}_{N-VP64} + \text{gRNA3}: \text{one base mismatch} \)

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

relative expression

FIG. 13F

\( \text{Cas9}_{N-VP64} + \text{gRNA3}: \text{two base mismatch} \)

match
FIG. 15A

gRNA target

...GTCCCCTCCACCCACAGTG GGG... 

repair donor 

|---|---|

target locus

|---|---|

DNA break stimulates HR
FIG. 15B-1

matches target sequence

AGCCGCCGUGUAAAAGUUC
UCGGGUCC

GUCCCUCCACCCACAGUG

relative rate of targeting

0 1 2 3
FIG. 15C

- gRNA target: \( ...GAGATGATGC\ldots\)
- Repair donor
- Target locus

DNA break stimulates HR

FIG. 15D-1

- Target sequence: \(5' \text{GAGAUGAUCGCCCUUCUUC} \text{AAGCUAGCUCCGUUAUCAA}3'\)

Relative rate of targeting:

- 0
- 1
- 2

- \(\text{GAGAUGAUCGCCCUUCUUC}\)
FIG. 16B-1

matches target sequence

5' - GGGGCCACUAGGGACAGGAU  AAGCUAGUCCGUUAUCAA

3' - UUUU
FIG. 16B-2

<table>
<thead>
<tr>
<th>Distance</th>
<th>Sequence</th>
<th>Relative Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>GGGGCCACUAGGGACAGGAU</td>
<td>1</td>
</tr>
<tr>
<td>1bp</td>
<td>GGGCCACUAGGGACAGGAU</td>
<td>1.5</td>
</tr>
<tr>
<td>2bp</td>
<td>GGCCACUAGGGACAGGAU</td>
<td>2</td>
</tr>
<tr>
<td>3bp</td>
<td>GCCACUAGGGACAGGAU</td>
<td>3</td>
</tr>
</tbody>
</table>

relative rate of targeting
FIG. 16C

gRNA target

GAGATGCACCCTTC

DNA break stimulates HR

target locus

repair donor
FIG. 16D-1

matches
target sequence

5' - GAGAUGAUCGCCCUUCUUC

3' - AAGGCUAGUCCGUUAUCAA
FIG. 16D-2

<table>
<thead>
<tr>
<th>Relative Rate of Targeting</th>
<th>GAGAUGCCUUCUUC</th>
<th>GAUGACUUUUCUUC</th>
<th>GAUCGCUUUCUUC</th>
<th>GCCCUUUCUUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2bp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5bp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9bp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIG. 17B

donor + Cas9 + gRNA

GTCCCTCCACCACAGT CAG
GTCCCTCCACCACAGT CAA
GTCCCTCCACCACAGT CGG
empty

% age rate of HR

0 0.5 1 1.5 2 2.5 3 3.5 4
FIG. 19A

Random hexamer library of binding sites
(TGTCNNNNACCC)

24bp barcode
(A/C/G)

Target Site

Shared Site

Min Promoter

Barcode

Tomato
Fig. 19B-2

TGTCTTTTATCCC  
mis-match

TGTGGGGGGACCC  
mis-match

Normalized expression

Sequence mismatches

0 1 2 3 4 5 6
FIG. 20A

<table>
<thead>
<tr>
<th>gRNA</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAVS1_s1</td>
<td>GGATCCTGTGTCCCGAGCT GGG</td>
</tr>
<tr>
<td>AAVS1_s2</td>
<td>GTTAATGGGCTCTGGTTCT GGG</td>
</tr>
<tr>
<td>AAVS1_s3</td>
<td>GGGGCCACTAGGGACAGGAT TGG</td>
</tr>
<tr>
<td>AAVS1_s4</td>
<td>CTTCCATGCTCTGATATG GGG</td>
</tr>
<tr>
<td>AAVS1_as1</td>
<td>TGGTCCAGCTCGGGGACAC AGG</td>
</tr>
<tr>
<td>AAVS1_as2</td>
<td>AGAACCAGAGCCACATTAAC CGG</td>
</tr>
<tr>
<td>AAVS1_as3</td>
<td>GTCACCAATCCTGTCCCTAG TGG</td>
</tr>
<tr>
<td>AAVS1_as4</td>
<td>AGACCCTATATCAGGAGACT AGG</td>
</tr>
</tbody>
</table>
FIG. 21C

as3 + s4

CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
CCCACAGTGGGCCACTA---------GG-GACAGGATTGTTGAC-AGAAAAGCCCAT---------------------------------ACCC
RNA-GUIDED TRANSCRIPTIONAL REGULATION

RELATED APPLICATION DATA

This application is a continuation of PCT application no. PCT/US2014/040468, 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 entirety.

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 complexed DNA 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 or with the 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 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, 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 certain aspects, 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 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.

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 which links the tracrRNA spacer 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 1 base truncation at the 5′ end of the spacer sequence. According to one aspect, the guide RNA includes a truncated spacer sequence having a 2 base truncation at the 5′ end of the spacer sequence. According to one aspect, the guide RNA includes a truncated spacer sequence having a 3 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 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. Accordingly 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 nucleic acids.

[0014] According to certain aspects, the tracr sequence may include between about 60 to about 500 nucleic acids. Accordingly to certain aspects, the tracr sequence may include between about 64 to about 500 nucleic acids. Accordingly to certain aspects, the tracr sequence may include between about 65 to about 500 nucleic acids. According to certain aspects, the tracr sequence may include between about 66 to about 500 nucleic acids. According to certain aspects, the tracr sequence may include between about 67 to about 500 nucleic acids. According to certain aspects, the tracr sequence may include between about 68 to about 500 nucleic acids. According to certain aspects, the tracr sequence may include between about 69 to about 500 nucleic acids. According to certain aspects, the tracr sequence may include between about 70 to about 500 nucleic acids. According to certain aspects, the tracr sequence may include between about 71 to about 500 nucleic acids.

[0015] According to certain aspects, the tracr sequence may include between about 60 to about 200 nucleic acids. Accordingly to certain aspects, the tracr sequence may include between about 64 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 65 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 66 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 67 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 68 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 69 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 70 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 71 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 72 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 73 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 74 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 75 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 76 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 77 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 78 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 79 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 80 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 81 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 82 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 83 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 84 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 85 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 86 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 87 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 88 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 89 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 90 to about 200 nucleic acids. According to certain aspects, the tracr sequence may include between about 91 to about 200 nucleic acids. Accordingly to certain aspects, the tracr sequence may include between about 92 to about 200 nucleic acids. Accordingly to certain aspects, the tracr sequence may include between about 93 to about 200 nucleic acids. Accordingly to certain aspects, the tracr sequence may include between about 94 to about 200 nucleic acids. Accordingly to certain aspects, the tracr sequence may include between about 95 to about 200 nucleic acids. Accordingly to certain aspects, the tracr sequence may include between about 96 to about 200 nucleic acids. Accordingly to certain aspects, the tracr sequence may include between about 97 to about 200 nucleic acids. Accordingly to certain aspects, the tracr sequence may include between about 98 to about 200 nucleic acids. Accordingly to certain aspects, the tracr sequence may include between about 99 to about 200 nucleic acids. Accordingly to certain aspects, the tracr 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. Accordingly 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. Accordingly 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. Accordingly 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. Accordingly to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid. Accordingly to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid to treat a disease or detrimental condition. Accordingly 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. Accordingly to one aspect, the one or more RNAs is a tracrRNA-cRNA 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. Accordingly 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 trncRNA-cRNA 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 trncRNA-cRNA 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 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 are on different strands of the
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 nucleases 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 tracrRNA-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 NO: 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 Cas9N, 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. FIG. 2D
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 is on 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 HNH 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 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 NGOs 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). 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.

[0067] FIG. 11A depicts designed guide RNAs. FIG. 11B depicts percentage rate of non-homologous end joining for various guide RNAs.

[0068] FIG. 12A depicts the Sox2 gene. FIG. 12B depicts the Nanog gene.

[0069] FIGS. 13A-13F depict the targeting landscape of two additional Cas9-gRNA complexes.

[0070] 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).

[0071] FIGS. 15A-15D depict gRNA2 (FIG. 15A-B) and gRNA3 (FIG. 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.

[0072] FIGS. 16A-16D depict a nuclease assay of two independent gRNA that were tested: gRNA1 (FIG. 16A-B) and gRNA3 (FIG. 16C-D) bearing truncations at the 5' end of the spacer. Sequences are set forth as SEQ ID NOs:132-134.

[0073] FIGS. 17A-17B depict a nuclease mediated HR assay that shows the PAM for the S. pyogenes Cas9 is NGOs and also NAG. Sequences are set forth as SEQ ID NOs:139-141.

[0074] 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.

[0075] FIGS. 19A-19C depict a comparison of TALE monomer specificity versus TALE protein specificity. Sequences are set forth as SEQ ID NOs:142-150.

[0076] FIGS. 20A-20B depict data related to off-set nicking. Sequences are set forth as SEQ ID NOs:151-158.

[0077] FIGS. 21A-21C depict off-set nicking and NHEJ profiles. Sequences are set forth as SEQ ID NOs:159-184.

DETAILED DESCRIPTION

[0078] 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 nuclease-null DNA binding protein. According to this aspect, the nuclease-null DNA binding protein may result from the alteration or modification of a DNA binding protein having nuclease activity. Such DNA binding proteins having nuclease activity are known to those of skill in the art, and include naturally occurring DNA binding proteins having nuclease 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.

Exemplary DNA binding proteins having nuclease activity function to nick or cut double stranded DNA. Such nuclease activity may result from the DNA binding protein having one or more polypeptide sequences exhibiting nuclease 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 nuclease activity known to those of skill in the art include the MerA-HINII 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-HINII 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 nuclease activity. Such alteration or modification includes altering one or more amino acids to inactivate the nuclease activity or the nucleosome domain. Such modification includes removing the polypeptide sequence or polypeptide sequences exhibiting nuclease activity, i.e. the nucleosome domain, such that the polypeptide sequence or polypeptide sequences exhibiting nuclease activity, i.e. nucleosome domain, are absent from the DNA binding protein. Other modifications to inactivate nuclease activity will be readily apparent to one of skill in the art based on the present disclosure. Accordingly, a nuclease-null DNA binding protein includes polypeptide sequences modified to inactivate nuclease activity or removal of a polypeptide sequence or sequences to inactivate nuclease activity. A nuclease-null DNA binding protein retains the ability to bind to DNA even though the nuclease 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 of all of the nucleosome sequences exhibiting nuclease 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 nuclease activity inactivated.

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 an RNA guided DNA binding protein nickase.

An exemplary DNA binding protein is an RNA guided DNA binding protein of a Type II CRISPR System which lacks nuclease activity. An exemplary DNA binding protein is a Cas9 protein nickase. An exemplary DNA binding protein is a Cas9 protein nickase. An exemplary DNA binding protein is a Cas9 protein nickase.
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 an additional aspect, Cas9 proteins are provided where the RuvC nuclease domain and the HNH1 nuclease domain are inactivated. According to an additional aspect, Cas9 proteins are provided insofar as the RuvC nuclease domain and the HNH1 nuclease domain are inactivated. According to an additional aspect, a Cas9 nuclease is provided where either the RuvC nuclease domain or the HNH1 nuclease domain is inactivated, thereby leaving the remaining nuclease domain active for nuclelease activity. In this manner, only one strand of the double stranded DNA is cut or nicked.

According to an additional aspect, nuclease null 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 337, 816-821 (2012). According to an additional aspect, the amino acids include D839 and N863. According to one aspect, one or more 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 D10, H840, D839 and H863 are substituted with alanine. According to one aspect, a Cas9 protein having one or more 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 nuclelease activity is absent or substantially absent within levels of detection. According to this aspect, nuclease activity for a Cas9N may be undetectable using known assays, i.e., below the level of detection of known assays.

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 be guided by the RNA. 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.

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 HNH1 nuclease 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.

[0088] 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 nuclelease activity: Cas9 wild-type; Cas9_{m1} which substitutes alanine for D10; Cas9_{m2} which substitutes alanine for D10 and alanine for I1840; Cas9_{m3} which substitutes alanine for D10, alanine for I1840, and alanine for D839; and Cas9_{m4} which substitutes alanine for D10, alanine for I1840, alanine for D839, and alanine for N863.

[0089] As shown in FIG. 4B, the Cas9 mutants: m3 and m4, and also their respective fusions with VP64 showed undetectable nuclease activity upon deep sequencing at targeted loci. The plots show the mutation 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.

[0090] 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 Cas9N or to the gRNA.

[0091] 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.

[0092] 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.

[0093] 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 FIG. 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.

[0094] To test 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.

[0095] 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 test 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.

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 Takahashi et al., Cell 131 861-872 (2007) hereby incorporated by reference in its entirety) or directly via qPCR of the endogenous genes. 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, methods of multiplex gene regulation are provided using the same or similar Cas9N.

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.

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 nuclelease activity, as with the naturally occurring Cas9 protein. Accordingly, such DNA binding systems need not have nuclelease activity inactivated. One exemplary DNA binding system is TALE. As a genome editing tool, usually TALE-FokI dimers are used, and for genome regulation TALE-VP64 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 promotor 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.

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).

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).

Results were confirmed using targeted experiments in a nuclease assay which is the subject of FIG. 10A-C directed to evaluating the targeting landscape of 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.

FIG. 8A-C is directed to high level specificity analysis processing flow 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 a 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. 21B 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.

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 nuclease mediated HR assay that the predicted PAM for the S. pyogenes Cas9 is NGG and also NAG.

[0108] 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 nickers 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.

[0109] FIG. 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 frameshifts 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 nicked 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.

[0110] FIG. 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.

[0111] 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 nuclease-null Cas9 protein or proteins and a transcriptional regulator protein or domain. Together, a guide RNA, a nuclease-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 nuclease-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.

[0112] 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 type. 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.

[0113] 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 1

Cas9 Mutants

Sequences homologous to Cas9 with known structures 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 toolkit.tuebingen.mp.gov/hhpred), 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 renders 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). This sequence alignment covers the previously reported mutation D10A which eliminates function of the RuvC domain in Cas9. To confirm this as an appropriate mutation, the metal binding residues were determined as before. In 4EP4, D7 helps to coordinate a magnesium ion. This position has sequence homology corresponding to Cas9 D10, confirming that this mutation helps remove metal binding, and thus catalytic activity from the Cas9 RuvC domain.

Example ii

Plasmid Construction

The Cas9 mutants were generated using the Quikchange kit (Agilent 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 oligo-nucleotides into the gRNA cloning vector (plasmid #41824). The vectors for the HR reporter assay involving a broken GFP were constructed by ligation PCR assembly of the GFP sequence bearing the stop codon and appropriate fragment assembled into the pEGFP 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 nucleic assays were as follows: 0.4×10⁶ 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 ~1×10⁶ cells was extracted using DNAeasy kit (Qiagen). For these FACS were conducted to amplify the targeting 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 NEHEJ efficiencies.

For transfections involving transcriptional activation assays: 0.4×10⁶ 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.
[0123] 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 Cas9N-VP64 plasmid, 2 μg gRNA and 0.25 μg of reporter library; or (2) 2 μg TALE-TF plasmid and 0.25 gm 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-performed using Superscript III (Invitrogen). Libraries for next-generation sequencing were generated by targeted per amplification of the transcript tags.

Example IV

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

[0124] 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:

[0125] 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 5 bp non-overlapping paired end reads on an Illumina HiSeq.

Construct Library Sequence Processing:

[0126] 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, 3^n 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 (see 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:

[0127] Alignment:

[0128] SeqPrep (downloaded from world wide website github.com/jsjohann/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, 3^n 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 to a separate binding site region.

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

[0129] Custom perl was used to generate these tables from the validated construct library sequences (FIG. 8A, 4^n 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 found 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, non-shared, 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:

[0130] 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 obtains from the fact that the number of separate aggregated tag counts for n mismatching positions grew with the number of combinations of mismatching positions (equal to

\[ \binom{n}{k} \]

and so dramatically increases with n; thus the averages of aggregated tag counts for different numbers n of mismatches (shown in FIGS. 2b, 2e, 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 18 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<sub>Δ</sub>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<sub>Δ</sub> protein. For the FIG. 12A SOX2 and FIG. 12B NANOG 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 FIG. 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 (FIG. 13A-C) and (FIG. 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 (FIG. 13B, 13E) and two base mismatch plots (FIG. 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 grey 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 FIG. 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 FIG. 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 (FIG. 15A-B) and gRNA3 (FIG. 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

Validation, 5' gRNA Truncations

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

Example X

Validation, S. pyogenes PAM

[0136] As shown in FIG. 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 190K Cas9 targets in human exons that had no alternate NGG targets sharing 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

Validation, TALE Mutations

[0137] Using a nuclease mediated HR assay (FIG. 18A-B) it was confirmed that 18-mer TALEs tolerate multiple mutations in their target sequences. As shown in FIG. 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. FIG. 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 214-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. FIG. 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] FIG. 20A-B shows data related to off-set nicking. In the context of genome-editing, off-set nickers 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 nicks, off-target single nick events will likely result in very low indel rates. Inducing off-set nicks 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 lacks 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 nicks to generate 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. The number of Sanger sequencing clones is highlighted above the bars, and the predicted overhang lengths are indicated below the corresponding x-axis legends.

Example XIV

Off-Set Nicking, NHEJ Profiles

[0141] FIG. 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 nicks 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 Targets for Endogenous Gene Regulation</th>
<th>gRNA Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targets in the REX1, OCT4, SOX2 and PAX6 promoters used in Cas9-gRNA mediated activation experiments are listed (SEQ ID NOs: 11-61).</td>
<td></td>
</tr>
<tr>
<td>gRNA</td>
<td>gRNA Target</td>
</tr>
<tr>
<td>REX1 1</td>
<td>ctcgacgagactacgcggtt agg</td>
</tr>
<tr>
<td>REX1 2</td>
<td>cctcgagccctccaaaaagtct agg</td>
</tr>
<tr>
<td>REX1 3</td>
<td>acgcttcgctcgcaggtca ggg</td>
</tr>
<tr>
<td>REX1 4</td>
<td>ccaggaactacgtaccacca ggg</td>
</tr>
</tbody>
</table>
### Table 1 - continued

<table>
<thead>
<tr>
<th>gRNA</th>
<th>gRNA Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>REX1 5</td>
<td>gcacaccccaacgcatcaaa tgg</td>
</tr>
<tr>
<td>REX1 6</td>
<td>aataaatctttcataagga agg</td>
</tr>
<tr>
<td>REX1 7</td>
<td>gctactggggaggtgaggc agg</td>
</tr>
<tr>
<td>REX1 8</td>
<td>tagctacagctcactattaa tgg</td>
</tr>
<tr>
<td>REX1 9</td>
<td>tcgctgtcacccccgcgtct cgg</td>
</tr>
<tr>
<td>REX1 10</td>
<td>tgggcaagaggtgaacgct cgg</td>
</tr>
<tr>
<td>OCT4 1</td>
<td>tcctctctctctccccgtctg tgg</td>
</tr>
<tr>
<td>OCT4 2</td>
<td>tcctctctctctccccgtctg tgg</td>
</tr>
<tr>
<td>OCT4 3</td>
<td>aatgcatgtgcgcgtgac tg</td>
</tr>
<tr>
<td>OCT4 4</td>
<td>ccctcaagctctttact tgg</td>
</tr>
<tr>
<td>OCT4 5</td>
<td>gacgtccaaatctctttact tgg</td>
</tr>
<tr>
<td>OCT4 6</td>
<td>gatgtcttggtggatggata agg</td>
</tr>
<tr>
<td>OCT4 7</td>
<td>cgcacccctctctccagttg agg</td>
</tr>
<tr>
<td>OCT4 8</td>
<td>tcttaacccccggagataot agg</td>
</tr>
<tr>
<td>OCT4 9</td>
<td>cacaagcagccagggattcgg agg</td>
</tr>
<tr>
<td>OCT4 10</td>
<td>tggcatgcctggagaaaaacac tgg</td>
</tr>
<tr>
<td>OCT4 11</td>
<td>tgaatgctgacgacatacaact agg</td>
</tr>
<tr>
<td>OCT4 12</td>
<td>ccctccagcagctcagttt cgg</td>
</tr>
<tr>
<td>OCT4 13</td>
<td>ccacaaaaaacagacctgca agg</td>
</tr>
<tr>
<td>OCT4 14</td>
<td>aagcttcgacctgattgatta gg</td>
</tr>
<tr>
<td>OCT4 15</td>
<td>atgtcgtctgagtgggtgag agg</td>
</tr>
<tr>
<td>OCT4 16</td>
<td>ggtcaactgggaaggggaaga tgg</td>
</tr>
<tr>
<td>OCT4 17</td>
<td>tgccgtctctcttgcaga tgg</td>
</tr>
<tr>
<td>OCT4 18</td>
<td>gcacagtcgccccagaggtt cgg</td>
</tr>
</tbody>
</table>

### Table 2

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

<table>
<thead>
<tr>
<th>Figure</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>0 1</td>
<td>1-samp</td>
<td>7.8E-05</td>
<td>**</td>
</tr>
<tr>
<td>1</td>
<td>2 2</td>
<td>2-samp</td>
<td>1.4E-06</td>
<td>***</td>
</tr>
<tr>
<td>2</td>
<td>3 3</td>
<td>2-samp</td>
<td>4.0E-01</td>
<td>***</td>
</tr>
<tr>
<td>3</td>
<td>4 4</td>
<td>2-samp</td>
<td>0</td>
<td>***</td>
</tr>
<tr>
<td>4</td>
<td>5 5</td>
<td>2-samp</td>
<td>0</td>
<td>***</td>
</tr>
<tr>
<td>5</td>
<td>6 6</td>
<td>2-samp</td>
<td>1.0E-217</td>
<td>***</td>
</tr>
<tr>
<td>6</td>
<td>7 7</td>
<td>2-samp</td>
<td>1.7E-43</td>
<td>***</td>
</tr>
<tr>
<td>7</td>
<td>8 8</td>
<td>2-samp</td>
<td>3.7E-02</td>
<td>N.S.</td>
</tr>
</tbody>
</table>
TABLE 2-continued

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>1-samp</th>
<th>2-samp</th>
<th>3-samp</th>
<th>4-samp</th>
<th>5-samp</th>
<th>6-samp</th>
<th>7-samp</th>
<th>8-samp</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a</td>
<td>0</td>
<td>1</td>
<td>1-samp</td>
<td>8.9E-01</td>
<td>N.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2-samp</td>
<td>1.9E-06</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2-samp</td>
<td>5.0E-147</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2-samp</td>
<td>0</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>2-samp</td>
<td>0</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2-samp</td>
<td>4.2E-62</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>2-samp</td>
<td>1.6E-03</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>2-samp</td>
<td>4.7E-01</td>
<td>N.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>0</td>
<td>1</td>
<td>1-samp</td>
<td>5.2E-02</td>
<td>N.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2-samp</td>
<td>2.8E-05</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2-samp</td>
<td>3.5E-21</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2-samp</td>
<td>1.4E-58</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>2-samp</td>
<td>8.3E-101</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2-samp</td>
<td>6.8E-94</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>2-samp</td>
<td>1.8E-61</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>2-samp</td>
<td>8.1E-24</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>0</td>
<td>1</td>
<td>1-samp</td>
<td>2.3E-18</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2-samp</td>
<td>2.4E-08</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2-samp</td>
<td>6.2E-54</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2-samp</td>
<td>4.0E-141</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>2-samp</td>
<td>1.9E-20</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2-samp</td>
<td>1.2E-03</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>2-samp</td>
<td>3.8E-05</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>2-samp</td>
<td>9.4E-01</td>
<td>N.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>0</td>
<td>1</td>
<td>1-samp</td>
<td>7.2E-03</td>
<td>N.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2-samp</td>
<td>5.0E-01</td>
<td>N.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2-samp</td>
<td>3.9E-84</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2-samp</td>
<td>8.5E-153</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>2-samp</td>
<td>8.6E-76</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2-samp</td>
<td>1.6E-03</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>2-samp</td>
<td>2.7E-01</td>
<td>N.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>2-samp</td>
<td>7.8E-02</td>
<td>N.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>0</td>
<td>1</td>
<td>1-samp</td>
<td>7.3E-01</td>
<td>N.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2-samp</td>
<td>2.4E-06</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2-samp</td>
<td>7.2E-140</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2-samp</td>
<td>0</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>2-samp</td>
<td>0</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2-samp</td>
<td>1.4E-72</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>2-samp</td>
<td>1.0E-03</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>0</td>
<td>1</td>
<td>1-samp</td>
<td>9.4E-02</td>
<td>N.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2-samp</td>
<td>5.2E-09</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2-samp</td>
<td>7.9E-86</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2-samp</td>
<td>2.9E-53</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>2-samp</td>
<td>3.5E-10</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>0</td>
<td>1</td>
<td>1-samp</td>
<td>1.3E-13</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2-samp</td>
<td>1.1E-04</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2-samp</td>
<td>3.7E-08</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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 levels have been indicated by boxplots in the figures indicated in the Figure column, where the boxes represent the distributions of these levels by numbers of mismatches from the target site. P-values were computed 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: *** for P < .0005/n, ** for P < .005/n, * for P < .05/n, and N.S. (Non-Significant) for P >= .05/n, where n is the number of comparisons.
<table>
<thead>
<tr>
<th>seed start position</th>
<th>Number position pairs both in seed</th>
<th>Number position pairs not both in seed</th>
<th>(-10\log P)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>171</td>
<td>19</td>
<td>3.11</td>
</tr>
<tr>
<td>3</td>
<td>153</td>
<td>37</td>
<td>1.46</td>
</tr>
<tr>
<td>4</td>
<td>136</td>
<td>54</td>
<td>2.01</td>
</tr>
<tr>
<td>5</td>
<td>120</td>
<td>70</td>
<td>3.34</td>
</tr>
<tr>
<td>6</td>
<td>105</td>
<td>85</td>
<td>5.65</td>
</tr>
<tr>
<td>7</td>
<td>91</td>
<td>91</td>
<td>7.34</td>
</tr>
<tr>
<td>8</td>
<td>78</td>
<td>112</td>
<td>6.61</td>
</tr>
<tr>
<td>9</td>
<td>66</td>
<td>124</td>
<td>7.10</td>
</tr>
<tr>
<td>10</td>
<td>55</td>
<td>135</td>
<td>9.72</td>
</tr>
<tr>
<td>11</td>
<td>45</td>
<td>145</td>
<td>9.83</td>
</tr>
<tr>
<td>12</td>
<td>36</td>
<td>154</td>
<td>10.44</td>
</tr>
<tr>
<td>13</td>
<td>28</td>
<td>162</td>
<td>10.72</td>
</tr>
<tr>
<td>14</td>
<td>21</td>
<td>169</td>
<td>8.97</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>175</td>
<td>5.61</td>
</tr>
<tr>
<td>16</td>
<td>10</td>
<td>180</td>
<td>3.34</td>
</tr>
<tr>
<td>17</td>
<td>6</td>
<td>184</td>
<td>2.26</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>187</td>
<td>1.16</td>
</tr>
</tbody>
</table>

Table 2(b) Statistical characterization of seed region in Figure 2D: \(-10\log(P\)-values) indicating the degree of separation between expression values for Cas9-gRNA and gRNA binding to target sequences with two mutations for those position pairs mutated within candidate seed regions at the 3'-end of the 20-bp target site vs. all other position pairs. The greatest separation, indicated by the largest \(-10\log(P\)-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 section “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

A. Sequences of the Cas9gRNAVP64 activator constructs based on the m4 mutant are displayed below. Three versions were constructed with the Cas9gRNAVP64 and Cas9gRNAVP64m4 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).

\(\text{>Cas9}_{m4}^{\text{VP64}}\)

\([0145]\)
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 MS2preN fusion protein format showing highest activity.

>MS2preN

>MS2preC

C. d’Tomato fluorescence based transcriptional activation reporter sequences are listed below (See control-TF target, gRNA targets, minCMV promoter and FLAG tag+ dTomato sequences are highlighted).

>TF Reporter 1
>TF Reporter 2

[0153]

<160> NUMBER OF SEQ ID NOS: 186
<210> SEQ ID NO 1
<211> LENGTH: 1369
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 1

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1  5  10  15

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

20  25  30

Lys Val Leu Gly Aan Thr Asp Arg His Ser Ile Lys Gly Asn Leu Ile

35  40  45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gin Leu Val Gin Thr Tyr
180 185 190
Asn Gin Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Asn Asp Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Glu Leu Ser Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Ala Gln Ile Gly Asp Gin Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gin Asp Leu Thr Leu Leu Lys
325 330 335
 Ala Leu Val Arg Gin Gin Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gin Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Gin Glu Glu Phe Tyr Lys Phe Asl Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gin Gin Ser Ile Pro His Gin Ile His Leu
405 410 415
Gly Gin Gin Leu Leu Asp Gin Gin Gin Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Gin Gin gin Glu Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
-continued

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Thr Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys 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
580 585 590 595
Thr Tyr His Asp Leu Leu Lys Ile Lys Asp Asp Phe Leu Asp
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 Glu Ser Gly Lys Thr Ile Leu Asp Phe Leu Asp Ser Asp Gly Phe
675 680 685
Ala Asn Arg 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 Glu Glu 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 Gin Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gin
755 760 765
Thr Thr Gin Lys Gly Gin Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gin Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gin 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 Asp His Ile Val Pro Gin Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
Gly 850 Ser 855 Asp 860 Asn Val Pro Ser Glu Glu Val Thr Arg Lys Asp Met Lys

Asn Tyr Trp Arg Gln Leu Leu Ann Ala Lys Leu Ile Thr Gln Arg Lys

Phe Asp Ann Leu Thr Lys Ala Glu Arg Gly Lys Leu Ser Glu Leu Asp

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

Lys His Val Ala Gln Ile Leu Ser Arg Ser Arg Met Asn Thr Gln Tyr Asp

Glu Asn Arg Leu Ile Arg Gly Val Lys Val Ile Thr Leu Lys Ser

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Glu Phe Tyr Lys Val Arg

Glu Ile Asn Tyr His His Ala His Asp Ala Tyr Leu Ann Ala Ala Val

Val Gly Thr Ala Leu Ile Lys Tyr Pro Lys Leu Glu Ser Glu Phe

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

Thr Gly Lys Val Arg Ser Lys Tyr Thr Arg Lys Met Ser Ile Ala

Lys Ser Glu Glu Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

Tyr Ser Asn Ile Met Asn Phe Phe Ile Thr Glu Ile Thr Ile Ala

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

Arg Lys Ile Val Leu Ser Met Asn Glu Gly Val Ann Ile Val Lys Lys Thr

Glu Val Gln Thr Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Tyr Asp Pro

Lys Tyr Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

Ser Val Lys Glu Lys Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

Phe Gly Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Lys

Phe Gly Leu Glu Ann Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

Glu Leu Glu Lys Gly Asn Lys Leu Ala Leu Pro Ser Lys Tyr Val

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
<table>
<thead>
<tr>
<th>Pro</th>
<th>Glu</th>
<th>Asp</th>
<th>Asn</th>
<th>Glu</th>
<th>Gln</th>
<th>Leu</th>
<th>Phe</th>
<th>Val</th>
<th>Glu</th>
<th>Gln</th>
<th>His</th>
<th>Lys</th>
</tr>
</thead>
<tbody>
<tr>
<td>1260</td>
<td>1255</td>
<td>1255</td>
<td>1260</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Tyr</td>
<td>Leu</td>
<td>Asp</td>
<td>Glu</td>
<td>Ile</td>
<td>Ile</td>
<td>Glu</td>
<td>Gln</td>
<td>Ile</td>
<td>Ser</td>
<td>Glu</td>
<td>Phe</td>
</tr>
<tr>
<td>1265</td>
<td>1270</td>
<td>1275</td>
<td>1280</td>
<td>1285</td>
<td>1290</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Val</td>
<td>Ile</td>
<td>Leu</td>
<td>Ala</td>
<td>Asp</td>
<td>Ala</td>
<td>Asn</td>
<td>Leu</td>
<td>Arg</td>
<td>Lys</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>1280</td>
<td>1285</td>
<td>1290</td>
<td>1300</td>
<td>1305</td>
<td>1310</td>
<td>1315</td>
<td>1320</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Asn</td>
<td>Lys</td>
<td>His</td>
<td>Arg</td>
<td>Asp</td>
<td>Lys</td>
<td>Pro</td>
<td>Ile</td>
<td>Arg</td>
<td>Glu</td>
<td>Gln</td>
<td>Ala</td>
</tr>
<tr>
<td>1295</td>
<td>1300</td>
<td>1305</td>
<td>1310</td>
<td>1315</td>
<td>1320</td>
<td>1325</td>
<td>1330</td>
<td>1335</td>
<td>1340</td>
<td>1345</td>
<td>1350</td>
<td>1355</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 2
<211> LENGTH: 4332
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<222> OTHER INFORMATION: VP64-activator construct
<400> SEQUENCE: 2

gcacaacctg agaagagaata ctcacattgg ctcagatcteg gcacaaacag cgtcgagtgg 60
gcgcagca cagcgcagttt cctcattgt cagcagcctc tcgaagcagc cggagagagc 120
gatccagcag gcataaagga gacagcactc aggcgctgcc tggagttcct cggagagagc 180
ggcagctgca ccgctgcagc acagcagtag caagcagcag cgttctgggt cggagagagc 240
gatccagcag gcatacactt ttagggtctg taagcagctg cggagagagc 300
catagctgg ttttggttgg gacagcagc cggagagagc 360
tttgggtaca ttttgggttcc ttcagtgagc ctagaagagc cggagagagc 420
ttccagatcag gttttttgcg gttttttgcg gttttttgcg gttttttgcg gttttttgcg 480
ggcacagcag gcagcagcag cggagagagc 540
ggcacagcag gcagcagcag cggagagagc 600
ggcacagcag gcagcagcag cggagagagc 660
tttttgtttc tttttttgcg gttttttgcg gttttttgcg gttttttgcg gttttttgcg 720
gttttttgcg gttttttgcg gttttttgcg gttttttgcg gttttttgcg gttttttgcg 780
gtcagcagcag gcagcagcag cggagagagc 840
tttttttgcg gttttttgcg gttttttgcg gttttttgcg gttttttgcg gttttttgcg 900
ggcacagcag gcagcagcag cggagagagc 960
ggcacagcag gcagcagcag cggagagagc 1020
ggcacagcag gcagcagcag cggagagagc 1080
tttttttgcg gttttttgcg gttttttgcg gttttttgcg gttttttgcg gttttttgcg 1140
tttttttgcg gttttttgcg gttttttgcg gttttttgcg gttttttgcg gttttttgcg 1200
tttttttgcg gttttttgcg gttttttgcg gttttttgcg gttttttgcg gttttttgcg 1260
-continued

tgcaacgta tctcagggc gcagagggtt ttctacctct ttttgaaga taacagggaa 1320
aagattgaga aamtcctac attctctgat cactactatg tagcctcccc cggccgggga 1380
aatctcagat tgggctgtta gactgcaaa tcagaagaga ccatactcct ctggaaacct 1440
gaggaagtg tggcttaagg ggcttctgcc ctcgctctca tgcaagggat gactacatct 1500
gatataaact tcgctcaagga aagaagttcttt ccataaacaact ctgtgcgtta ccgagaacct 1560
acagtttata aagagctctac caagctcaca taagtcacct aagggatag agaagccagca 1620
tttctgtcttg gagaagcagaa gaaagctatc tgggacttctc tttcaagagca aacacgcga 1680
gtacgtctga cagctcttcag aagaagcata ttcacaaaga tttcagtttt gacagtctttt 1740
aataacgagc gagttggagga tcgcttccag gactcctgg gaagaagtca caagctctctg 1800
aatattaatt aagacacggag ctctctggaac atgagagaga aagagccacact aagttgaggg 1860
attgtctca ctctccagtt gttttgagagt agggagatga tttcagaaag cttgaaacct 1920
tgcctactc tctcactcag caaagttcagتا aacagcttcg aaggggctgcg atatacagga 1980
tgggggcggc tgtcaggaaga acgtatcacaag ggaggtccag cagacaagagag aggaaagaca 2040
atcctgctat ttttactcgt ctgcagatttgc gcacagcgga aacctcatgca gttgtgcctat 2100
gatgacacttc tcaaccttaa agaggacatc cagaaagcac aagttaccttg caggggggg 2160
agttcttcaag acgcatacct tatacttgcag cttgacccag ctataaaaag gggaaatactg 2220
ccagccggtta aagctcttgga tgaacctgttc aataagttgg agagcctaa gccgcgaga 2280
agttgtctat cagagccgccc aagagctcagc aagcctcagc aaagctaatc 2340
gaaaggatga agaggattga agagggatata aagagacaggg cttcccaaat ccttaaggaa 2400
cacccagtgg taaaacaccca gtcctgagaata gagatctctt acctctctag ataagctcagc 2460
ggcagggaca tgtacgcttg tccagaaactg gcacatcact gcctctcctga ttcagcctgtg 2520
getgctatcg tgcgcctccggg ttttcttaaa gttgattctca ttgataaatag agtttgctag 2580
agatccgata aagctagagg agagatcagtc aagctcccccct cagagaagac tgtcagaaaa 2640
atgaaaraatt tttcgcggcca gttgcttgaac gcacaagctgc tcaacacaacg gaagttctag 2700
aatctgacta aggctgacaag agtgggcttg tcttgaggtgg ataagccgag cttctctaaa 2760
aggctagcttg tggacccagc ccagatcacc aagcaagtttt ccacaaattt cagttcagcc 2820
atgaaacaca aagctgctgga aatagccagca tctgatcag agggagagtct tattactctg 2880
aagctcaagct tgtctctcag ttcagaaag gcctttcgtg tttataaggt gcagaggcac 2940
aacaatttcc aacatggccga tggattgcag tcagttctgacag tcagctggagc tcaactttacc 3000
aaaaaaattcc caaagcctttg aactgacattt gtttaagggag actataaagc gtctagtggt 3060
aggaagatga tgtcaagactc tggacggagaa ataggcagac cccacgctaaa ggagttcct 3120
ttagcagcat ttagatcttt ttcagagacc gagaattcagc tggccaaattt aagagattg 3180
agacgccca ctttacagaa aacagggaaa acagagaaaa tcgcttgagga caaggttagg 3240
gattctcggga cagttgccggc gtttgctgcc atggctggagg gcacatcggt taagaacacc 3300
gaagactaca cggaggggctt ctcacaaagaga agatctcctt gggcagggag aagagcaaca 3360
cgtgacgcc gcacaaggggt tggaggcacc aagagaagcg gcggattgaa tgtccctaca 3420
gtgcctttaca gtgtgacctcct tgtggtcagaa gttcagaaag gcggagttcaaaa aacactctaaa 3480
agagttcagc acagtcctttgt cacacactac atggagcgat cagggaccta aaaaacgcc 3540
-continued

```
<210> SEQ ID NO 3
<211> LENGTH: 4365
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: VP64-activator construct

<400> SEQUENCE: 3

gcaacactgg ccagaaagaa ggaggaagtc ggagggggga tggacaaagga tataccatt
  60

ggctgctaat tgcggacaa acaagctgggc agggccgctga atgggacaaa tataaaggt 120
ggcaagagaa aaatcaacgct tgcctgctgct gcaagctgct gcaagctgct 180

gagcagaggg gacatgctgct gcaagctgct gcaagctgct gcaagctgct 240

gtggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 300

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 360

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 420

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 480

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 540

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 600

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 660

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 720

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 780

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 840

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 900

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 960

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 1020

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 1080

tggactcag cccctctctg caaatctctct gctgcctgct gcaagctgct gcaagctgct 1140
```

agccagagg aattttacaa atttattaag ccacatcttg aaaaaatgga cggcacccgag 1200
gagctctgct taaacctgaaa cagagaagat cttggagcga aacagcgcac tttagcaaat 1260
ggaaagcattt ccacacaggt tcaocctgggc gaactgcaag atactctcag ggaggaagag 1320
gattttcacc cctttttgaa agataacaggg gaaagaaggg aagaaatcct cacttttgcag 1380
atactctcact atgtaggccct ctctggccgg gaaatctcga gattgctgtg gatgacctgc 1440
aatatcagaa agacaccataac tccctggaac ttggaggaag tgcgtgctaa gggggcctct 1500
goecagcttct tccatcgaagg gattgactaa ttgtgatataa atotgctata cagaaaggtg 1560
cctctcactaact cttcttcgctg tgaagagtac tttcaagttt taacaagacct cacoaaggtc 1620
aatatacgctg caaagagggt gagaagaacca gcccacttgt ctggagacga gaaagaagat 1680
atctctgagccc tctccctccga gacggacccg aaagttaccg tgaacagct ctgaagaagac 1740
tatatttcaaaa agatttgaagtt ttctgtccttt gttgaaatca ggcagaggtc ggtgctctcc 1800
aacaatttcc tggagaggtt tctggtctct cttcccagct tctaatggac gacagcttctt 1860
gccagaggag agaacaagga cattccttgac gcacatgcttc tccctcttacct tttgtttgaa 1920
agtggagga tggattgaac aacggttggata actactgtcct aacccttgca gaccaaaagt 1980
atgaaacagct ctaaagggcgg cctataataca ggtgtggtgg ggctgtgcaaa aacacggtact 2040
aagctttgcagc gacagacgc ggagttctgg aacatcctgg atttcttttata gctctgtgta 2100
cttggcaacc gcacactcttac gccttggtgctt ctttaaccttca taagggagac 2160
atcccaagaa ccaatcttctt cccgggaggg gacagcttcttc aggcaacoat cgtgaaactt 2220
gcaggtgcct cagcatacga aaggggaata ctgcagacgcg tttaagtgctgt ggtagaactc 2280
gtcacaagtg tggagacca acagcagcag aatacggtta tggagagttt ggcacaggaac 2340
caaactctgc aagaggggca gagaacaagtt aagggagaga ggcagaggtc gggaaggggt 2400
atataaagac tgggtgcttca ccaatctttta tcggaccaac ccaacagggc aaggtccttc 2460
aatgaaagagct tcacacgcttt ctccacgcat aacgggagg gacagtgactg gatgccgga 2520
cctggcctatc atotgtgtctcg cgactacagc gttcgtgctta gctgtgccca gctttttcctc 2580
aagagattt ctagtgtataa aaggtggtt gacagacgct taaactcgg tagggaagtg 2640
gataacgcttc ctccagaga aggtgcaag aatggattata atatgggcgc gcagcttcttg 2700
aagcgttcaag gcgtgcaag aacggtacgc aagaggttgg aagttctgta aagagggrggc 2760
cgtgctgttg tcgaaaaaag cggcttccatca aaaaggggcttg tgtggaagac aacgccagtc 2820
accaagcaac tggcccataat tctcacttgg cagctacgca ccaagtgaactc gagaaaaagt 2880
aaactgggatt cagaggtgtaa aagttattact ctggagccct gccagtgttctc agatttcgaa 2940
aagjactctttgattttattaggtgctggaggactcacttccacaaattcctagtcgtggg 3000	taacagcgctt cagctggaact cactgtccattt ccacaaactct tggagttcgg 3060
tttgtatcc gcagctatcag aagttcgatgt ggaggtcataa tgggtcacaa gttcagcagc 3120
gaaacaacttc ggtggcaccctt ccacttttcgctcactattata gttatatgga tttttcctcg 3180
aacagagatc aagctggcaaaa gttgagttcag aacgcagcgc aagaaatatc gcaaatagc 3240
gaaagggagcttacgtgctgt cagcagatcg gggtaggtcg gaaggtctgg 3300
tccatgctgctctgctacttc cgcgtctgtgtt aaacaaagac aacacccgatg gacaagggag 3360
gaaagttccttccogaaaagaagagcagttcagcagcgcgaaagtcggcctgc3420
cccaagaat acggcggatt cgattctct acagtgcctt acagttctc gggtgcgct 3480
aagctgagag aaggaagac taaaaaacct aaagcagtc aggaagctg ggctacaca 3540
atcatggagc gataagctt gaaaaaac ccatacagct ttctgagge gaagaagat 3600
aaagggcaga aaagacgtct cactattagct cttcctccga actcttctct ttgacctgaa 3660
aacgcgcgaa acagactgtg cgccgtgcg ccagaagttga ccagcgcgaa 3720
tctgcgcctta aaactgtaaa tttcttgatg ctgccagccc actatgaaga gctacaaaggg 3780
ttcgccagaa ataatgagca gaagcagctg ttctgagaac aacaaacaac atacattgat 3840
gagcatctcg acaaatctag cgaatctcct aaagatcgca ccctgcgcgca cgctacacct 3900
gataaggct ttctcggta caataacgac aggctataac ccctcaggg gcacagaaga 3940
aacattatc acgtgtctac tctgaccaac ttgggcgtgc ctcggcgcct caagattcct 4020
gacacacca tagacagaaa gcaggtaccc ctcacaaaggg aggtcaaggg ccgaacactg 4090
atccatagt caatactgagc gctctatgaa aacaaactgg acctctctcc aacctgtgaga 4160
gacacgcagg ctgacaccaaa gaagagaggg aagttccgag cccgaggttcc ccggagcgcct 4200
gacgctagg acgattttga tctgtgattg ccggagcttg ccggctgcctg caggatggac 4260
tctgacatgc ttcgctgagga tggccctggat gaacttggac cccaaacttg cggagctgac 4320
gcccttgatg atttgaacct gcacagtcgt gcattactta gatga 4365

<210> SEQ ID NO 4
<211> LENGTH: 4425
<212> TYPE: DNA
<220> ORGANISM: Artificial
<222> PRIMARY:
<223> OTHER INFORMATION: VP64-activator construct
<400> SEQUENCE: 4

gcacaacatgg acacagactg ctccatgtgg ctgtcctgag gcacaaacag cggctgcgtg 60
gcgtctacta cggagcagta aagctgccg agcacaataa tcaagattct cgggcaatac 120
gatcggccaga gcataaaaga gaaacattcg gggcgctccc ttggtgactc cggggagacg 180
gccgaaggca cgccgctcag aagaagatac ggagcgagat atccggcagc aaagaacggg 240
atcgtctcag tgccccagat cttaaggaat gagaagctgatagt gatccagag ctctctctct 300
cattaggctg aggctgcctt tgcgggagag gcaatggaga 360
	tggcaatac ttcgctgcag gctggctctc catgctgagct acaccaaacg ctccacaca 420
	acggagacttg cttgcggctgc actcctacat ggttcagctc tctgagctgc tctgagctgc 480
gcgcatacg tcaacctcg cggacactcc tctctagag gggaagttgg gcacacaaac 540
	agggagtctg gcacaaactt tatcactact ttctaaactcct ttgtgcgtgct cctaatctaa 600
	aacgcgacgga acttactgtac ctgcgctgcg cggggaacaac ccaccaacgg 660

tccggcgcgc tcgaacactc ctcctcgaggg gcagagagga ccgccttgctt 720

gtgaatgtta ctcctgtgcgt actcggtgtg accccacact ttaaatctaat ctgctgctgtg 780

gcggagagcg ccctgccctatt ccagactgact agtagacgct ccgaacactg 840

tggggctaga gcggacgcct gtaagcagtag tttttctgg ggcaacagaa ccttcgcaac 900

gcactgctgg gtagttgact tggagcagtc aacagcggta ccacacaaagc tcgctgccgtg 960

gcactgctgg ccacagagcata ctacagagcagt ccacacaaagc tcgctgccgtg 1020
gtagacagc aactgcttgta gaagtcacag gaaattttct tggagcagtc taaaaatggc
1080
taccgcgagt aacatgcaac cgggacacg cacgaggaat tttacaamtt tattaagccc
1140
aatattggaa aataagagac aagcagagag atgtctgtta aagtttaacag aagagatctg
1200
ttgccacaaag acgacacatttg caacatccag acgatcctcaca agatgctgga
1260
tgcaagccta tttgctagagg gcagagcaggt ttttcacccct ttgtgaaaga taacagggaa
1320
aagattgaga aatcctcagc attccggtata cctctactatgc tagggcccccag cggcgcggga
1380
aatccagat tggcttggtac gactgcaaa tcagagaaga ccctcatcttc cttgagaacct
1440
gaggaagccg tggcataaggg gctcttctgcc cctgctccca tcgataagagc gactaacttct
1500
gataaaaata cttgctaaag aaaaagttcct cttaaacaact cttgctgta egaagtaacct
1560
acagttaata acggccttac caagttcaaa ttcgcttcag aagagtgtaga aagcgcgaca
1620
tttgctgtag gcagcagcagaa gaacgctata cttggcaacctct cttcccagc gaaagtgatg
1680
igaacagcagc aagcggctta aagagcatcag tttgaaagcggtagttgactaacct
1740
aaaatcatta aggagcctca gctcctccctg gaggctctca cgtctccgag aagagctgcct
1800
atctctccga cctctctcttg gtttgtaagag tggagcatga ttgaggaacg cttgaaactc
1860
atctctctca cttctctcttg gtttgtaagag tggagcatga ttgaggaacg cttgaaactc
1920
ttacgtctca cttctctcttg gttgtagactt aagagcgtca gaggcgcctg atataacgga
1980
tgggggcgcct cttgcaagga aactcctacat ggggcctagc agaaggacagtt tgaaagcaca
2040
atctctctca cttctctcttg gttgtagactt aagagcgtca gaggcgcctg atataacgga
2100
gatgactctc tcacctctgaa gaggacacca caagaaacacag aatgctctcc gcaaggggggc
2160
agctctcagc agcgagcatcgc tattctgcaag gtagggccca gtagaagattc aagagcgata
2220
caagcggtta aagcctgtagc tcgaacctgtc aaagaatagc gaagcgacaa gaggccgaa
2280
atctctctca cttctctcttg gttgtagactt aagagcgtca gaggcgcctg atataacgga
2340
caagaggtga aagagctgta aagagggattaa aagaacgtgg gttcctcaaat ctttaagga
2400
caccagcttg aaaaaaacaga gtcctgaag aagagctgct cattctacct gcagggcggc
2460
ggcggacattc tcagcagctg caagatcatcgt gtctctcaagc ttcagagctg
2520
gttcagctgc tcacccctgagatcttttcagttgtgcaattt agtggagtttgca
2580
agatcggata aagagctgta aagagggattaa aagaacgtgg gttcctcaaat ctttaagga
2640
atgaaaaattc atgggctgagc gggagctgac gcacaacctga caacacacagc agatgcaggt
2700
aatctctca aagcctgtagc tcgaacctgtc aaagaatagc gaagcgacaa gaggccgaa
2760
agcagcggcct ctcagcagctgccagcttcagcagcagcagc gcttctccggag atataacgga
2820
atgaaacaca aagcctgtagc tcgaacctgtc aaagaatagc gaagcgacaa gaggccgaa
2880
aagtctacag tcagctctcag tttcagagaa gcttttcaagt tcattagatt gaggagatc
2940
aacaatttac ccgtgcctga tattagctac tggagagctac ggtagggcag cggccatttc
3000
aaaaatattc ccaaggtctga atctctatttt gttatggcagc attcatagat gtagatttga
3060
agaagaaaag tcagcagcttca ctagcagggta atagcagcggg ccacaacctc gattttcttt
3120	tacagcatt tattgaaatct tttcagagaa gcttttcaagt tcattagatt gaggagatc
3180
agaagaaaag tcagcagcttca ctagcagggta atagcagcggg ccacaacctc gattttcttt
3240
gatgctcggc agtgcctgctac gttgcctcagc tggagcagtct taaaagagccc 3300
gaagtacaga cggaggtct ctccaaagaa agtatcttcc cggagaaggaa cagcagacaag 3360
tctgatcgcg cgaaaaaaga ttggagcccc aagaaatagc gcyggatcgg ttcctctaca 3420
gtctgattaag tggtaaggct gtggagccaa gtggagaaag ggaaagttca aaaaactttaa 3480
agcgtagaac aagtctctttc ataccacaac tggcagcgaa cagcagacaca aaaaaaaccct 3540
atcgcgtttc taagcaggaa aggtataaa aagttcaca aaaaacccct gcttcttcttt 3600
cccaagtac cacccagtaga gctgtaaaac gcggcagggc ggtggtcagc tagtgtgcggc 3660
gagtgccaga aagggtagaa gctgggacct cgctccttaa acagccatg tttgatcttg 3720
gcccgccacat cgaagatgac ccggctcata cggcagatg atgccgctag cagcagctgc 3780
gtgggaaacac cccaacacta gctttgtgac ataccgcggg aaccagcacg aaccgtgacc 3840
agagtgacac tcggcagacgc taaactcgat aagttgcttt cggctcatac ttagccacgg 3900
gataagccac tcggccgcag ggcgaatagg aattaccact agtttactct gacaaacttgg 3960
gggcgccgct aagccctcaa gtactccgac acaccatagc acagaagagc gttacaactt 4020
acaaagagtg tctggagaccc caccctgctt cttaggcttc aatgggggct ctgacgaca 4080
agagatacag gctctctgcgt ctctcagacgc gcgccagcag gcagcagctgc 4140
agagtgacac gcggctgccag aagctgctggc gatattcttc ggttattcht gttggtagcct 4200
agagatacag gctctctgcgt ctctcagacgc gcgccagcag gcagcagctgc 4260
agagatacag gctctctgcgt ctctcagacgc gcgccagcag gcagcagctgc 4320
agagatacag gctctctgcgt ctctcagacgc gcgccagcag gcagcagctgc 4380
agagatacag gctctctgcgt ctctcagacgc gcgccagcag gcagcagctgc 4425

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

<400> SEQUENCE: 5
ccaccaatgg acctagaaaa aagggagaga ggctggagcc gcctcttatt gttcttact 60
ltactcaggt cgtctctgcg gacactagcc gcttcatctgc gccccaagcgc 120
aettogtcaa gggattacgt gatagttcata cttgcttaca cacagcagcgc 180
taacctctag cgtctctgcg gacactagcc gcttcatctgc gccccaagcgc 240
tgcctaaag ggctctctgg tctcttttct attgactaa atggcactaa aacccttaca attttgcaca 300
cgattcctgc gtctcttttatt gttactggg ctcctctaa aaggggagaga ggctggagcc gcctcttatt gttcttact 360
ecgcctctgc gacactagcc gcttcatctgc gccccaagcgc 420
tgcctctgc gacactagcc gcttcatctgc gccccaagcgc 480
agcctctgcgt cgtctctgcg gacactagcc gcttcatctgc gccccaagcgc 540
agcctctgcgt cgtctctgcg gacactagcc gcttcatctgc gccccaagcgc 607

<210> SEQ ID NO 6
<211> LENGTH: 661
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: MS2-activator construct
<400> SEQUENCE: 6
gccacacttg gcacataggaa aaagaggaag gtggcgcccg cttctagaaat ggcttttaaec
60

<210> SEQ ID NO: 7
<211> LENGTH: 557
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: MS2-activator construct
<220> FEATURE:
<221> NAME/KEY: mioc.feature
<222> LOCATION: (320)...(339)
<223> OTHER INFORMATION: wherein N is G, A, T or C
<400> SEQUENCE: 7
tgtatcaaaa actgagccttt aataggaaca attcatagtga cttgatcccg tacaaggtgc
60
gggccaggaag agggccatatt tccctagatt cttctcatatt tgcatatagc atacaaggtc
120
gtttagagaga taatttgaat tattttagaat gtaaadacaa agatatagc aaaaaactag
180
tgacttagaga aatsaatatt tcttttgatt tttgcagatt taaaattag tttaaatgat
240
gacttacata tgcctagtgg aacctggagatt tttttttttttt ttaatatcttg
300
tggsaaagac gaaacacgcn mnmmmnmmn mnmmmmnmn mnmmmnmmn mnmmmmcmn agaaatagca
360
agttataata aggttaggct gttactaatg tttagaagtt gcaacgagtgc gggtgcctgctg
420
agggcgatgc tagaataaatg gggggcgctg tgtatgtagc gcgtgctgcttg tagttgagaat
480
tctaatggct atctttaggct aagagcgtcacc aagagacgcc ctagatgagcc tggtsaatgatc
540
agaatattttctcagac
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
tagggataac aggctaatag tgttcccccc acccacaaggt gggggagaggt aggggtgtac
60
ggttgggaggg ctataaaagc agagcctgctt tagtgaaccc tccatagcgcc tggagaatcc
120
gccacatgg actacagga tgcacgagat aaaaactcgc gttggagact ggctttccac
180
tgagacaagg gcagggaggt catcaaaagag ttcagtcgct tcaggtgctgc catggagggc 240
ccagcagcc gcacagcaggg tgaatgcgag ggctggaggc agggcagccc ctgacgaggc 300
acacacaggg ccacagttgaa ggtgaccaag ggcagccccgc tggtgctgctgcttgggacac 360
tgtgtcgcac tgcgtatatga gcgggctgcc gcgtcaagctg agcgcccccgc gcacatccccc 420
gattacagag acgctgtctct cccgaggggc tcgaagttggg agcgctgtgt gaacactgag 480
ggagggctgc tgcctactgct gaccagggac tcctccccgc aggggagggc gctgctactac 540
aagggtagaag gtcgagggcc caactcccccc cccaagggcc ccgtaattgca gaagaagac 600
agtgggctggg agcctccgctc gagcggctctg taaccccgagc aggggctgtgt gaagggggag 660
atacagcgct cccgtggaagtt gaagagcgcc gcctcactac tggagcggctc tgaacccagt 720
tactggtgcc gcacctccccgc gcggctactac acgtggaacac caaggggagc 780
atacctcccc caacagggaggtc cccagccactgc tcgggcaagct gcagagccgct cgggggcgc 840
cacgactcgt tctgtaaagcc caggagccag cttgtacaaag taa 882

<210> SEQ ID NO 9
<211> LENGTH: 982
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Activation reporter construct

<400> SEQUENCE: 9

tagggataac agggtaatag tgggtgcact agggcaaggga tcggtgaggt aggggtgac 60
gtgggaggc ctataatacg aggtctgttt tgcgtgaccc tcagatcgcc tggagathtc 120
gcaccactgtc actacagggg ccagagcagat ccagccgctc tgcgagccct gcgcgatgac 190
gttgaagcag gcaggagagg catcacaaga ggtcgggcctc tcaaggtgcct ctcgagggcc 240
ttcctgaacgc gcacagcaggt cagagttgcag gcgggggagt gacggcggcc ctgcaggggc 300
acacagacgg ccacagctgaa ggtgaccaag gcgggcccccc gcgcctctgc gcgggaccctc 360
tctgcgcacattgccag cggtcagcaag agcagccccc gcacgctcccc 420
gattacagag acgctgtctct cccgaggggc tcgaagttggg agcgctgtgt gaacactgag 480
ggagggctgc tgcctactgct gaccagggac tcctccccgc aggggagggc gctgctactac 540
aagggtagaag gtcgagggcc caactcccccc cccaagggcc ccgtaattgca gaagaagac 600
agtgggctggg agcctccgctc gagcggctctg taaccccgagc aggggctgtgt gaagggggag 660
atacagcgct cccgtggaagtt gaagagcgcc gcctcactac tggagcggctc tgaacccagt 720
tactggtgcc gcacctccccgc gcggctactac acgtggaacac caaggggagc 780
atacctcccc caacagggaggtc cccagccactgc tcgggcaagct gcagagccgct cgggggcgc 840
cacgactcgt tctgtaaagcc caggagccag cttgtacaaag taa 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:
<221> NAME/KEY: misc_feature
<222> LOCATION: (22)...(44)
<223> OTHER INFORMATION: wherein N is G, A, T or C
<220> FEATURE:  
<221> NAME/KEY: misc_feature  
<222> LOCATION: (154)...(177)  
<223> OTHER INFORMATION: wherein N is G, A, T or C

<400> SEQUENCE: 10  
taggataac  agggaatag  tnnnnnnnn  nnnnnnnnn  nnnncaggt  aggcgtgtac  60  
ggtagggaggc  ctctataagc  agagctcgg  ttaggaaccg  tcaagacgcc  tggagaattc  120  
gccaacatcgg  actacaagga  tgaagacgat  aaannnnnnn  nnnnnnnnn  nnnnnnact  180  
tcgcgtggcg  gactgggttc  cacecttgac  aaggccaggg  agtcgtacaa  agatgccatg  240  
cgytcaagg  tgccagctggaa  gggtccatcg  aacgccacag  agttccagt  cgagggcgag  300  
gggaggggccc  gcccctcaaga  gggccacacc  acgccaaagc  tgaaggtgac  caagggcggc  360  
cccctgcccc  tgcgctgggga  catcctgtcc  ccccatgtcga  tgcagcggctc  caagggctac  420  
gtgaagccac  cgcgcgacat  ccocggatct  aagagctgtt  ccttcacccga  ggggctcaag  480  
tggagagcgc  tggatgaactt  cgaggacggc  ggtctggtga  cggcagcca  gcgtcctccc  540  
ctgccgaggc  ggcacgctgt  ctacacaggtc  agatagcgcc  gcaccaacct  ccoccccgac  600  
ggccccctaa  tcgccagggag  gacactggggc  tggagggctc  ccoccgaggc  ccgtgaccct  660  
cggagcccggc  tggctgaggg  cgagatccac  cagcccccct  agctggaagc  cggcccacac  720  
tacctgcttg  aagtcagcagc  catctacagt  gccaagagaag  ccttgcaactct  ccocggctac  780  
tactaagtgg  ccccaacagt  gcacatacct  tccacecaagc  agagactaccc  cactcgctgaa  840  
cagttgacgc  gcttccaggg  cccocacacac  ctctctgctg  aagcatggga  cagagttgac  900  
asatgaaat  tc  912

<210> SEQ ID NO 11  
<211> LENGTH: 23  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 11  
tgcgcggatc  actgcgcgggtg  agg  23

<210> SEQ ID NO 12  
<211> LENGTH: 23  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 12  
tctgcgcctc  cactactgctg  agg  23

<210> SEQ ID NO 13  
<211> LENGTH: 23  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 13  
aagctgtacct  ctgcagatca  ggg  23
<FEATURE>
<OTHER INFORMATION: Target probe>
<SEQUENCE: 20>
cggcgagag agtggagcgg ggg 23

<SEQ ID NO: 21>
<LENGTH: 23>
<TYPE: DNA>
<ORGANISM: Artificial>
<FEATURE:>
<OTHER INFORMATION: Target probe>
<SEQUENCE: 21>
tctctcttcct tctcctggtc tgg 23

<SEQ ID NO: 22>
<LENGTH: 23>
<TYPE: DNA>
<ORGANISM: Artificial>
<FEATURE:>
<OTHER INFORMATION: Target probe>
<SEQUENCE: 22>
tctctgcctaa gcccctggag agg 23

<SEQ ID NO: 23>
<LENGTH: 23>
<TYPE: DNA>
<ORGANISM: Artificial>
<FEATURE:>
<OTHER INFORMATION: Target probe>
<SEQUENCE: 23>
aatgcagttgt cocagtcggtc tgg 23

<SEQ ID NO: 24>
<LENGTH: 23>
<TYPE: DNA>
<ORGANISM: Artificial>
<FEATURE:>
<OTHER INFORMATION: Target probe>
<SEQUENCE: 24>
cecagccccc ctaagtctgctt ggg 23

<SEQ ID NO: 25>
<LENGTH: 23>
<TYPE: DNA>
<ORGANISM: Artificial>
<FEATURE:>
<OTHER INFORMATION: Target probe>
<SEQUENCE: 25>
gagtcctaat cctctttttact agg 23

<SEQ ID NO: 26>
<LENGTH: 23>
<TYPE: DNA>
<ORGANISM: Artificial>
<FEATURE:>
<OTHER INFORMATION: Target probe>
<SEQUENCE: 26>
gagtgcgg atttggtgta 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
cagcaacctca tcctccagtg agg

<210> SEQ ID NO 28
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe
<400> SEQUENCE: 28
tctaaaaacc agggaatcat ggg

<210> SEQ ID NO 29
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe
<400> SEQUENCE: 29
cacagagcag cagggatcc agg

<210> SEQ ID NO 30
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe
<400> SEQUENCE: 30
gatggcaagc tgagaaasc cgg

<210> SEQ ID NO 31
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe
<400> SEQUENCE: 31
tgaaatgac gcatacaatt agg

<210> SEQ ID NO 32
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe
<400> SEQUENCE: 32
ccagtcctca ccttgccctc tgg
<210> SEQ ID NO 33
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 33
ccagagaaaa cacaccctga 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
aaggggtgag cacttgttta ggg
23

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

<400> SEQUENCE: 35
atgctctgag ttgggttggag agg
23

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

<400> SEQUENCE: 36
ggcctcagtga agggaagta ggg
23

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

<400> SEQUENCE: 37
tggcagtccta ccttgttgaaga tgg
23

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

<400> SEQUENCE: 38
ggcaacagtc cagaggtcttg tgg
23

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

<400> SEQUENCE: 39

taaaaatataaaaaactaaca 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

ctctggtggtt acgtgcacg tag ggg

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

<400> SEQUENCE: 41

ggccagaggt caagctaggt ggg

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

<400> SEQUENCE: 42

cagcagcagaa accttctcta ggg

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

<400> SEQUENCE: 43

gttgaargaa gacagtctag tgg

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

<400> SEQUENCE: 44

taagaaagaa gcaagatcag tgg

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

<400> SEQUENCE: 45

taagaaacactgcag tgg
tgaaagctaa gagaggagag 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
tgacacacca aacctgtcac tgg 23

<210> SEQ ID NO 47
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe
<400> SEQUENCE: 47
ttaaaccact tctttcgaag agg 23

<210> SEQ ID NO 48
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe
<400> SEQUENCE: 48
gtgctgccc gcgtgcctct ggg 23

<210> SEQ ID NO 49
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe
<400> SEQUENCE: 49
cttccccggc cttccccgcgc ggg 23

<210> SEQ ID NO 50
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe
<400> SEQUENCE: 50
caaaaacccg cagcgagact ggg 23

<210> SEQ ID NO 51
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe
<400> SEQUENCE: 51
agggagccgc gcgcgcgtat tgg 23
<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 cacacgagat ggg

<210> SEQ ID NO 53
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 53
gasaagccta aagagccaga ggg

<210> SEQ ID NO 54
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 54
tagtagaattt ccaaaccttc agg

<210> SEQ ID NO 55
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 55
tccgctcrg ttcgccaggg tgg

<210> SEQ ID NO 56
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 56
cagacaccaca ccacccgagg tgg

<210> SEQ ID NO 57
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 57
tccaatattt tgggattac agg

<210> SEQ ID NO 58
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220>  FEATURE:
<223>  OTHER INFORMATION: Target probe

<400>  SEQUENCE: 58

tgatttaaa gttggaacg tgg

<210>  SEQ ID NO 59
<211>  LENGTH: 23
<212>  TYPE: DNA
<213>  ORGANISM: Artificial
<220>  FEATURE:
<223>  OTHER INFORMATION: Target probe

<400>  SEQUENCE: 59

tctagttcc cacctagtct ggg

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

<400>  SEQUENCE: 60

gattaaatga gattccacaa ggg

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

<400>  SEQUENCE: 61

cgccagggg ggtggtccta agg

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

<400>  SEQUENCE: 62

gtccccctca cccacagtg ggg

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

<400>  SEQUENCE: 63

ggggccacta ggacaggat tgg

<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
taatactttt atctgtcccc ttccacccac agtgggggcca ctagggcag gattgggtgac
agaasagccoc
<210> SEQ ID NO 65
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 65
ggggccccaga gggacaggt 20
<210> SEQ ID NO 66
<211> LENGTH: 80
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Guide RNA
<400> SEQUENCE: 66
guuuaagcg ugaauagcg aaguuasaau aagggcagcu uguuaucaac uugaaaaagu 60
gggccccagcu gggcuuuuu 80
<210> SEQ ID NO 67
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 67
gtccctccca cccacagtg cag 23
<210> SEQ ID NO 68
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 68
gtccctccca cccacagtg cca 23
<210> SEQ ID NO 69
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 69
gtccctccca cccacagtg cgg 23
<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
tgtccctcc aocccacagt ggggccacta gggacaggat tggtgacaga aa 52

<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
tgtccccccc aocccacagt ggggccacta gggacaggat tggtgacaga aa 52

aaaaccctcc aocccacagt ggggccacta gggacaggat tggtgacaga 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 aocccacagt ggggccacta gggacaggat tggtgacaga 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
tgtccctcc tttttcacgt ggggccacta gggacaggat tggtgacaga 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
cacggggtg tgtccctccc tgg 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
ggtgcccatc tgtgtcagagc tgg 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
cccatctgg tgtagtgga 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
ggccacaagt tcagctgtgctcgg23

<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 actcaccsectcgg23

<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
cgcaagtttc tctgcaccacctcgg23

<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
cggcaagct gccgacgccctcgg23

<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
gacaggag ggcaccacccctcgg23

<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
gcctcaccg tgcacacgaatcgg23

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

SEQ ID NO 83
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 83
ggccggacac gctgaacctg tgg

SEQ ID NO 84
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 84
taacaggta atgtcagggc cgg

SEQ ID NO 85
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 85
aggtgagctc ttattgcggt agg

SEQ ID NO 86
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 86
cctcaggttc agcttgcggt agg

SEQ ID NO 87
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 87
ggccagggc agcttgcggt tgg

SEQ ID NO 88
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 88
gagatgatcg cccctttctc tgg

SEQ ID NO 89
LENGTH: 20
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 89

-continued

```
gagatgacct ccctttcttc
<210> SEQ ID NO 90
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 90

gagatgacct gcgggtttcttc
<210> SEQ ID NO 91
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 91

gtccctccca 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

gagatgacct gcgggtttcttc tgg
<210> SEQ ID NO 93
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence
<400> SEQUENCE: 93

guccccucca cccccacagu
<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

guccccucca cccccacagu
<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

guccccucca cccccacagag
```

<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 102

guccuccuca gcacacagug 20

<210> SEQ ID NO 103
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 103

guccuccuca gcacacagug 20

<210> SEQ ID NO 104
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 104

guccuccuca ccacacagug 20

<210> SEQ ID NO 105
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 105

guccccucşa ccacacagus 20

<210> SEQ ID NO 106
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 106

guccuccucca ccacacagac 20

<210> SEQ ID NO 107
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 107

guccuccucca ccacacacag 20

<210> SEQ ID NO 108
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 108
guccuccua coccugagug
<210> SEQ ID NO 109
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence
<400> SEQUENCE: 109

guccuccua cggacagug
<210> SEQ ID NO 110
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence
<400> SEQUENCE: 110

guccuccua ggccacagug
<210> SEQ ID NO 111
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence
<400> SEQUENCE: 111

guccuccgu ccccacagug
<210> SEQ ID NO 112
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence
<400> SEQUENCE: 112

ggggccacta ggacagaggt ggg
<210> SEQ ID NO 113
<211> LENGTH: 23
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence
<400> SEQUENCE: 113

gagaugauc cccuucuc
<210> SEQ ID NO 114
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence
<400> SEQUENCE: 114

gagaugauc cccuucuug
<210> SEQ ID NO 115
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 115

gagaugacgccuucuac 20

<210> SEQ ID NO 116
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 116

gagaugacgccuucuac 20

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

<400> SEQUENCE: 117

gagaugacgccuucuuc 20

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

<400> SEQUENCE: 118

gagaugacgccuacuuc 20

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

<400> SEQUENCE: 119

gagaugacgccuacuuc 20

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

<400> SEQUENCE: 120

gagaugacgccuucuuc 20

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

<400> SEQUENCE: 121
gagaauag cgcucuucu

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

<400> SEQUENCE: 122
gagaauag cgcucuucu

<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
gagaauag cgcucuucu

<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
gagaaucc cccucuucu

<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
gagaauagg cccucuucu

<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
gagaauag cccucuucag

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

<400> SEQUENCE: 128
gagaugauc cccccacuucc

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

<400> SEQUENCE: 129
gagaugauc cccccacuucc

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

<400> SEQUENCE: 130
gagaugauc cccccacuucc

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

<400> SEQUENCE: 131
gagaugauc cccccacuucc

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

<400> SEQUENCE: 132
gagatgatcg cccccctctt tgg

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

<400> SEQUENCE: 133
gggpccacua ggacaggau
<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

gggccacuag ggcacaggau 19

<210> SEQ ID NO 135
<211> LENGTH: 16
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 135

ggccacuagg ggcagau 18

<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 gcaaggau 17

<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

gagaugauc ccccwucuwuc 20

<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 ccuucucu 18

<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

gauggcocu ucuuc 15

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

<400> SEQUENCE: 140

gccucuucu c

<210> SEQ ID NO 141
<211> LENGTH: 11
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURES:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 141

gtccccctcca cccccagtg c

<210> SEQ ID NO 142
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURES:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 142
tgccnnnnnn accc

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

<400> SEQUENCE: 143
tgtaaaaaaa accc

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

<400> SEQUENCE: 144
tgcgcccccc accc

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

<400> SEQUENCE: 145
tgtaaaaaaa accc

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

SEQUENCE: 146
tgctggggg acc 14

SEQ ID NO 147
LENGTH: 14
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 147
tgctcccccc acc 14

SEQ ID NO 148
LENGTH: 14
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 148
tgctttttttt acc 14

SEQ ID NO 149
LENGTH: 14
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 149
tgccccccccc acc 14

SEQ ID NO 150
LENGTH: 14
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 150
tgcttttttt acc 14

SEQ ID NO 151
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 151
ggatctctgt tccccgagct ggg 23

SEQ ID NO 152
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Target oligonucleotide sequence

SEQUENCE: 152
gtzaatgagg cctgtgtctt ggg

<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

ggggcccacta gggacaggat 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

cctcctagtc tcgtgataatt 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
tgtcccagc tcgggacac 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
	agacccagcg cccattacc 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

gtsscaatct ctgtccttag 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
	agaccccaata tcaggagact 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

gggatctctgt gtcoccagac tgggaccacc ttatatcccc agggccggttt aatgtggctc 60
tggttcctgg tacct 75

<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

gggatctctgt gtcoccagac tgggaccacc ttatatcccc agggccggttt aatgtgtttc 60
tggtactt 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

gggatctctgt gtcoccagac tgggaccacc ttatatcccc agggcagggc cgttggacc 60
accttatatt cccccgggag ggcgggttaa tgtggtcttg gtcttgtggtctt 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

gggatctctgt gtcoccctctg gttctgggt 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

gggatctctgt gtcoccagac tgggaccacc ttatatcttg ggtactt 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

gggatctctgt ggtactt 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
agggccggtt aatgtggtcg tggttctggg tacattttattc tgtcccccctc accccacagt 60
gggggcacta gggacaggat tggtgacaga aaa 93

<210> SEQ ID NO 166
<211> LENGTH: 93
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 166
agggccggtt aatgaatgtg gctctggttgc tgggtacttt tatctgtcccc cttccacccca 60
cagttggggcc actagcagaca aaa 93

<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
agggccgttt aatgtggtcct tggttctgggt tacattttattc tgtcccccag tgggccccct 60
gatrgtgtgca agaaaaa 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
agggccgttt caggattgtg gacgaaaaaa 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
agggccggtt aatgtgccga ttggtgacag 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
agggccggtt aatgtgctc tggttctggg tacttttattc tgcocccgat tggtgacaga 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
agggccggtt aatgtgctc tggttctggg tacttttattc tgcocccctcc accccacaagt 60
gggcacagaa tgtgtgacag 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
agggccggtt aatgtgtga cagaaaa 27

<210> SEQ ID NO 173
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 173
agggccggtt aatgtgccct tggttctggg tactttttattc tgcocccctcc accccagggg 60
acagtcctgc cctccaccgc cagggacagg atgtgtgaca 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
agggccggtt aatgtgctc tggttctggg tactttttattc tgcocccctcc accactagg 60
acagaggattg 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
cacacagttg ggcaatagg gacaggattg tgtagagaaa agccccatac 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

ccacagttg ggccactacc cc 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

ccacagttg ggccactagt agaaaagccc cactcctagg cttcctcct ctttaggctt 60
ctctctctcct agtctcctga tttggtct ascccc 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

ccacagttg ggccactagg gacaggattg gtgacagaaa agcoccatac cttggcctctc 60
tctctctctgcgt tttgctctaa cccccc 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

ccacagttg ggccacccctt aggctcctct cttcctagtc tctcgtattt ggctctaaacc 60
tttgctctcctc 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

ccacagttg ggccactagt gatattgggt ctaacccc 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

ccacagttg ggccactagg gacaggattg gtgacagaaa agcoccatac cttggcctctc 60
tctctctctgcgt tttgctctaa cccccc 94
I. A method of altering a target nucleic acid in a cell comprising:

providing to the cell two or more RNAs with each RNA being complementary to an adjacent site in the target nucleic acid,

providing to the cell a Cas9 protein nickase and being guided by the two or more RNAs, and

wherein the two or more RNAs and the Cas9 protein nickase co-localize to the DNA target nucleic acid and nick the target nucleic acid resulting in two or more adjacent nicks.

2. The method of claim 1 wherein the two or more RNAs are provided to the cell by introducing into the cell a first foreign nucleic acid encoding the two or more RNAs,
wherein the Cas9 protein nickase is provided to the cell by introducing into the cell a second foreign nucleic acid encoding the Cas9 protein, and wherein the two or more RNAs and the Cas9 protein nickase are expressed and wherein the Cas9 protein nickase co-localizes with the two or more RNAs to the target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks.

3. The method of claim 1 wherein the two or more adjacent nicks are on the same strand of the double stranded DNA.

4. The method of claim 1 wherein the two or more adjacent nicks are on the same strand of the double stranded DNA and result in homologous recombination.

5. The method of claim 1 wherein the two or more adjacent nicks are on different strands of the double stranded DNA.

6. The method of claim 1 wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks.

7. The method of claim 1 wherein 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.

8. The method of claim 1 wherein the two or more adjacent nicks are on different strands of the double stranded DNA and are offset with respect to one another.

9. The method of claim 1 wherein 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.

10. The method of claim 1 wherein 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.

11. The method of claim 1 further including 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.

12. The method of claim 1 wherein the target nucleic acid is genomic DNA, mitochondrial DNA, viral DNA or exogenous DNA.

13. A method of altering a viral DNA target nucleic acid in a cell comprising

providing to the cell two or more RNAs with each RNA being complementary to an adjacent site in the viral DNA target nucleic acid,

providing to the cell a Cas9 protein nickase and being guided by the two or more RNAs, and wherein the two or more RNAs and the Cas9 protein nickase co-localize to the viral DNA target nucleic acid and nick the viral DNA target nucleic acid resulting in two or more adjacent nicks.

14. The method of claim 13 wherein the two or more RNAs are provided to the cell by introducing into the cell a first foreign nucleic acid encoding the two or more RNAs, wherein the Cas9 protein nickase is provided to the cell by introducing into the cell a second foreign nucleic acid encoding the Cas9 protein, and wherein the two or more RNAs and the Cas9 protein nickase are expressed and wherein the Cas9 protein nickase co-localizes with the two or more RNAs to the viral DNA target nucleic acid and nicks the viral DNA target nucleic acid resulting in two or more adjacent nicks.

15. The method of claim 13 wherein the two or more adjacent nicks are on the same strand of the double stranded DNA.

16. The method of claim 13 wherein the two or more adjacent nicks are on the same strand of the double stranded DNA and result in homologous recombination.

17. The method of claim 13 wherein the two or more adjacent nicks are on different strands of the double stranded DNA.

18. The method of claim 13 wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks.

19. The method of claim 13 wherein 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.

20. The method of claim 13 wherein the two or more adjacent nicks are on different strands of the double stranded DNA and are offset with respect to one another.

21. The method of claim 13 wherein 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.

22. The method of claim 13 wherein 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.

23. The method of claim 13 further including providing to the cell 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.

24. The method of claim 2 further including 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.