US 20140356959A1

(19) **United States**

(12) **Patent Application Publication**

CHURCH et al.

(10) **Pub. No.: US 2014/0356959 A1**

(43) **Pub. Date: Dec. 4, 2014**

(54) **RNA-GUIDED TRANSCRIPTIONAL REGULATION**

(71) Applicant: **President and Fellows of Harvard College**, Cambridge, MA (US)

(72) Inventors: **George M. CHURCH**, Brookline, MA (US); **Prashant G. MALI**, Somerville, MA (US); **Kevin M. ESVELT**, Cambridge, MA (US)

(21) Appl. No.: **14/319,289**

(22) Filed: **Jun. 30, 2014**

Related U.S. Application Data

(63) Continuation of application No. PCT/US2014/040868, filed on Jun. 4, 2014.

(60) Provisional application No. 61/830,787, filed on Jun. 4, 2013.

Publication Classification

(51) **Int. Cl.**

C12N 15/113 (2006.01)

C12N 15/85 (2006.01)

(52) **U.S. Cl.**

CPC ***C12N 15/113*** (2013.01); ***C12N 15/85*** (2013.01); *C12N 2310/11* (2013.01)

USPC **435/455**; 435/375; 435/255.1; 435/419; 435/468; 435/471; 435/366

(57) **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.

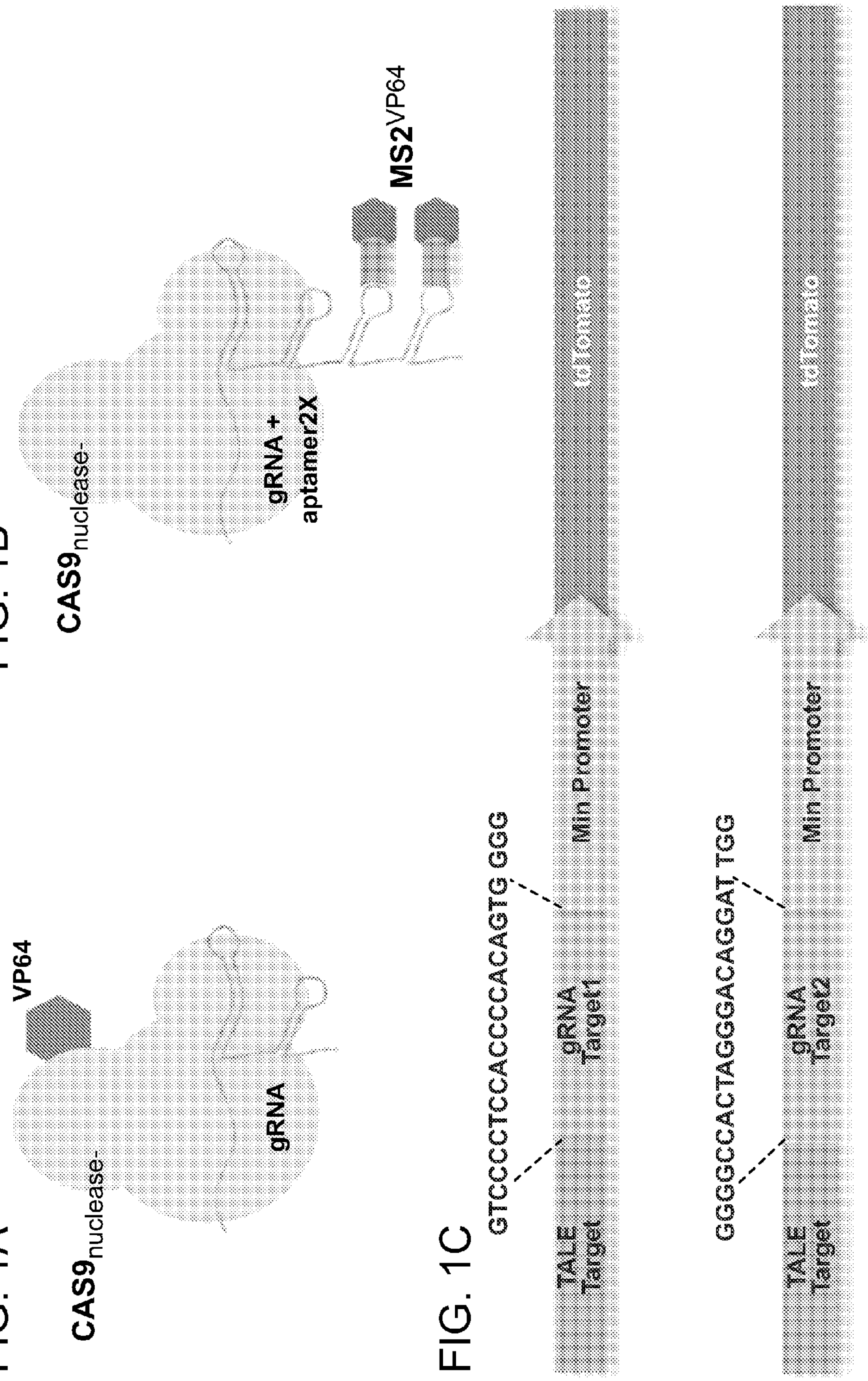


FIG. 1A

FIG. 1B

FIG. 1C

FIG. 1D-1

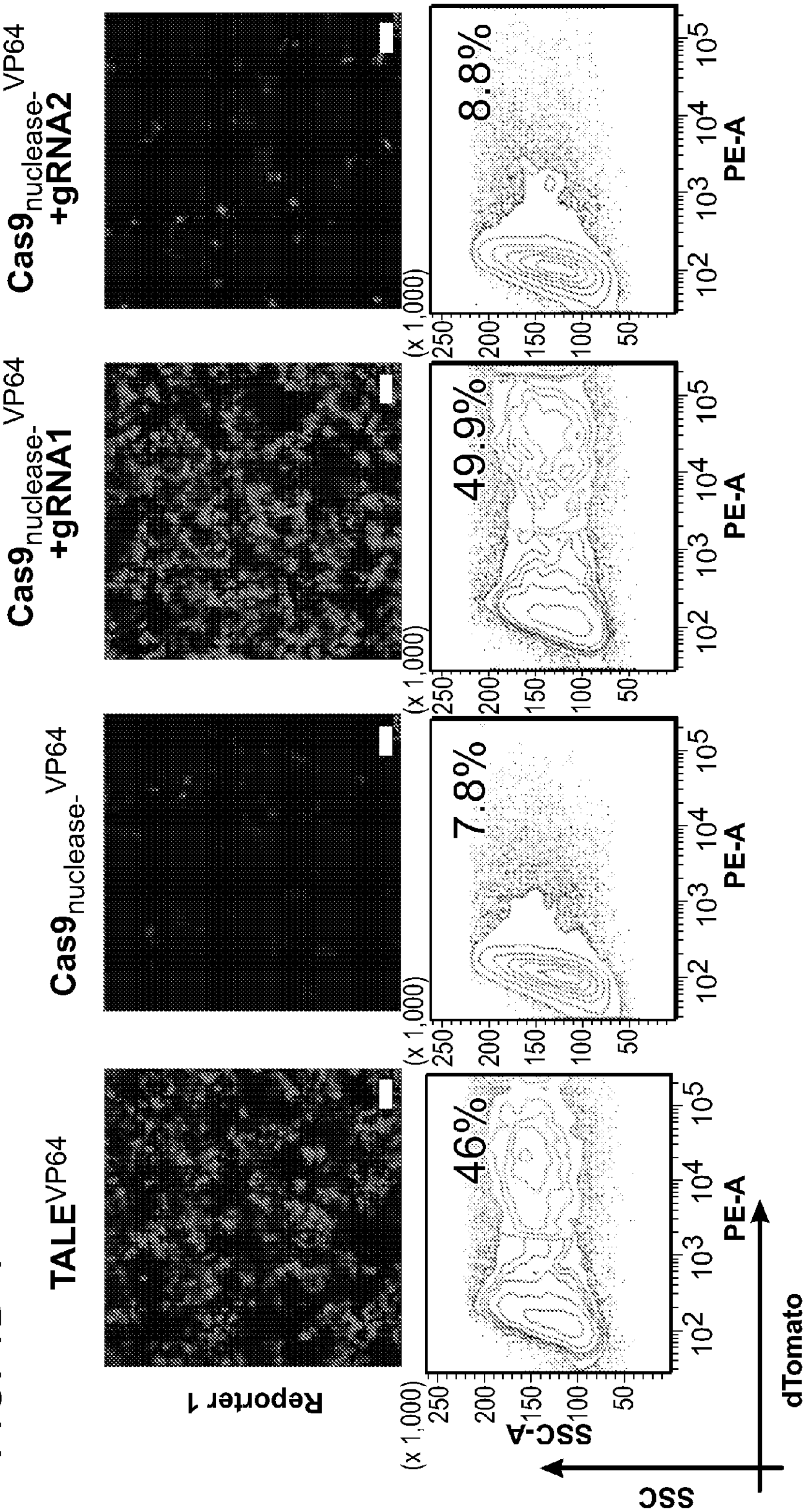
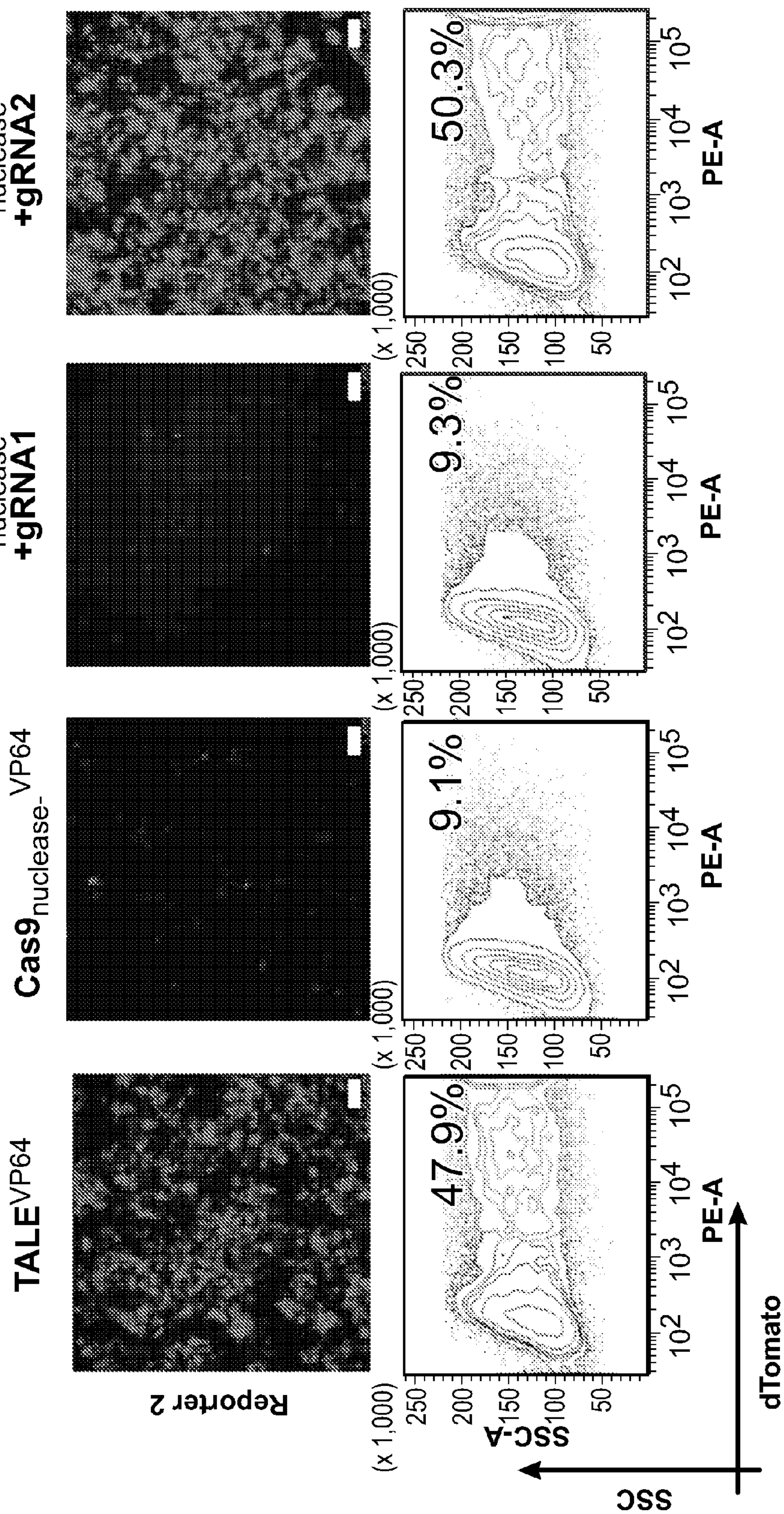

TALE^VP64
Cas9nuclease-^VP64
Cas9nuclease-^VP64 +gRNA1
Cas9nuclease-^VP64 +gRNA2
Reporter 1
46%
7.8%
49.9%
8.8%
SSC-A
(x 1,000)
250
200
150
100
50
PE-A
10^2
10^3
10^4
10^5
SSC
dTomato

FIG. 1D-2



TALE^{VP64}
Cas9$_{nuclease-}$^{VP64}
Cas9$_{nuclease-}$^{VP64} +gRNA1
Cas9$_{nuclease-}$^{VP64} +gRNA2
Reporter 2
47.9%
9.1%
9.3%
50.3%
SSC-A
(x 1,000)
250
200
150
100
50
PE-A
10^2
10^3
10^4
10^5
SSC
dTomato

FIG. 1E-1

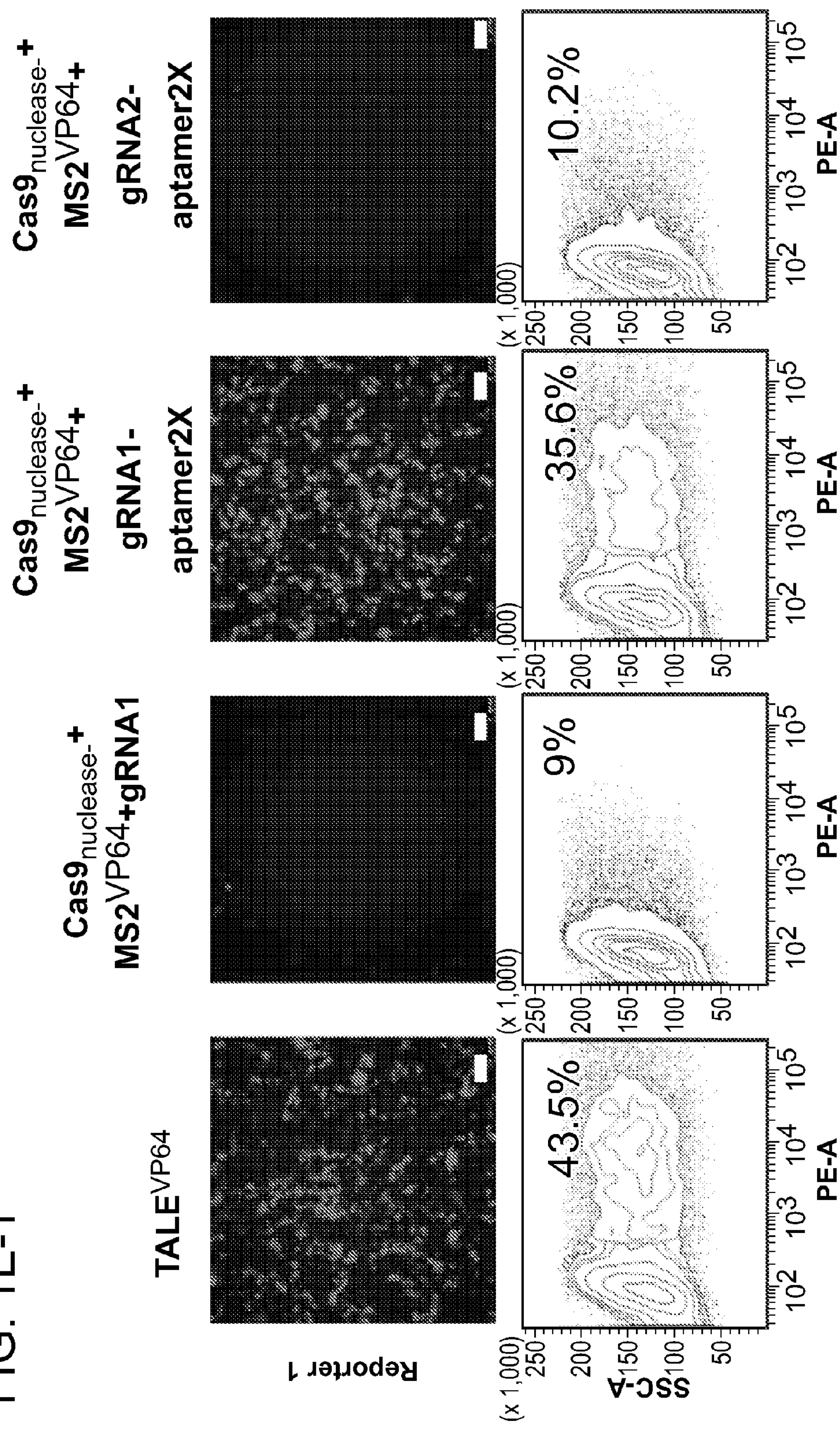
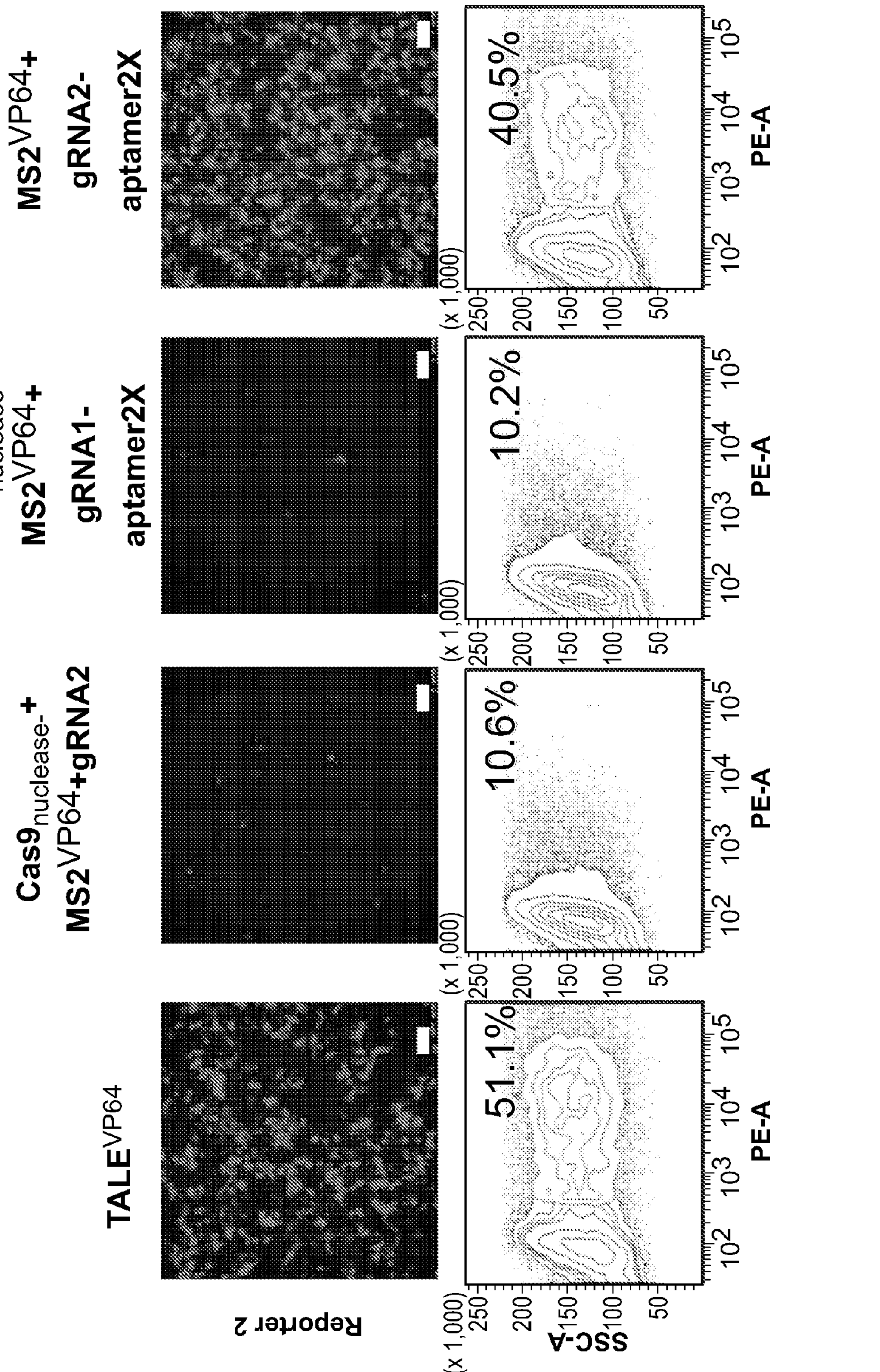
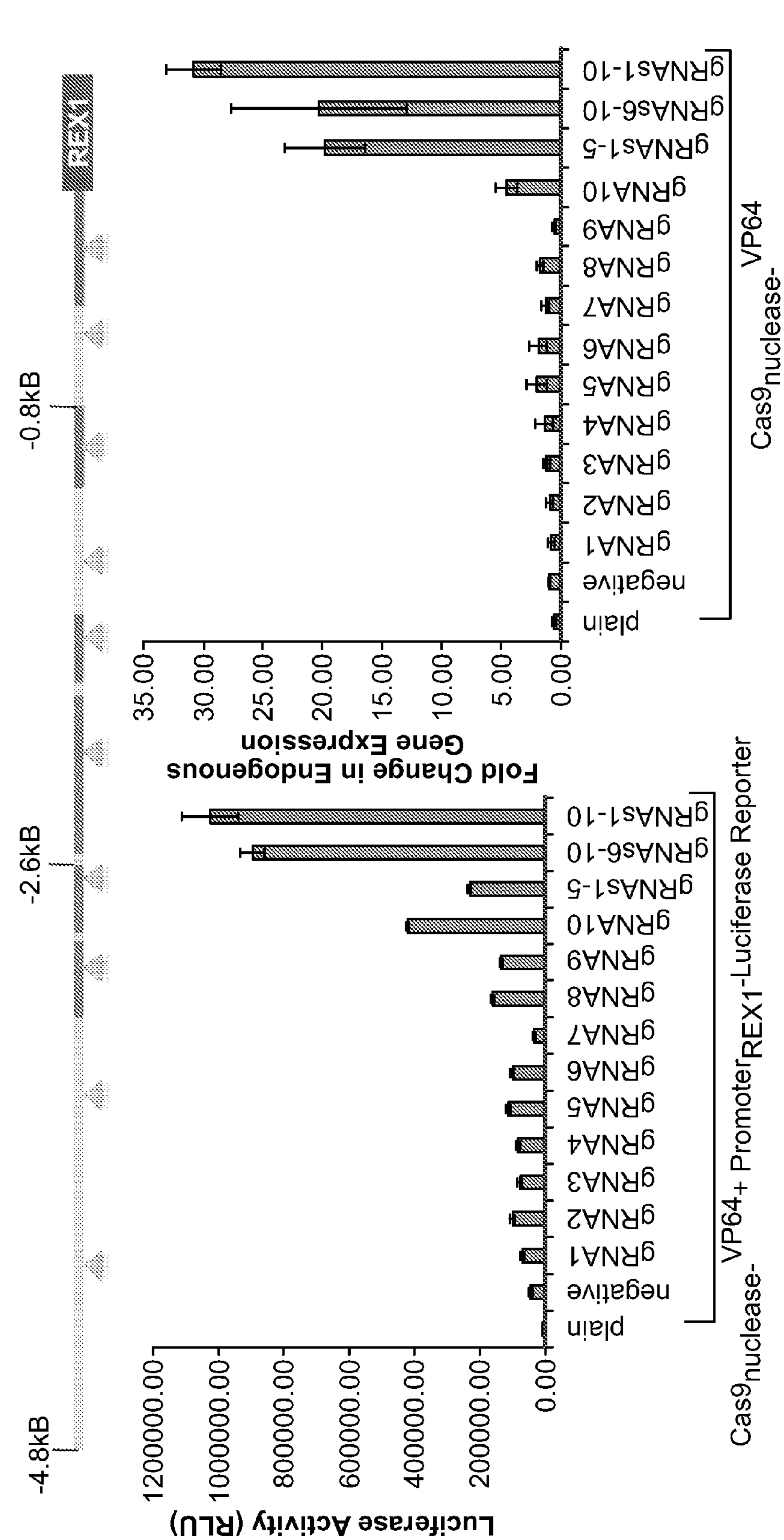

TALE^{VP64}
Cas9$_{nuclease-}$+ MS2^{VP64}+gRNA1
Cas9$_{nuclease-}$+ MS2^{VP64}+ gRNA1- aptamer2X
Cas9$_{nuclease-}$+ MS2^{VP64}+ gRNA2- aptamer2X
Reporter 1
SSC-A
(x 1,000)
250
200
150
100
50
10^2
10^3
10^4
10^5
PE-A
43.5%
9%
35.6%
10.2%

FIG. 1E-2



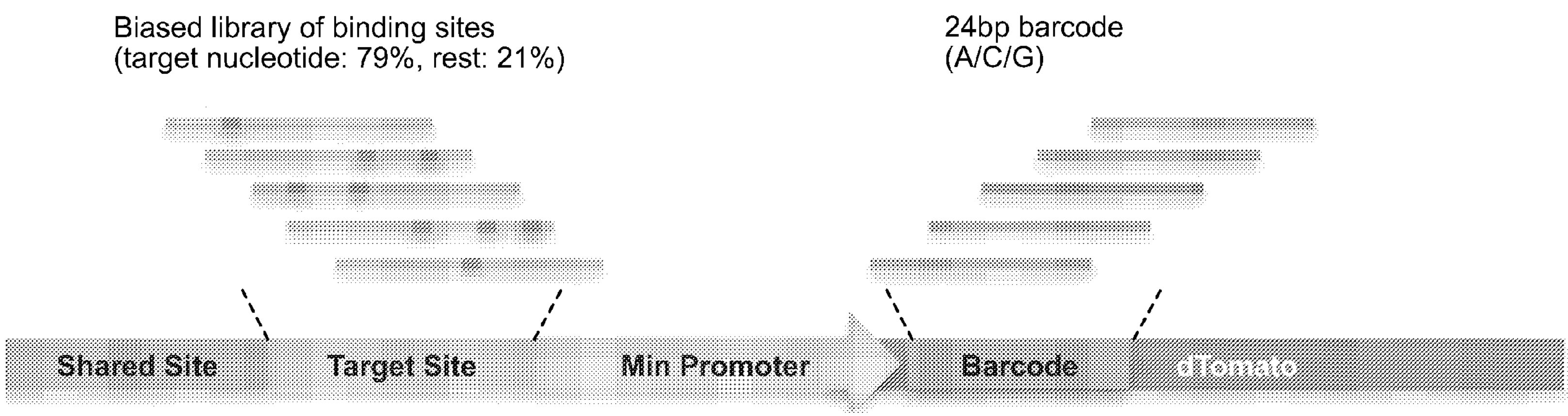
TALE^{VP64}
Cas9$_{nuclease-}$+MS2^{VP64}+gRNA2
Cas9$_{nuclease-}$+MS2^{VP64}+gRNA1-aptamer2X
Cas9$_{nuclease-}$+MS2^{VP64}+gRNA2-aptamer2X
Reporter 2
SSC-A
(x 1,000)
250
200
150
100
50
51.1%
10.6%
10.2%
40.5%
10^2
10^3
10^4
10^5
PE-A

FIG. 1F



-4.8kB
-2.6kB
-0.8kB
REX1
Luciferase Activity (RLU)
1200000.00
1000000.00
800000.00
600000.00
400000.00
200000.00
0.00
plain
negative
gRNA1
gRNA2
gRNA3
gRNA4
gRNA5
gRNA6
gRNA7
gRNA8
gRNA9
gRNA10
gRNAs1-5
gRNAs6-10
gRNAs1-10
Cas9nuclease-VP64+ PromoterREX1-Luciferase Reporter
Fold Change in Endogenous Gene Expression
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
Cas9nuclease-VP64

FIG. 2A



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

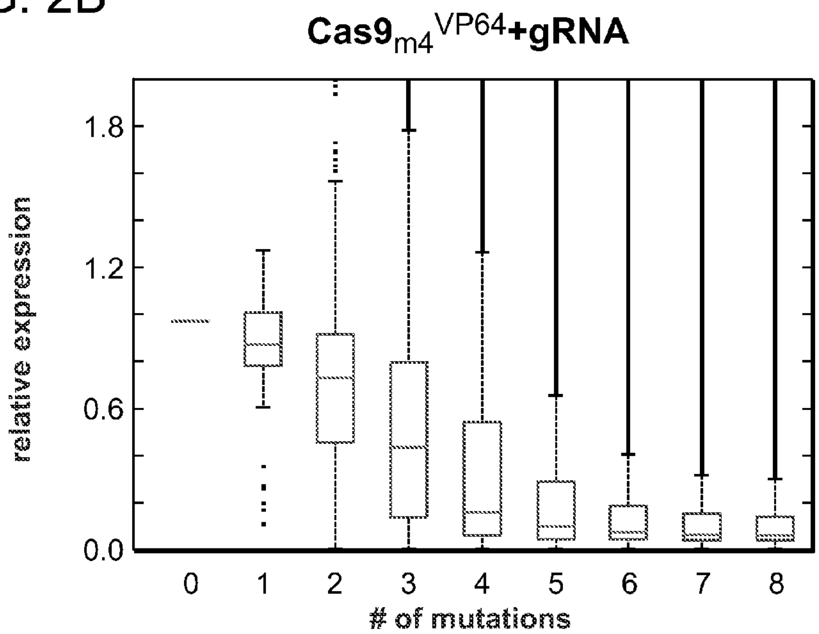

Cas9$_{m4}$^{VP64}+gRNA
relative expression
1.8
1.2
0.6
0.0
0
1
2
3
4
5
6
7
8
# of mutations

FIG. 2C

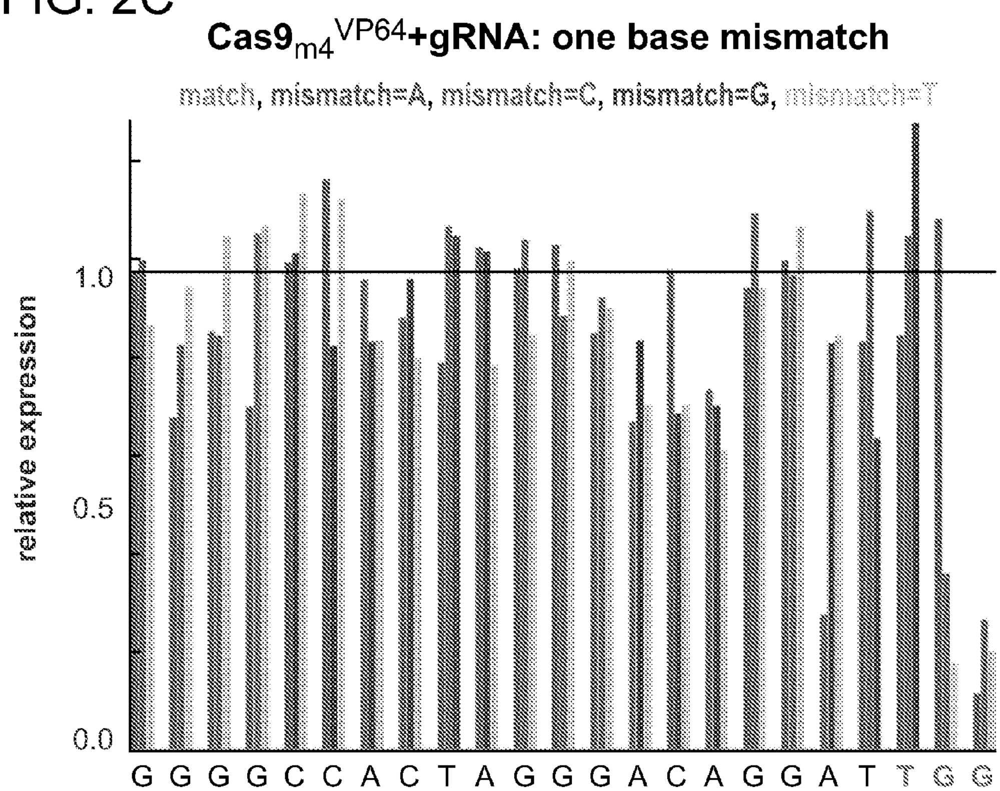

Cas9$_{m4}$^{VP64}+gRNA: one base mismatch
match, mismatch=A, mismatch=C, mismatch=G, mismatch=T
relative expression
1.0
0.5
0.0
G G G G C C A C T A G G G A C A G G A T T G G

FIG. 2D

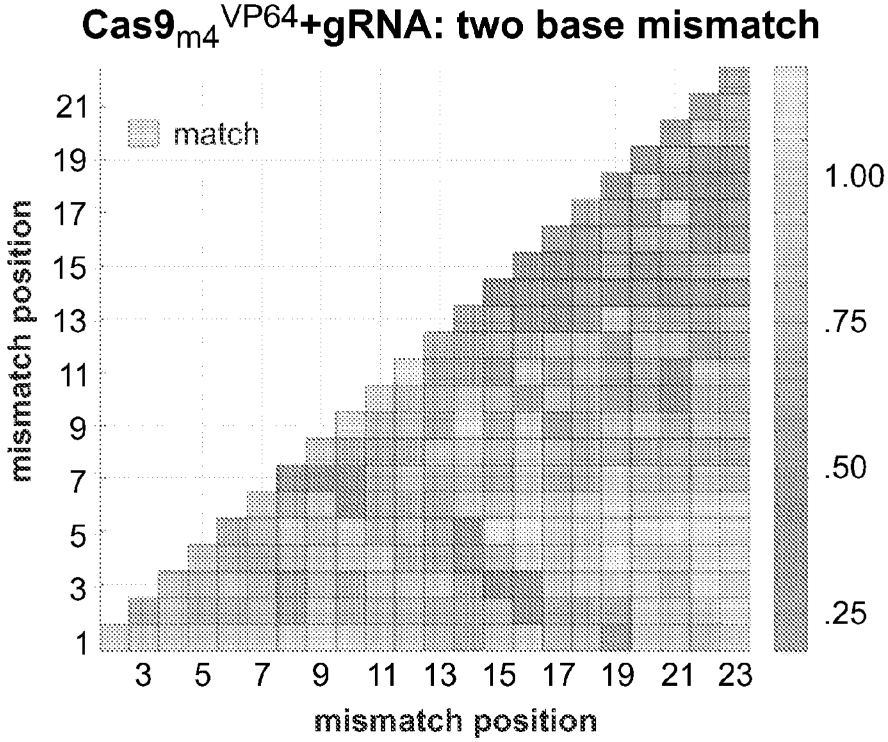

Cas9m4VP64+gRNA: two base mismatch
match
mismatch position
21
19
17
15
13
11
9
7
5
3
1
3
5
7
9
11
13
15
17
19
21
23
mismatch position
1.00
.75
.50
.25

FIG. 2E

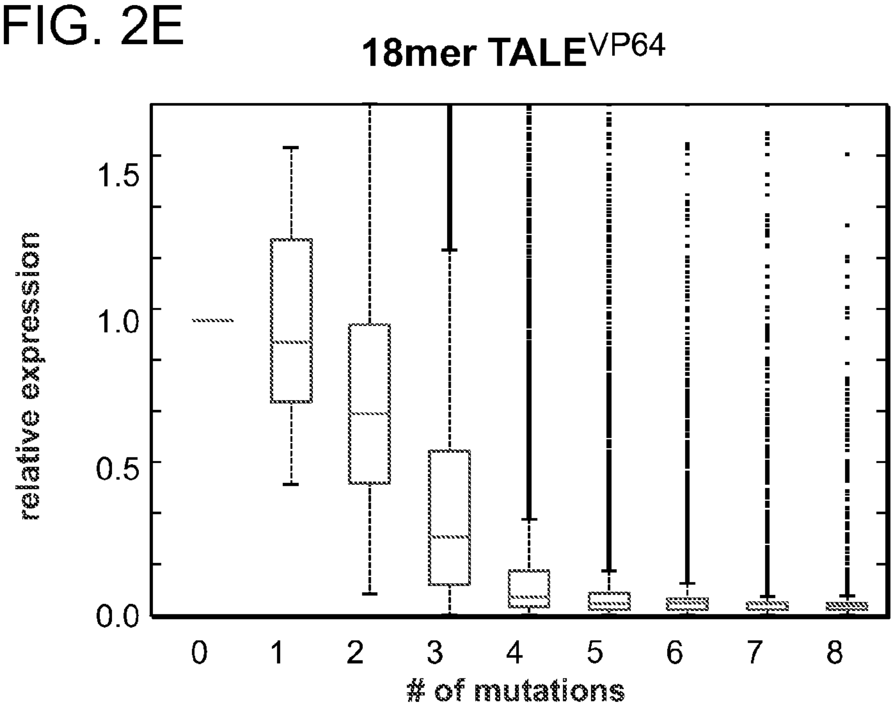

18mer TALEVP64
relative expression
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7
8
# of mutations

FIG. 2F

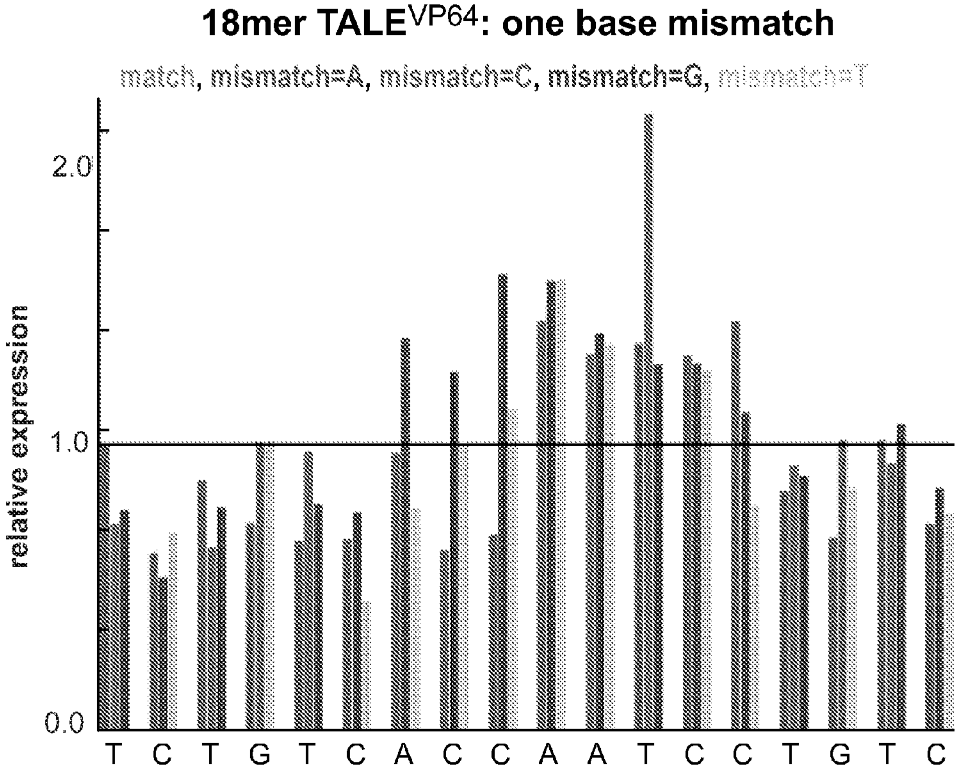

18mer TALE^{VP64}: one base mismatch
match, mismatch=A, mismatch=C, mismatch=G, mismatch=T
relative expression
2.0
1.0
0.0
T C T G T C A C C A A T C C T G T C

FIG. 2G

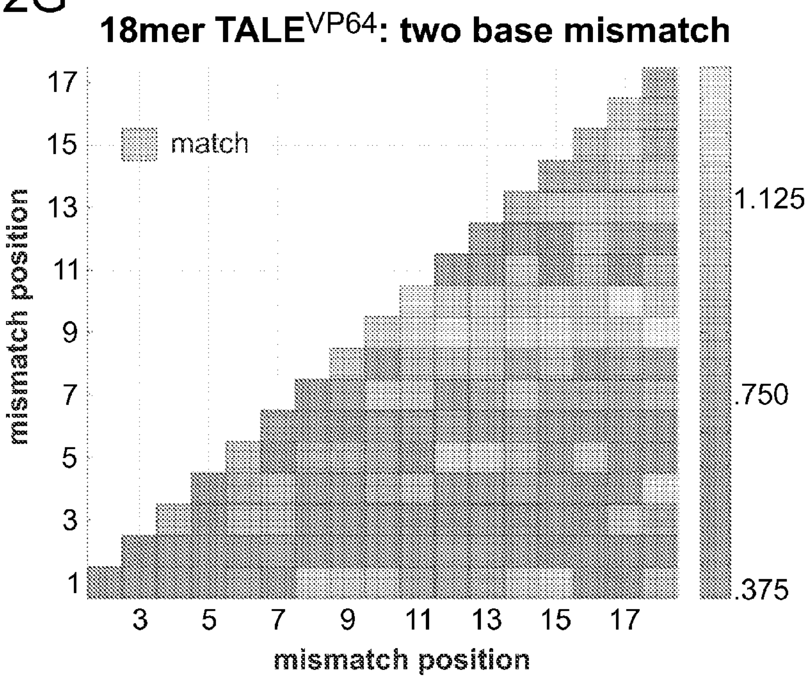

18mer TALE^{VP64}: two base mismatch
17
15
13
11
9
7
5
3
1
match
mismatch position
3 5 7 9 11 13 15 17
mismatch position
1.125
.750
.375

FIG. 3A

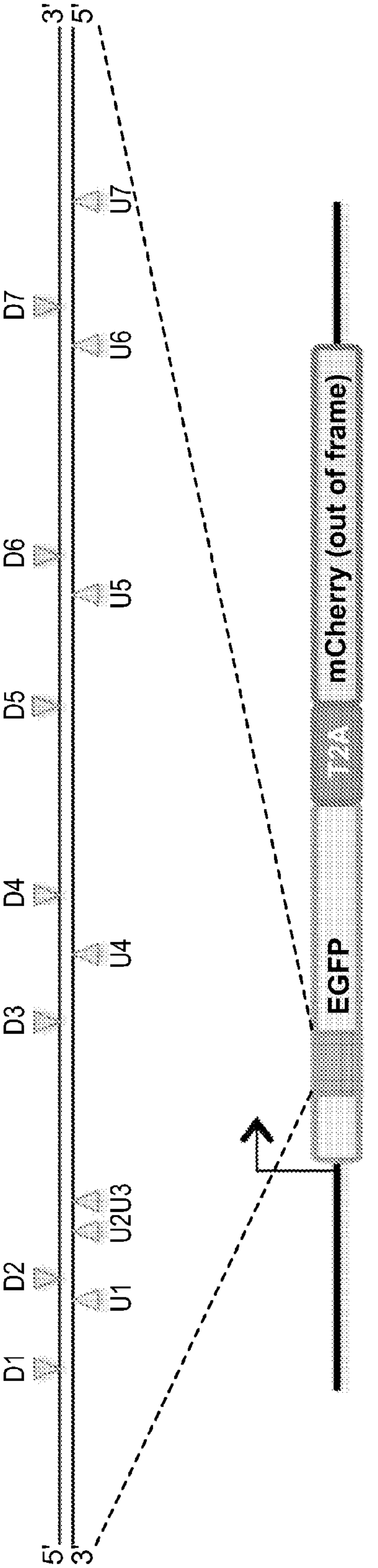
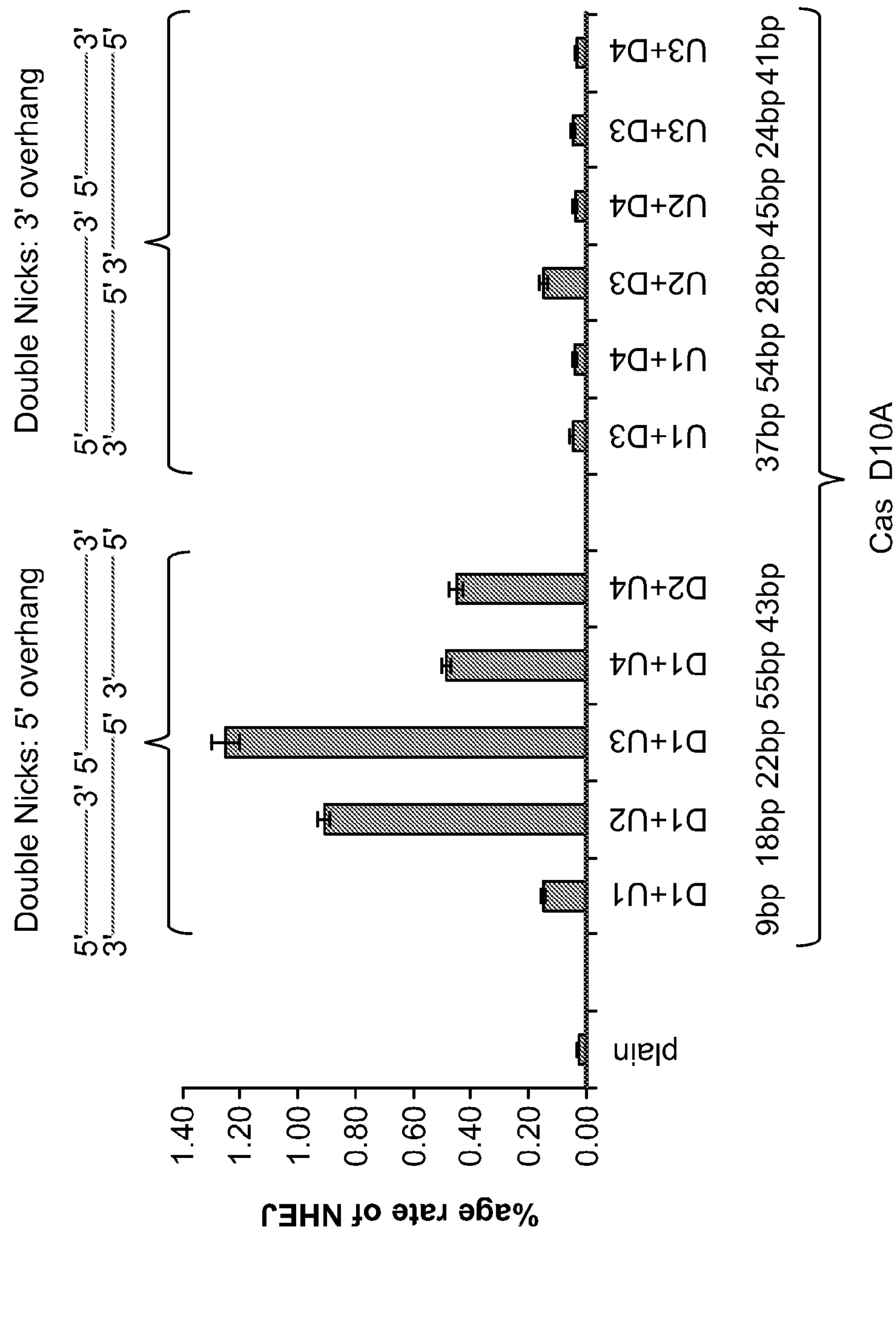
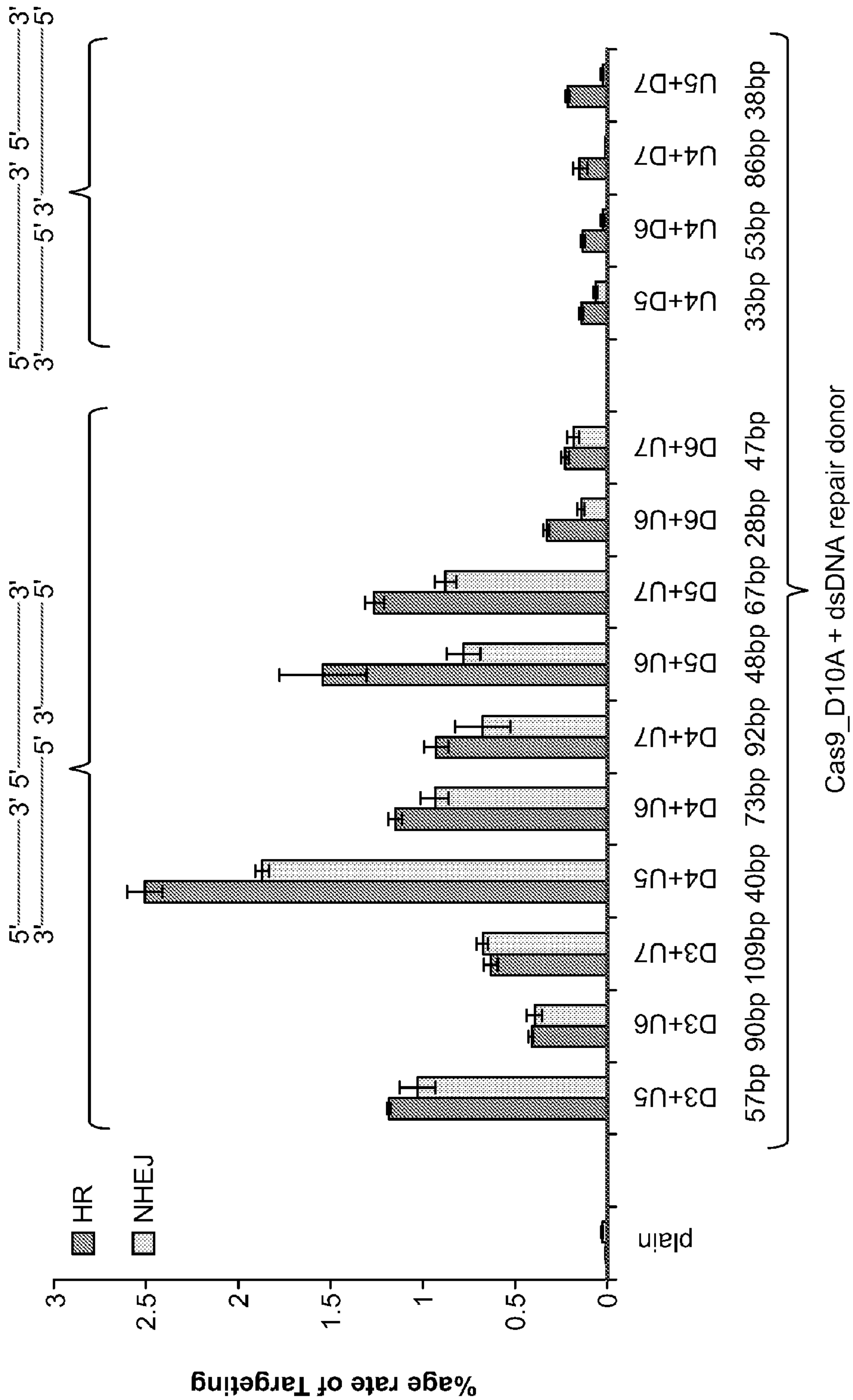

5'
3'
3'
5'
D1
D2
D3
D4
D5
D6
D7
U1
U2U3
U4
U5
U6
U7
EGFP
T2A
mCherry (out of frame)

FIG. 3B



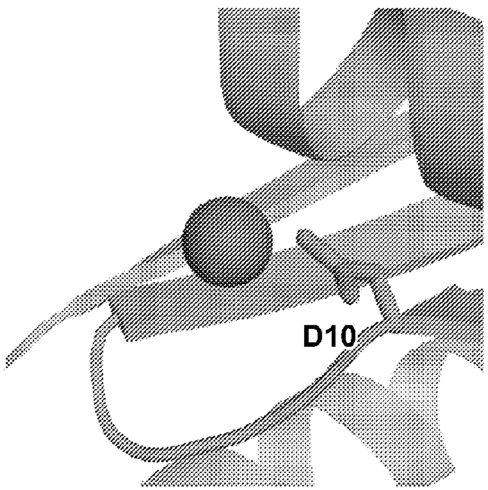
Double Nicks: 5' overhang
Double Nicks: 3' overhang
%age rate of NHEJ
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
plain
D1+U1
D1+U2
D1+U3
D1+U4
D2+U4
U1+D3
U1+D4
U2+D3
U2+D4
U3+D3
U3+D4
9bp 18bp 22bp 55bp 43bp
37bp 54bp 28bp 45bp 24bp 41bp
Cas_D10A

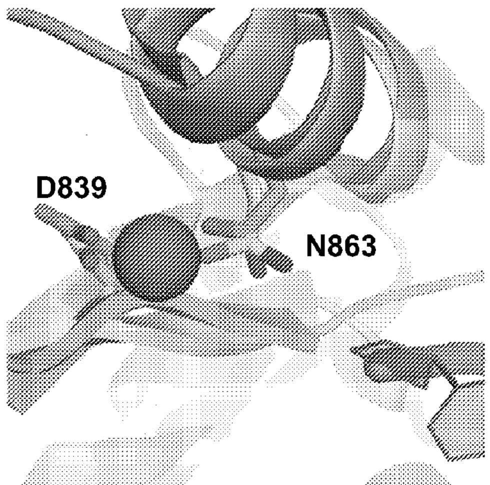
FIG. 3C



HR
NHEJ
%age rate of Targeting
0
0.5
1
1.5
2
2.5
3
plain
D3+U5
D3+U6
D3+U7
D4+U5
D4+U6
D4+U7
D5+U6
D5+U7
D6+U6
D6+U7
U4+D5
U4+D6
U4+D7
U5+D7
57bp 90bp 109bp 40bp 73bp 92bp 48bp 67bp 28bp 47bp
33bp 53bp 86bp 38bp
Double Nicks: 5' overhang
5'——3' 5'——3'
3'——5' 3'——5'
Double Nicks: 3' overhang
5'——5' 3'——3'
3'——3' 5'——5'
Cas9_D10A + dsDNA repair donor

FIG. 4A





Name	Mutations
Cas9	wild-type
$Cas9_{m1}$	D10A
$Cas9_{m2}$	D10A+H840A
$Cas9_{m3}$	D10A+D839A+H840A
$Cas9_{m4}$	D10A+D839A+H840A+N863A

FIG. 4B

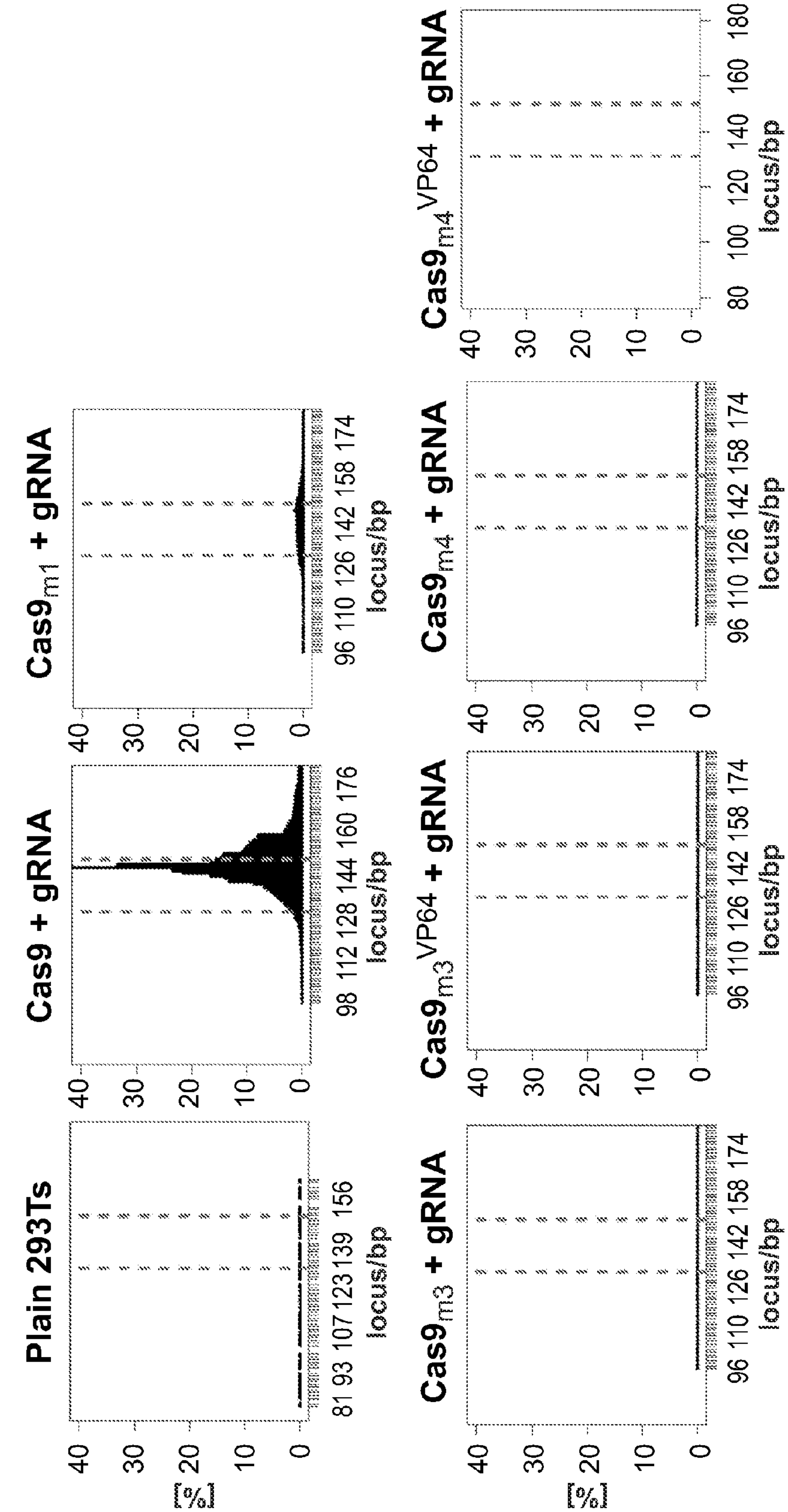
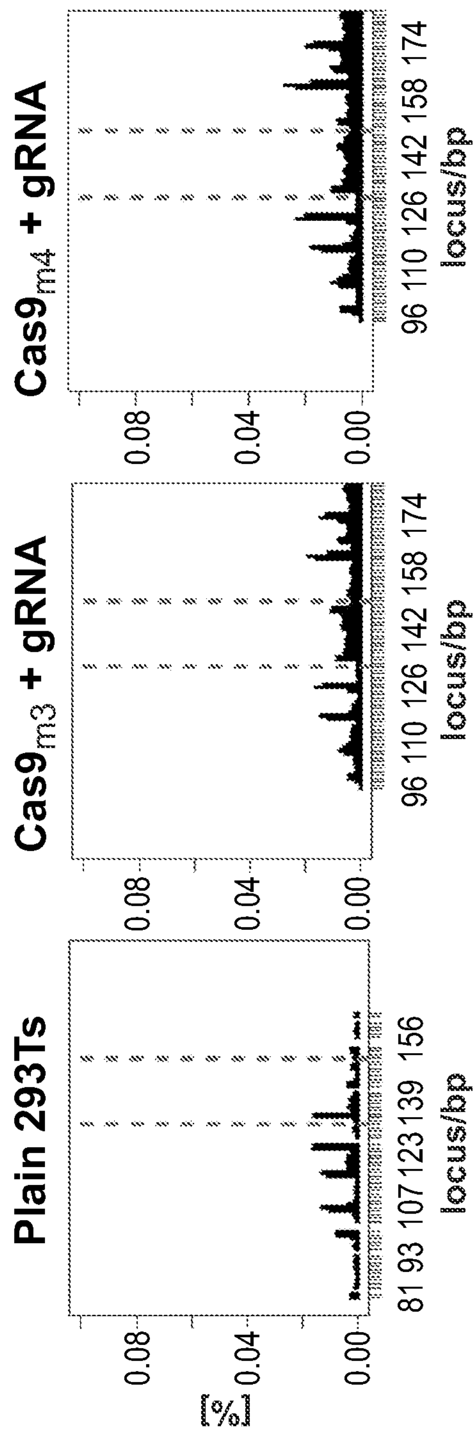

Plain 293Ts
Cas9 + gRNA
Cas9m1 + gRNA
Cas9m3 + gRNA
Cas9m3VP64 + gRNA
Cas9m4 + gRNA
Cas9m4VP64 + gRNA
[%]
locus/bp

FIG. 4C



Plain 293Ts
Cas9m3 + gRNA
Cas9m4 + gRNA
[%]
0.08
0.04
0.00
81 93 107 123 139 156
96 110 126 142 158 174
locus/bp

FIG. 5A

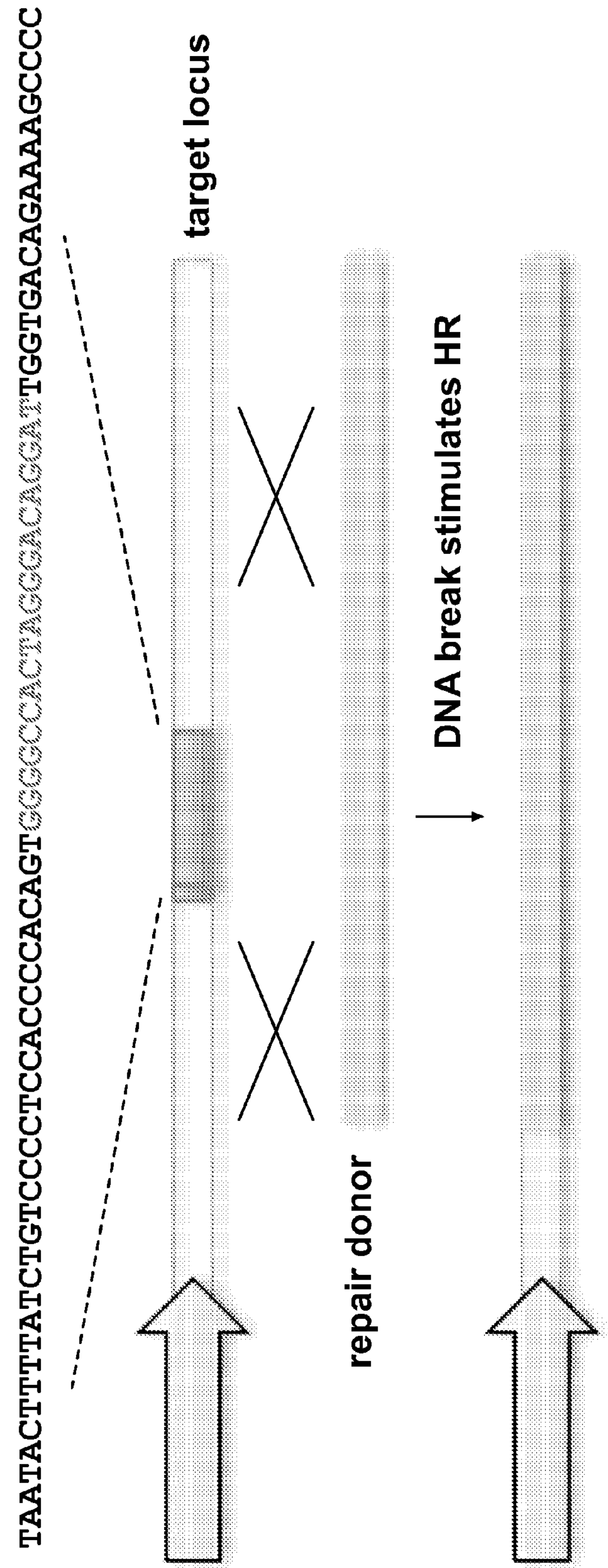
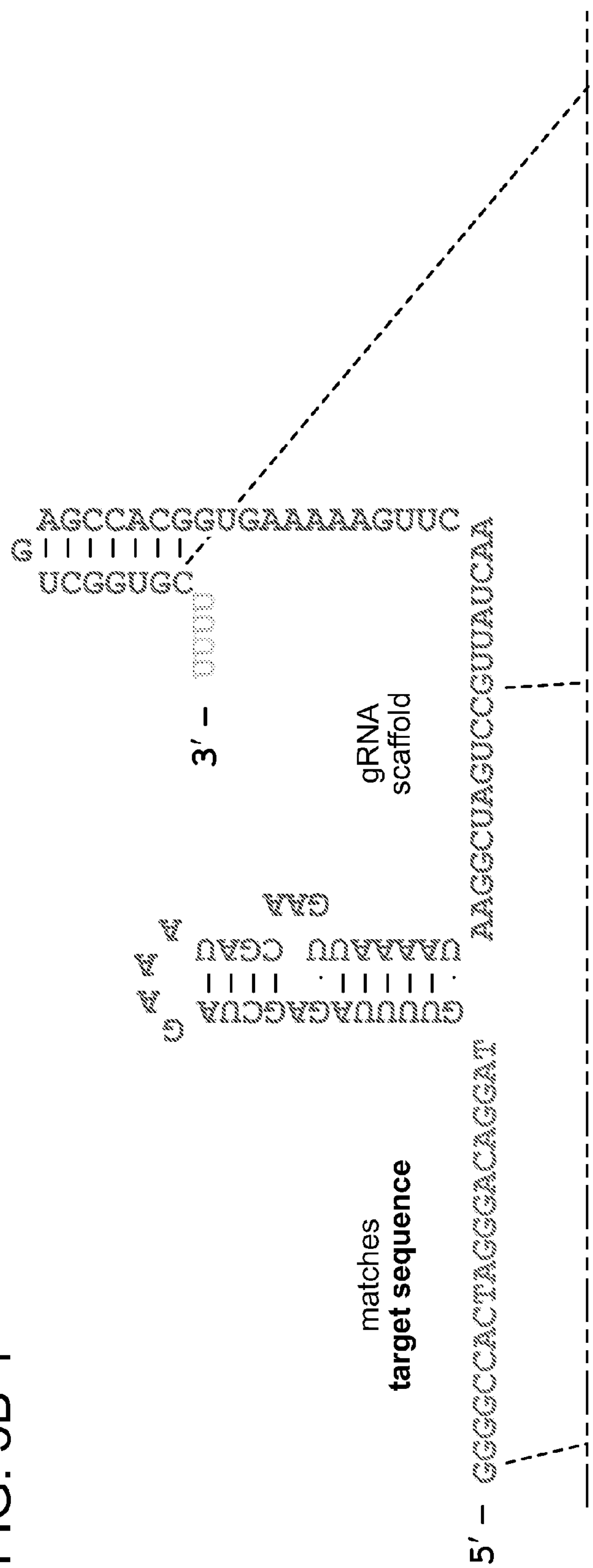

gRNA Target
TAATACTTTTATCTGTCCCCTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAAAGCCCC
target locus
repair donor
DNA break stimulates HR

FIG. 5B-1


5′ –
GGGGCCACTAGGGACAGGAT
matches
target sequence
GUUUUAGAGCUA
gRNA
scaffold
AAGGCUAGUCCGUUAUCAA
CUUGAAAAAGUGGCACCGA
G
UCGGUGC
3′ –

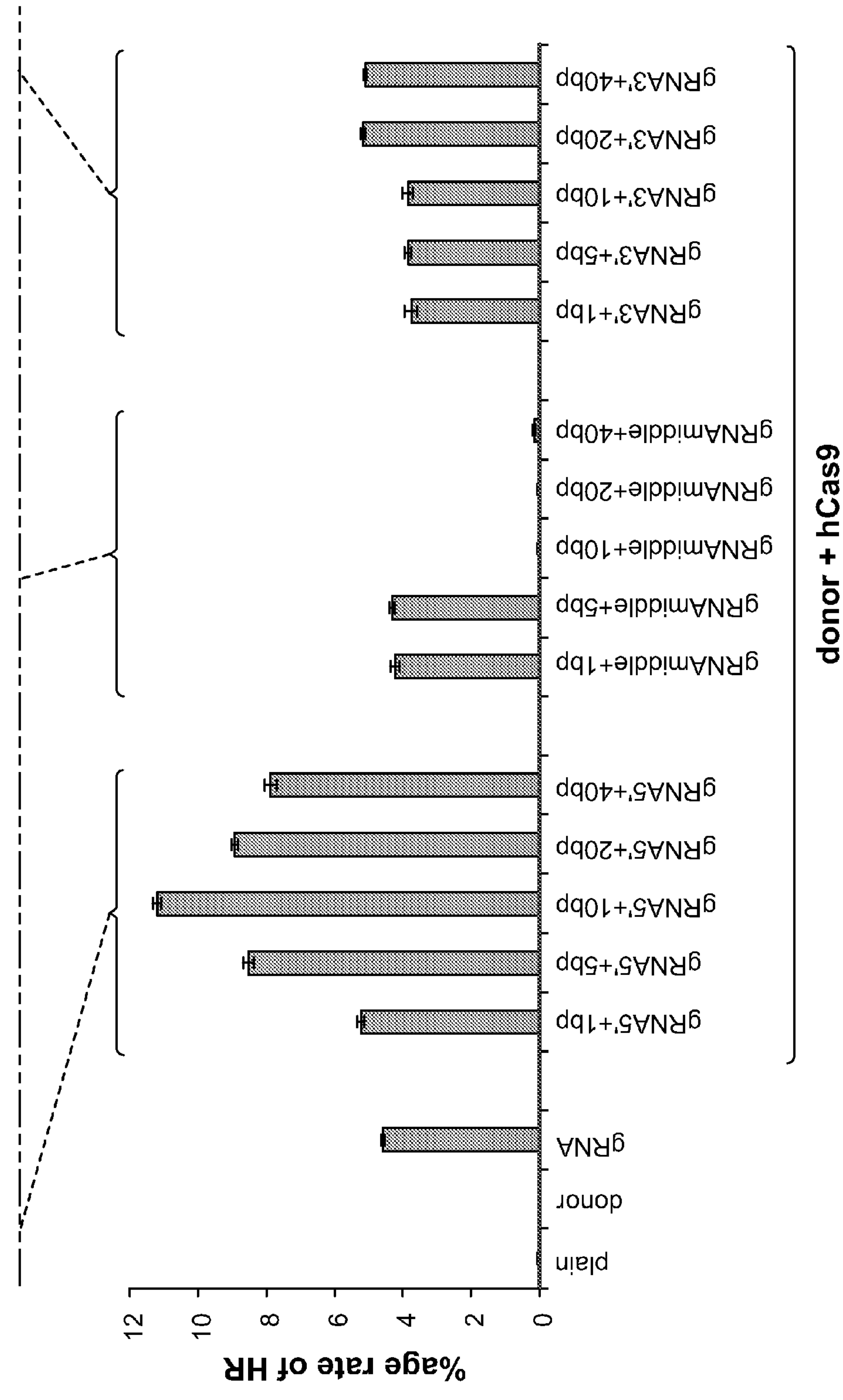

%age rate of HR
0
2
4
6
8
10
12
plain
donor
gRNA
gRNA5'+1bp
gRNA5'+5bp
gRNA5'+10bp
gRNA5'+20bp
gRNA5'+40bp
gRNAmiddle+1bp
gRNAmiddle+5bp
gRNAmiddle+10bp
gRNAmiddle+20bp
gRNAmiddle+40bp
gRNA3'+1bp
gRNA3'+5bp
gRNA3'+10bp
gRNA3'+20bp
gRNA3'+40bp
donor + hCas9

FIG. 5B-2

FIG. 6A

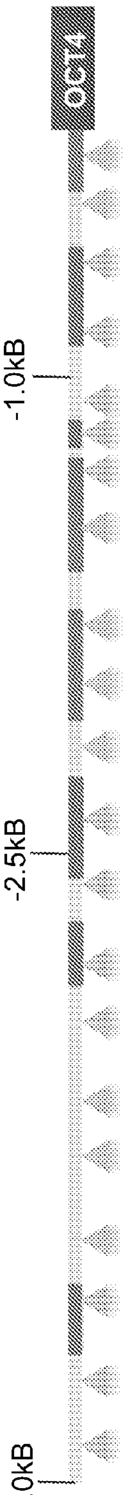

-5.0kB
-2.5kB
-1.0kB
OCT4

FIG. 6B

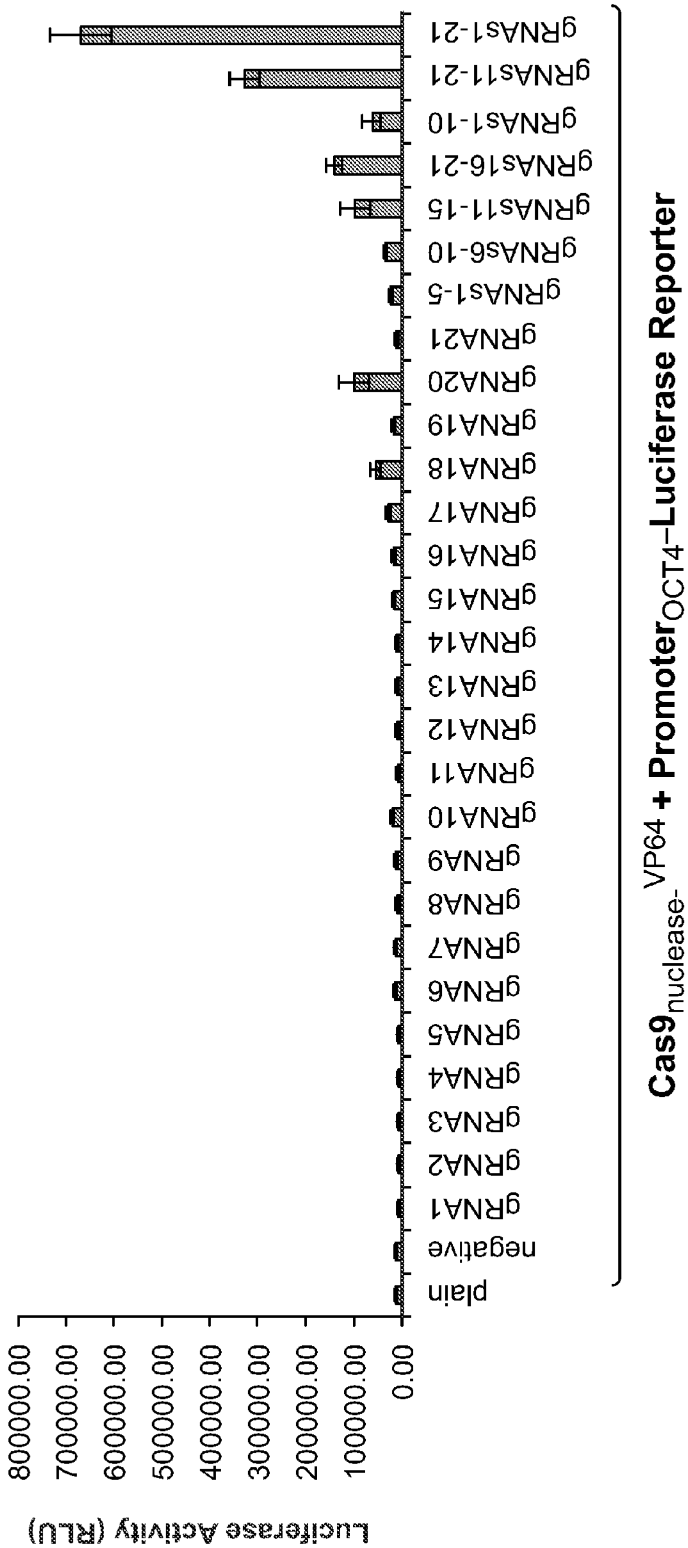
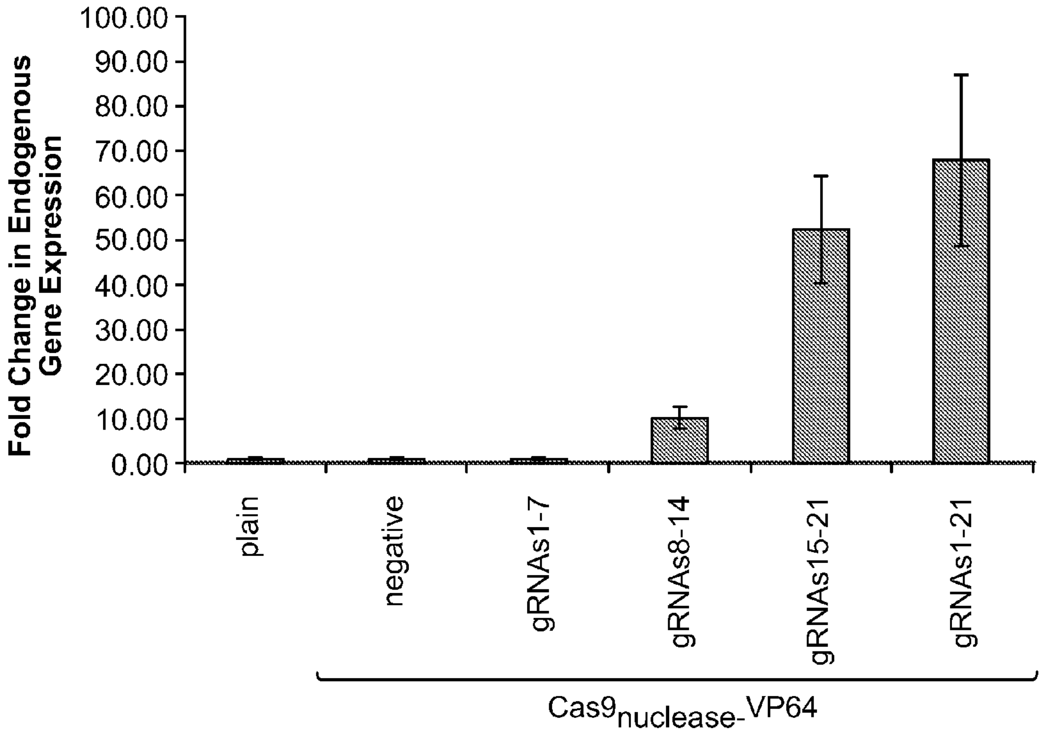

Luciferase Activity (RLU)
800000.00
700000.00
600000.00
500000.00
400000.00
300000.00
200000.00
100000.00
0.00
plain
negative
gRNA1
gRNA2
gRNA3
gRNA4
gRNA5
gRNA6
gRNA7
gRNA8
gRNA9
gRNA10
gRNA11
gRNA12
gRNA13
gRNA14
gRNA15
gRNA16
gRNA17
gRNA18
gRNA19
gRNA20
gRNA21
gRNAs1-5
gRNAs6-10
gRNAs11-15
gRNAs16-21
gRNAs1-10
gRNAs11-21
gRNAs1-21
Cas9nuclease-VP64 + PromoterOCT4–Luciferase Reporter

FIG. 6C



Fold Change in Endogenous Gene Expression
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
plain
negative
gRNAs1-7
gRNAs8-14
gRNAs15-21
gRNAs1-21
Cas9nuclease-VP64

FIG. 7A

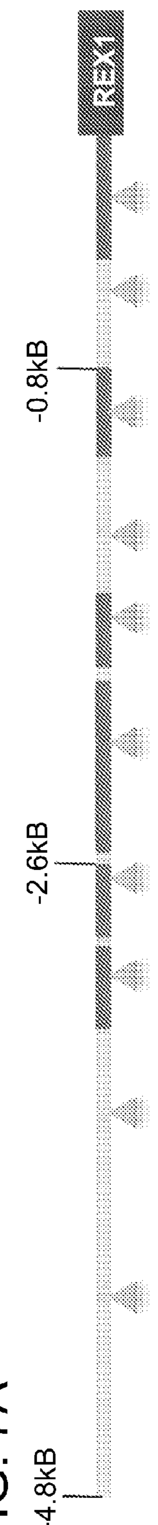

-4.8kB
-2.6kB
-0.8kB
REX1

FIG. 7B

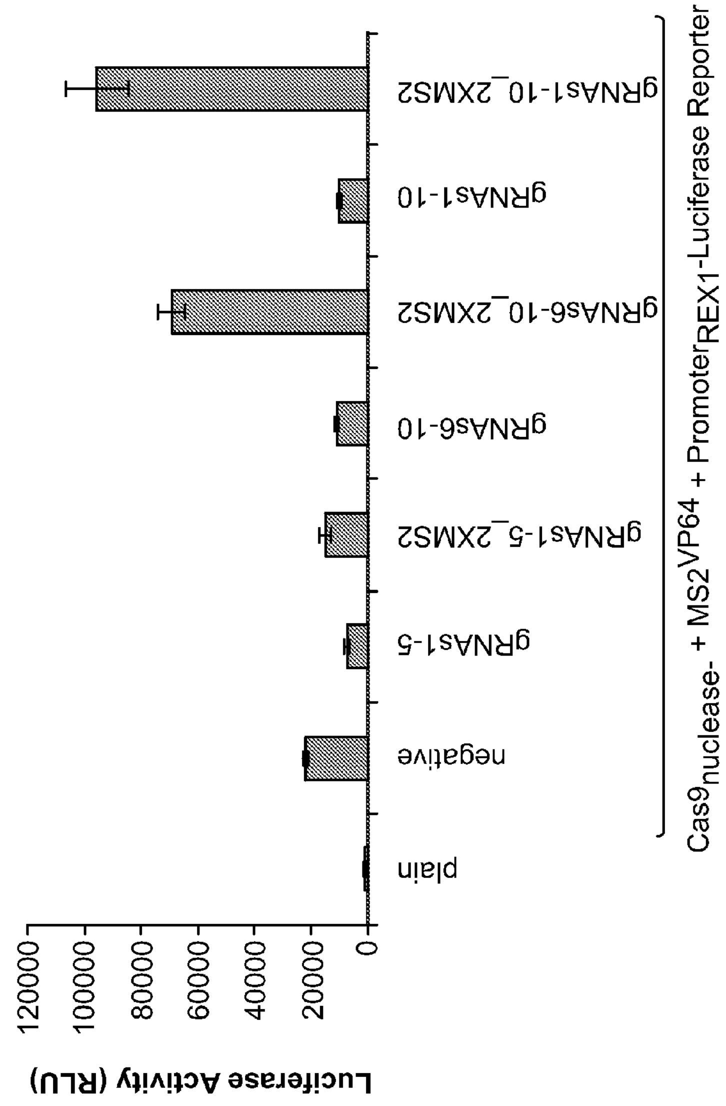
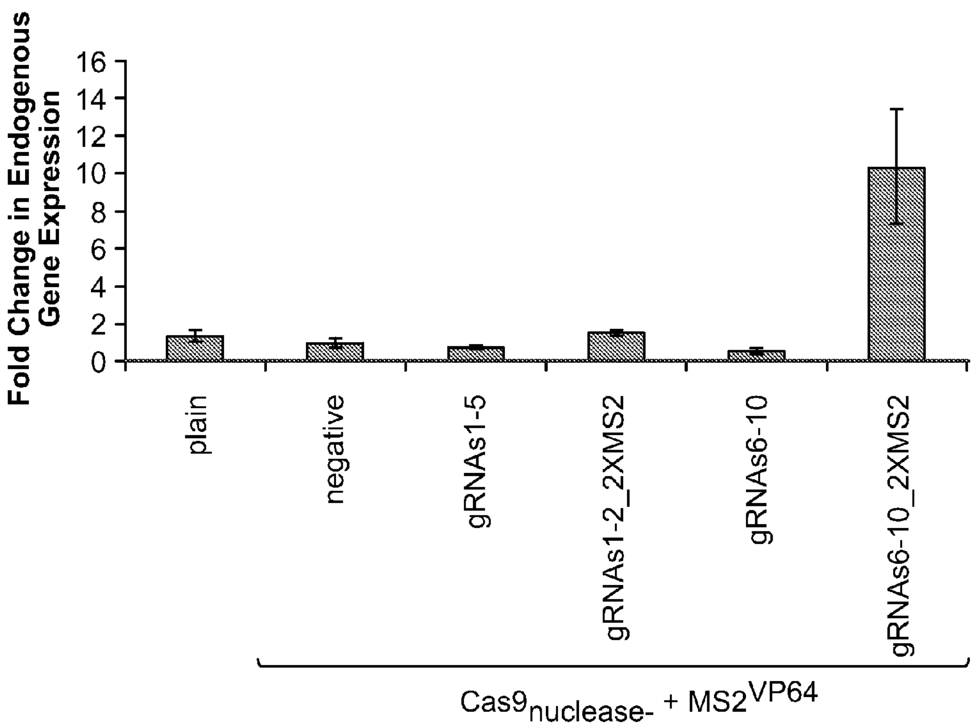
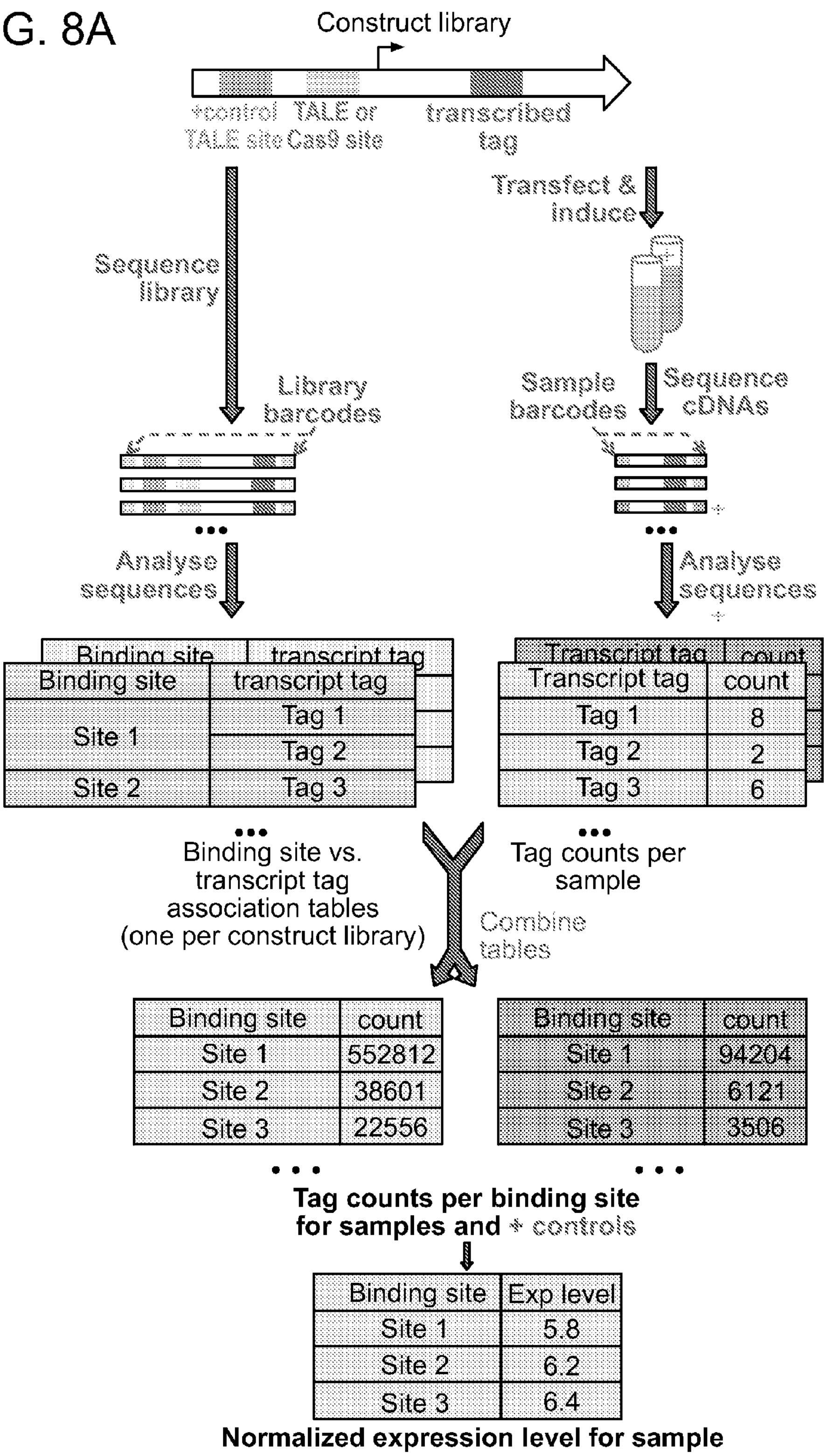

Luciferase Activity (RLU)
120000
100000
80000
60000
40000
20000
0
plain
negative
gRNAs1-5
gRNAs1-5_2XMS2
gRNAs6-10
gRNAs6-10_2XMS2
gRNAs1-10
gRNAs1-10_2XMS2
Cas9nuclease- + MS2VP64 + PromoterREX1-Luciferase Reporter

FIG. 7C



Fold Change in Endogenous Gene Expression
16
14
12
10
8
6
4
2
0
plain
negative
gRNAs1-5
gRNAs1-2_2XMS2
gRNAs6-10
gRNAs6-10_2XMS2
Cas9nuclease- + MS2VP64

FIG. 8A



Construct library
+control TALE site
TALE or Cas9 site
transcribed tag
Sequence library
Transfect & induce
Library barcodes
Sample barcodes
Sequence cDNAs
Analyse sequences
Analyse sequences
Binding site
transcript tag
Site 1
Tag 1
Tag 2
Site 2
Tag 3
Transcript tag
count
Tag 1
8
Tag 2
2
Tag 3
6
Binding site vs. transcript tag association tables (one per construct library)
Tag counts per sample
Combine tables
Binding site
count
Site 1
552812
Site 2
38601
Site 3
22556
Binding site
count
Site 1
94204
Site 2
6121
Site 3
3506
Tag counts per binding site for samples and + controls
Binding site
Exp level
Site 1
5.8
Site 2
6.2
Site 3
6.4
Normalized expression level for sample

FIG. 8B

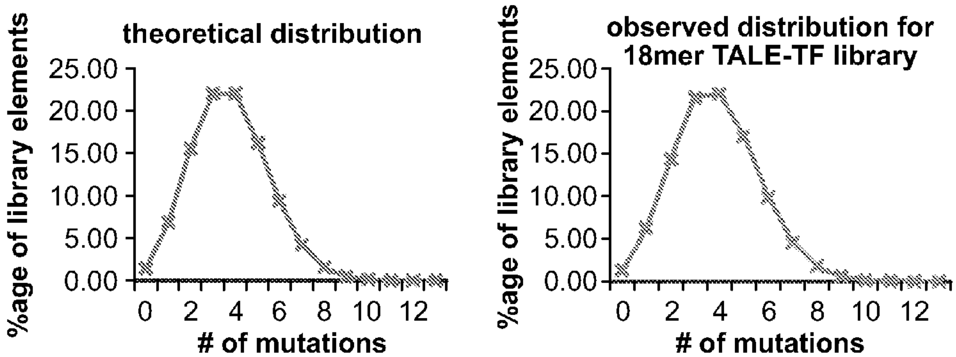

theoretical distribution
%age of library elements
25.00
20.00
15.00
10.00
5.00
0.00
0 2 4 6 8 10 12
# of mutations
observed distribution for
18mer TALE-TF library
%age of library elements
25.00
20.00
15.00
10.00
5.00
0.00
0 2 4 6 8 10 12
# of mutations

FIG. 8C

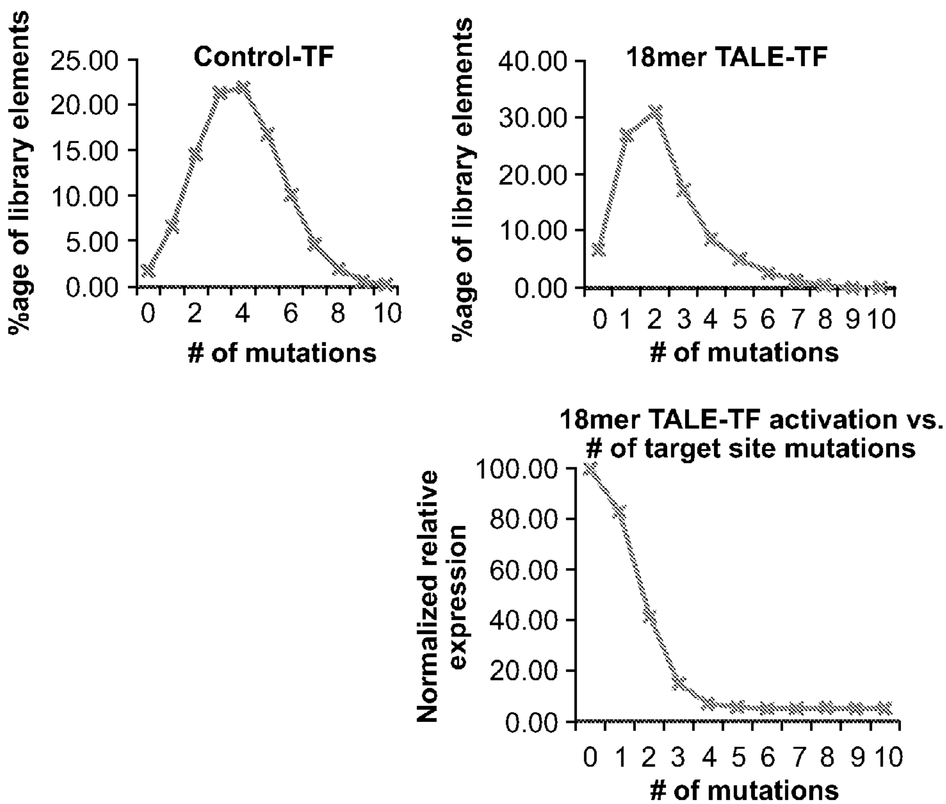
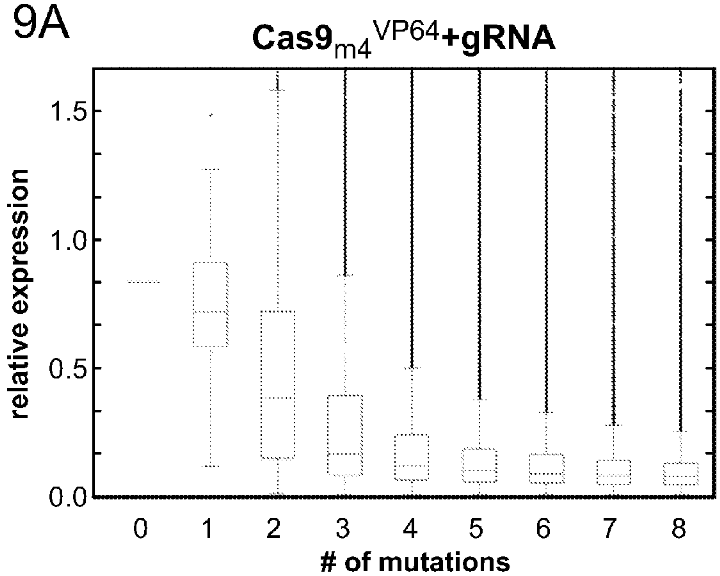
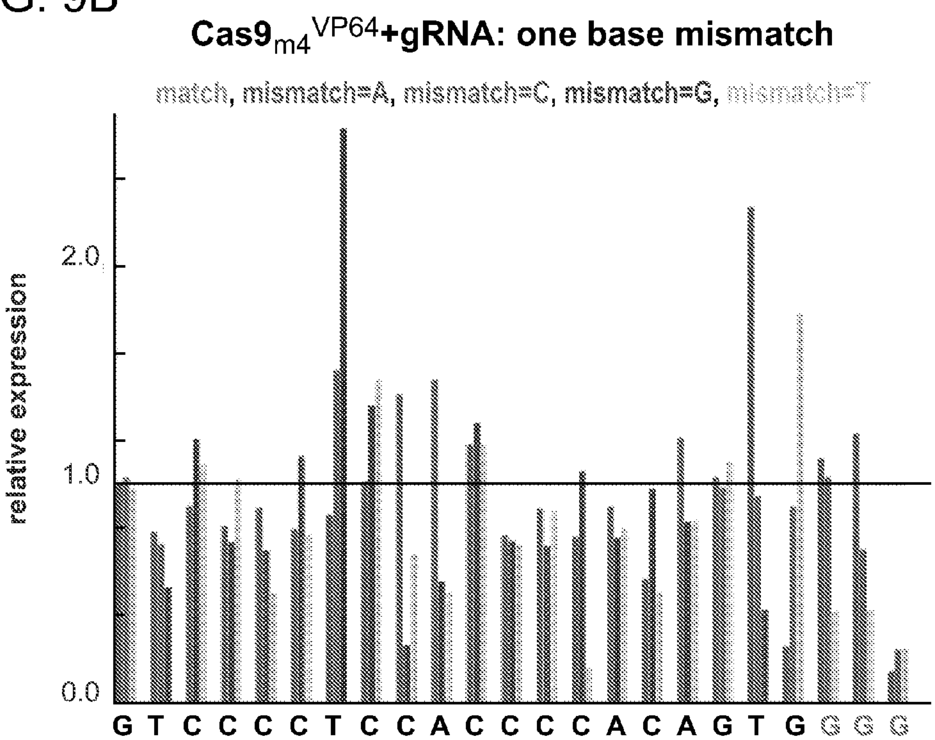

Control-TF
%age of library elements
25.00
20.00
15.00
10.00
5.00
0.00
0 2 4 6 8 10
# of mutations
18mer TALE-TF
%age of library elements
40.00
30.00
20.00
10.00
0.00
0 1 2 3 4 5 6 7 8 9 10
# of mutations
18mer TALE-TF activation vs.
# of target site mutations
Normalized relative
expression
100.00
80.00
60.00
40.00
20.00
0.00
0 1 2 3 4 5 6 7 8 9 10
# of mutations

FIG. 9A



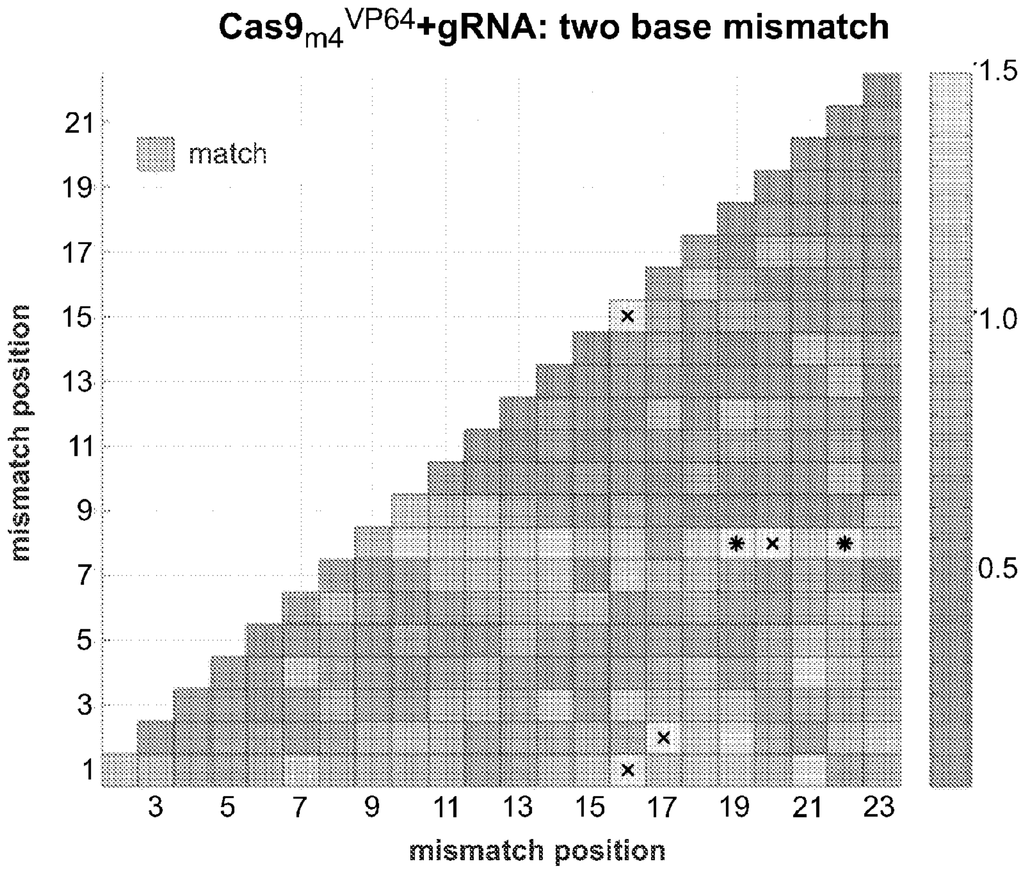
Cas9$_{m4}$^{VP64}+gRNA
relative expression
0.0
0.5
1.0
1.5
0
1
2
3
4
5
6
7
8
# of mutations

FIG. 9B



Cas9$_{m4}$^{VP64}+gRNA: one base mismatch
match, mismatch=A, mismatch=C, mismatch=G, mismatch=T
relative expression
0.0
1.0
2.0
G T C C C C T C C A C C C A C A G T G G G G

FIG. 9C



Cas9$_{m4}$^{VP64}+gRNA: two base mismatch
match
mismatch position
mismatch position
1
3
5
7
9
11
13
15
17
19
21
23
1.5
1.0
0.5

FIG. 9D

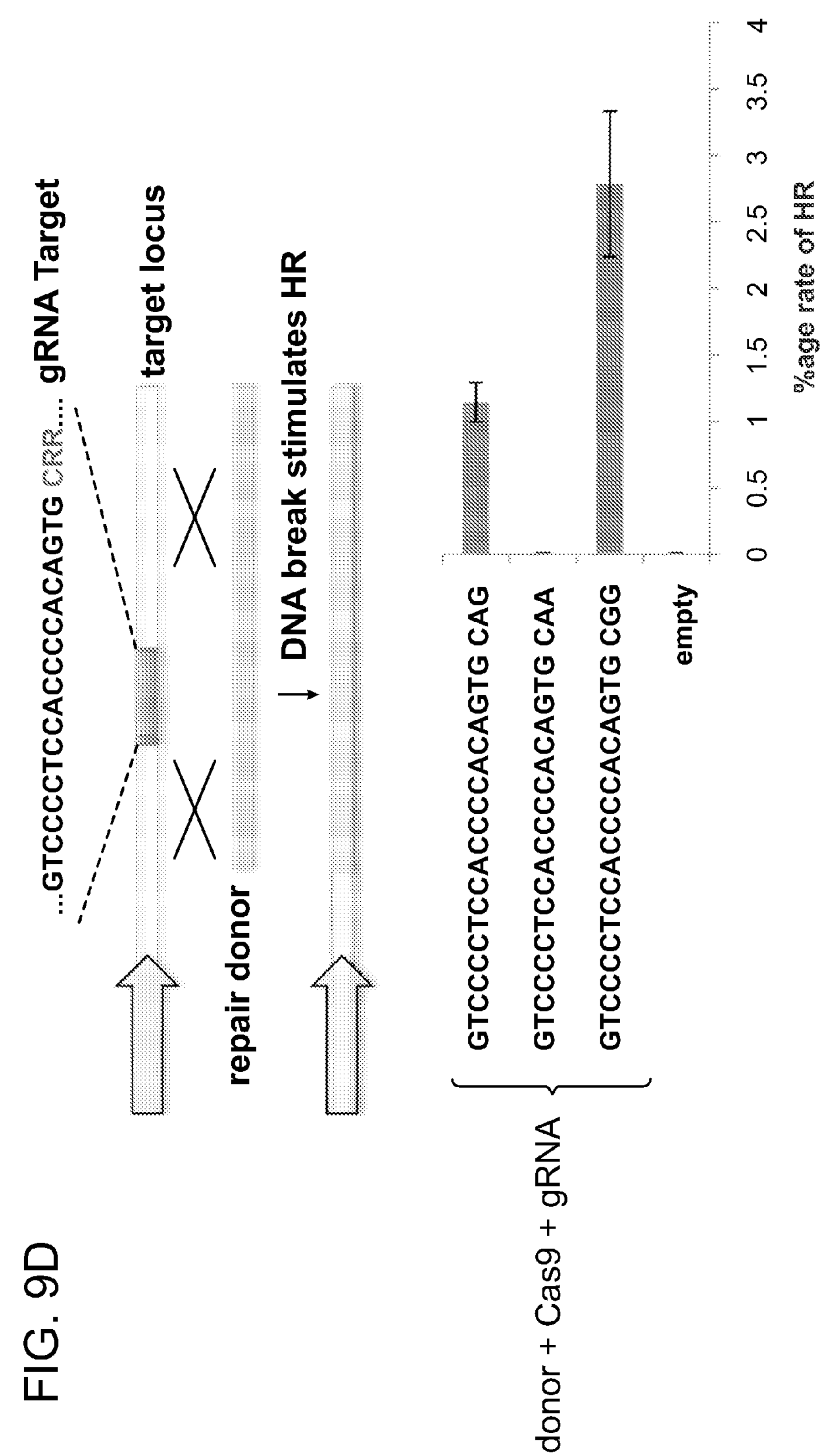

...GTCCCCTCCACCCCACAGTG CRR.... gRNA Target
target locus
repair donor
DNA break stimulates HR
donor + Cas9 + gRNA
GTCCCCTCCACCCCACAGTG CAG
GTCCCCTCCACCCCACAGTG CAA
GTCCCCTCCACCCCACAGTG CGG
empty
0
0.5
1
1.5
2
2.5
3
3.5
4
%age rate of HR

FIG. 10A-1

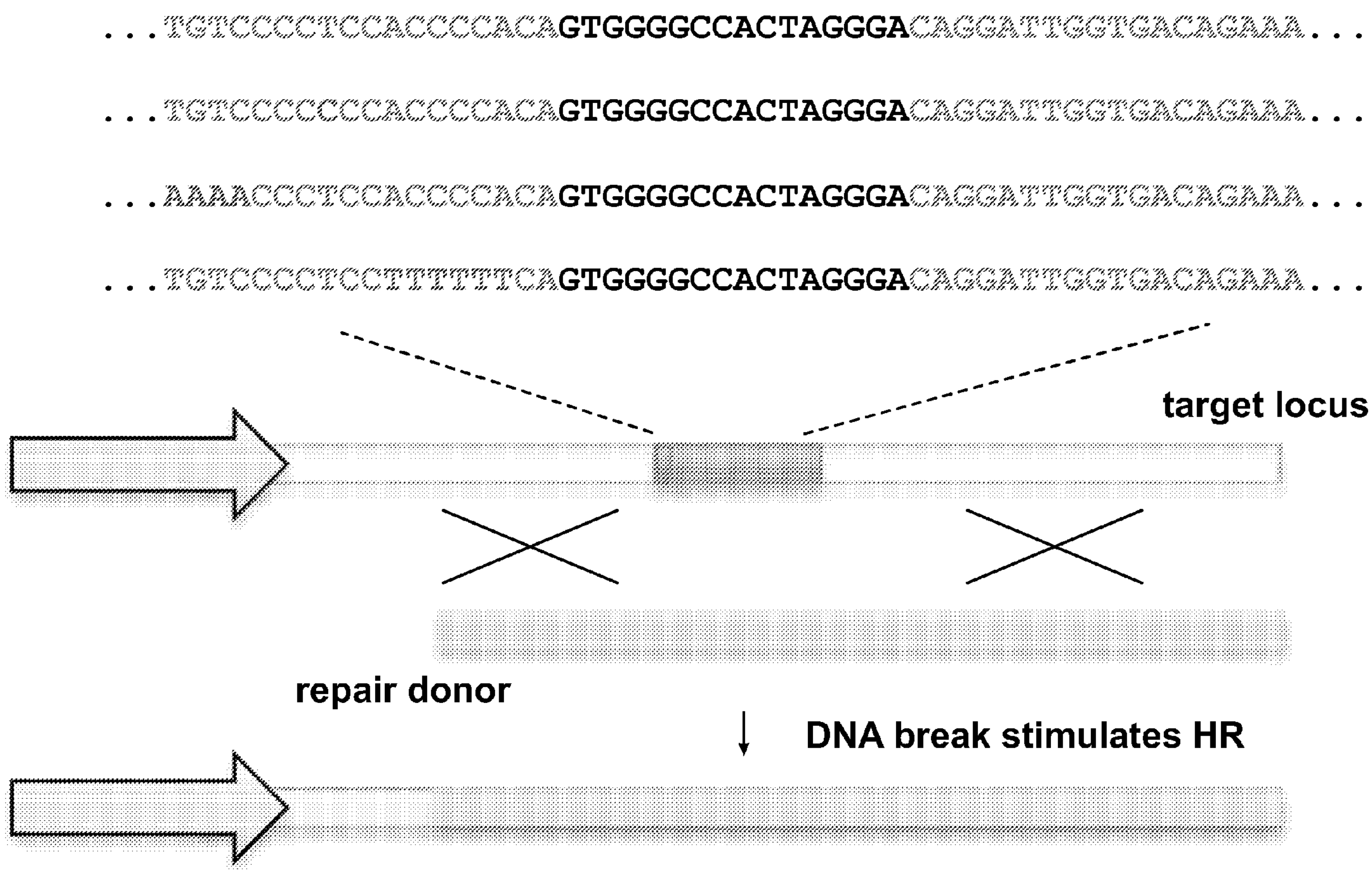

...TGTCCCCTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
...TGTCCCCCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
...AAAACCCTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
...TGTCCCCTCCTTTTTCAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
target locus
repair donor
DNA break stimulates HR

FIG. 10A-2

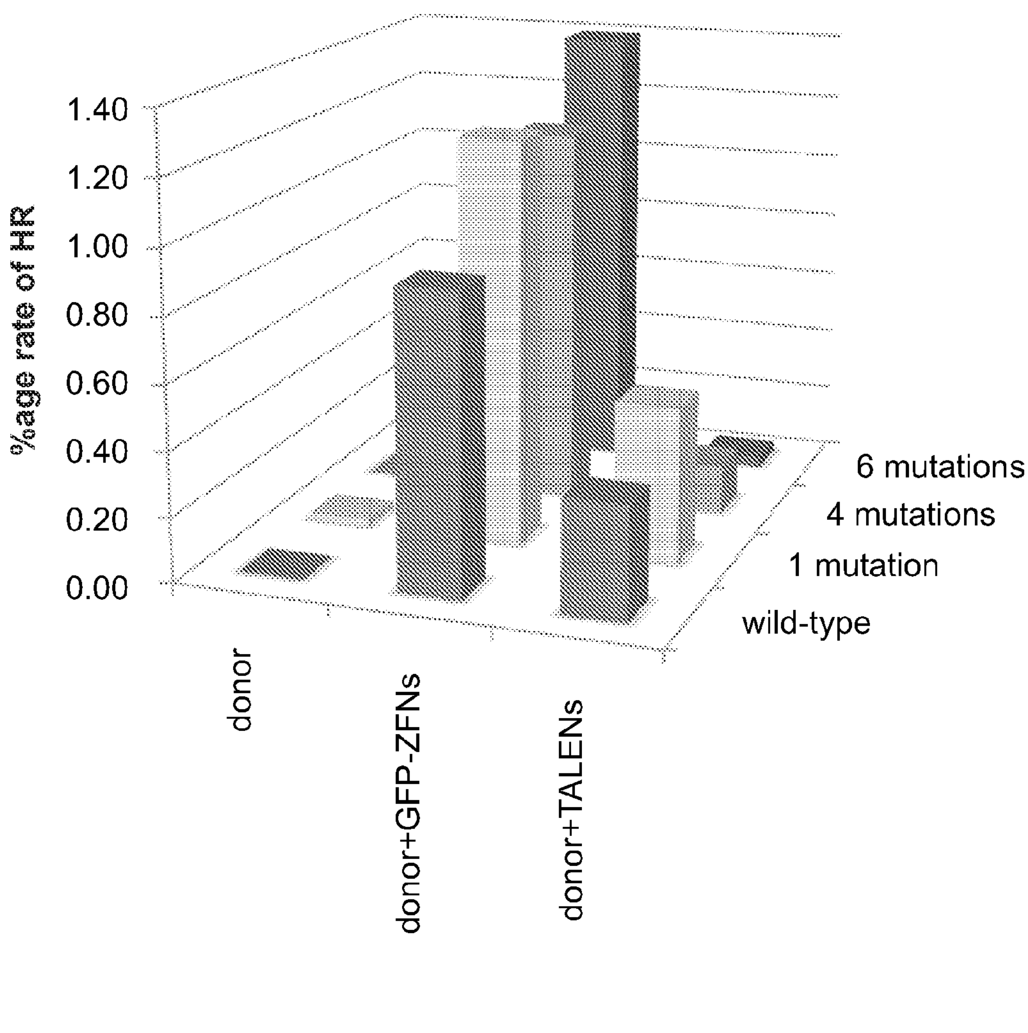

1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
%age rate of HR
6 mutations
4 mutations
1 mutation
wild-type
donor
donor+GFP-ZFNs
donor+TALENs

FIG. 10B

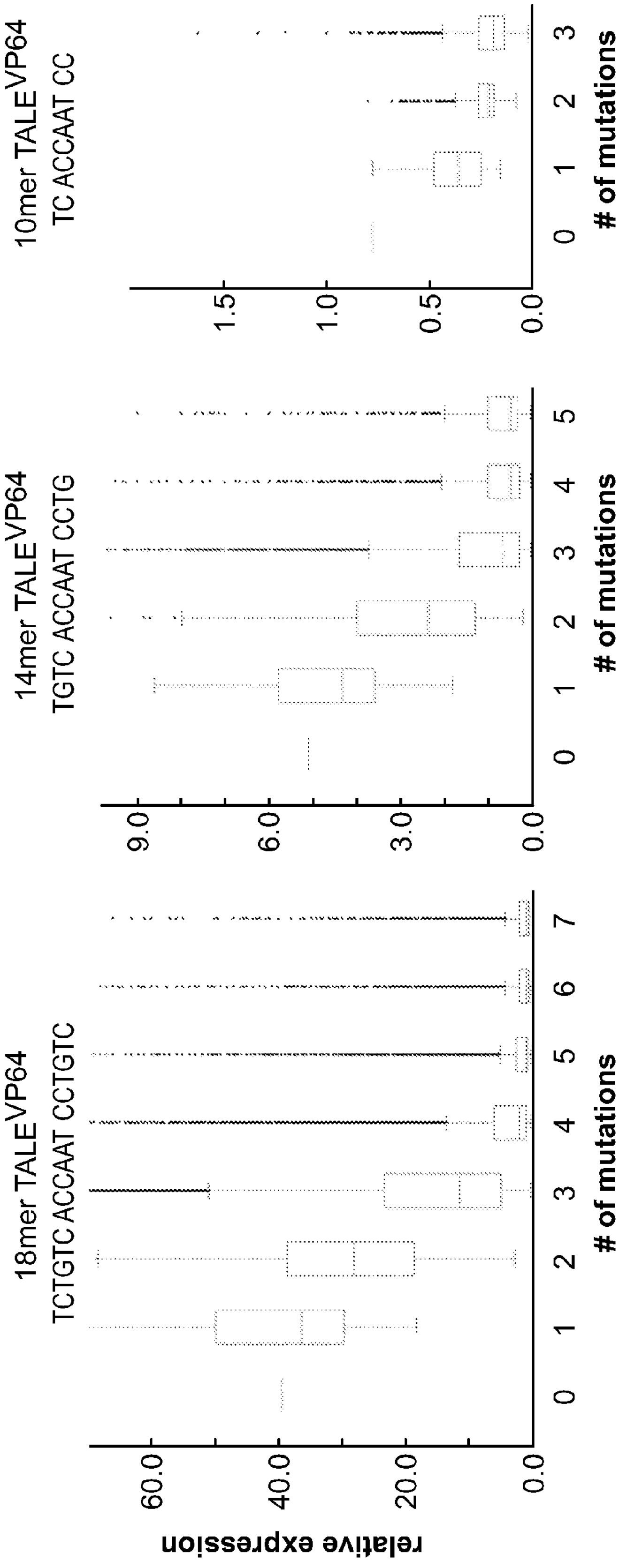

18mer TALE VP64
TCTGTC ACCAAT CCTGTC
14mer TALE VP64
TGTC ACCAAT CCTG
10mer TALE VP64
TC ACCAAT CC
relative expression
60.0
40.0
20.0
0.0
9.0
6.0
3.0
0.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7
# of mutations

FIG. 10C

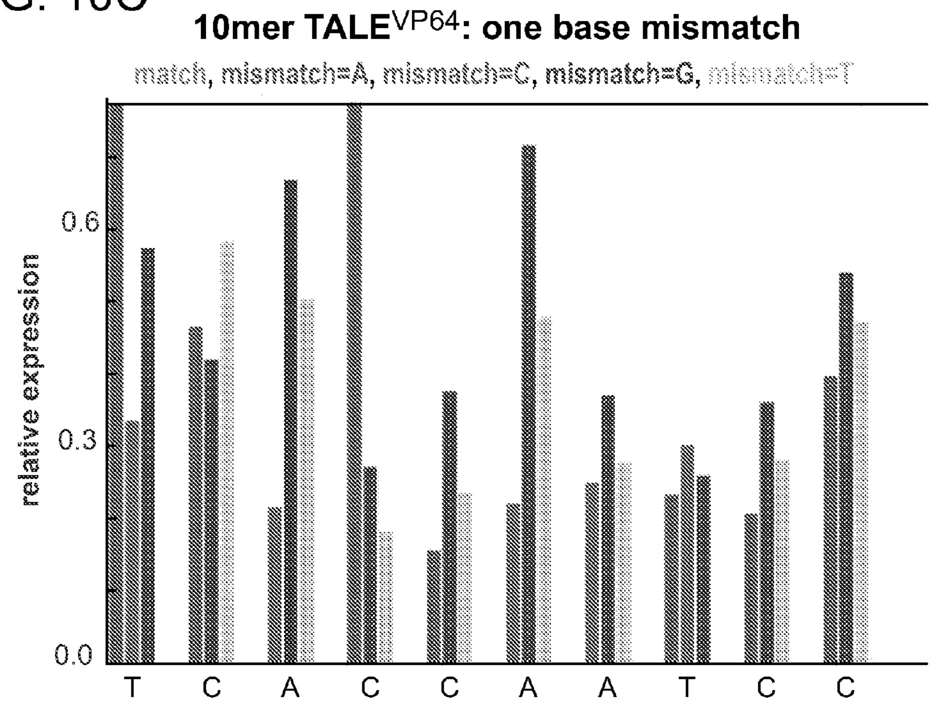


FIG. 10D

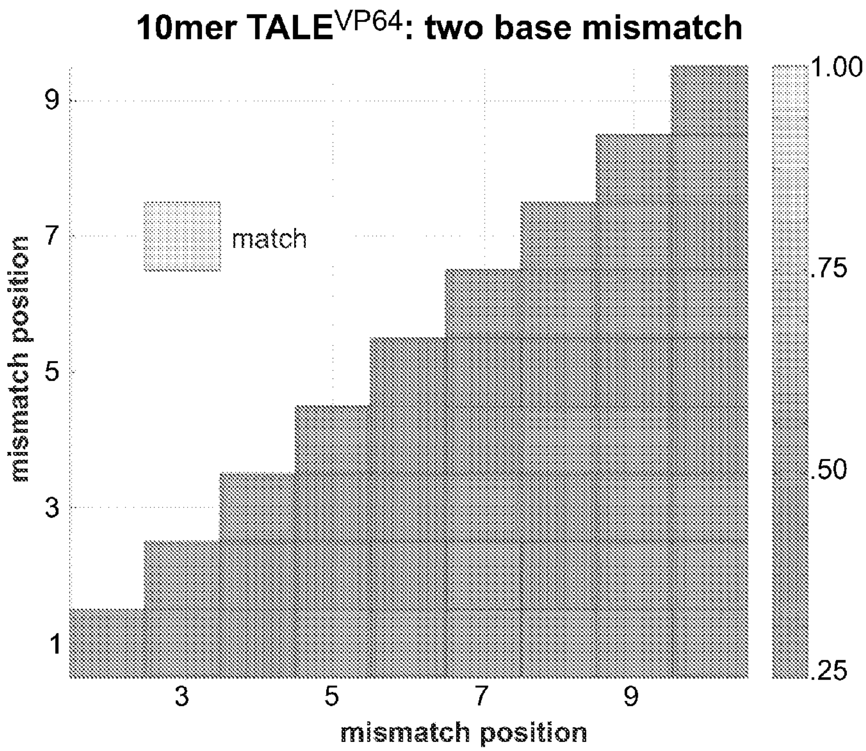


FIG. 11A

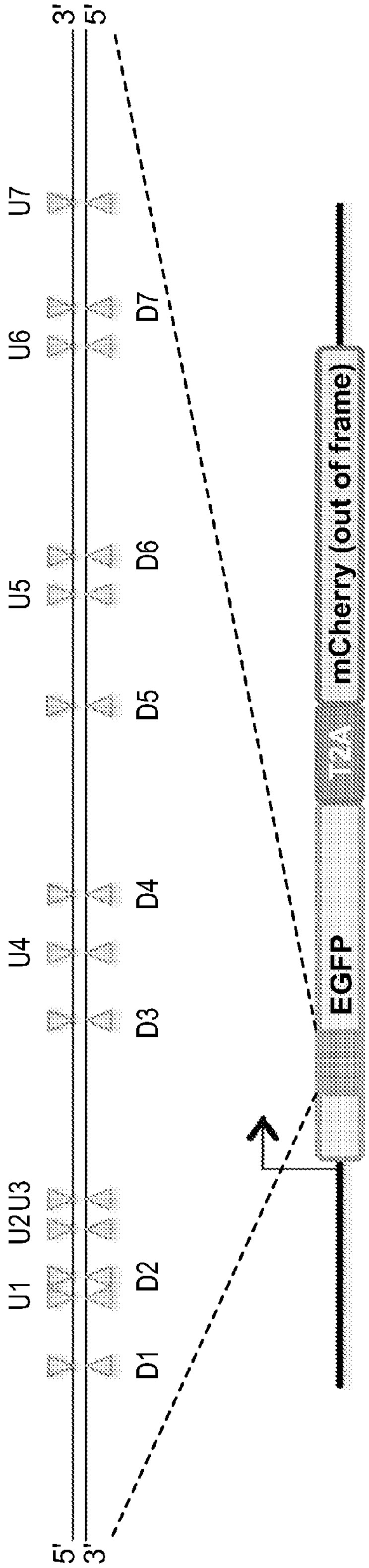

5'
3'
3'
5'
U1
U2
U3
U4
U5
U6
U7
D1
D2
D3
D4
D5
D6
D7
EGFP
T2A
mCherry (out of frame)

FIG. 11B

gRNA Target Sequence	
	plain
CACCGGGGTGGTGCCCATCC TGG	U1
GGTGCCCATCCTGGTCGAGC TGG	U2
CCCATCCTGGTCGAGCTGGA CGG	U3
GGCCACAAGTTCAGCGTGTC CGG	U4
CGCAAATAAGAGCTCACCTA CGG	U5
CTGAAGTTCATCTGCACCAC CGG	U6
CCGGCAAGCTGCCCGTGCCC TGG	U7
GACCAGGATGGGCACCACCC CGG	D1
GCCGTCCAGCTCGACCAGGA TGG	D2
GGCCGGACACGCTGAACTTG TGG	D3
TAACAGGGTAATGTCGAGGC CGG	D4
AGGTGAGCTCTTATTTGCGT AGG	D5
CTTCAGGGTCAGCTTGCCGT AGG	D6
GGGCACGGGCAGCTTGCCGG TGG	D7

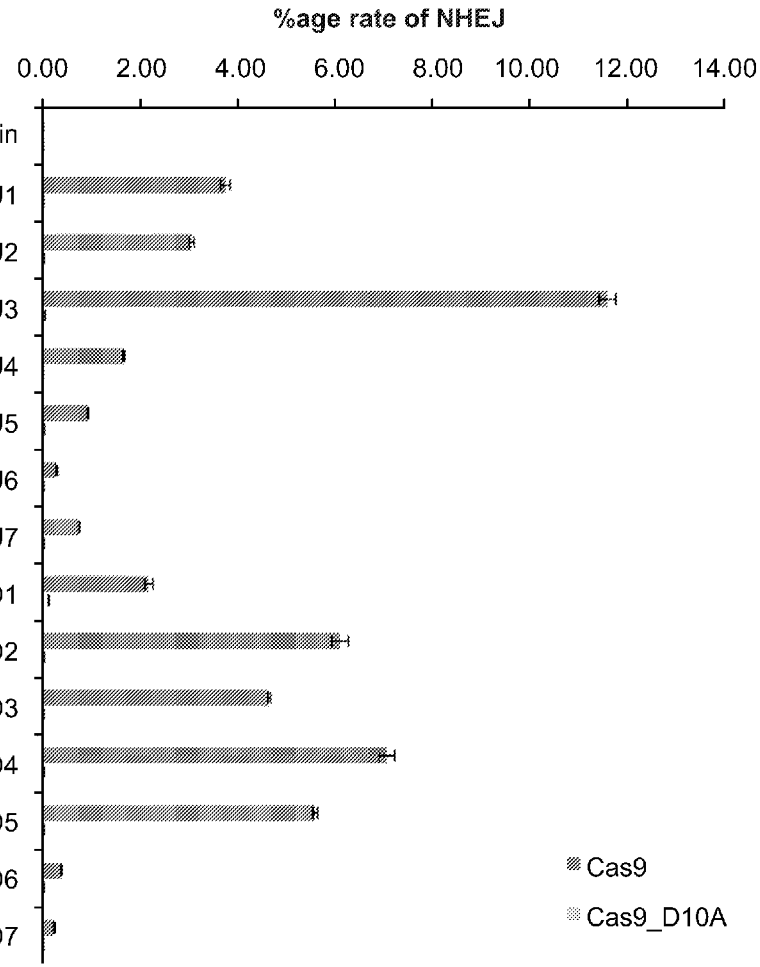


FIG. 12A

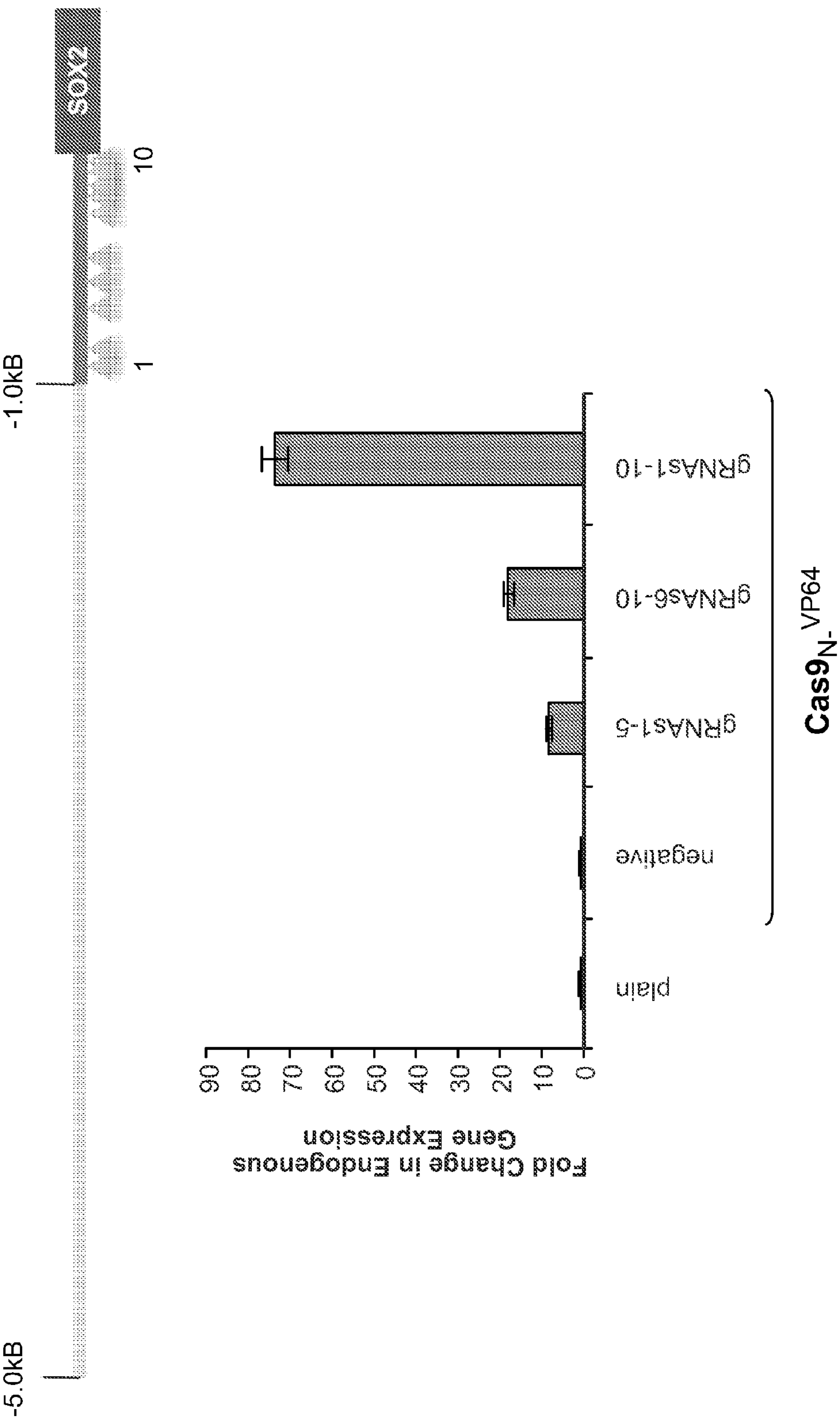
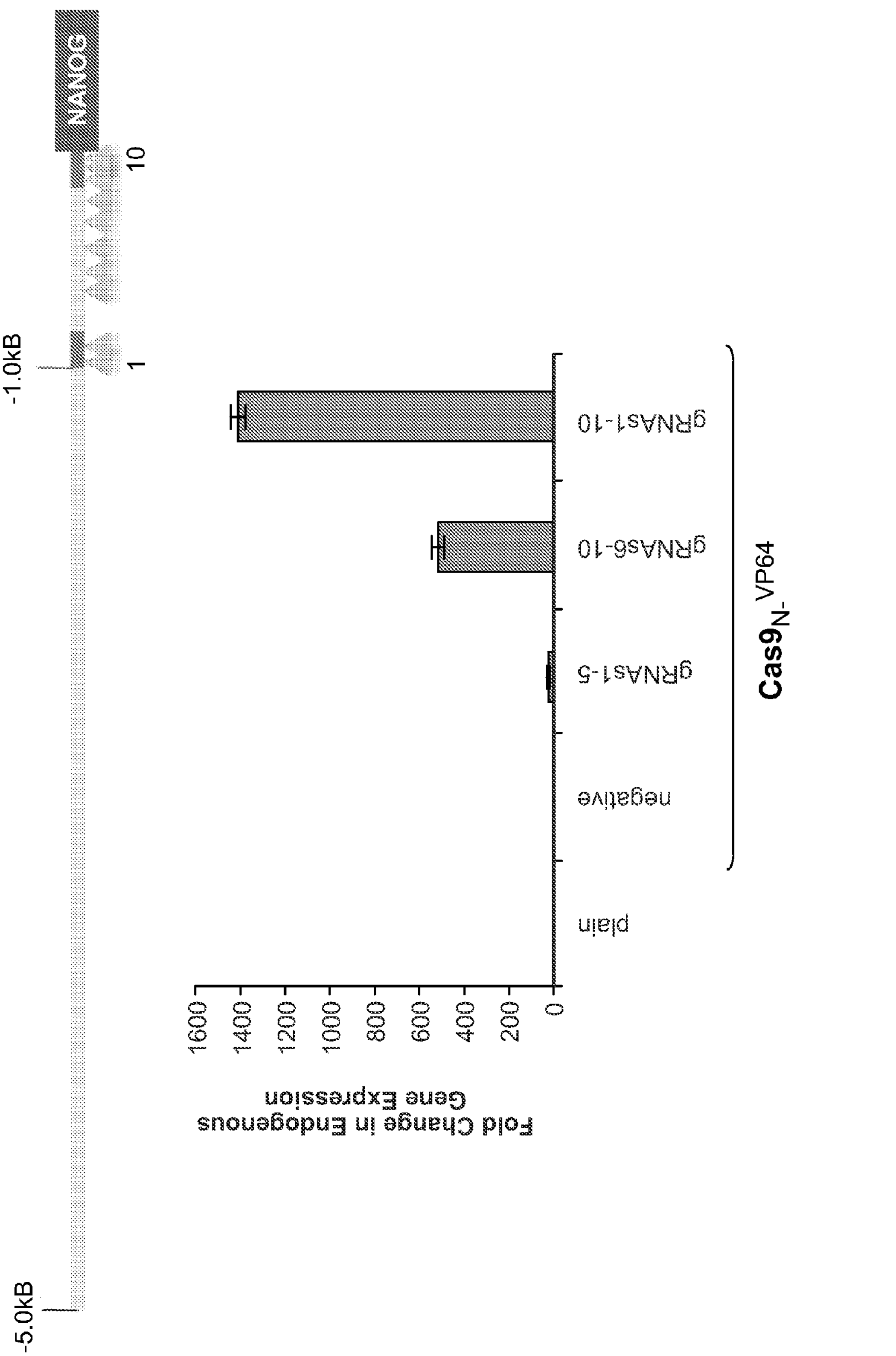

-5.0kB
-1.0kB
1
10
SOX2
Fold Change in Endogenous Gene Expression
0
10
20
30
40
50
60
70
80
90
plain
negative
gRNAs1-5
gRNAs6-10
gRNAs1-10
Cas9N-VP64

FIG. 12B



-5.0kB
-1.0kB
1
10
NANOG
Fold Change in Endogenous Gene Expression
0
200
400
600
800
1000
1200
1400
1600
plain
negative
gRNAs1-5
gRNAs6-10
gRNAs1-10
Cas9N-VP64

FIG. 13A

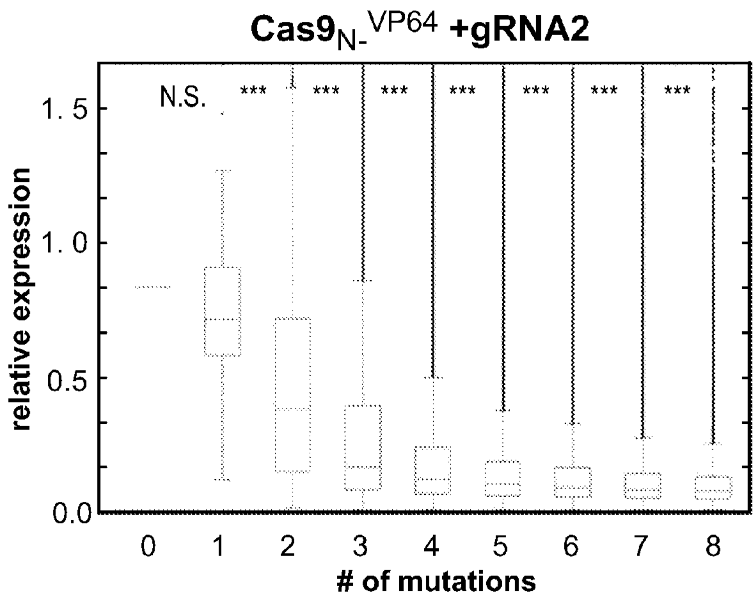


FIG. 13B

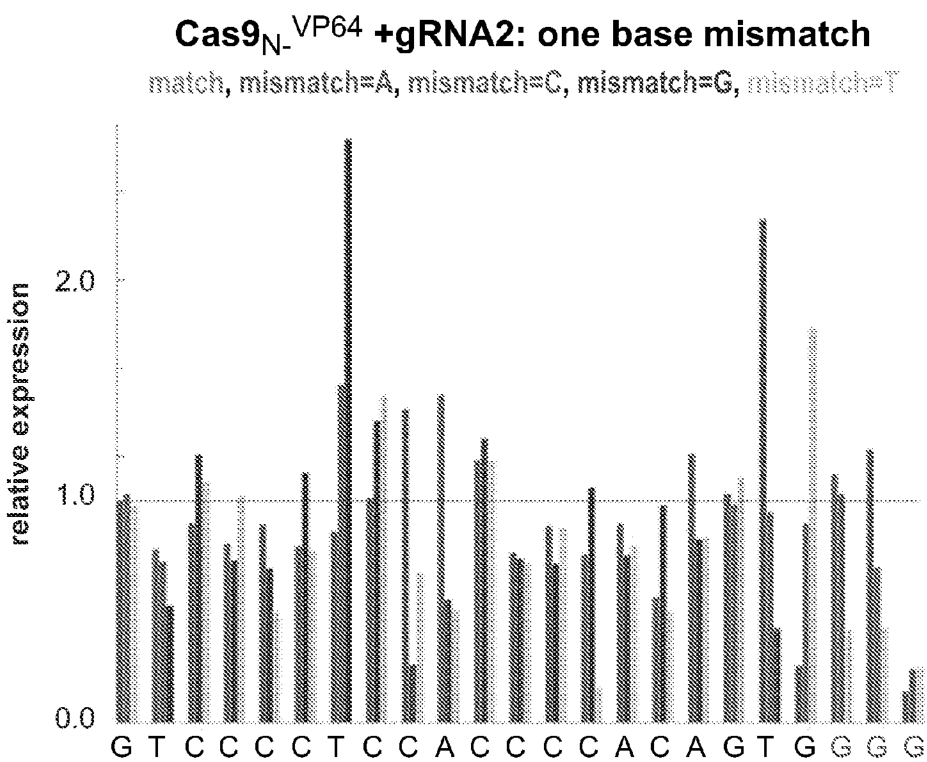


FIG. 13C

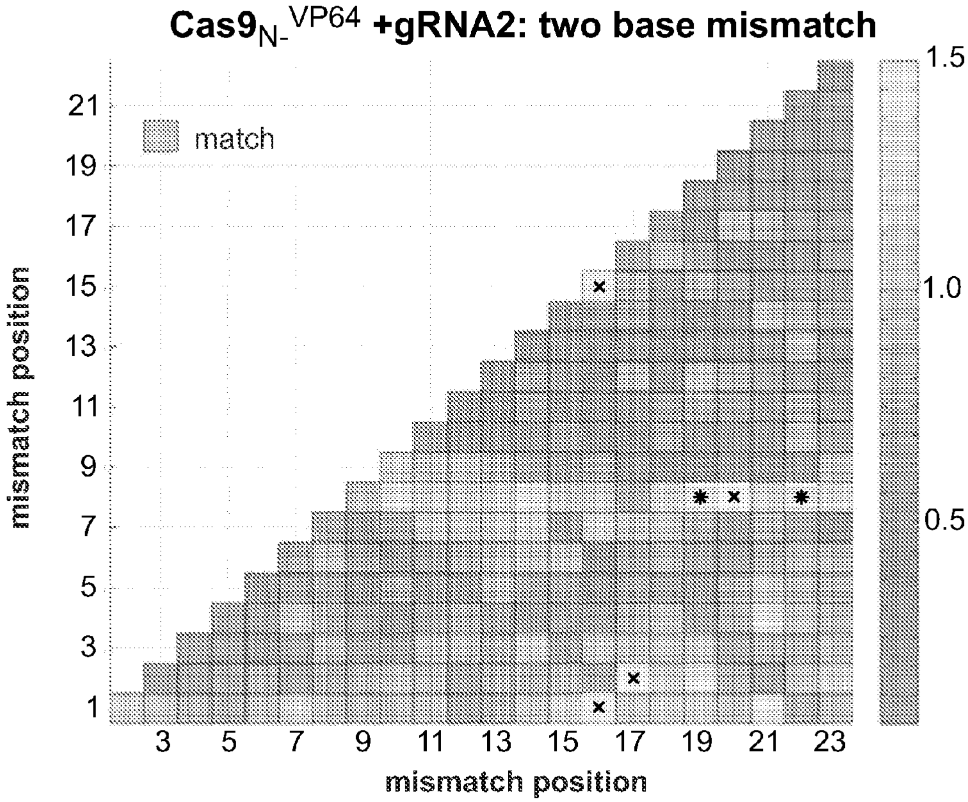


FIG. 13D

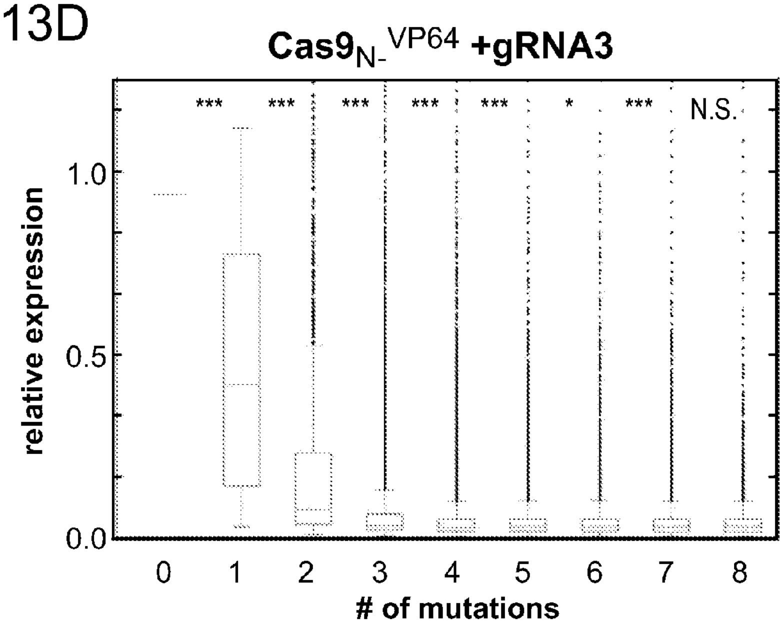


FIG. 13E

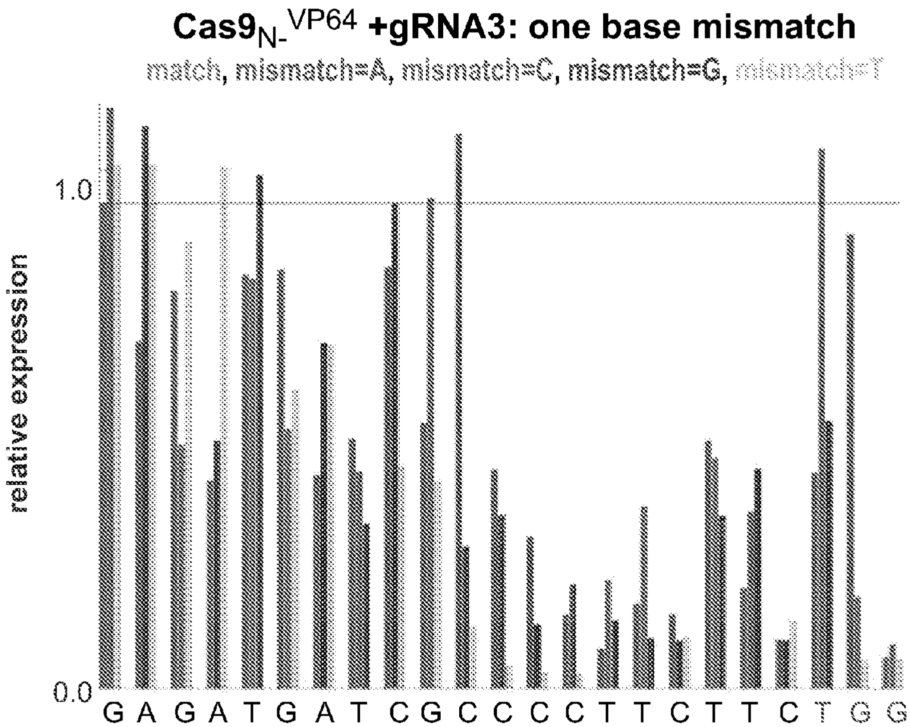


FIG. 13F

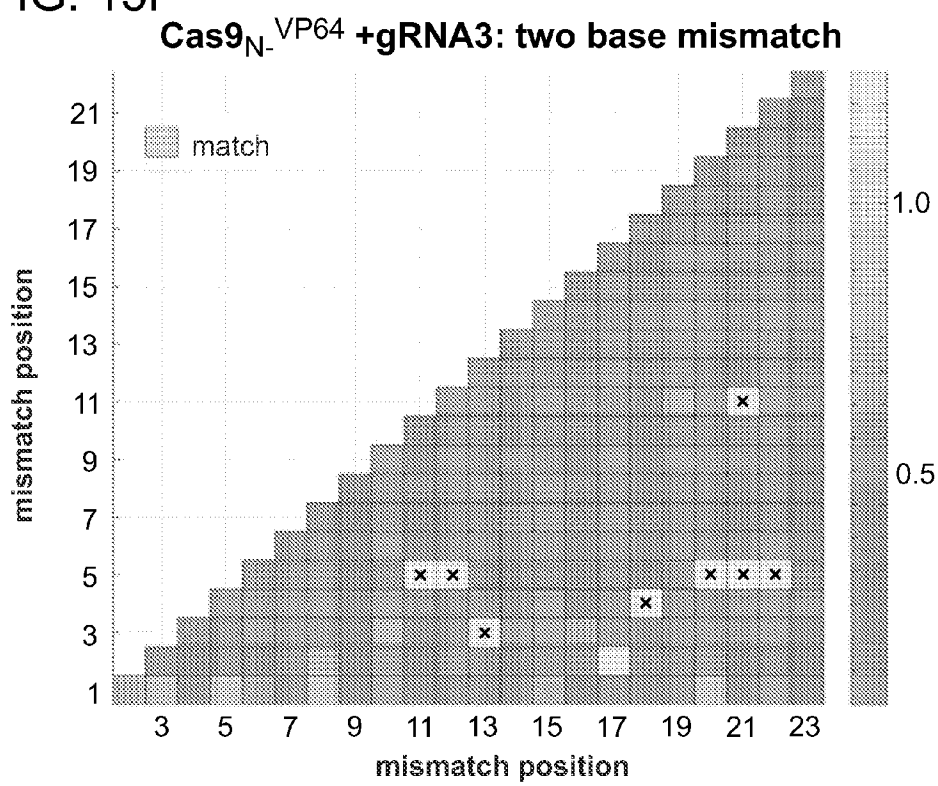


FIG. 14A

```
Target   : GAGATGATCGCCCCTTCTTC TGG
gRNA3    : GAGATGATCGCCCCTTCTTC
gRNA3mut: GTGATGACCGGCCGTTCTTC
```

FIG. 14B

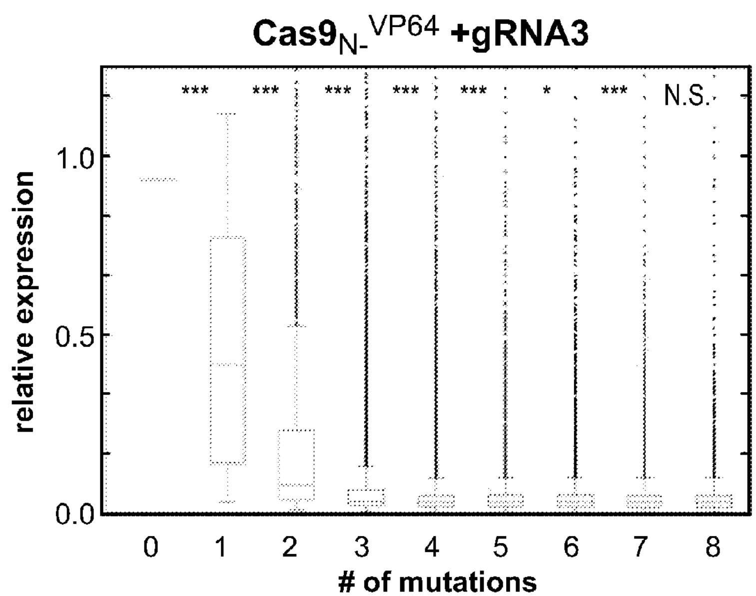


FIG. 14C

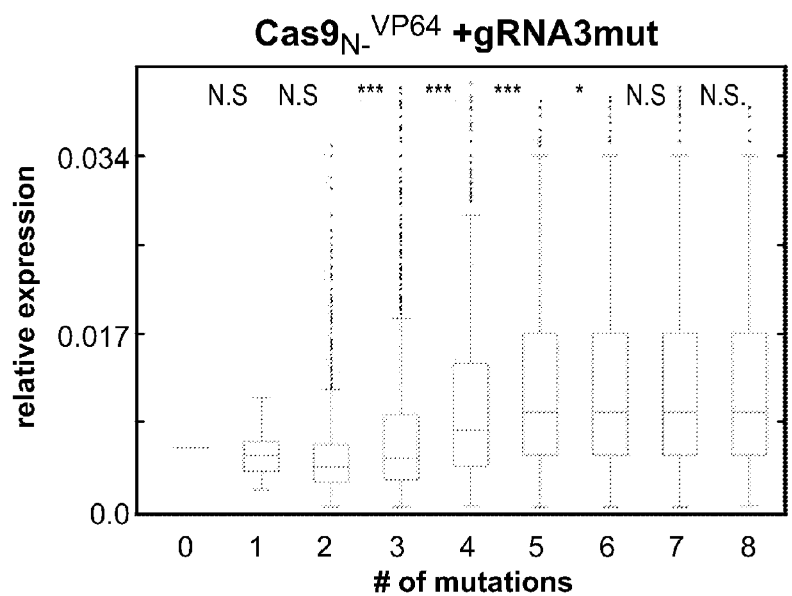


FIG. 15A

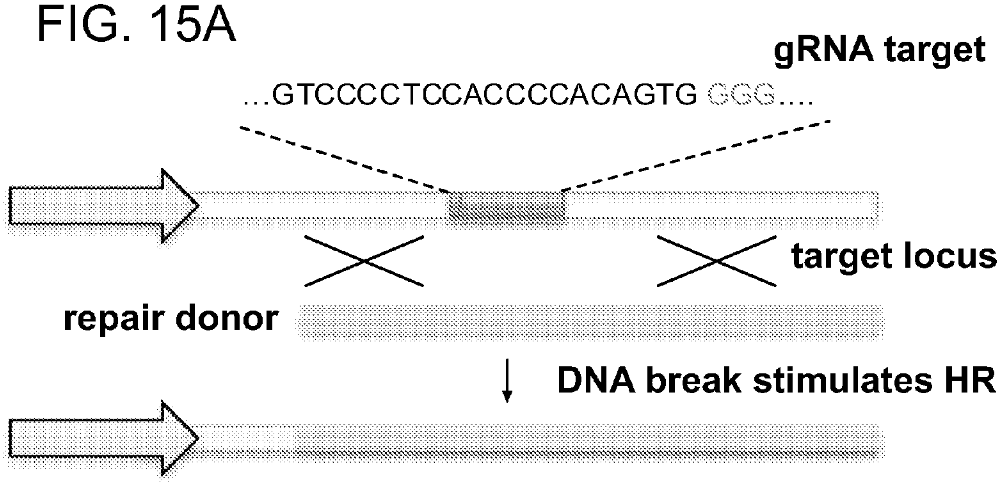
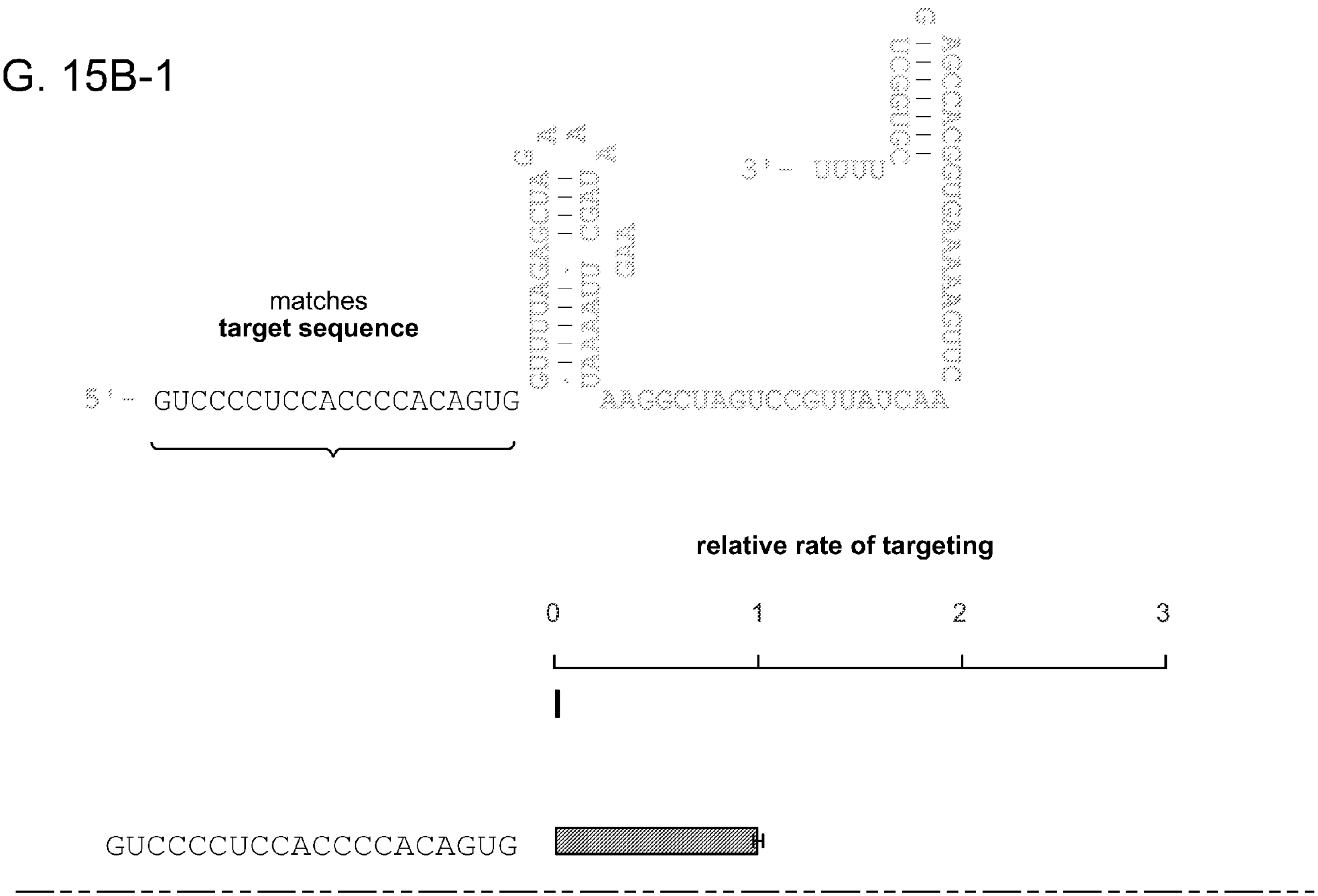

gRNA target
...GTCCCCTCCACCCCACAGTG GGG....
target locus
repair donor
DNA break stimulates HR

FIG. 15B-1



matches
target sequence
5' - GUCCCCUCCACCCCACAGUG
AAGGCUAGUCCGUUAUCAA
AGCCACGGUGAAAAAGUUC
UCGGUGC
3' - UUUU
relative rate of targeting
0
1
2
3
GUCCCCUCCACCCCACAGUG

FIG. 15B-2

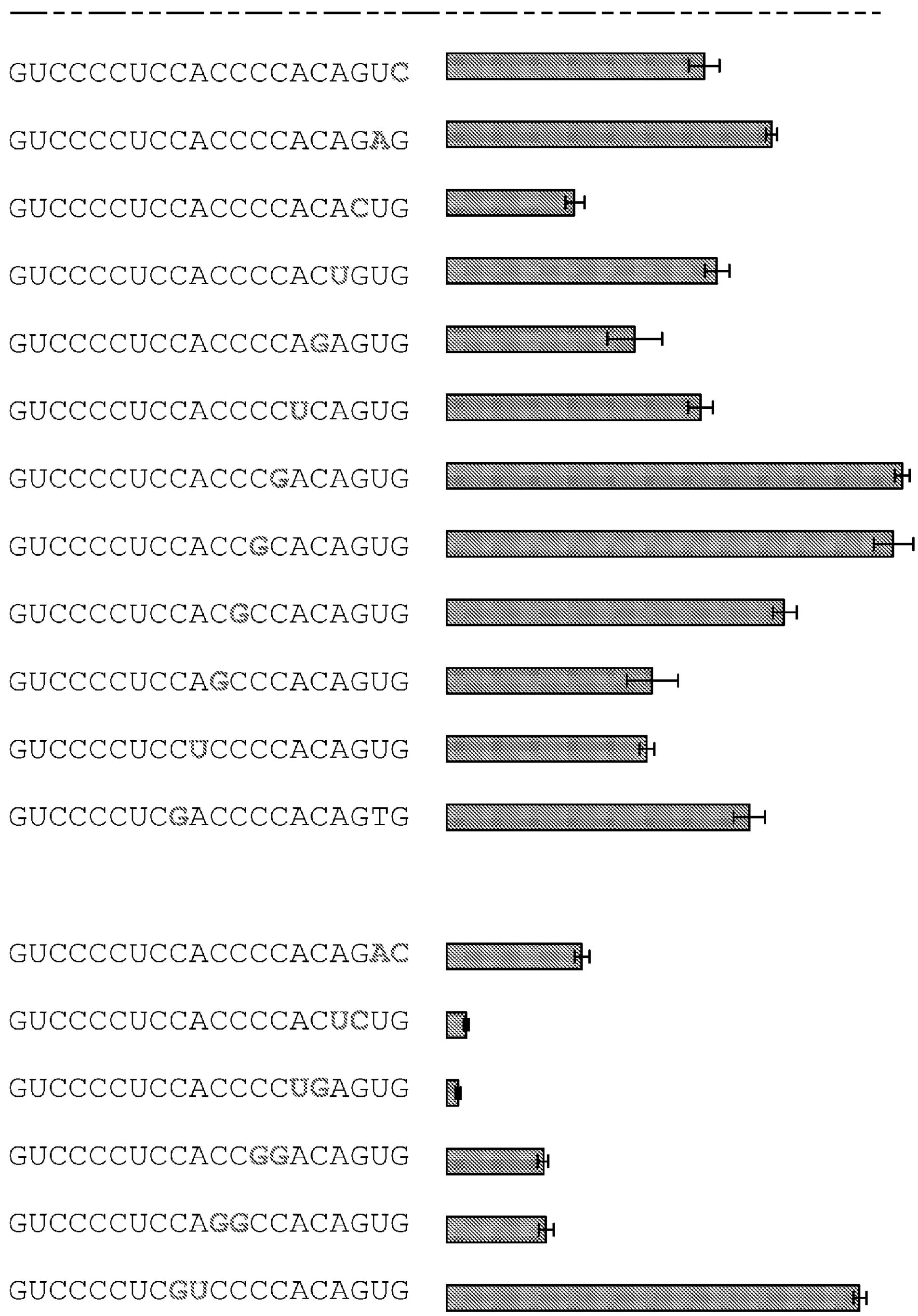

GUCCCCUCCACCCCACAGUC
GUCCCCUCCACCCCACAGAG
GUCCCCUCCACCCCACACUG
GUCCCCUCCACCCCACUGUG
GUCCCCUCCACCCCAGAGUG
GUCCCCUCCACCCCUCAGUG
GUCCCCUCCACCCGACAGUG
GUCCCCUCCACCGCACAGUG
GUCCCCUCCACGCCACAGUG
GUCCCCUCCAGCCCACAGUG
GUCCCCUCCUCCCCACAGUG
GUCCCCUCGACCCCACAGTG
GUCCCCUCCACCCCACAGAC
GUCCCCUCCACCCCACUCUG
GUCCCCUCCACCCCUGAGUG
GUCCCCUCCACCGGACAGUG
GUCCCCUCCAGGCCACAGUG
GUCCCCUCGUCCCCACAGUG

FIG. 15C

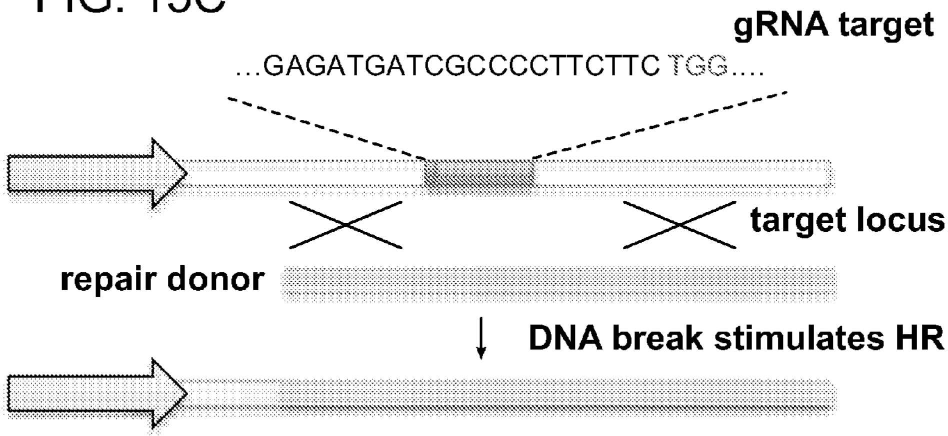


FIG. 15D-1

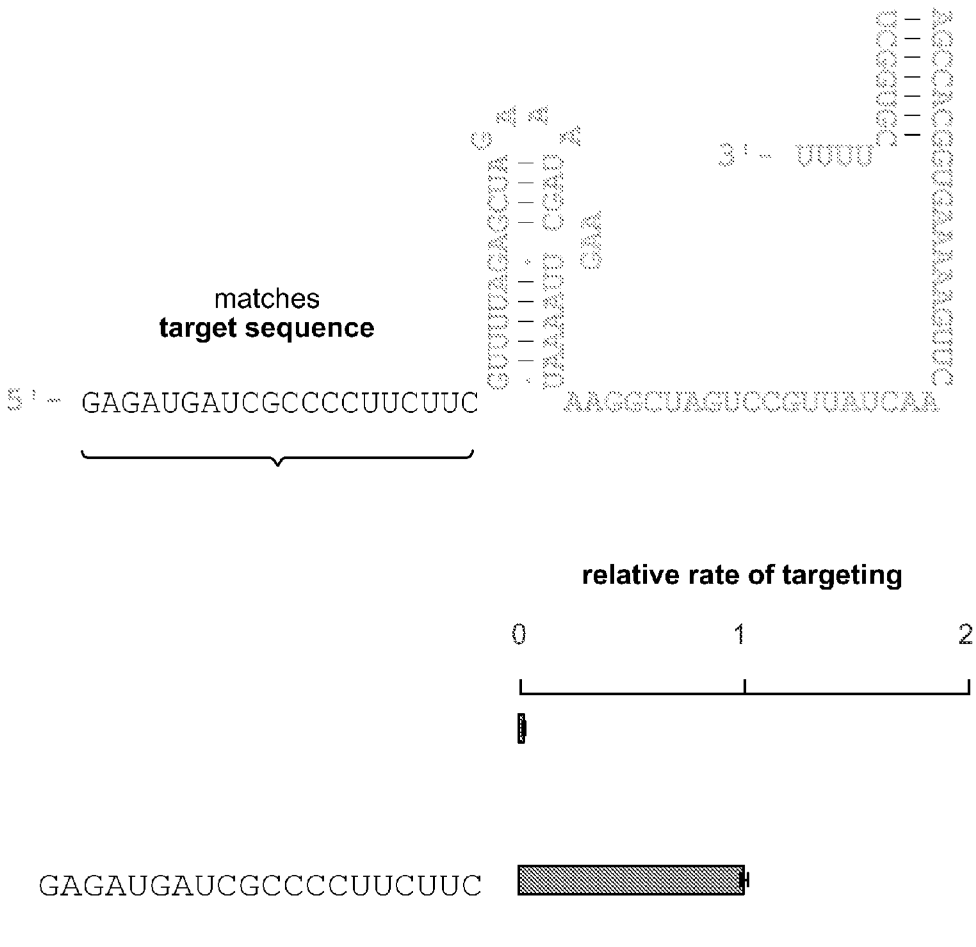
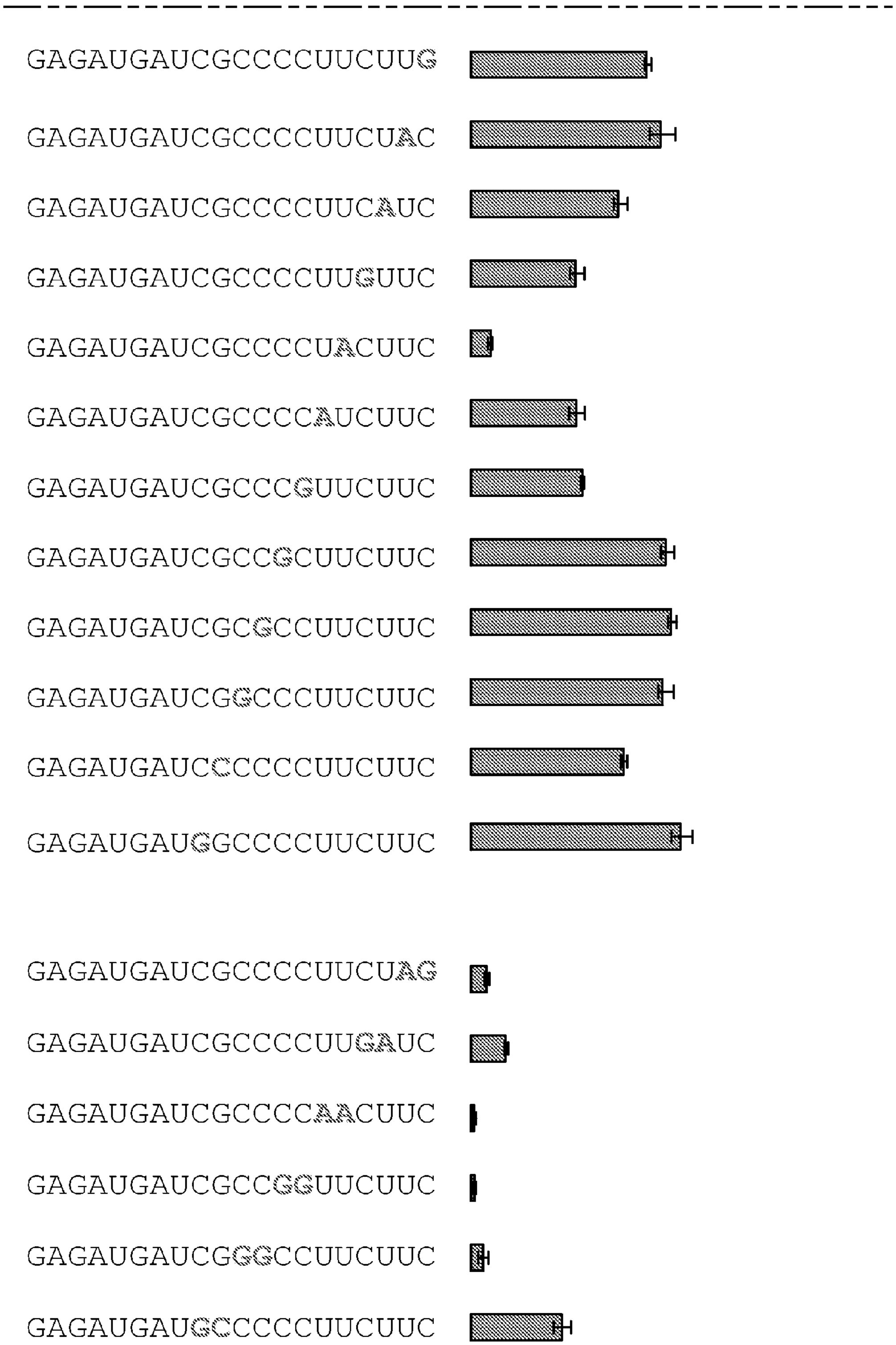


FIG. 15D-2


GAGAUGAUCGCCCCUUCUUG
GAGAUGAUCGCCCCUUCUAC
GAGAUGAUCGCCCCUUCAUC
GAGAUGAUCGCCCCUUGUUC
GAGAUGAUCGCCCCUACUUC
GAGAUGAUCGCCCCAUCUUC
GAGAUGAUCGCCCGUUCUUC
GAGAUGAUCGCCGCUUCUUC
GAGAUGAUCGCGCCUUCUUC
GAGAUGAUCGGCCCUUCUUC
GAGAUGAUCCCCCCUUCUUC
GAGAUGAUGGCCCCUUCUUC
GAGAUGAUCGCCCCUUCUAG
GAGAUGAUCGCCCCUUGAUC
GAGAUGAUCGCCCCAACUUC
GAGAUGAUCGCCGGUUCUUC
GAGAUGAUCGGGCCUUCUUC
GAGAUGAUGCCCCCUUCUUC

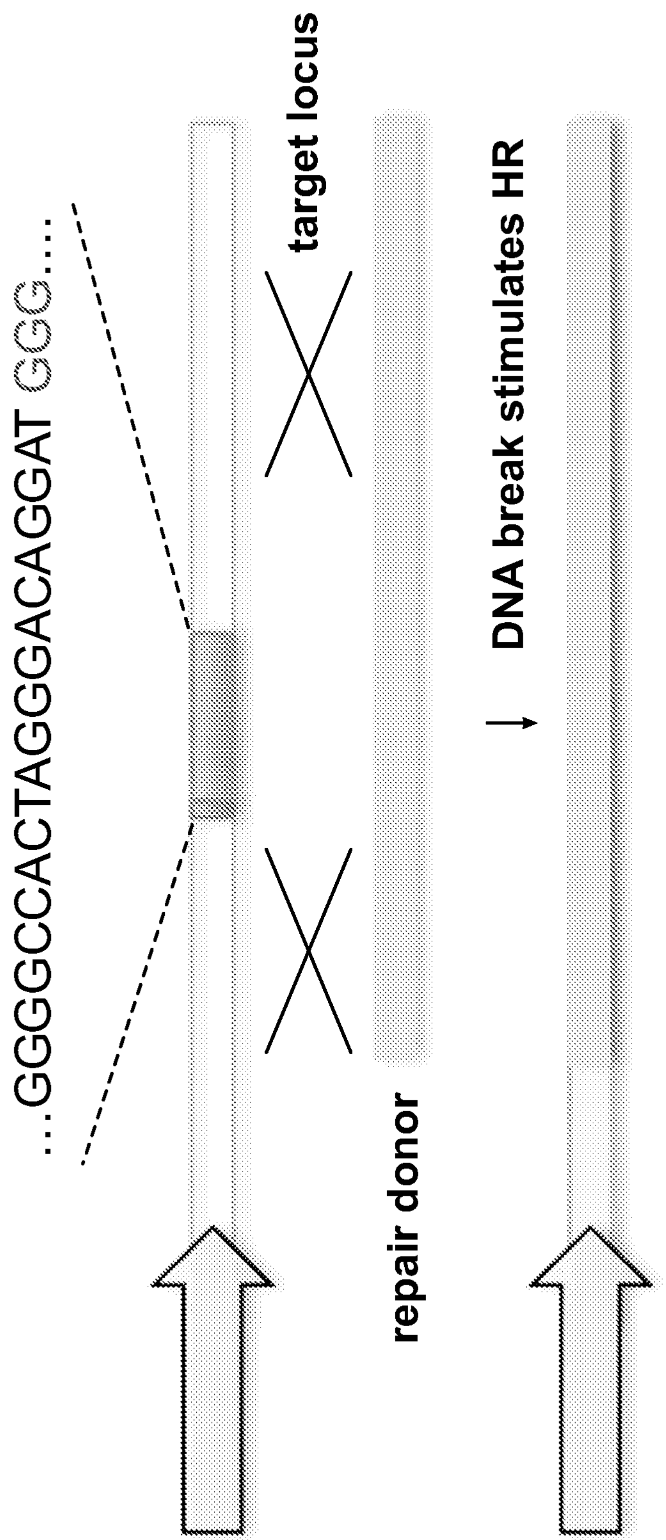

gRNA target
...GGGGCCACTAGGGACAGGAT GGG....
target locus
repair donor
DNA break stimulates HR

FIG. 16A

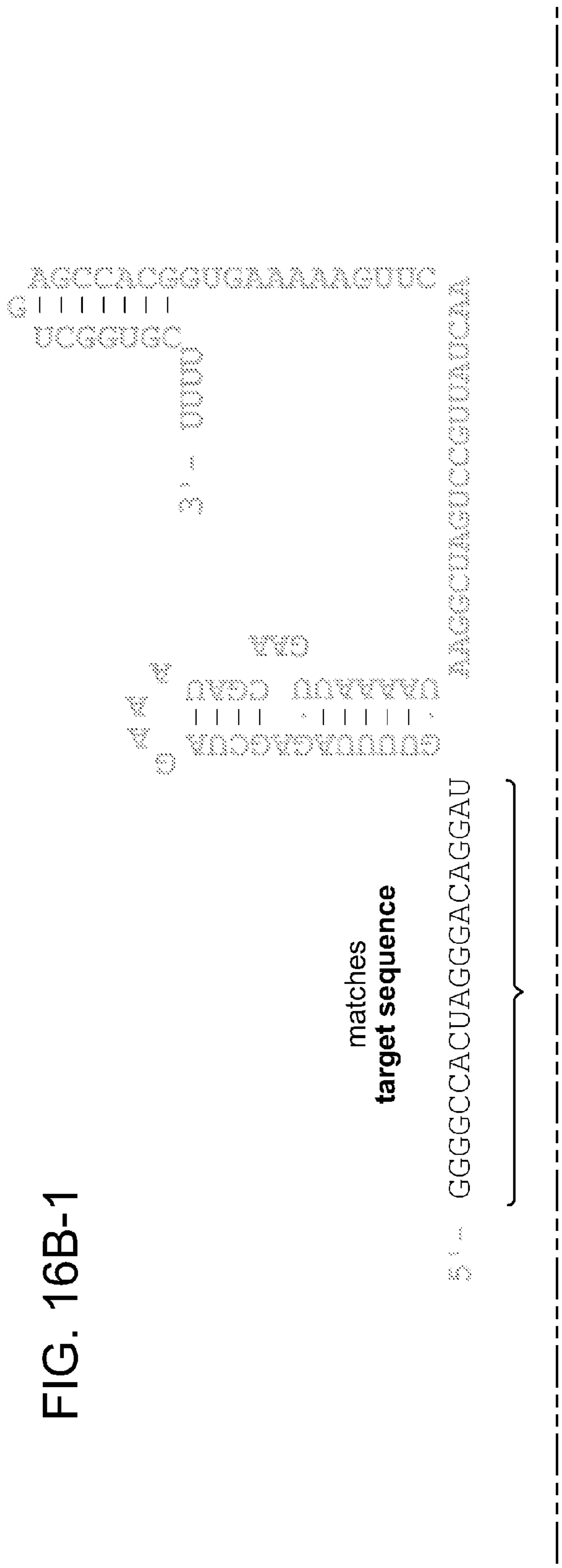


FIG. 16B-1

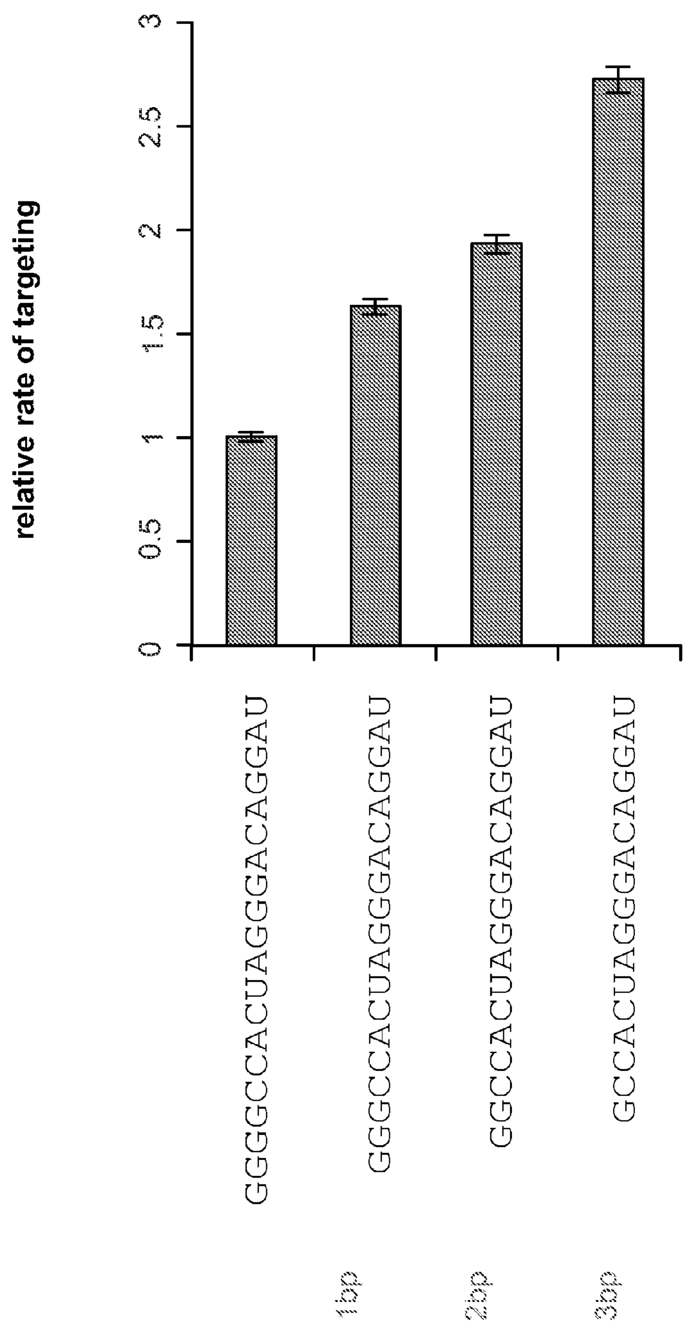
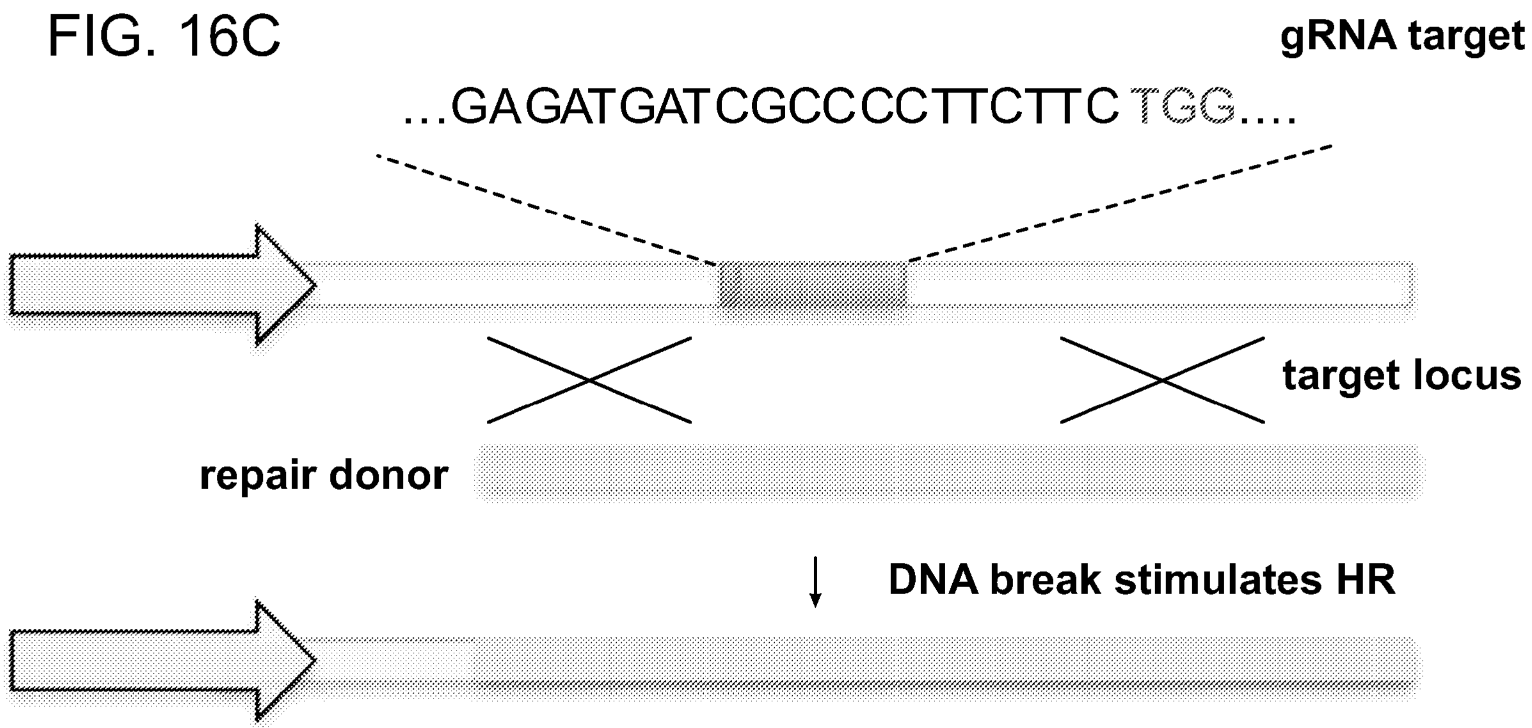

relative rate of targeting
0
0.5
1
1.5
2
2.5
3
GGGGCCACUAGGGACAGGAU
1bp
GGGCCACUAGGGACAGGAU
2bp
GGCCACUAGGGACAGGAU
3bp
GCCACUAGGGACAGGAU

FIG. 16B-2

FIG. 16C


gRNA target
...GAGATGATCGCCCCTTCTTC TGG....
target locus
repair donor
DNA break stimulates HR

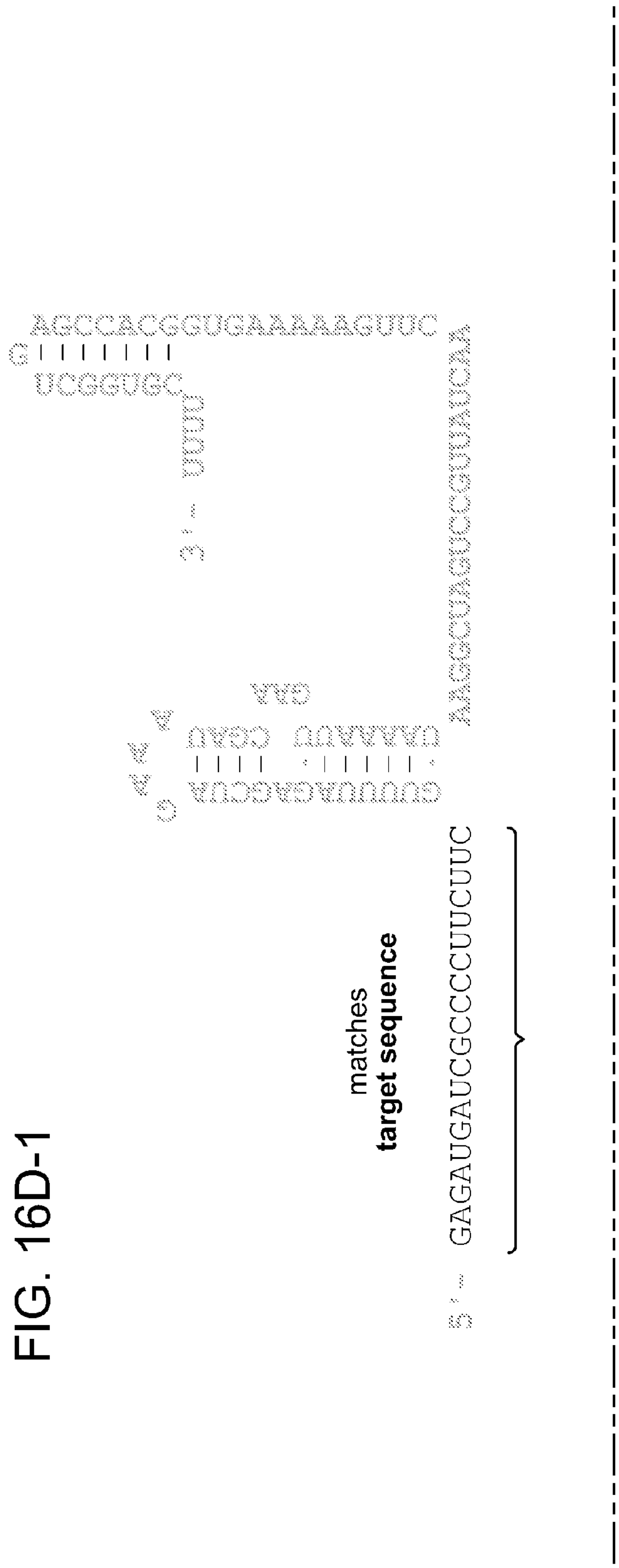


FIG. 16D-1

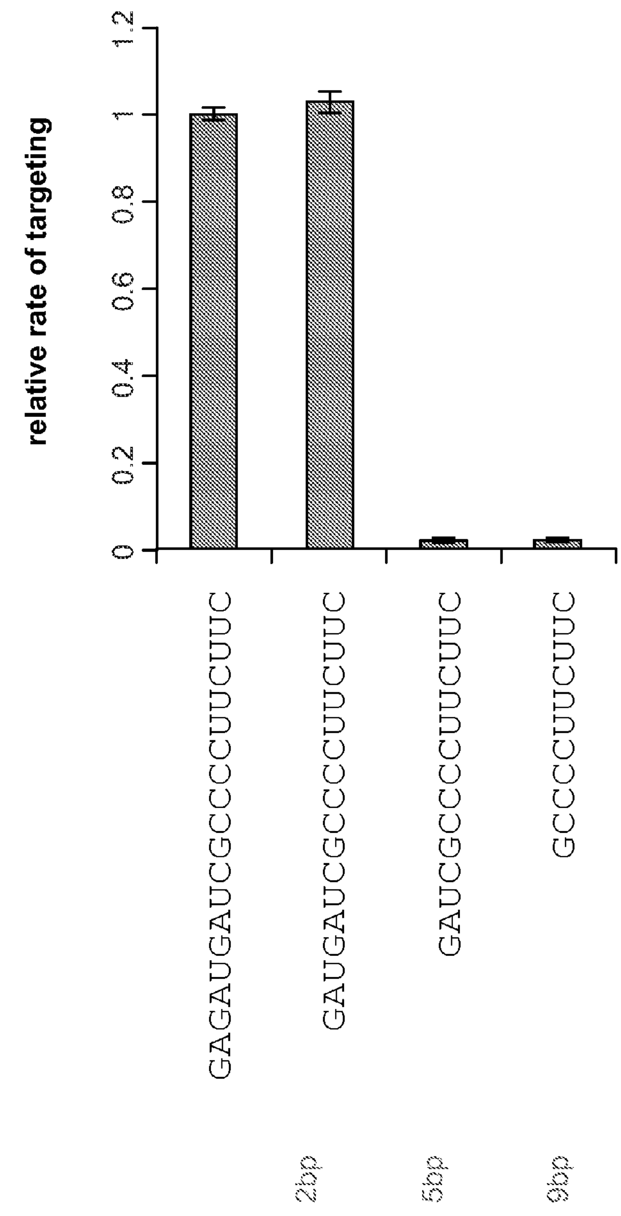

relative rate of targeting
0
0.2
0.4
0.6
0.8
1
1.2
GAGAUGAUCGCCCCUUCUUC
2bp
GAUGAUCGCCCCUUCUUC
5bp
GAUCGCCCCUUCUUC
9bp
GCCCCUUCUUC

FIG. 16D-2

FIG. 17A

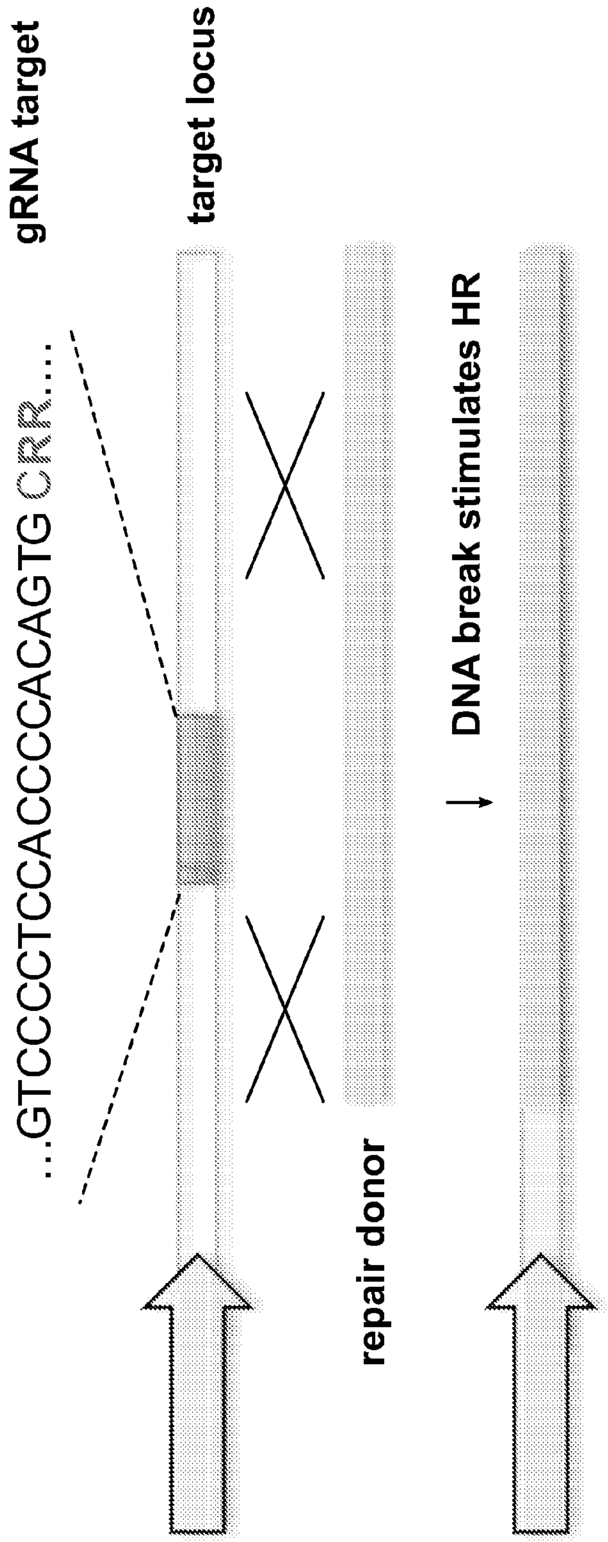
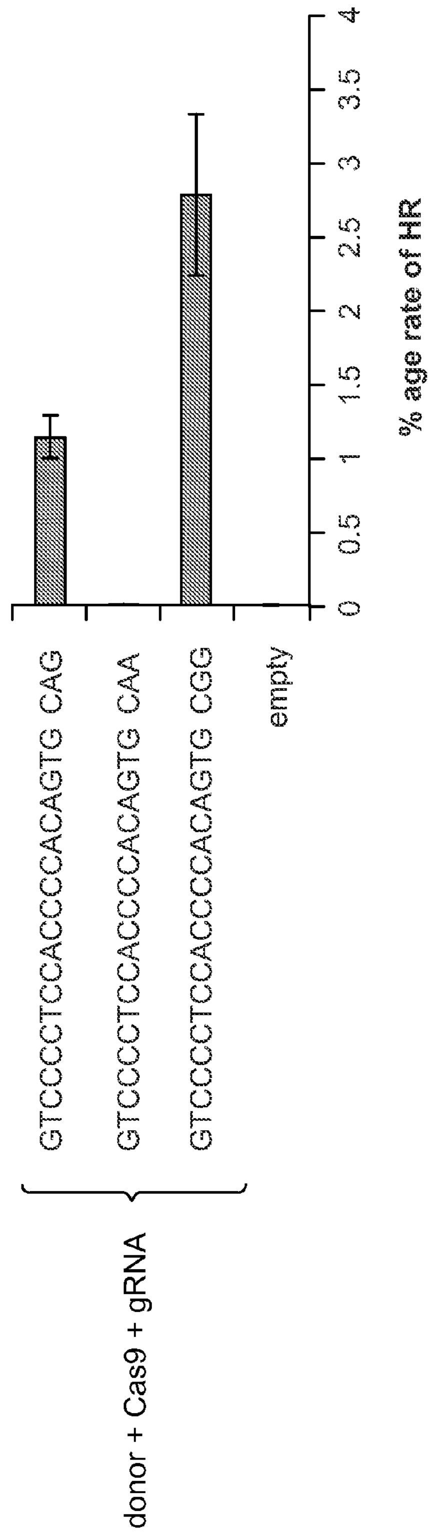

...GTCCCCTCCACCCCACAGTG CRR....
gRNA target
target locus
repair donor
DNA break stimulates HR

FIG. 17B



donor + Cas9 + gRNA
GTCCCCTCCACCCCACAGTG CAG
GTCCCCTCCACCCCACAGTG CAA
GTCCCCTCCACCCCACAGTG CGG
empty
0
0.5
1
1.5
2
2.5
3
3.5
4
% age rate of HR

FIG. 18A

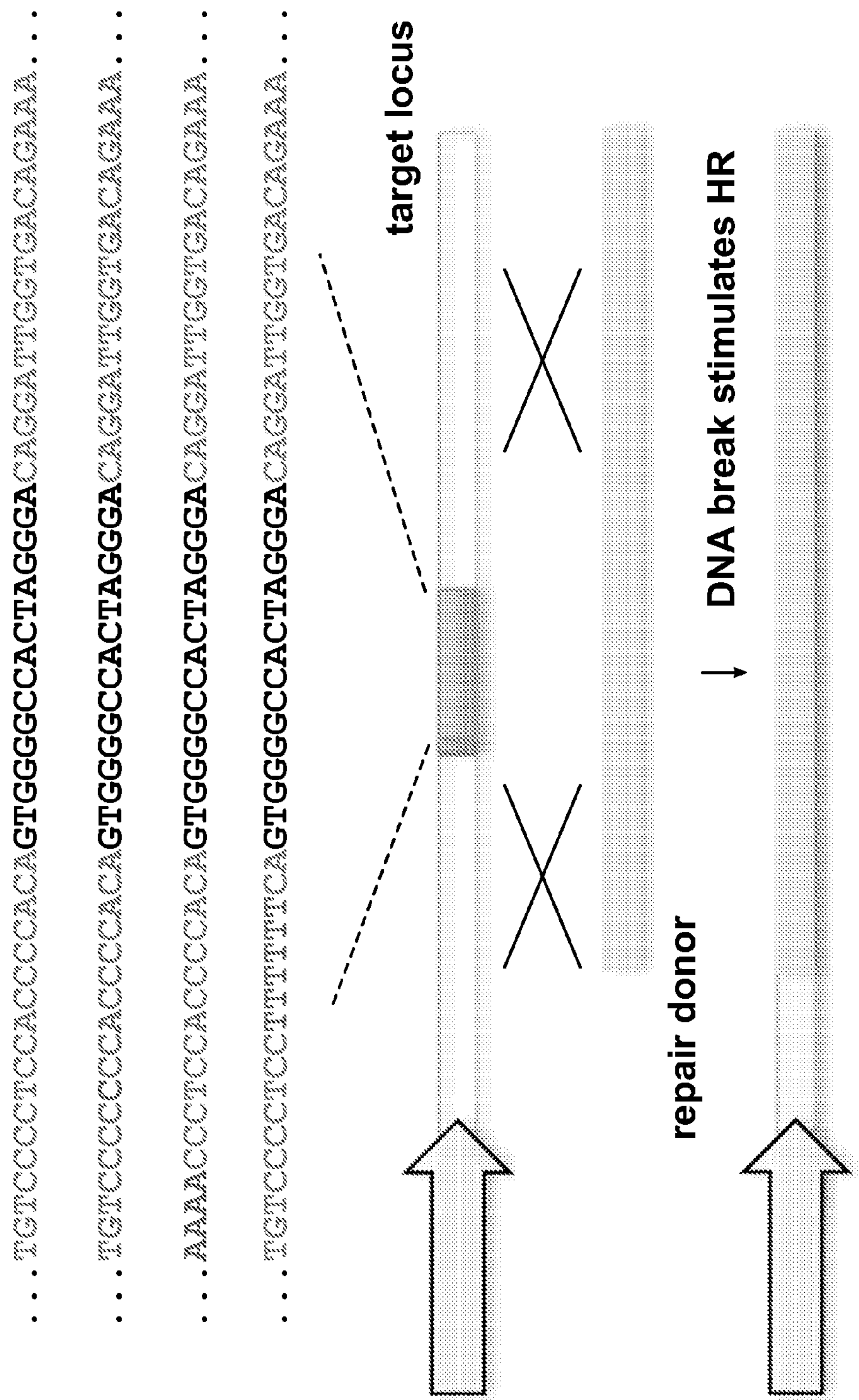

...TGTCCCCTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
...TGTCCCCCCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
...AAAACCCTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
...TGTCCCCTCCTTTTTCAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
target locus
repair donor
DNA break stimulates HR

FIG. 18B

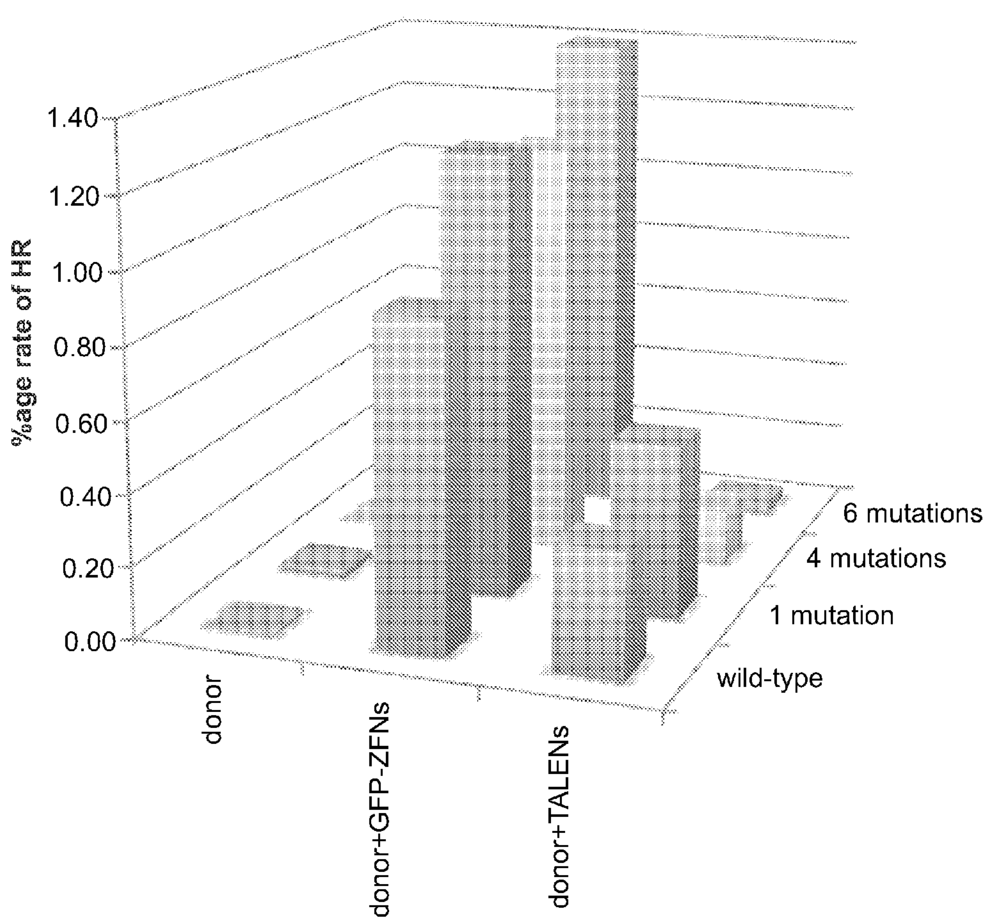

1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
%age rate of HR
6 mutations
4 mutations
1 mutation
wild-type
donor
donor+GFP-ZFNs
donor+TALENs

FIG. 19A

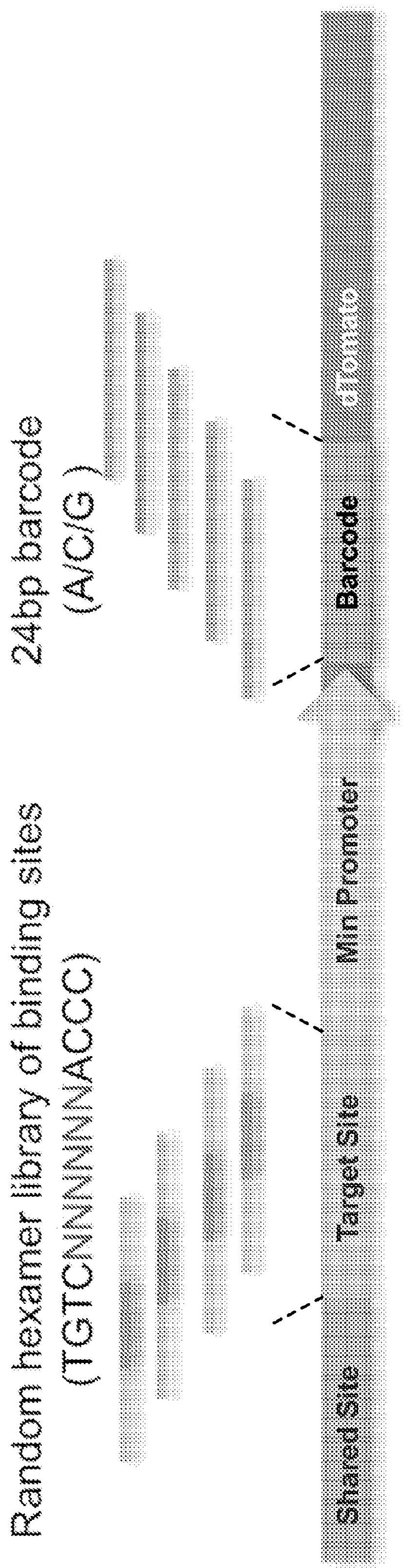

Random hexamer library of binding sites
(TGTCNNNNNNACCC)
24bp barcode
(A/C/G)
Shared Site
Target Site
Min Promoter
Barcode
dTomato

FIG. 19B-1

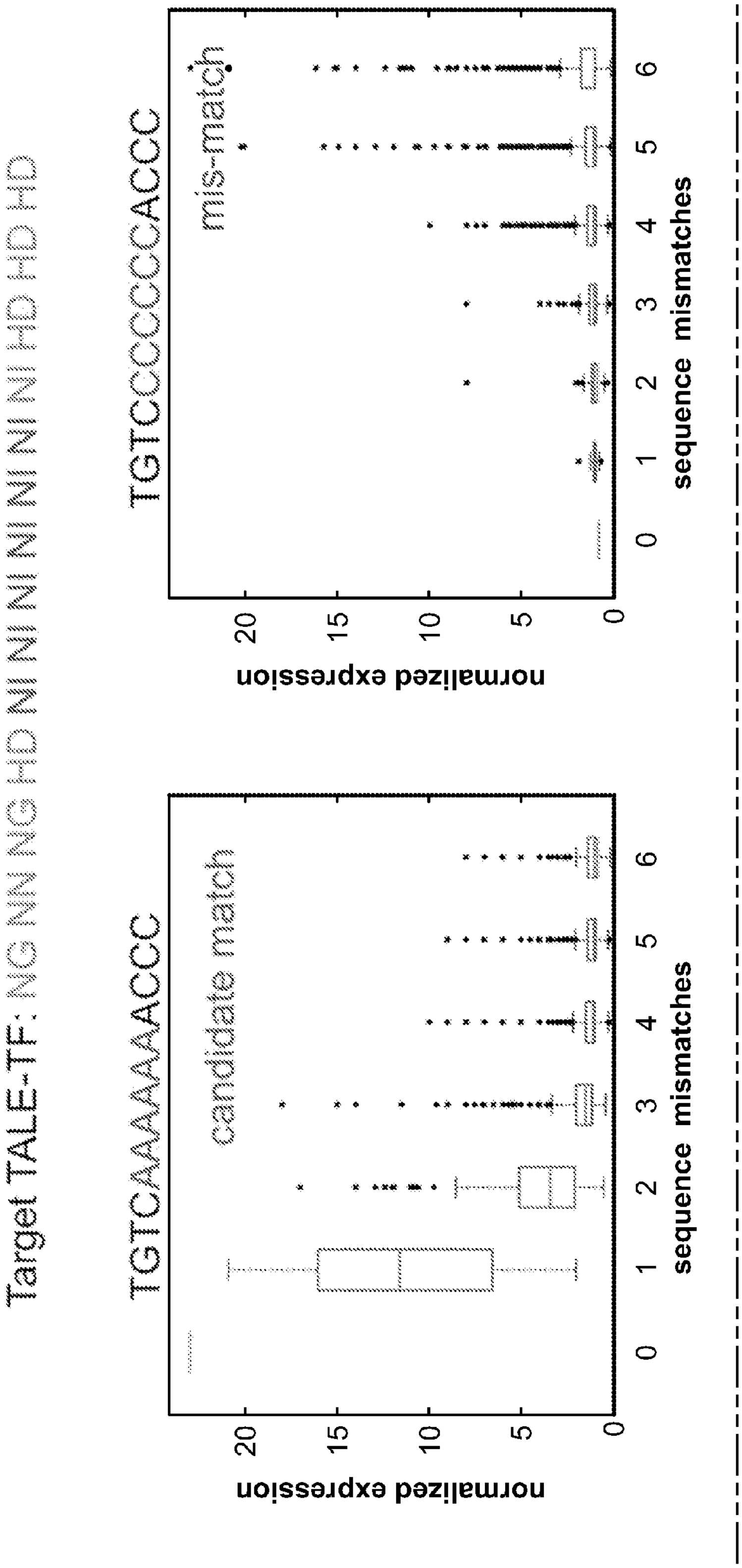
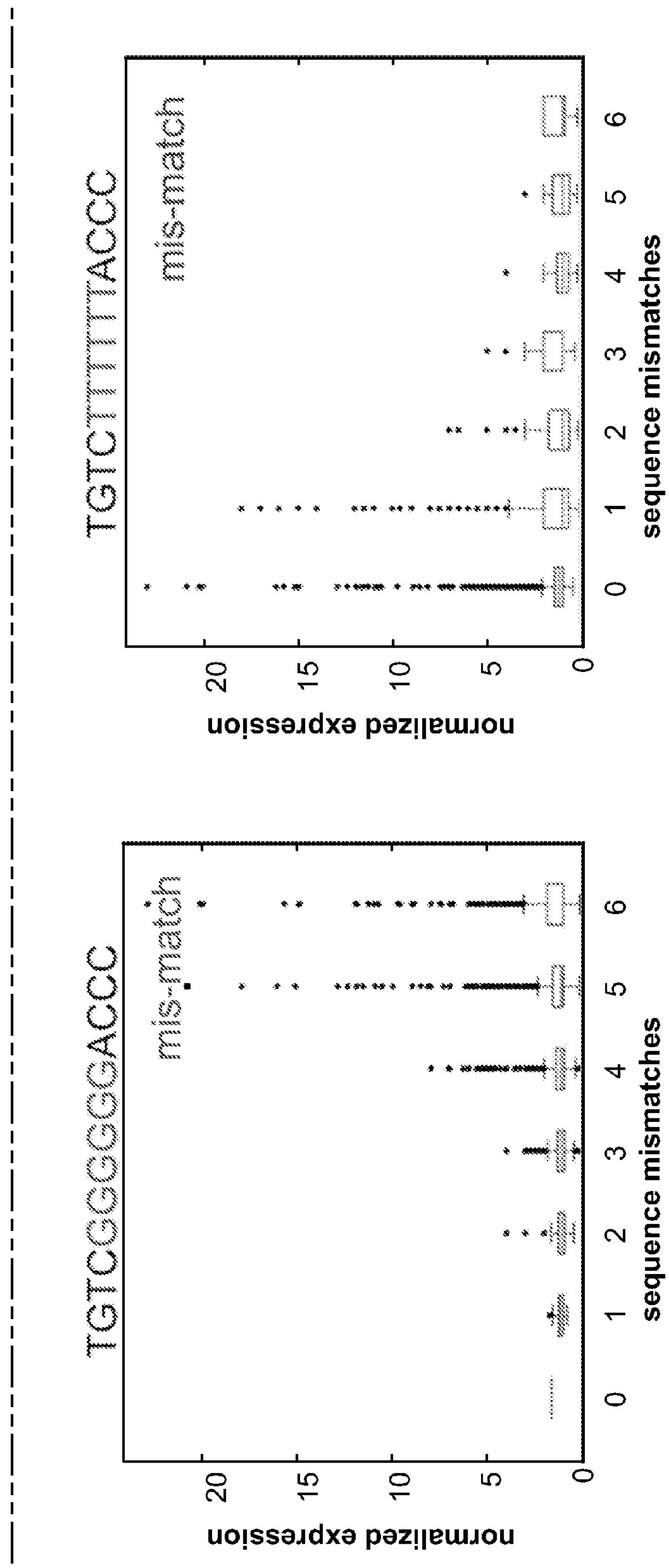

Target TALE-TF: NG NN NG HD NI NI NI NI NI NI HD HD HD
TGTCAAAAAACCC
candidate match
normalized expression
sequence mismatches
0
5
10
15
20
1
2
3
4
6
TGTCCCCCCCACCC
mis-match
normalized expression
sequence mismatches

FIG. 19B-2



TGTCGGGGGGACCC
mis-match
normalized expression
sequence mismatches
0
5
10
15
20
0
1
2
3
4
5
6
TGTCTTTTTTACCC
mis-match
normalized expression
sequence mismatches
0
5
10
15
20
0
1
2
3
4
5
6

FIG. 19C-1

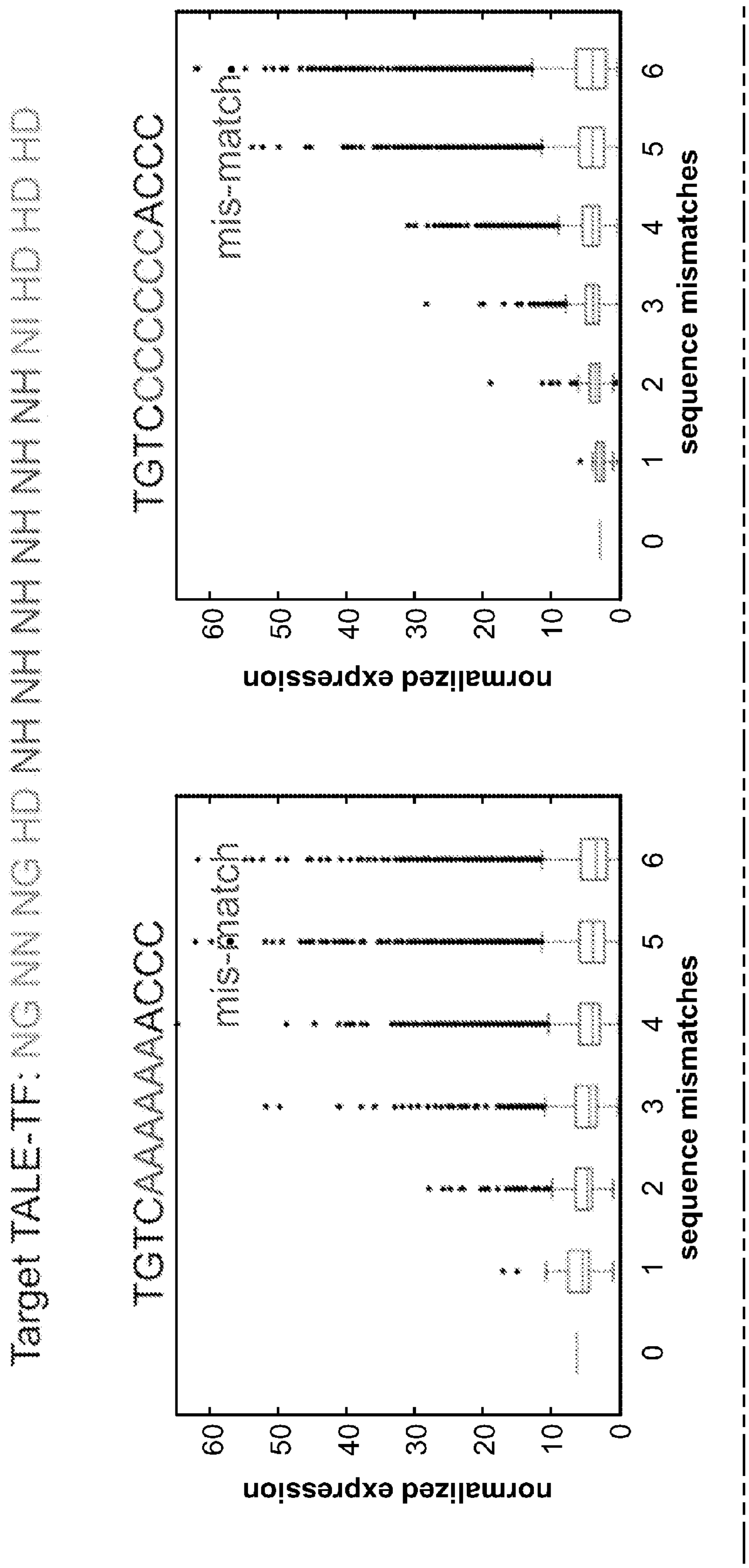

Target TALE-TF: NG NN NG HD NH NH NH NH NH NH NI HD HD HD
TGTCAAAAAAACCC
TGTCCCCCCCACCC
mis-match
mis-match
normalized expression
normalized expression
sequence mismatches
sequence mismatches
0
10
20
30
40
50
60
0
1
2
3
4
5
6

FIG. 19C-2

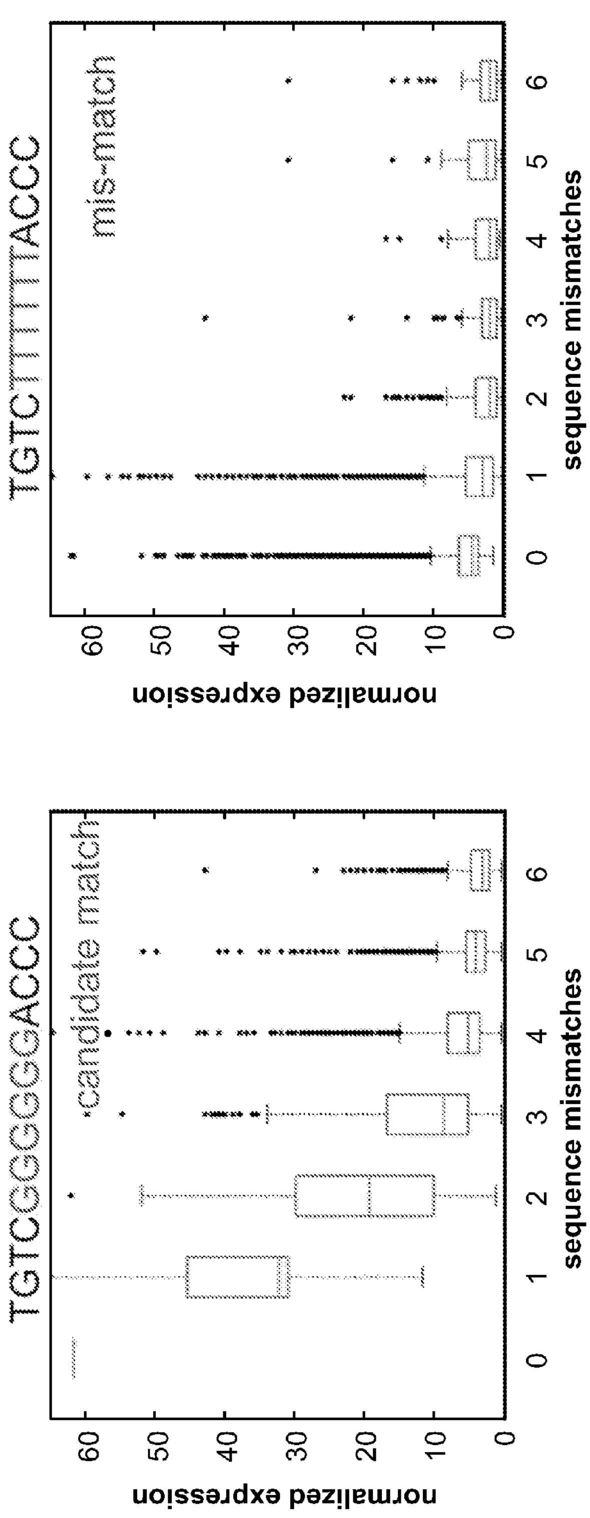
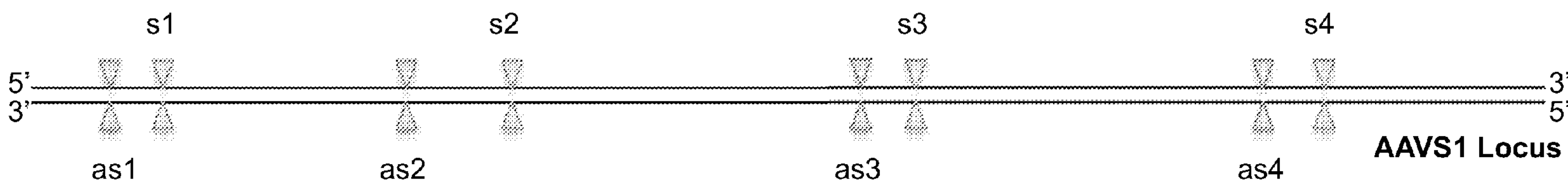

TGTCGGGGGACCC
candidate match
normalized expression
0 10 20 30 40 50 60
sequence mismatches
0 1 2 3 4 5 6
TGTCTTTTTACCC
mis-match
normalized expression
0 10 20 30 40 50 60
sequence mismatches
0 1 2 3 4 5 6

FIG. 20A



gRNA	Sequence	
AAVS1_s1	GGATCCTGTGTCCCCGAGCT	GGG
AAVS1_s2	GTTAATGTGGCTCTGGTTCT	GGG
AAVS1_s3	GGGGCCACTAGGGACAGGAT	TGG
AAVS1_s4	CTTCCTAGTCTCCTGATATT	GGG
AAVS1_as1	TGGTCCCAGCTCGGGGACAC	AGG
AAVS1_as2	AGAACCAGAGCCACATTAAC	CGG
AAVS1_as3	GTCACCAATCCTGTCCCTAG	TGG
AAVS1_as4	AGACCCAATATCAGGAGACT	AGG

FIG. 20B

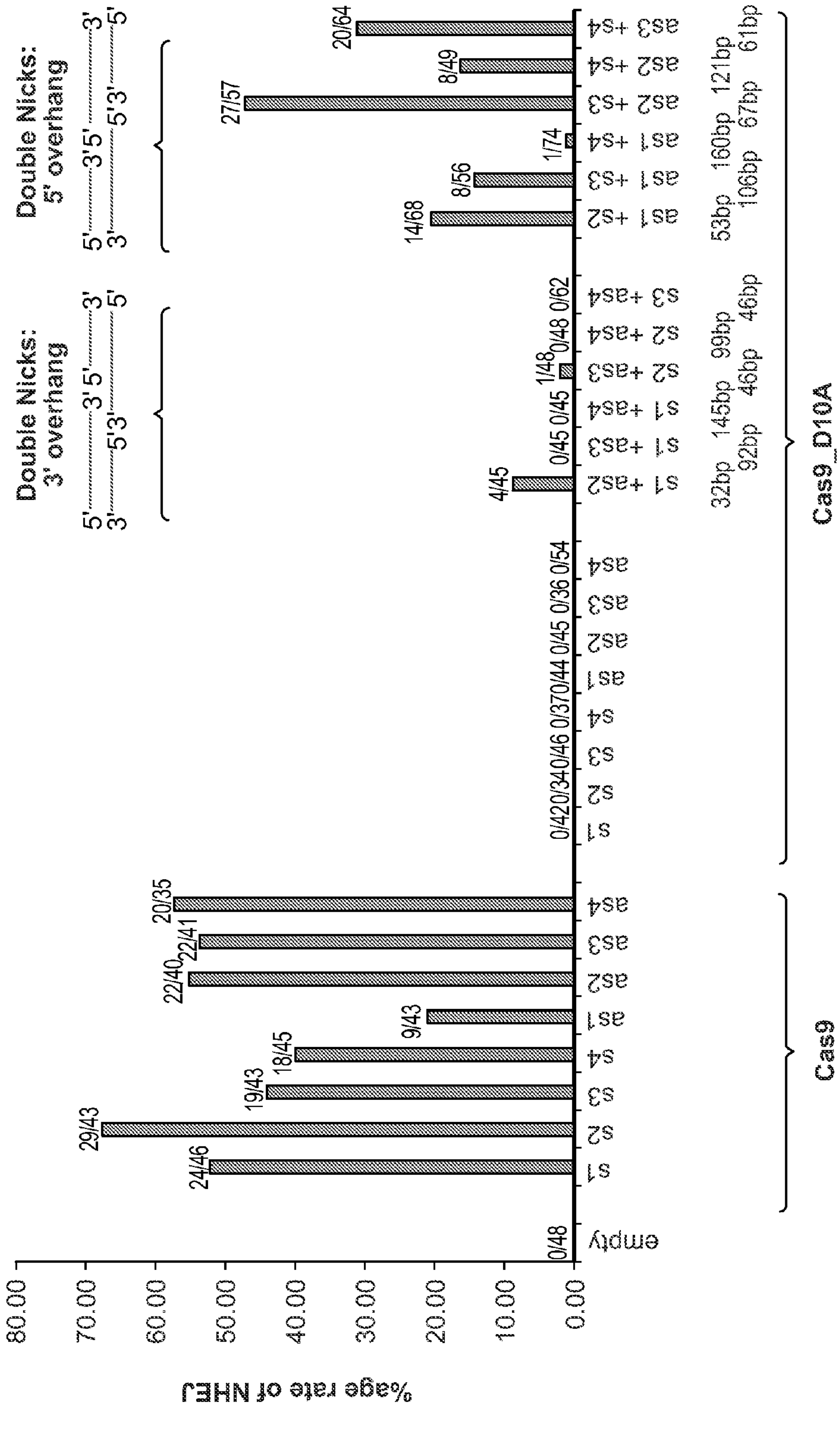

%age rate of NHEJ
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
empty
s1
s2
s3
s4
as1
as2
as3
as4
0/48
24/46
29/43
19/43
18/45
9/43
22/40
22/41
20/35
Cas9
0/42
0/34
0/46
0/37
0/44
0/45
0/36
0/54
s1 +as2
s1 +as3
s1 +as4
s2 +as3
s2 +as4
s3 +as4
4/45
0/45
0/45
1/48
0/48
0/62
32bp
92bp
145bp
46bp
99bp
46bp
Double Nicks:
3' overhang
5'-3'5'-3'
3'-5'3'-5'
as1 +s2
as1 +s3
as1 +s4
as2 +s3
as2 +s4
as3 +s4
14/68
8/56
1/74
27/57
8/49
20/64
53bp
106bp
160bp
67bp
121bp
61bp
Double Nicks:
5' overhang
5'-3'5'-3'
3'-5'3'-5'
Cas9_D10A

FIG. 21A

as1 + s2

```
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGG------TTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGCAGGGCGGGTTGGACCACCTTATATTCCCAGGGCAGGGCCGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCG---T----C----T---------GG---------------------------------------------TT-------CT--GG----G--TACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGT--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATA------------------------------------------------------------------TTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATC-T-TGT--------------------------------------------------------------------------------------------------GGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
```

FIG. 21B

as2 + s3

```
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAAA
AGGGCCGGTTAATGAATGTGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--A--------------GACAGAAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCC-----------------------------------------GATTGGTGACAGAAAA
AGGGCCGGTT-------------------------------------------------------------------------------CAGGATTGGTGACAGAAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAAA
AGGGCCGGTTAATG----TGGC----------------------------------------------------------------------GATTGGTGACAGAAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCC-----------------------------------------GATTGGTGACAGAAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGG--------------ACAGGATTGGTGACAAAAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAAA
AGGGCCGGTTAATG-----------------------------------------------------------------------------------TGGTGACAGAAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCAGGGACAGTCTGTCCCCTCCACCCCAGGGACAGGATTGGTGACAGAAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACC--A--------------CC---ACT--AGGGACAGGATTGGTGACAGAAAA
```

FIG. 21C

as3 + s4

```
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCAT-----------------------------------------ACCCC
CCCACAGTGGGGCCACTA---------------------------------------------------------------------------------ACCCC
CCCACAGTGGGGCCACTA--------GTAGAACAGCCCCATCCTTAGGCCTCCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCAC------------------------------------------CCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------G----------------------------------------------------------TGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AAAAAAGCCCCATCCTTACGCCTCCTCCTTCCTAGTCTCCTGATATTGGGGCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GAC------------------------------AGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTAGGGACAGGGG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCC---------------------------ACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
```

RNA-GUIDED TRANSCRIPTIONAL REGULATION

RELATED APPLICATION DATA

[0001] This application is a continuation of PCT application no. PCT/US2014/040868, designating the United States and filed Jun. 4, 2014; which claims the benefit U.S. Provisional Patent Application No. 61/830,787 filed on Jun. 4, 2013; each of which are hereby incorporated by reference in their entireties.

STATEMENT OF GOVERNMENT INTERESTS

[0002] 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

[0003] Bacterial and archaeal CRISPR-Cas systems rely on short guide RNAs in complex with Cas proteins to direct degradation of complementary sequences present within invading foreign nucleic acid. See Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602-607 (2011); Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109, E2579-2586 (2012); Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821 (2012); Sapranauskas, R. et al. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli. Nucleic acids research* 39, 9275-9282 (2011); and Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics* 45, 273-297 (2011). A recent in vitro reconstitution of the *S. pyogenes* type II CRISPR system demonstrated that crRNA ("CRISPR RNA") fused to a normally trans-encoded tracrRNA ("trans-activating CRISPR RNA") is sufficient to direct Cas9 protein to sequence-specifically cleave target DNA sequences matching the crRNA. Expressing a gRNA homologous to a target site results in Cas9 recruitment and degradation of the target DNA. See H. Deveau et al., Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus. Journal of Bacteriology* 190, 1390 (February 2008).

SUMMARY

[0004] Aspects of the present disclosure are directed to a complex of a guide RNA, a DNA binding protein and a double stranded DNA target sequence. According to certain aspects, DNA binding proteins within the scope of the present disclosure include a protein that forms a complex with the guide RNA and with the guide RNA guiding the complex to a double stranded DNA sequence wherein the complex binds to the DNA sequence. This aspect of the present disclosure may be referred to as co-localization of the RNA and DNA binding protein to 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.

[0005] 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.

[0006] According to one aspect, the foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein further encodes the transcriptional regulator protein or domain fused to the RNA guided nuclease-null DNA binding protein. According to one aspect, the foreign nucleic acid encoding one or more RNAs further encodes a 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.

[0007] 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.

[0008] 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.

[0009] 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.

[0010] 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 tracer mate sequence. The guide RNA may also include a tracr sequence, a portion of which hybridizes to the tracr mate sequence. The guide RNA may also include a linker nucleic acid sequence which links the tracer mate sequence and the tracr sequence to produce the tracrRNA-crRNA fusion. The spacer sequence binds to target DNA, such as by hybridization.

[0011] 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 a 1 to 4 base truncation at the 5' end of the spacer sequence.

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

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

[0014] According to certain aspects, the tracr sequence may include between about 60 to about 500 nucleic acids. According to certain aspects, the tracr sequence may include between about 64 to about 500 nucleic acids. According 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 80 to about 500 nucleic acids. According to certain aspects, the tracr sequence may include between about 90 to about 500 nucleic acids. According to certain aspects, the tracr sequence may include between about 100 to about 500 nucleic acids.

[0015] According to certain aspects, the tracr sequence may include between about 60 to about 200 nucleic acids. According 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 80 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 100 to about 200 nucleic acids.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[0043] According to the methods of altering a DNA target nucleic acid, the two or more adjacent nicks are on the same strand of the double stranded DNA. According to one aspect, the two or more adjacent nicks are on the same strand of the double stranded DNA and result in homologous recombination. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA and 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 nuclease domain and being guided by the two or more RNAs, and wherein the two or more RNAs and the at least one Cas9 protein nickase are expressed and wherein the at least one Cas9 protein nickase co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks, and wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks resulting in fragmentation of the target nucleic acid thereby preventing expression of the target nucleic acid.

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

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

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

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

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

[0052] According to one aspect, the two or more RNAs are guide RNAs. According to one aspect, the two or more RNAs are 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. **1**A and FIG. **1**B are schematics of RNA-guided transcriptional activation. FIG. **1**C is a design of a reporter construct (SEQ ID NOs:62 and 63). FIG. **1**D shows data demonstrating that Cas9N-VP64 fusions display RNA-guided transcriptional activation as assayed by both fluorescence-activated cell sorting (FACS) and immunofluorescence assays (IF). FIG. **1**E 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. **1**F depicts data demonstrating transcriptional induction by individual gRNAs and multiple gRNAs.

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

[0058] FIG. **3**A depicts a schematic of a guide RNA design. FIG. **3**B 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. **3**C 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. **4**A 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. **4**B depicts data showing undetectable nuclease activity for Cas9 mutants m3 and m4, and also their respective fusions with VP64. FIG. **4**C is a higher-resolution examination of the data in FIG. **4**B.

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

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

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

[0063] FIG. **8**A depicts in schematic a high level specificity analysis processing flow for calculation of normalized expression levels. FIG. **8**B 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. **8**C 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. **9**A depicts data for analysis of the targeting landscape of a Cas9-gRNA complex showing tolerance to 1-3 mutations in its target sequence. FIG. **9**B 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. **9**C 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. **9**D depicts data from a nuclease mediated HR assay confirming that the predicted PAM for the *S. pyogenes* Cas9 is NGG and also NAG (SEQ ID NOs:67-69).

[0065] FIG. **10**A 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. **10**B depicts data from analysis of the targeting landscape of TALEs of 3 different sizes (18-mer, 14-mer and 10-mer). FIG. **10**C depicts data for 10-mer TALEs show near single-base mismatch resolution. FIG. **10**D depicts heat plot data for 10-mer TALEs show near single-base mismatch resolution.

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

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

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

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

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

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

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

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

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

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

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

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

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

DETAILED DESCRIPTION

[0079] Embodiments of the present disclosure are based on the use of DNA binding proteins to co-localize transcriptional regulator proteins or domains to DNA in a manner to regulate a target nucleic acid. Such DNA binding proteins are readily known to those of skill in the art to bind to DNA for various purposes. Such DNA binding proteins may be naturally occurring. DNA binding proteins included within the scope of the present disclosure include those which may be guided by RNA, referred to herein as guide RNA. According to this aspect, the guide RNA and the RNA guided DNA binding protein form a co-localization complex at the DNA. According to certain aspects, the DNA binding protein may be a 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.

[0080] 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 nuclease 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 McrA-HNH nuclease related domain and the RuvC-like nuclease domain. Accordingly, exemplary DNA binding proteins are those that in nature contain one or more of the McrA-HNH nuclease related domain and the RuvC-like nuclease 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 nuclease domain. Such modification includes removing the polypeptide sequence or polypeptide sequences exhibiting nuclease activity, i.e. the nuclease domain, such that the polypeptide sequence or polypeptide sequences exhibiting nuclease activity, i.e. nuclease 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. The 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 or all of the nuclease 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 nuclease sequences exhibiting nuclease activity inactivated.

[0081] According to one aspect, a DNA binding protein having two or more nuclease domains may be modified or altered to inactivate all but one of the nuclease 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.

[0082] 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 nuclease-null Cas9 protein. An exemplary DNA binding protein is a Cas9 protein nickase.

[0083] In *S. pyogenes*, Cas9 generates a blunt-ended double-stranded break 3 bp upstream of the protospacer-adjacent motif (PAM) via a process mediated by two catalytic domains in the protein: an HNH domain that cleaves the complementary strand of the DNA and a RuvC-like domain that cleaves the non-complementary strand. See Jinke et al., *Science* 337, 816-821 (2012) hereby incorporated by reference in its entirety. Cas9 proteins are known to exist in many Type II CRISPR systems including the following as identified in the supplementary information to Makarova et al., Nature Reviews, Microbiology, Vol. 9, June 2011, pp. 467-477: *Methanococcus maripaludis* C7; *Corynebacterium diphtheriae; Corynebacterium efficiens* YS-314; *Corynebacterium glutamicum* ATCC 13032 Kitasato; *Corynebacterium glutamicum* ATCC 13032 Bielefeld; *Corynebacterium glutamicum* R; *Corynebacterium kroppenstedtii* DSM 44385; *Mycobacterium abscessus* ATCC 19977; *Nocardia farcinica* IFM10152; *Rhodococcus erythropolis* PR4; *Rhodococcus jostii* RHA1; *Rhodococcus opacus* B4 uid36573; *Acidothermus cellulolyticus* 11B; *Arthrobacter chlorophenolicus* A6; *Kribbella flavida* DSM 17836 uid43465; *Thermomonospora curvata* DSM 43183; *Bifidobacterium dentium* Bd1; *Bifidobacterium longum* DJO10A; *Slackia heliotrinireducens* DSM 20476; *Persephonella marina* EX H1; *Bacteroides fragilis* NCTC 9434; *Capnocytophaga ochracea* DSM 7271; *Flavobacterium psychrophilum* JIP02 86; *Akkermansia muciniphila* ATCC BAA 835; *Roseiflexus castenholzii* DSM 13941; *Roseiflexus* RS1; *Synechocystis* PCC6803; *Elusimicrobium minutum* Pei191; uncultured Termite group 1 bacterium phylotype Rs D17; *Fibrobacter succinogenes* S85; *Bacillus cereus* ATCC 10987; *Listeria innocua; Lactobacillus casei; Lactobacillus rhamnosus* GG; *Lactobacillus salivarius* UCC118; *Streptococcus agalactiae* A909; *Streptococcus agalactiae* NEM316; *Streptococcus agalactiae* 2603; *Streptococcus dysgalactiae equisimilis* GGS 124; *Streptococcus equi zooepidemicus* MGCS10565; *Streptococcus gallolyticus* UCN34 uid46061; *Streptococcus gordonii Challis* subst CH1; *Streptococcus mutans* NN2025 uid46353; *Streptococcus mutans; Streptococcus pyogenes* M1 GAS; *Streptococcus pyogenes* MGAS5005; *Streptococcus pyogenes* MGAS2096; *Streptococcus pyogenes* MGAS9429; *Streptococcus pyogenes* MGAS10270; *Streptococcus pyogenes* MGAS6180; *Streptococcus pyogenes* MGAS315; *Streptococcus pyogenes* SSI-1; *Streptococcus pyogenes* MGAS10750; *Streptococcus pyogenes* NZ131; *Streptococcus thermophiles* CNRZ1066; *Streptococcus thermophiles* LMD-9; *Streptococcus thermophiles* LMG 18311; *Clostridium botulinum* A3 Loch Maree; *Clostridium botulinum* B Eklund 17B; *Clostridium botulinum* Ba4 657; *Clostridium botulinum* F Langeland; *Clostridium cellulolyticum* H10; *Finegoldia magna* ATCC 29328; *Eubacterium rectale* ATCC 33656; *Mycoplasma gallisepticum; Mycoplasma* mobile 163K; *Mycoplasma penetrans; Mycoplasma synoviae* 53; *Streptobacillus moniliformis* DSM 12112; *Bradyrhizobium* BTAi1; *Nitrobacter hamburgensis* X14; *Rhodopseudomonas palustris* BisB 18; *Rhodopseudomonas palustris* B is B5; *Parvibaculum lavamentivorans* DS-1; *Dinoroseobacter shibae* DFL 12; *Gluconacetobacter diazotrophicus* Pal 5 FAPERJ; *Gluconacetobacter diazotrophicus* Pal 5 JGI; *Azospirillum* B510 uid46085; *Rhodospirillum rubrum* ATCC 11170; *Diaphorobacter* TPSY uid29975; *Verminephrobacter eiseniae* EF01-2; *Neisseria meningitides*

053442; *Neisseria meningitides* alpha14; *Neisseria meningitides* Z2491; *Desulfovibrio salexigens* DSM 2638; *Campylobacter jejuni doylei* 269 97; *Campylobacter jejuni* 81116; *Campylobacter jejuni; Campylobacter lari* RM2100; *Helicobacter hepaticus; Wolinella succinogenes; Tolumonas auensis* DSM 9187; *Pseudoalteromonas atlantica* T6c; *Shewanella pealeana* ATCC 700345; *Legionella pneumophila* Paris; *Actinobacillus succinogenes* 130Z; *Pasteurella multocida; Francisella tularensis novicida* U112; *Francisella tularensis holarctica; Francisella tularensis* FSC 198; *Francisella tularensis tularensis; Francisella tularensis* WY96-3418; and *Treponema denticola* ATCC 35405. Accordingly, aspects of the present disclosure are directed to a Cas9 protein present in a Type II CRISPR system, which has been rendered nuclease null or which has been rendered a nickase as described herein.

[0084] The Cas9 protein may be referred by one of skill in the art in the literature as Csn1. The *S. pyogenes* Cas9 protein sequence that is the subject of experiments described herein is shown below. See Deltcheva et al., *Nature* 471, 602-607 (2011) hereby incorporated by reference in its entirety.

```
                                            (SEQ ID NO: 1)
MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE

                    -continued

DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK

PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ

SITGLYETRIDLSQLGGD-
```

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

[0086] According to an additional aspect, 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 or all of D10, H840, D839 and H863 are substituted with an amino acid which reduces, substantially eliminates or eliminates nuclease activity. According to one aspect, one or more or all of D10, H840, D839 and H863 are substituted with alanine. According to one aspect, a Cas9 protein having one or more or all of D10, H840, D839 and H863 substituted with an amino acid which reduces, substantially eliminates or eliminates nuclease activity, such as alanine, is referred to as a nuclease-null Cas9 or Cas9N and exhibits reduced or eliminated nuclease activity, or nuclease activity is absent or substantially absent within levels of detection. 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.

[0087] According to one aspect, the nuclease null Cas9 protein includes homologs and orthologs thereof which retain the ability of the protein to bind to the DNA and 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 an RNA guided DNA binding protein.

[0088] According to one aspect, the nuclease null Cas9 protein includes the sequence as set forth for naturally occurring Cas9 from *S. pyogenes* excepting the protein sequence of the RuvC nuclease domain and the HNH 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.

[0089] For purposes of the present disclosure, FIG. **4**A depicts metal coordinating residues in known protein structures with homology to Cas9. Residues are labeled based on position in Cas9 sequence. Left: RuvC structure, PDB ID: 4EP4 (blue) position D7, which corresponds to D10 in the Cas9 sequence, is highlighted in a Mg-ion coordinating position. Middle: Structures of HNH endonuclease domains from PDB IDs: 3M7K (orange) and 4H9D (cyan) including a coordinated Mg-ion (gray sphere) and DNA from 3M7K (purple). Residues D92 and N113 in 3M7K and 4H9D positions D53 and N77, which have sequence homology to Cas9 amino acids D839 and N863, are shown as sticks. Right: List of mutants made and analyzed for nuclease activity: Cas9 wild-type; $Cas9_{m1}$ which substitutes alanine for D10; $Cas9_{m2}$ which substitutes alanine for D10 and alanine for H840; $Cas9_{m3}$ which substitutes alanine for D10, alanine for H840, and alanine for D839; and $Cas9_{m4}$ which substitutes alanine for D10, alanine for H840, alanine for D839, and alanine for N863.

[0090] As shown in FIG. **4**B, 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. **4**C is a higher-resolution examination of the data in FIG. **4**B and confirms that the mutation landscape shows comparable profile as unmodified loci.

[0091] 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.

[0092] 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.

[0093] 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.

[0094] The gRNA tethers capable of transcriptional regulation were constructed by identifying which regions of the gRNA will tolerate modifications by inserting random sequences into the gRNA and assaying for Cas9 function. gRNAs bearing random sequence insertions at either the 5' end of the crRNA portion or the 3' end of the tracrRNA portion of a chimeric gRNA retain functionality, while insertions into the tracrRNA scaffold portion of the chimeric gRNA result in loss of function. See FIGS. **5**A-B summarizing gRNA flexibility to random base insertions. FIG. **5**A is a schematic of a homologous recombination (HR) assay to determine Cas9-gRNA activity. As shown in FIG. **5**B, 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.

[0095] To attach VP64 to the gRNA, two copies of the MS2 bacteriophage coat-protein binding RNA stem-loop were appended to the 3' end of the gRNA. See Fusco et al., *Current Biology: CB*13, 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.

[0096] FIG. **1**A is a schematic of RNA-guided transcriptional activation. As shown in FIG. **1**A, 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. **1**B, 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. **1**C shows design of reporter constructs used to assay transcriptional activation. The two reporters bear distinct gRNA target sites, and share a control TALE-TF target site. As shown in FIG. 1D, Cas9N-VP64 fusions display RNA-guided transcriptional activation as assayed by both fluorescence-activated cell sorting (FACS) and immunofluorescence assays (IF). Specifically, while the control TALE-TF activated both reporters, the Cas9N-VP64 fusion activates reporters in a gRNA sequence specific manner. As shown in FIG. 1E, gRNA sequence-specific transcriptional activation from reporter constructs only in the presence of all 3 components: Cas9N, MS2-VP64 and gRNA bearing the appropriate MS2 aptamer binding sites was observed by both FACS and IF.

[0097] According to certain aspects, methods are provided for regulating endogenous genes using Cas9N, one or more gRNAs and a transcriptional regulatory protein or domain. According to one aspect, an endogenous gene can be any desired gene, referred to herein as a target gene. According to one exemplary aspect, genes target for regulation included ZFP42 (REX1) and POU5F1 (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.

[0098] FIGS. 6A-C is directed to RNA-guided OCT4 regulation using Cas9N-VP64. As shown in FIG. 6A, **21** gRNAs targeting a ~5 kb stretch of DNA upstream of the transcription start site were designed for the OCT4 gene. The DNase hypersensitive sites are highlighted in green. FIG. 6B shows transcriptional activation using a promoter-luciferase reporter construct. FIG. 6C shows transcriptional activation directly via qPCR of the endogenous genes. While introduction of individual gRNAs modestly stimulated transcription, multiple gRNAs acted synergistically to stimulate robust multi-fold transcriptional activation.

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

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

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

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

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

[0104] As shown in FIG. **2**B, 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. **2**C, 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. **2**D, introduction of 2 base mismatches significantly impairs the Cas9-gRNA complex activity, however only when these are localized to the 8-10 bases nearer the 3' end of the gRNA target sequence (in the heat plot the target sequence positions are labeled from 1-23 starting from the 5' end).

[0105] The mutational tolerance of another widely used genome editing tool, TALE domains, was determined using the transcriptional specificity assay described herein. As shown in FIG. **2**E, 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. **2**F, the 18-mer TALE is, similar to the Cas9-gRNA complexes, largely insensitive to single base mismatched in its target. As shown in FIG. **2**G, introduction of 2 base mismatches significantly impairs the 18-mer TALE activity. TALE activity is more sensitive to mismatches nearer the 5' end of its target sequence (in the heat plot the target sequence positions are labeled from 1-18 starting from the 5' end).

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

[0107] FIGS. **8**A-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. **8**A, 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 (3^{rd} level, left) to establish which tags co-occur with binding sites, resulting in an association table of binding sites vs. transcribed tags (4^{th} 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 transfected into a cell population and a set of different Cas9/gRNA or TALE transcription factors are induced in samples of the populations (2^{nd} 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 (3^{rd} and 4^{th} 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 (5^{th} 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. **2**B and **2**E, and in FIG. **9**A and FIG. **10**B. Not covered in this overall process flow are several levels of filtering for erroneous tags, for tags not associable with a construct library, and for tags apparently shared with multiple binding sites. FIG. **8**B 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. **8**C 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. **8**B, while the distribution is skewed to the left for the non-control TALE sample because sites with fewer mismatches induce higher expression levels. Below: Computing the relative enrichment between these by dividing the tag counts obtained for the target-TF by those obtained for the control-TF reveals the average expression level versus the number of mutations in the target site.

[0108] These results are further reaffirmed by specificity data generated using a different Cas9-gRNA complex. As shown in FIG. **9**A, a different Cas9-gRNA complex is tolerant to 1-3 mutations in its target sequence. As shown in FIG. **9**B, the Cas9-gRNA complex is also largely insensitive to point mutations, except those localized to the PAM sequence. As shown in FIG. **9**C, 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. **9**D, it was confirmed using a nuclease mediated HR assay that the predicted PAM for the *S. pyogenes* Cas9 is NGG and also NAG.

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

[0110] FIGS. **3**A-C is directed to multiplex off-set nicking and methods for reducing the off-target binding with the guide RNAs. As shown in FIG. **3**A, 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 nicks 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. **3**B, 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. **3**C, generating 3' overhangs also favors the ratio of HR over NHEJ events, but the total number of HR events is significantly lower than when a 5' overhang is generated.

[0111] FIGS. **11**A-B is directed to Cas9D10A nickase mediated NHEJ. As shown in FIG. **11**A, 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. **11**B, it was observed that unlike the wild-type Cas9 which results in DSBs and robust NHEJ across all targets, most nicks (using the Cas9D10A mutant) seldom result in NHEJ events. All 14 sites are located within a contiguous 200 bp stretch of DNA and over 10-fold differences in targeting efficiencies were observed.

[0112] According to certain aspects, methods are described herein of modulating expression of a target nucleic acid in a cell that include introducing one or more, two or more or a plurality of foreign nucleic acids into the cell. The foreign nucleic acids introduced into the cell encode for a guide RNA or guide RNAs, a 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.

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

[0114] Target nucleic acids include any nucleic acid sequence to which a co-localization complex as described herein can be useful to either regulate or nick. Target nucleic acids include genes. For purposes of the present disclosure, DNA, such as double stranded DNA, can include the target nucleic acid and a co-localization complex can bind to or otherwise co-localize with the DNA at or adjacent or near the target nucleic acid and in a manner in which the co-localization complex may have a desired effect on the target nucleic acid. Such target nucleic acids can include endogenous (or naturally occurring) nucleic acids and exogenous (or foreign) nucleic acids. One of skill based on the present disclosure will readily be able to identify or design guide RNAs and Cas9 proteins which co-localize to a DNA including a target nucleic acid. One of skill will further be able to identify transcriptional regulator proteins or domains which likewise co-localize to a DNA including a target nucleic acid. DNA includes genomic DNA, mitochondrial DNA, viral DNA or exogenous DNA.

[0115] 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.

[0116] 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.

[0117] 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.

[0118] The following examples are set forth as being representative of the present disclosure. These examples are not to be construed as limiting the scope of the present disclosure as these and other equivalent embodiments will be apparent in view of the present disclosure, figures and accompanying claims.

Example I

Cas9 Mutants

[0119] Sequences homologous to Cas9 with known structure were searched to identify candidate mutations in Cas9 that could ablate the natural activity of its RuvC and HNH domains. Using HHpred (world wide website toolkit.tuebingen.mpg.de/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 IDs: 3M7K and 4H9D respectively). These proteins were examined to find residues involved in magnesium ion coordination. The corresponding residues were then identified in the sequence alignment to Cas9. Two Mg-coordinating side-chains in each structure were identified that aligned to the same amino acid type in Cas9. They are 3M7K D92 and N113, and 4H9D D53 and N77. These residues corresponded to Cas9 D839 and N863. It was also reported that mutations of Pad residues D92 and N113 to alanine rendered the nuclease catalytically deficient. The Cas9 mutations D839A and N863A were made based on this analysis. Additionally, HHpred also predicts homology between Cas9 and the N-terminus of a *Thermus thermophilus* RuvC (PDB ID: 4EP4). 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

[0120] 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 oligonucleotides into the gRNA cloning vector (plasmid #41824). The vectors for the HR reporter assay involving a broken GFP were constructed by fusion PCR assembly of the GFP sequence bearing the stop codon and appropriate fragment assembled into the EGIP 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

[0121] 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% CO_2 in a humidified incubator.

[0122] Transfections involving nuclease assays were as follows: 0.4×10^6 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 $\sim1\times10^6$ cells was extracted using DNAeasy kit (Qiagen). For these PCR was 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 NHEJ efficiencies.

[0123] For transfections involving transcriptional activation assays: 0.4×10^6 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-2XMS2aptamer 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.

[0124] For transfections involving transcriptional activation assays for specificity profile of Cas9-gRNA complexes and TALEs: 0.4×10^6 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 μg of reporter library; or (3) 2 μg control-TF plasmid and 0.25 μg of reporter library. Cells were harvested 24 hrs post transfection (to avoid the stimulation of reporters being in saturation mode). Total RNA extraction was performed using RNAeasy-plus kit (Qiagen), and standard RT-per 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

[0125] The high-level logic flow for this process is depicted in FIG. **8**A, and additional details are given here. For details on construct library composition, see FIGS. **8**A (level 1) and **8**B.

Sequencing:

[0126] For Cas9 experiments, construct library (FIG. **8**A, level 3, left) and reporter gene cDNA sequences (FIG. **8**A, level 3, right) were obtained as 150 bp overlapping paired end reads on an Illumina MiSeq, while for TALE experiments, corresponding sequences were obtained as 51 bp non-overlapping paired end reads on an Illumina HiSeq.

Construct Library Sequence Processing:

[0127] 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. **8**A, 3[rd] 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. **8**A, 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:

[0128] Alignment: SeqPrep (downloaded from world wide website github.com/jstjohn/SeqPrep) was first used to merge the overlapping read pairs to the 79 bp common segment, after which novoalign (version above) was used to align these 79 bp common segments as unpaired single reads to a set of reference sequences (see FIG. **8**A, 3[rd] level, right) in which (as for the construct library sequencing) the 24 bp degenerate transcript tag was specified as Ns while the sample barcodes were explicitly provided. Both TALE and Cas9 cDNA sequence regions corresponded to the same 63 bp regions of cDNA flanked by pairs of 8 bp sample barcode sequences. Validity checking: The same conditions were applied as for construct library sequencing (see above) except that: (a) Here, due prior SeqPrep merging of read pairs, validity processing did not have to filter for unique alignments of both reads in a read pair but only for unique alignments of the merged reads. (b) Only transcript tags appeared in the cDNA sequence reads, so that validity processing only applied these tag regions of the reference sequences and not also to a separate binding site region.

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

[0129] Custom perl was used to generate these tables from the validated construct library sequences (FIG. **8**A, 4[th] 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, nonshared, 2+ tags; and the least secure (but most abundant) class comprising all tags regardless of safety, sharing, or number of occurrences.

Computation of Normalized Expression Levels:

[0130] Custom perl code was used to implement the steps indicated in FIG. **8**A, 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. **8**C). 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

$$\binom{L}{n}3^n\Big),$$

3″), and so dramatically increases with n; thus the averages of aggregated tag counts for different numbers n of mismatches (shown in FIGS. **2***b*, **2***e*, and in FIGS. **9**A and **10**B) 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 $Cas9_{N}$-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 Cas9N– protein. For the FIG. **12**A SOX2 and FIG. **12**B 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 FIGS. **12**A-B, two additional genes, SOX2 and NANOG, were regulated via sgRNAs targeting within an upstream ~1 kb stretch of promoter DNA. The sgRNAs proximal to the transcriptional start site resulted in robust gene activation.

Example VI

Evaluating the Landscape of Targeting by Cas9-gRNA Complexes

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

Example VII

Validations, Specificity of Reporter Assay

[0133] As shown in FIGS. **14**A-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. **14**A: 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. **14**B: It was confirmed that this assay was specific for the gRNA being evaluated (data re-plotted from FIG. **13**D), 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. **13**A, **13**D), specifically, sgRNA2 tolerates up to 3 mismatches and sgRNA3 only up to 1. The greatest sensitivity to mismatches was localized to the 3' end of the spacer, albeit mismatches at other positions were also observed to affect activity.

Example VIII

Validations, Single and Double-Base gRNA Mismatches

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

Example IX

Validations, 5' gRNA Truncations

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

Example X

Validations, *S. pyogenes* PAM

[0136] As shown in FIGS. **17**A-B, it was confirmed using a nuclease mediated HR assay that the PAM for the *S. pyogenes* Cas9 is NGG and also NAG. Data are means+/−SEM (N=3). According to an additional investigation, a generated set of about 190K Cas9 targets in human exons that had no alternate NGG targets 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

Validations, TALE Mutations

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

Example XII

TALE Monomer Specificity Versus TALE Protein Specificity

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

Example XIII

Off-Set Nicking, Native Locus

[0139] FIGS. **20**A-B shows data related to off-set nicking. In the context of genome-editing, off-set nicks 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. **20**A: 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 nicks the complementary strand, different two-way combinations of the gRNAs was used to induce a range of programmed 5' or 3' overhangs. FIG. **20**B: 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] FIGS. **21**A-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. **3**B). 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. **3**C).

Example XV

[0142]

TABLE 1

gRNA Targets for Endogenous Gene Regulation Targets in the REX1, OCT4, SOX2 and NANOG promoters used in Cas9-gRNA mediated activation experiments are listed and set forth as SEQ ID NOs: 11-61.

gRNA Name	gRNA Target
REX1 1	ctggcggatcactcgcggtt agg
REX1 2	cctcggcctccaaaagtgct agg
REX1 3	acgctgattcctgcagatca ggg
REX1 4	ccaggaatacgtatccacca ggg
REX1 5	gccacacccaagcgatcaaa tgg
REX1 6	aaataatacattctaaggta agg
REX1 7	gctactggggaggctgaggc agg
REX1 8	tagcaatacagtcacattaa tgg
REX1 9	ctcatgtgatccccccgtct cgg
REX1 10	ccgggcagagagtgaacgcg cgg
OCT4 1	ttccttccctctcccgtgct tgg
OCT4 2	tctctgcaaagcccctggag agg
OCT4 3	aatgcagttgccgagtgcag tgg
OCT4 4	cctcagcctcctaaagtgct ggg
OCT4 5	gagtccaaatcctctttact agg
OCT4 6	gagtgtctggatttgggata agg
OCT4 7	cagcacctcatctcccagtg agg
OCT4 8	tctaaaacccagggaatcat ggg
OCT4 9	cacaaggcagccagggatcc agg
OCT4 10	gatggcaagctgagaaacac tgg
OCT4 11	tgaaatgcacgcatacaatt agg
OCT4 12	ccagtccagacctggccttc tgg
OCT4 13	cccagaaaaacagaccctga agg
OCT4 14	aagggttgagcacttgttta ggg
OCT4 15	atgtctgagttttggttgag agg
OCT4 16	ggtcccttgaaggggaagta ggg
OCT4 17	tggcagtctactcttgaaga tgg
OCT4 18	ggcacagtgccagaggtctg tgg
OCT4 19	taaaaataaaaaaactaaca ggg
OCT4 20	tctgtgggggacctgcactg agg
OCT4 21	ggccagaggtcaaggctagt ggg
SOX2 1	cacgaccgaaacccttctta cgg
SOX2 2	gttgaatgaagacagtctag tgg
SOX2 3	taagaacagagcaagttacg tgg
SOX2 4	tgtaaggtaagagaggagag cgg
SOX2 5	tgacacaccaactcctgcac tgg
SOX2 6	tttacccacttccttcgaaa agg
SOX2 7	gtggctggcaggctggctct ggg
SOX2 8	ctcccccggcctcccccgcg cgg
SOX2 9	caaaacccggcagcgaggct ggg
SOX2 10	aggagccgccgcgcgctgat tgg
NANOG 1	cacacacacccacacgagat ggg
NANOG 2	gaagaagctaaagagccaga ggg
NANOG 3	atgagaatttcaataacctc agg
NANOG 4	tcccgctctgttgcccaggc tgg
NANOG 5	cagacacccaccaccatgcg tgg
NANOG 6	tcccaatttactgggattac agg
NANOG 7	tgatttaaaagttggaaacg tgg
NANOG 8	tctagttccccacctagtct ggg
NANOG 9	gattaactgagaattcacaa ggg
NANOG 10	cgccaggaggggtgggtcta agg

Example XVI

[0143]

TABLE 2

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

a					
FIG.	Expression level comparison: mutations vs. mutations		t-test	P-value	Symbol
2b	0	1	1-samp	7.8E-05	**
	1	2	2-samp	1.4E-06	***
	2	3	2-samp	4.0E-61	***
	3	4	2-samp	0	***
	4	5	2-samp	0	***
	5	6	2-samp	1.0E-217	***
	6	7	2-samp	1.7E-43	***
	7	8	2-samp	3.7E-02	N.S.
2e	0	1	1-samp	8.9E-01	N.S.
	1	2	2-samp	1.9E-06	***
	2	3	2-samp	5.0E-147	***
	3	4	2-samp	0	***
	4	5	2-samp	0	***
	5	6	2-samp	4.2E-62	***
	6	7	2-samp	1.6E-03	*
	7	8	2-samp	4.7E-01	N.S.
S7a	0	1	1-samp	5.2E-02	N.S.
	1	2	2-samp	2.8E-05	***
	2	3	2-samp	3.5E-21	***
	3	4	2-samp	1.4E-58	***
	4	5	2-samp	8.3E-101	***
	5	6	2-samp	6.8E-94	***
	6	7	2-samp	1.8E-61	***
	7	8	2-samp	8.1E-24	***
S7d and S8d	0	1	1-samp	2.3E-18	***
	1	2	2-samp	2.4E-08	***
	2	3	2-samp	6.2E-54	***
	3	4	2-samp	4.0E-141	***
	4	5	2-samp	1.9E-20	***
	5	6	2-samp	1.2E-03	*
	6	7	2-samp	3.8E-05	***
	7	8	2-samp	9.4E-01	N.S.
S8c	0	1	1-samp	7.2E-03	N.S.
	1	2	2-samp	5.0E-01	N.S.
	2	3	2-samp	3.9E-84	***
	3	4	2-samp	8.5E-153	***
	4	5	2-samp	8.6E-76	***
	5	6	2-samp	1.6E-03	*
	6	7	2-samp	7.1E-01	N.S.
	7	8	2-samp	7.8-02	N.S.
S13a (left)	0	1	1-samp	7.3E-01	N.S.
	1	2	2-samp	2.4E-06	***
	2	3	2-samp	7.2E-140	***
	3	4	2-samp	0	***
	4	5	2-samp	0	***
	5	6	2-samp	1.0E-72	***
	6	7	2-samp	4.0E-03	*
S13a (middle)	0	1	1-samp	9.4E-02	N.S.
	1	2	2-samp	5.2E-09	***
	2	3	2-samp	7.9E-86	***
	3	4	2-samp	2.9E-53	***
	4	5	2-samp	3.5E-10	***
S13a (right)	0	1	1-samp	1.3E-13	***
	1	2	2-samp	1.1E-04	***
	2	3	2-samp	3.7E-08	***

TABLE 2-continued

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

b			
seed start position	Number postion pairs		-log10 P-value
	both in seed	not both in seed	
2	171	19	3.11
3	153	37	1.46
4	136	54	2.01
5	120	70	3.34
6	105	85	5.65
7	91	99	7.34
8	78	112	6.61
9	66	124	7.10
10	55	135	9.72
11	45	145	9.83
12	36	154	10.44
13	28	162	10.72
14	21	169	8.97
15	15	175	5.61
16	10	180	3.34
17	6	184	2.26
18	3	187	1.16

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 FIGS. indicated in the FIG. 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, * P < .05/n, and N.S. (Non-Significant) for P >= .05/n, where n is the number of comparison.
Table 2(b) Statistical characterization of seed region in FIG. 2D: log10 (P-values) indicating the degree of separation between expression values for Cas9N VP64 + gRNA binding to target sequences with two mutations for those position pairs mutated within candidate seed regions at the 3' end of the 20bp target site vs. all other position pairs. The greatest separation, indicated by the largest -log10 (P-values) (highlighted above), is found in the last 8-9bp 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

[0144] A. Sequences of the $Cas9_{N}$-VP64 activator constructs based on the m4 mutant are displayed below. Three versions were constructed with the $Cas9_{m4}^{VP64}$ and $Cas9_{m4}^{VP64}$N fusion protein formats showing highest activity. Corresponding vectors for the m3 and m2 mutants (FIG. **4**A) were also constructed (NLS and VP64 domains are highlighted).

$Cas9_{m4}^{VP64}$

[0145]

(SEQ ID NO: 2)
gccaccATGGACAAGAAGTACTCCATTGGGCTCGCTATCGGCACAAACAG
CGTCGGCTGGGCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAAT
TCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAAGAACCTCATT
GGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGCGGCTCAA
AAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACC
TGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTC
CATAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAGCACGAGCG

-continued

CCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGT
ACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAG
GCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCG
GGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCG
ACAAACTCTTTATCCAACTGGTTCAGACTTACAATCAGCTTTTCGAAGAG
AACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAG
GCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGG
AGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTG
ACCCCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCA
ACTGAGCAAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGA
TCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCAGAC
GCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGC
TCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCACCAAGACT
TGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAAGTACAAG
GAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGG
CGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAA
AAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTG
TTGCGCAAACAGCGCACTTTCGACAATGGAAGCATCCCCCACCAGATTCA
CCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCCCT
TTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCGGATA
CCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGAT
GACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACTTCGAGGAAGTCG
TGGATAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTT
GATAAAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGTA
CGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAG
AAGGGATGAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATC
GTGGACCTCCTCTTCAAGACGAACCGGAAAGTTACCGTGAAACAGCTCAA
AGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCG
GAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTG
AAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACAT
TCTTGAGGACATTGTCCTCACCCTTACGTTGTTTGAAGATAGGGAGATGA
TTGAAGAACGCTTGAAAACTTACGCTCATCTCTTCGACGACAAAGTCATG
AAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAAA
ACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATT
TTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCAT
GATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGG
CCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCCCAG
CTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTC
AAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCG

-continued

AGAGAACCAAACTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGA
AGAGGATTGAAGAGGGTATAAAAGAACTGGGGTCCCAAATCCTTAAGGAA
CACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTACTA
CCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATC
GGCTCTCCGACTACGACGTGGCTGCTATCGTGCCCCAGTCTTTTCTCAAA
GATGATTCTATTGATAATAAAGTGTTGACAAGATCCGATAAAgcTAGAGG
GAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATT
ATTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGAT
AATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGG
CTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGG
CCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAA
CTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGA
TTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACC
ACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATC
AAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGT
GTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGG
CCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACC
GAGATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAAC
AAACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTCGCGA
CAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGTTAAAAAGACC
GAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAA
CAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACG
GCGGATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAA
GTGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGG
CATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTTC
TCGAGGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTT
CCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGC
TAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAAT
ACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCT
CCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTA
CCTTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCC
TCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGG
GATAAGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTTACTCT
GACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATAG
ACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATT
CATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCT
CGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGAGGCCA
GCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTG
GGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGC

-continued

CCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATT
TCGACCTGGACATGCTGATTAACTCTAGATGA

$Cas9_{m4}^{VP64}$N Sequences

[0146]

(SEQ ID NO: 3)
gccaccATGCCCAAGAAGAAGAGGAAGGTGGGAAGGGGGATGGACAAGAA
GTACTCCATTGGGCTCGCTATCGGCACAAACAGCGTCGGCTGGGCCGTCA
TTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAAT
ACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGA
CTCCGGGGAGACGGCCGAAGCCACGCGGCTCAAAAGAACAGCACGGCGCA
GATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGAGATCTTTAGT
AATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTC
CTTTTTGGTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCA
ATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCAACCATATATCAT
CTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTTGAT
CTATCTCGCGCTGGCGCATATGATCAAATTTCGGGGACACTTCCTCATCG
AGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAA
CTGGTTCAGACTTACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATC
CGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAATCCCGGC
GGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGGCCTG
TTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATC
TAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCT
ACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCA
GACCTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTGA
TATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTA
TGATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCC
CTTGTCAGACAGCAACTGCCTGAGAAGTACAAGGAAATTTTCTTCGATCA
GTCTAAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAGGAGG
AATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCACCGAG
GAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCAC
TTTCGACAATGGAAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACG
CTATCCTCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACAGG
GAAAAGATTGAGAAAATCCTCACATTTCGGATACCCTACTATGTAGGCCC
CCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATCAGAAG
AGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTCT
GCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAA
CGAAAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTT
ATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCA

-continued

GCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAA

GACGAACCGGAAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAA

AGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTC

AACGCATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGACAA

GGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCC

TCACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAA

ACTTACGCTCATCTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGCG

CCGATATACAGGATGGGGGCGGCTGTCAAGAAAACTGATCAATGGGATCC

GAGACAAGCAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGATGGA

TTTGCCAACCGGAACTTCATGCAGTTGATCCATGATGACTCTCTCACCTT

TAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTC

ACGAGCACATCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATA

CTGCAGACCGTTAAGGTCGTGGATGAACTCGTCAAAGTAATGGGAAGGCA

TAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACTACCC

AGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGT

ATAAAAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACAC

CCAGCTTCAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGG

ACATGTACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGACTACGAC

GTGGCTGCTATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAA

TAAAGTGTTGACAAGATCCGATAAAgcTAGAGGGAAGAGTGATAACGTCC

CCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTGCTG

AACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAGGCTGA

ACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGC

TTGTTGAGACACGCCAGATCACCAAGCACGTGGCCCAAATTCTCGATTCA

CGCATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAA

AGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTCAGAAAGGACTTTC

AGTTTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCC

TACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATCCCAAGCT

TGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAA

TGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTC

TTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGGCCAA

TGGAGAGATTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAG

AAATCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTG

TCCATGCCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCGGAGG

CTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCTGATCG

CACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCCT

ACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTC

TAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGC

GATCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAAAGGATAT

-continued

AAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTCTT

TGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGAGCTGC

AGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTAT

CTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCA

GAAGCAGCTGTTCGTGGAACAACACAAACACTACCTTGATGAGATCATCG

AGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTC

GATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGA

GCAGGCAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGC

CTGCAGCCTTCAAGTACTTCGACACCACCATAGACAGAAAGCGGTACACC

TCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAATTACGGG

GCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGG

CTGACCCCAAGAAGAAGAGGAAGGTGGAGGCCAGCGGTTCCGGACGGGCT

GACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGA

TGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACC

TCGACATGCTCGGCAGTGACGCCCTTGATGATTTCGACCTGGACATGCTG

ATTAACTCTAGATGA

Cas$9_{m4}{}^{VP64}$C

[0147]

(SEQ ID NO: 4)

gccaccATGGACAAGAAGTACTCCATTGGGCTCGCTATCGGCACAAACAG

CGTCGGCTGGGCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAAT

TCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAAGAACCTCATT

GGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGCGGCTCAA

AAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACC

TGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTC

CATAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAGCACGAGCG

CCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGT

ACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAG

GCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTCG

GGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCG

ACAAACTCTTTATCCAACTGGTTCAGACTTACAATCAGCTTTTCGAAGAG

AACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAG

GCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGG

AGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTG

ACCCCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCA

ACTGAGCAAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGA

TCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTCAGAC

GCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGC

-continued

TCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCACCAAGACT
TGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAAGTACAAG
GAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGG
CGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAA
AAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTG
TTGCGCAAACAGCGCACTTTCGACAATGGAAGCATCCCCCACCAGATTCA
CCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCCCT
TTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCGGATA
CCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGAT
GACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACTTCGAGGAAGTCG
TGGATAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTT
GATAAAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGTA
CGAGTACTTCACAGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAG
AAGGGATGAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATC
GTGGACCTCCTCTTCAAGACGAACCGGAAAGTTACCGTGAAACAGCTCAA
AGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCG
GAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTG
AAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACAT
TCTTGAGGACATTGTCCTCACCCTTACGTTGTTTGAAGATAGGGAGATGA
TTGAAGAACGCTTGAAAACTTACGCTCATCTCTTCGACGACAAAGTCATG
AAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAAA
ACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATT
TTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCAT
GATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGG
CCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCCCAG
CTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTC
AAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCG
AGAGAACCAAACTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGA
AGAGGATTGAAGAGGGTATAAAAGAACTGGGGTCCCAAATCCTTAAGGAA
CACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTACTA
CCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATC
GGCTCTCCGACTACGACGTGGCTGCTATCGTGCCCCAGTCTTTTCTCAAA
GATGATTCTATTGATAATAAAGTGTTGACAAGATCCGATAAAgcTAGAGG
GAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATT
ATTGGCGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGAT
AATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGG
CTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGCACGTGG
CCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAA
CTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGA

-continued

TTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACC
ACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATC
AAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGT
GTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGG
CCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACC
GAGATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAAC
AAACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTCGCGA
CAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGTTAAAAAGACC
GAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAA
CAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACG
GCGGATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAA
GTGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGG
CATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTTC
TCGAGGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTT
CCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGC
TAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAAT
ACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCT
CCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTA
CCTTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCC
TCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGG
GATAAGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTTACTCT
GACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATAG
ACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATT
CATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCT
CGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGAGGCCA
GCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTG
GGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGC
CCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATT
TCGACCTGGACATGCTGATTAACTCTAGAGCGGCCGCAGATCCAAAAAAG
AAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAA
GAGAAAGGTAGATACGGCCGCATAG

B. Sequences of the MS2-activator constructs and corresponding gRNA backbone vector with 2×MS2 aptamer domains is provided below (NLS, VP64, gRNA spacer, and MS2-binding RNA stem loop domains are highlighted). Two versions of the former were constructed with the $MS2_{VP64}N$ fusion protein format showing highest activity.

$MS2_{VP64}N$

[0148]

(SEQ ID NO: 5)

```
gccaccATGGGACCTAAGAAAAAGAGGAAGGTGGCGGCCGCTTCTAGAAT
GGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCG
ACGTGACTGTCGCCCCAAGCAACTTCGCTAACGGGATCGCTGAATGGATC
AGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTCA
GAGCTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAG
GCGCCTGGCGTTCGTACTTAAATATGGAACTAACCATTCCAATTTTCGCC
ACGAATTCCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAA
AGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACTCCGGCATCTACG
AGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGAT
ATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTC
GGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTG
ATGATTTCGACCTGGACATGCTGATTAACTCTAGATGA
```

$MS2_{VP64}C$

[0149]

(SEQ ID NO: 6)

```
gccaccATGGGACCTAAGAAAAAGAGGAAGGTGGCGGCCGCTTCTAGAAT
GGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCG
ACGTGACTGTCGCCCCAAGCAACTTCGCTAACGGGATCGCTGAATGGATC
AGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTCA
GAGCTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAG
GCGCCTGGCGTTCGTACTTAAATATGGAACTAACCATTCCAATTTTCGCC
ACGAATTCCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAA
AGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACTCCGGCATCTACG
AGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGAT
ATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTC
GGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTG
ATGATTTCGACCTGGACATGCTGATTAACTCTAGAGCGGCCGCAGATCCA
AAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAA
AAAGAAGAGAAAGGTAGATACGGCCGCATAG
```

$gRNA_{2XMS2}$

(SEQ ID NO: 7)

```
TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGG
TACCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATT
TGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACT
GTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATT
```

-continued

```
TCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATA
TGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTG
TGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCT
AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG
GCACCGAGTCGGTGCTCTGCAGGTCGACTCTAGAAAACATGAGGATCACC
CATGTCTGCAGTATTCCCGGGTTCATTAGATCCTAAGGTACCTAATTGCC
TAGAAAACATGAGGATCACCCATGTCTGCAGGTCGACTCTAGAAATTTTT
TCTAGAC
```

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

TF Reporter 1

[0150]

(SEQ ID NO: 8)

```
TAGGGATAACAGGGTAATAGTGTCCCCTCCACCCCACAGTGGGGCGAGGT
AGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCG
TCAGATCGCCTGGAGAATTCgccaccatgGACTACAAGGATGACGACGAT
AAAACTTCCGGTGGCGGACTGGGTTCCACCGTGAGCAAGGGCGAGGAGGT
CATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGAACG
GCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC
ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGC
CTGGGACATCCTGTCCCCCCAGTTCATGTACGGCTCCAAGGCGTACGTGA
AGCACCCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGC
TTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGTCTGGTGACCGT
GACCCAGGACTCCTCCCTGCAGGACGGCACGCTGATCTACAAGGTGAAGA
TGCGCGGCACCAACTTCCCCCCCGACGGCCCCGTAATGCAGAAGAAGACC
ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCT
GAAGGGCGAGATCCACCAGGCCCTGAAGCTGAAGGACGGCGGCCACTACC
TGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCAACTGCCC
GGCTACTACTACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGA
CTACACCATCGTGGAACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGT
TCCTGTACGGCATGGACGAGCTGTACAAGTAA
```

TF Reporter 2

[0151]

(SEQ ID NO: 9)

```
TAGGGATAACAGGGTAATAGTGGGGCCACTAGGGACAGGATTGGCGAGGT
AGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCG
TCAGATCGCCTGGAGAATTCgccaccatgGACTACAAGGATGACGACGAT
```

-continued

AAAACTTCCGGTGGCGGACTGGGTTCCACCGTGAGCAAGGGCGAGGAGGT

CATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGAACG

GCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGC

ACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGC

CTGGGACATCCTGTCCCCCCAGTTCATGTACGGCTCCAAGGCGTACGTGA

AGCACCCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGC

TTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGTCTGGTGACCGT

GACCCAGGACTCCTCCCTGCAGGACGGCACGCTGATCTACAAGGTGAAGA

TGCGCGGCACCAACTTCCCCCCCGACGGCCCCGTAATGCAGAAGAAGACC

ATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCT

GAAGGGCGAGATCCACCAGGCCCTGAAGCTGAAGGACGGCGGCCACTACC

TGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCAACTGCCC

GGCTACTACTACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGA

CTACACCATCGTGGAACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGT

TCCTGTACGGCATGGACGAGCTGTACAAGTAA

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

Specificity Reporter Libraries

[0152]

(SEQ ID NO: 10)

TAGGGATAACAGGGTAATAGTNNNNNNNNNNNNNNNNNNNNNNNCGAGGT

AGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCG

TCAGATCGCCTGGAGAATTCgccaccatgGACTACAAGGATGACGACGAT

AAANNNNNNNNNNNNNNNNNNNNNNNNACTTCCGGTGGCGGACTGGGTTC

CACCGTGAGCAAGGGCGAGGAGGTCATCAAAGAGTTCATGCGCTTCAAGG

TGCGCATGGAGGGCTCCATGAACGGCCACGAGTTCGAGATCGAGGGCGAG

GGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGAC

CAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCA

TGTACGGCTCCAAGGCGTACGTGAAGCACCCCGCCGACATCCCCGATTAC

AAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTT

CGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCTGCAGGACGG

CACGCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCCCGACG

GCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGC

CTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACCAGGCCCTGAA

GCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGACCATCTACATGG

CCAAGAAGCCCGTGCAACTGCCCGGCTACTACTACGTGGACACCAAGCTG

GACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCG

CTCCGAGGGCCGCCACCACCTGTTCCTGTACGGCATGGACGAGCTGTACA

AGTAAGAATTC

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 186

<210> SEQ ID NO 1
<211> LENGTH: 1368
<212> TYPE: PRT
<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 1

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

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

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

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

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

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

-continued

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
```

-continued

```
        515                 520                 525

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

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

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

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

-continued

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

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

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

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

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

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

-continued

```
Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                 1315                 1320

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

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

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


<210> SEQ ID NO 2
<211> LENGTH: 4332
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: VP64-activator construct

<400> SEQUENCE: 2

gccaccatgg acaagaagta ctccattggg ctcgctatcg gcacaaacag cgtcggctgg     60

gccgtcatta cggacgagta caaggtgccg agcaaaaaat tcaaagttct gggcaatacc    120

gatcgccaca gcataaagaa gaacctcatt ggcgccctcc tgttcgactc cggggagacg    180

gccgaagcca cgcggctcaa aagaacagca cggcgcagat atacccgcag aaagaatcgg    240

atctgctacc tgcaggagat ctttagtaat gagatggcta aggtggatga ctctttcttc    300

cataggctgg aggagtcctt tttggtggag gaggataaaa agcacgagcg ccacccaatc    360

tttggcaata tcgtggacga ggtggcgtac catgaaaagt acccaaccat atatcatctg    420

aggaagaagc ttgtagacag tactgataag gctgacttgc ggttgatcta tctcgcgctg    480

gcgcatatga tcaaatttcg gggacacttc ctcatcgagg gggacctgaa cccagacaac    540

agcgatgtcg acaaactctt tatccaactg gttcagactt acaatcagct tttcgaagag    600

aacccgatca acgcatccgg agttgacgcc aaagcaatcc tgagcgctag gctgtccaaa    660

tcccggcggc tcgaaaacct catcgcacag ctccctgggg agaagaagaa cggcctgttt    720

ggtaatctta tcgccctgtc actcgggctg acccccaact ttaaatctaa cttcgacctg    780

gccgaagatg ccaagcttca actgagcaaa gacacctacg atgatgatct cgacaatctg    840

ctggcccaga tcggcgacca gtacgcagac ctttttttgg cggcaaagaa cctgtcagac    900

gccattctgc tgagtgatat tctgcgagtg aacacggaga tcaccaaagc tccgctgagc    960

gctagtatga tcaagcgcta tgatgagcac caccaagact tgactttgct gaaggccctt   1020

gtcagacagc aactgcctga gaagtacaag gaaattttct tcgatcagtc taaaaatggc   1080

tacgccggat acattgacgg cggagcaagc caggaggaat tttacaaatt tattaagccc   1140

atcttggaaa aaatggacgg caccgaggag ctgctggtaa agcttaacag agaagatctg   1200

ttgcgcaaac agcgcacttt cgacaatgga agcatccccc accagattca cctgggcgaa   1260

ctgcacgcta tcctcaggcg gcaagaggat ttctacccct ttttgaaaga taacagggaa   1320

aagattgaga aaatcctcac atttcggata ccctactatg taggcccccт cgcccgggga   1380

aattccagat tcgcgtggat gactcgcaaa tcagaagaga ccatcactcc ctggaacttc   1440

gaggaagtcg tggataaggg ggcctctgcc cagtccttca tcgaaaggat gactaacttt   1500

gataaaaatc tgcctaacga aaaggtgctt cctaaacact ctctgctgta cgagtacttc   1560

acagtttata acgagctcac caaggtcaaa tacgtcacag aagggatgag aaagccagca   1620
```

-continued

```
ttcctgtctg gagagcagaa gaaagctatc gtggacctcc tcttcaagac gaaccggaaa   1680
gttaccgtga aacagctcaa agaagactat ttcaaaaaga ttgaatgttt cgactctgtt   1740
gaaatcagcg gagtggagga tcgcttcaac gcatccctgg gaacgtatca cgatctcctg   1800
aaaatcatta aagacaagga cttcctggac aatgaggaga acgaggacat tcttgaggac   1860
attgtcctca cccttacgtt gtttgaagat agggagatga ttgaagaacg cttgaaaact   1920
tacgctcatc tcttcgacga caaagtcatg aaacagctca agaggcgccg atatacagga   1980
tgggggcggc tgtcaagaaa actgatcaat gggatccgag acaagcagag tggaaagaca   2040
atcctggatt ttcttaagtc cgatggattt gccaaccgga acttcatgca gttgatccat   2100
gatgactctc tcacctttaa ggaggacatc cagaaagcac aagtttctgg ccagggggac   2160
agtcttcacg agcacatcgc taatcttgca ggtagcccag ctatcaaaaa gggaatactg   2220
cagaccgtta aggtcgtgga tgaactcgtc aaagtaatgg gaaggcataa gcccgagaat   2280
atcgttatcg agatggcccg agagaaccaa actacccaga agggacagaa gaacagtagg   2340
gaaaggatga agaggattga agagggtata aaagaactgg ggtcccaaat ccttaaggaa   2400
cacccagttg aaaacaccca gcttcagaat gagaagctct acctgtacta cctgcagaac   2460
ggcagggaca tgtacgtgga tcaggaactg gacatcaatc ggctctccga ctacgacgtg   2520
gctgctatcg tgccccagtc ttttctcaaa gatgattcta ttgataataa agtgttgaca   2580
agatccgata aagctagagg gaagagtgat aacgtcccct cagaagaagt tgtcaagaaa   2640
atgaaaaatt attggcggca gctgctgaac gccaaactga tcacacaacg gaagttcgat   2700
aatctgacta aggctgaacg aggtggcctg tctgagttgg ataaagccgg cttcatcaaa   2760
aggcagcttg ttgagacacg ccagatcacc aagcacgtgg cccaaattct cgattcacgc   2820
atgaacacca agtacgatga aaatgacaaa ctgattcgag aggtgaaagt tattactctg   2880
aagtctaagc tggtctcaga tttcagaaag gactttcagt tttataaggt gagagagatc   2940
aacaattacc accatgcgca tgatgcctac ctgaatgcag tggtaggcac tgcacttatc   3000
aaaaaatatc ccaagcttga atctgaattt gtttacggag actataaagt gtacgatgtt   3060
aggaaaatga tcgcaaagtc tgagcaggaa ataggcaagg ccaccgctaa gtacttcttt   3120
tacagcaata ttatgaattt tttcaagacc gagattacac tggccaatgg agagattcgg   3180
aagcgaccac ttatcgaaac aaacggagaa acaggagaaa tcgtgtggga caagggtagg   3240
gatttcgcga cagtccggaa ggtcctgtcc atgccgcagg tgaacatcgt taaaaagacc   3300
gaagtacaga ccggaggctt ctccaaggaa agtatcctcc cgaaaaggaa cagcgacaag   3360
ctgatcgcac gcaaaaaaga ttgggacccc aagaaatacg gcggattcga ttctcctaca   3420
gtcgcttaca gtgtactggt tgtggccaaa gtggagaaag ggaagtctaa aaaactcaaa   3480
agcgtcaagg aactgctggg catcacaatc atggagcgat caagcttcga aaaaaacccc   3540
atcgactttc tcgaggcgaa aggatataaa gaggtcaaaa aagacctcat cattaagctt   3600
cccaagtact ctctctttga gcttgaaaac ggccggaaac gaatgctcgc tagtgcgggc   3660
gagctgcaga aaggtaacga gctggcactg ccctctaaat acgttaattt cttgtatctg   3720
gccagccact atgaaaagct caaagggtct cccgaagata atgagcagaa gcagctgttc   3780
gtggaacaac acaaacacta ccttgatgag atcatcgagc aaataagcga attctccaaa   3840
agagtgatcc tcgccgacgc taacctcgat aaggtgcttt ctgcttacaa taagcacagg   3900
```

-continued

```
gataagccca tcagggagca ggcagaaaac attatccact tgtttactct gaccaacttg   3960

ggcgcgcctg cagccttcaa gtacttcgac accaccatag acagaaagcg gtacacctct   4020

acaaaggagg tcctggacgc cacactgatt catcagtcaa ttacggggct ctatgaaaca   4080

agaatcgacc tctctcagct cggtggagac agcagggctg accccaagaa gaagaggaag   4140

gtggaggcca gcggttccgg acgggctgac gcattggacg attttgatct ggatatgctg   4200

ggaagtgacg ccctcgatga ttttgacctt gacatgcttg gttcggatgc ccttgatgac   4260

tttgacctcg acatgctcgg cagtgacgcc cttgatgatt tcgacctgga catgctgatt   4320

aactctagat ga                                                        4332
```

<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

```
gccaccatgc ccaagaagaa gaggaaggtg ggaaggggga tggacaagaa gtactccatt     60

gggctcgcta tcggcacaaa cagcgtcggc tgggccgtca ttacggacga gtacaaggtg    120

ccgagcaaaa aattcaaagt tctgggcaat accgatcgcc acagcataaa gaagaacctc    180

attggcgccc tcctgttcga ctccggggag acggccgaag ccacgcggct caaaagaaca    240

gcacggcgca gatatacccg cagaaagaat cggatctgct acctgcagga gatctttagt    300

aatgagatgg ctaaggtgga tgactctttc ttccataggc tggaggagtc ctttttggtg    360

gaggaggata aaaagcacga gcgccaccca atctttggca atatcgtgga cgaggtggcg    420

taccatgaaa agtacccaac catatatcat ctgaggaaga agcttgtaga cagtactgat    480

aaggctgact tgcggttgat ctatctcgcg ctggcgcata tgatcaaatt tcggggacac    540

ttcctcatcg agggggacct gaacccagac aacagcgatg tcgacaaact ctttatccaa    600

ctggttcaga cttacaatca gcttttcgaa gagaacccga tcaacgcatc cggagttgac    660

gccaaagcaa tcctgagcgc taggctgtcc aaatcccggc ggctcgaaaa cctcatcgca    720

cagctccctg gggagaagaa gaacggcctg tttggtaatc ttatcgccct gtcactcggg    780

ctgaccccca actttaaatc taacttcgac ctggccgaag atgccaagct tcaactgagc    840

aaagacacct acgatgatga tctcgacaat ctgctggccc agatcggcga ccagtacgca    900

gaccttttt tggcggcaaa gaacctgtca gacgccattc tgctgagtga tattctgcga    960

gtgaacacgg agatcaccaa agctccgctg agcgctagta tgatcaagcg ctatgatgag   1020

caccaccaag acttgacttt gctgaaggcc cttgtcagac agcaactgcc tgagaagtac   1080

aaggaaattt tcttcgatca gtctaaaaat ggctacgccg gatacattga cggcggagca   1140

agccaggagg aattttacaa atttattaag cccatcttgg aaaaaatgga cggcaccgag   1200

gagctgctgg taaagcttaa cagagaagat ctgttgcgca aacagcgcac tttcgacaat   1260

ggaagcatcc cccaccagat tcacctgggc gaactgcacg ctatcctcag gcggcaagag   1320

gatttctacc cctttttgaa agataacagg gaaaagattg agaaaatcct cacatttcgg   1380

ataccctact atgtaggccc cctcgcccgg ggaaattcca gattcgcgtg gatgactcgc   1440

aaatcagaag agaccatcac tccctggaac ttcgaggaag tcgtggataa gggggcctct   1500
```

-continued

```
gcccagtcct tcatcgaaag gatgactaac tttgataaaa atctgcctaa cgaaaaggtg   1560
cttcctaaac actctctgct gtacgagtac ttcacagttt ataacgagct caccaaggtc   1620
aaatacgtca cagaagggat gagaaagcca gcattcctgt ctggagagca gaagaaagct   1680
atcgtggacc tcctcttcaa gacgaaccgg aaagttaccg tgaaacagct caaagaagac   1740
tatttcaaaa agattgaatg tttcgactct gttgaaatca gcggagtgga ggatcgcttc   1800
aacgcatccc tgggaacgta tcacgatctc ctgaaaatca ttaaagacaa ggacttcctg   1860
gacaatgagg agaacgagga cattcttgag gacattgtcc tcacccttac gttgtttgaa   1920
gatagggaga tgattgaaga acgcttgaaa acttacgctc atctcttcga cgacaaagtc   1980
atgaaacagc tcaagaggcg ccgatataca ggatgggggc ggctgtcaag aaaactgatc   2040
aatgggatcc gagacaagca gagtggaaag acaatcctgg attttcttaa gtccgatgga   2100
tttgccaacc ggaacttcat gcagttgatc catgatgact ctctcacctt taaggaggac   2160
atccagaaag cacaagtttc tggccagggg gacagtcttc acgagcacat cgctaatctt   2220
gcaggtagcc cagctatcaa aaagggaata ctgcagaccg ttaaggtcgt ggatgaactc   2280
gtcaaagtaa tgggaaggca taagcccgag aatatcgtta tcgagatggc ccgagagaac   2340
caaactaccc agaagggaca gaagaacagt agggaaagga tgaagaggat tgaagagggt   2400
ataaaagaac tggggtccca aatccttaag gaacacccag ttgaaaacac ccagcttcag   2460
aatgagaagc tctacctgta ctacctgcag aacggcaggg acatgtacgt ggatcaggaa   2520
ctggacatca atcggctctc cgactacgac gtggctgcta tcgtgcccca gtcttttctc   2580
aaagatgatt ctattgataa taaagtgttg acaagatccg ataaagctag agggaagagt   2640
gataacgtcc cctcagaaga agttgtcaag aaaatgaaaa attattggcg gcagctgctg   2700
aacgccaaac tgatcacaca acggaagttc gataatctga ctaaggctga acgaggtggc   2760
ctgtctgagt tggataaagc cggcttcatc aaaaggcagc ttgttgagac acgccagatc   2820
accaagcacg tggcccaaat tctcgattca cgcatgaaca ccaagtacga tgaaaatgac   2880
aaactgattc gagaggtgaa agttattact ctgaagtcta agctggtctc agatttcaga   2940
aaggactttc agttttataa ggtgagagag atcaacaatt accaccatgc gcatgatgcc   3000
tacctgaatg cagtggtagg cactgcactt atcaaaaaat atcccaagct tgaatctgaa   3060
tttgtttacg gagactataa agtgtacgat gttaggaaaa tgatcgcaaa gtctgagcag   3120
gaaataggca aggccaccgc taagtacttc ttttacagca atattatgaa tttttttcaag   3180
accgagatta cactggccaa tggagagatt cggaagcgac cacttatcga aacaaacgga   3240
gaaacaggag aaatcgtgtg ggacaagggt agggatttcg cgacagtccg gaaggtcctg   3300
tccatgccgc aggtgaacat cgttaaaaag accgaagtac agaccggagg cttctccaag   3360
gaaagtatcc tcccgaaaag gaacagcgac aagctgatcg cacgcaaaaa agattgggac   3420
cccaagaaat acggcggatt cgattctcct acagtcgctt acagtgtact ggttgtggcc   3480
aaagtggaga aagggaagtc taaaaaactc aaaagcgtca aggaactgct gggcatcaca   3540
atcatggagc gatcaagctt cgaaaaaaac cccatcgact ttctcgaggc gaaaggatat   3600
aaagaggtca aaaaagacct catcattaag cttcccaagt actctctctt tgagcttgaa   3660
aacggccgga aacgaatgct cgctagtgcg ggcgagctgc agaaaggtaa cgagctggca   3720
ctgccctcta aatacgttaa tttcttgtat ctggccagcc actatgaaaa gctcaaaggg   3780
```

-continued

tctcccgaag ataatgagca gaagcagctg ttcgtggaac aacacaaaca ctaccttgat 3840

gagatcatcg agcaaataag cgaattctcc aaaagagtga tcctcgccga cgctaacctc 3900

gataaggtgc tttctgctta caataagcac agggataagc ccatcaggga gcaggcagaa 3960

aacattatcc acttgtttac tctgaccaac ttgggcgcgc ctgcagcctt caagtacttc 4020

gacaccacca tagacagaaa gcggtacacc tctacaaagg aggtcctgga cgccacactg 4080

attcatcagt caattacggg gctctatgaa acaagaatcg acctctctca gctcggtgga 4140

gacagcaggg ctgaccccaa gaagaagagg aaggtggagg ccagcggttc cggacgggct 4200

gacgcattgg acgattttga tctggatatg ctgggaagtg acgccctcga tgattttgac 4260

cttgacatgc ttggttcgga tgcccttgat gactttgacc tcgacatgct cggcagtgac 4320

gcccttgatg atttcgacct ggacatgctg attaactcta gatga 4365

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

<400> SEQUENCE: 4

gccaccatgg acaagaagta ctccattggg ctcgctatcg gcacaaacag cgtcggctgg 60

gccgtcatta cggacgagta caaggtgccg agcaaaaaat tcaaagttct gggcaatacc 120

gatcgccaca gcataaagaa gaacctcatt ggcgccctcc tgttcgactc cggggagacg 180

gccgaagcca cgcggctcaa aagaacagca cggcgcagat atacccgcag aaagaatcgg 240

atctgctacc tgcaggagat ctttagtaat gagatggcta aggtggatga ctctttcttc 300

cataggctgg aggagtcctt tttggtggag gaggataaaa agcacgagcg ccacccaatc 360

tttggcaata tcgtggacga ggtggcgtac catgaaaagt acccaaccat atatcatctg 420

aggaagaagc ttgtagacag tactgataag gctgacttgc ggttgatcta tctcgcgctg 480

gcgcatatga tcaaatttcg gggacacttc ctcatcgagg gggacctgaa cccagacaac 540

agcgatgtcg acaaactctt tatccaactg gttcagactt acaatcagct tttcgaagag 600

aacccgatca acgcatccgg agttgacgcc aaagcaatcc tgagcgctag gctgtccaaa 660

tcccggcggc tcgaaaacct catcgcacag ctccctgggg agaagaagaa cggcctgttt 720

ggtaatctta tcgccctgtc actcgggctg acccccaact ttaaatctaa cttcgacctg 780

gccgaagatg ccaagcttca actgagcaaa gacacctacg atgatgatct cgacaatctg 840

ctggcccaga tcggcgacca gtacgcagac ctttttttgg cggcaaagaa cctgtcagac 900

gccattctgc tgagtgatat tctgcgagtg aacacggaga tcaccaaagc tccgctgagc 960

gctagtatga tcaagcgcta tgatgagcac caccaagact tgactttgct gaaggccctt 1020

gtcagacagc aactgcctga gaagtacaag gaaattttct tcgatcagtc taaaaatggc 1080

tacgccggat acattgacgg cggagcaagc caggaggaat tttacaaatt tattaagccc 1140

atcttggaaa aaatggacgg caccgaggag ctgctggtaa agcttaacag agaagatctg 1200

ttgcgcaaac agcgcacttt cgacaatgga agcatccccc accagattca cctgggcgaa 1260

ctgcacgcta tcctcaggcg gcaagaggat ttctacccct tttgaaaga taacagggaa 1320

aagattgaga aaatcctcac atttcggata ccctactatg taggcccccт cgcccgggga 1380

-continued

```
aattccagat tcgcgtggat gactcgcaaa tcagaagaga ccatcactcc ctggaacttc   1440
gaggaagtcg tggataaggg ggcctctgcc cagtccttca tcgaaaggat gactaacttt   1500
gataaaaatc tgcctaacga aaggtgcttt cctaaacact ctctgctgta cgagtacttc   1560
acagtttata acgagctcac caaggtcaaa tacgtcacag aagggatgag aaagccagca   1620
ttcctgtctg gagagcagaa gaaagctatc gtggacctcc tcttcaagac gaaccggaaa   1680
gttaccgtga aacagctcaa agaagactat ttcaaaaaga ttgaatgttt cgactctgtt   1740
gaaatcagcg gagtggagga tcgcttcaac gcatccctgg gaacgtatca cgatctcctg   1800
aaaatcatta aagacaagga cttcctggac aatgaggaga acgaggacat tcttgaggac   1860
attgtcctca cccttacgtt gtttgaagat agggagatga ttgaagaacg cttgaaaact   1920
tacgctcatc tcttcgacga caaagtcatg aaacagctca agaggcgccg atatacagga   1980
tgggggcggc tgtcaagaaa actgatcaat gggatccgag acaagcagag tggaaagaca   2040
atcctggatt ttcttaagtc cgatggattt gccaaccgga acttcatgca gttgatccat   2100
gatgactctc tcacctttaa ggaggacatc cagaaagcac aagtttctgg ccagggggac   2160
agtcttcacg agcacatcgc taatcttgca ggtagcccag ctatcaaaaa gggaatactg   2220
cagaccgtta aggtcgtgga tgaactcgtc aaagtaatgg gaaggcataa gcccgagaat   2280
atcgttatcg agatggcccg agagaaccaa actacccaga agggacagaa gaacagtagg   2340
gaaaggatga agaggattga agagggtata aaagaactgg ggtcccaaat ccttaaggaa   2400
cacccagttg aaaacaccca gcttcagaat gagaagctct acctgtacta cctgcagaac   2460
ggcagggaca tgtacgtgga tcaggaactg gacatcaatc ggctctccga ctacgacgtg   2520
gctgctatcg tgccccagtc ttttctcaaa gatgattcta ttgataataa agtgttgaca   2580
agatccgata aagctagagg gaagagtgat aacgtcccct cagaagaagt tgtcaagaaa   2640
atgaaaaatt attggcggca gctgctgaac gccaaactga tcacacaacg gaagttcgat   2700
aatctgacta aggctgaacg aggtggcctg tctgagttgg ataaagccgg cttcatcaaa   2760
aggcagcttg ttgagacacg ccagatcacc aagcacgtgg cccaaattct cgattcacgc   2820
atgaacacca agtacgatga aaatgacaaa ctgattcgag aggtgaaagt tattactctg   2880
aagtctaagc tggtctcaga tttcagaaag gactttcagt tttataaggt gagagagatc   2940
aacaattacc accatgcgca tgatgcctac ctgaatgcag tggtaggcac tgcacttatc   3000
aaaaaatatc ccaagcttga atctgaattt gtttacggag actataaagt gtacgatgtt   3060
aggaaaatga tcgcaaagtc tgagcaggaa ataggcaagg ccaccgctaa gtacttcttt   3120
tacagcaata ttatgaattt tttcaagacc gagattacac tggccaatgg agagattcgg   3180
aagcgaccac ttatcgaaac aaacggagaa acaggagaaa tcgtgtggga caagggtagg   3240
gatttcgcga cagtccggaa ggtcctgtcc atgccgcagg tgaacatcgt taaaaagacc   3300
gaagtacaga ccggaggctt ctccaaggaa agtatcctcc cgaaaaggaa cagcgacaag   3360
ctgatcgcac gcaaaaaaga ttgggacccc aagaaatacg gcggattcga ttctcctaca   3420
gtcgcttaca gtgtactggt tgtggccaaa gtggagaaag ggaagtctaa aaaactcaaa   3480
agcgtcaagg aactgctggg catcacaatc atggagcgat caagcttcga aaaaaacccc   3540
atcgactttc tcgaggcgaa aggatataaa gaggtcaaaa aagacctcat cattaagctt   3600
cccaagtact ctctctttga gcttgaaaac ggccggaaac gaatgctcgc tagtgcgggc   3660
```

-continued

```
gagctgcaga aaggtaacga gctggcactg ccctctaaat acgttaattt cttgtatctg   3720

gccagccact atgaaaagct caaagggtct cccgaagata atgagcagaa gcagctgttc   3780

gtggaacaac acaaacacta ccttgatgag atcatcgagc aaataagcga attctccaaa   3840

agagtgatcc tcgccgacgc taacctcgat aaggtgcttt ctgcttacaa taagcacagg   3900

gataagccca tcagggagca ggcagaaaac attatccact tgtttactct gaccaacttg   3960

ggcgcgcctg cagccttcaa gtacttcgac accaccatag acagaaagcg gtacacctct   4020

acaaaggagg tcctggacgc cacactgatt catcagtcaa ttacggggct ctatgaaaca   4080

agaatcgacc tctctcagct cggtggagac agcagggctg accccaagaa gaagaggaag   4140

gtggaggcca gcggttccgg acgggctgac gcattggacg attttgatct ggatatgctg   4200

ggaagtgacg ccctcgatga ttttgacctt gacatgcttg gttcggatgc ccttgatgac   4260

tttgacctcg acatgctcgg cagtgacgcc cttgatgatt tcgacctgga catgctgatt   4320

aactctagag cggccgcaga tccaaaaaag aagagaaagg tagatccaaa aaagaagaga   4380

aaggtagatc caaaaaagaa gagaaaggta gatacggccg catag                   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

```
ccaccatggg acctaagaaa aagaggaagg tggcggccgc ttctagaatg gcttctaact     60

ttactcagtt cgttctcgtc gacaatggcg gaactggcga cgtgactgtc gccccaagca    120

acttcgctaa cgggatcgct gaatggatca gctctaactc gcgttcacag gcttacaaag    180

taacctgtag cgttcgtcag agctctgcgc agaatcgcaa atacaccatc aaagtcgagg    240

tgcctaaagg cgcctggcgt tcgtacttaa atatggaact aaccattcca attttcgcca    300

cgaattccga ctgcgagctt attgttaagg caatgcaagg tctcctaaaa gatggaaacc    360

cgattccctc agcaatcgca gcaaactccg gcatctacga ggccagcggt tccggacggg    420

ctgacgcatt ggacgatttt gatctggata tgctgggaag tgacgccctc gatgattttg    480

accttgacat gcttggttcg gatgcccttg atgactttga cctcgacatg ctcggcagtg    540

acgcccttga tgatttcgac ctggacatgc tgattaactc tagatga                  587
```

<210> SEQ ID NO 6
<211> LENGTH: 681
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: MS2-activator construct

<400> SEQUENCE: 6

```
gccaccatgg gacctaagaa aaagaggaag gtggcggccg cttctagaat ggcttctaac     60

tttactcagt tcgttctcgt cgacaatggc ggaactggcg acgtgactgt cgccccaagc    120

aacttcgcta acgggatcgc tgaatggatc agctctaact cgcgttcaca ggcttacaaa    180

gtaacctgta gcgttcgtca gagctctgcg cagaatcgca aatacaccat caaagtcgag    240

gtgcctaaag gcgcctggcg ttcgtactta aatatggaac taaccattcc aattttcgcc    300
```

-continued

```
acgaattccg actgcgagct tattgttaag gcaatgcaag gtctcctaaa agatggaaac    360

ccgattccct cagcaatcgc agcaaactcc ggcatctacg aggccagcgg ttccggacgg    420

gctgacgcat tggacgattt tgatctggat atgctgggaa gtgacgccct cgatgatttt    480

gaccttgaca tgcttggttc ggatgccctt gatgactttg acctcgacat gctcggcagt    540

gacgcccttg atgatttcga cctggacatg ctgattaact ctagagcggc cgcagatcca    600

aaaaagaaga gaaaggtaga tccaaaaaag aagagaaagg tagatccaaa aaagaagaga    660

aaggtagata cggccgcata g                                               681
```

```
&lt;210&gt; SEQ ID NO 7
&lt;211&gt; LENGTH: 557
&lt;212&gt; TYPE: DNA
&lt;213&gt; ORGANISM: Artificial
&lt;220&gt; FEATURE:
&lt;223&gt; OTHER INFORMATION: MS2-activator construct
&lt;220&gt; FEATURE:
&lt;221&gt; NAME/KEY: misc_feature
&lt;222&gt; LOCATION: (320)..(339)
&lt;223&gt; OTHER INFORMATION: wherein N is G, A, T or C

&lt;400&gt; SEQUENCE: 7

tgtacaaaaa agcaggcttt aaaggaacca attcagtcga ctggatccgg taccaaggtc     60

gggcaggaag agggcctatt tcccatgatt ccttcatatt tgcatatacg atacaaggct    120

gttagagaga taattagaat taatttgact gtaaacacaa agatattagt acaaaatacg    180

tgacgtagaa agtaataatt tcttgggtag tttgcagttt taaaattatg ttttaaaatg    240

gactatcata tgcttaccgt aacttgaaag tatttcgatt tcttggcttt atatatcttg    300

tggaaaggac gaaacaccgn nnnnnnnnnn nnnnnnnnng ttttagagct agaaatagca    360

agttaaaata aggctagtcc gttatcaact tgaaaaagtg gcaccgagtc ggtgctctgc    420

aggtcgactc tagaaaacat gaggatcacc catgtctgca gtattcccgg gttcattaga    480

tcctaaggta cctaattgcc tagaaaacat gaggatcacc catgtctgca ggtcgactct    540

agaaattttt tctagac                                                    557
```

```
&lt;210&gt; SEQ ID NO 8
&lt;211&gt; LENGTH: 882
&lt;212&gt; TYPE: DNA
&lt;213&gt; ORGANISM: Artificial
&lt;220&gt; FEATURE:
&lt;223&gt; OTHER INFORMATION: Activation reporter construct

&lt;400&gt; SEQUENCE: 8

tagggataac agggtaatag tgtcccctcc accccacagt ggggcgaggt aggcgtgtac     60

ggtgggaggc ctatataagc agagctcgtt tagtgaaccg tcagatcgcc tggagaattc    120

gccaccatgg actacaagga tgacgacgat aaaacttccg gtggcggact gggttccacc    180

gtgagcaagg gcgaggaggt catcaaagag ttcatgcgct tcaaggtgcg catggagggc    240

tccatgaacg gccacgagtt cgagatcgag ggcgagggcg agggccgccc ctacgagggc    300

acccagaccg ccaagctgaa ggtgaccaag ggcggccccc tgcccttcgc ctgggacatc    360

ctgtcccccc agttcatgta cggctccaag gcgtacgtga agcaccccgc cgacatcccc    420

gattacaaga agctgtcctt ccccgagggc ttcaagtggg agcgcgtgat gaacttcgag    480

gacggcggtc tggtgaccgt gacccaggac tcctccctgc aggacggcac gctgatctac    540
```

-continued

```
aaggtgaaga tgcgcggcac caacttcccc cccgacggcc ccgtaatgca gaagaagacc    600

atgggctggg aggcctccac cgagcgcctg taccccccgcg acggcgtgct gaagggcgag    660

atccaccagg ccctgaagct gaaggacggc ggccactacc tggtggagtt caagaccatc    720

tacatggcca agaagcccgt gcaactgccc ggctactact acgtggacac caagctggac    780

atcacctccc acaacgagga ctacaccatc gtggaacagt acgagcgctc cgagggccgc    840

caccacctgt tcctgtacgg catggacgag ctgtacaagt aa                        882
```

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

<400> SEQUENCE: 9

tagggataac agggtaatag tggggccact agggacagga ttggcgaggt aggcgtgtac     60

ggtgggaggc ctatataagc agagctcgtt tagtgaaccg tcagatcgcc tggagaattc    120

gccaccatgg actacaagga tgacgacgat aaaacttccg gtggcggact gggttccacc    180

gtgagcaagg gcgaggaggt catcaaagag ttcatgcgct tcaaggtgcg catggagggc    240

tccatgaacg gccacgagtt cgagatcgag ggcgagggcg agggccgccc ctacgagggc    300

acccagaccg ccaagctgaa ggtgaccaag ggcggccccc tgcccttcgc ctgggacatc    360

ctgtccccccc agttcatgta cggctccaag gcgtacgtga agcaccccgc cgacatcccc    420

gattacaaga agctgtcctt ccccgagggc ttcaagtggg agcgcgtgat gaacttcgag    480

gacggcggtc tggtgaccgt gacccaggac tcctccctgc aggacggcac gctgatctac    540

aaggtgaaga tgcgcggcac caacttcccc cccgacggcc ccgtaatgca gaagaagacc    600

atgggctggg aggcctccac cgagcgcctg taccccccgcg acggcgtgct gaagggcgag    660

atccaccagg ccctgaagct gaaggacggc ggccactacc tggtggagtt caagaccatc    720

tacatggcca agaagcccgt gcaactgccc ggctactact acgtggacac caagctggac    780

atcacctccc acaacgagga ctacaccatc gtggaacagt acgagcgctc cgagggccgc    840

caccacctgt tcctgtacgg catggacgag ctgtacaagt aa                        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

tagggataac agggtaatag tnnnnnnnnn nnnnnnnnnn nnnncgaggt aggcgtgtac     60

ggtgggaggc ctatataagc agagctcgtt tagtgaaccg tcagatcgcc tggagaattc    120

gccaccatgg actacaagga tgacgacgat aaannnnnnn nnnnnnnnnn nnnnnnnact    180
```

-continued

```
tccggtggcg gactgggttc caccgtgagc aagggcgagg aggtcatcaa agagttcatg    240

cgcttcaagg tgcgcatgga gggctccatg aacggccacg agttcgagat cgagggcgag    300

ggcgagggcc gcccctacga gggcacccag accgccaagc tgaaggtgac caagggcggc    360

cccctgccct tcgcctggga catcctgtcc ccccagttca tgtacggctc caaggcgtac    420

gtgaagcacc ccgccgacat ccccgattac aagaagctgt ccttccccga gggcttcaag    480

tgggagcgcg tgatgaactt cgaggacggc ggtctggtga ccgtgaccca ggactcctcc    540

ctgcaggacg gcacgctgat ctacaaggtg aagatgcgcg gcaccaactt cccccccgac    600

ggccccgtaa tgcagaagaa gaccatgggc tgggaggcct ccaccgagcg cctgtacccc    660

cgcgacggcg tgctgaaggg cgagatccac caggccctga agctgaagga cggcggccac    720

tacctggtgg agttcaagac catctacatg gccaagaagc ccgtgcaact gcccggctac    780

tactacgtgg acaccaagct ggacatcacc tcccacaacg aggactacac catcgtggaa    840

cagtacgagc gctccgaggg ccgccaccac ctgttcctgt acggcatgga cgagctgtac    900

aagtaagaat 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

ctggcggatc actcgcggtt 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

cctcggcctc caaaagtgct 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

acgctgattc ctgcagatca ggg                                             23


<210> SEQ ID NO 14
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 14

ccaggaatac gtatccacca ggg                                             23


<210> SEQ ID NO 15
```

-continued

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

<400> SEQUENCE: 15

gccacaccca agcgatcaaa tgg 23

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

<400> SEQUENCE: 16

aaataataca ttctaaggta agg 23

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

<400> SEQUENCE: 17

gctactgggg aggctgaggc agg 23

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

<400> SEQUENCE: 18

tagcaataca gtcacattaa tgg 23

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

<400> SEQUENCE: 19

ctcatgtgat ccccccgtct cgg 23

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

<400> SEQUENCE: 20

ccgggcagag agtgaacgcg cgg 23

<210> SEQ ID NO 21
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 21

ttccttccct ctcccgtgct tgg 23

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

<400> SEQUENCE: 22

tctctgcaaa gcccctggag agg 23

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

<400> SEQUENCE: 23

aatgcagttg ccgagtgcag tgg 23

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

<400> SEQUENCE: 24

cctcagcctc ctaaagtgct ggg 23

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

<400> SEQUENCE: 25

gagtccaaat cctctttact agg 23

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

<400> SEQUENCE: 26

gagtgtctgg atttgggata agg 23

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

<400> SEQUENCE: 27

-continued

cagcacctca tctcccagtg agg 23

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

<400> SEQUENCE: 28

tctaaaaccc agggaatcat ggg 23

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

<400> SEQUENCE: 29

cacaaggcag ccagggatcc agg 23

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

<400> SEQUENCE: 30

gatggcaagc tgagaaacac tgg 23

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

<400> SEQUENCE: 31

tgaaatgcac gcatacaatt agg 23

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

<400> SEQUENCE: 32

ccagtccaga cctggccttc tgg 23

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

<400> SEQUENCE: 33

cccagaaaaa cagaccctga agg 23

<210> SEQ ID NO 34

-continued

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

<400> SEQUENCE: 34

aagggttgag 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

atgtctgagt tttggttgag 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

ggtcccttga aggggaagta 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

tggcagtcta ctcttgaaga 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

ggcacagtgc cagaggtctg 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

taaaaataaa aaaactaaca ggg 23

<210> SEQ ID NO 40
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 40

tctgtggggg acctgcactg agg 23

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

<400> SEQUENCE: 41

ggccagaggt caaggctagt ggg 23

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

<400> SEQUENCE: 42

cacgaccgaa acccttctta cgg 23

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

<400> SEQUENCE: 43

gttgaatgaa gacagtctag tgg 23

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

<400> SEQUENCE: 44

taagaacaga gcaagttacg tgg 23

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

<400> SEQUENCE: 45

tgtaaggtaa 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

-continued

tgacacacca actcctgcac 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

tttacccact tccttcgaaa 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

gtggctggca ggctggctct 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

ctcccccggc ctcccccgcg cgg 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

caaaacccgg cagcgaggct 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

aggagccgcc gcgcgctgat 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 23

<210> SEQ ID NO 53

-continued

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

<400> SEQUENCE: 53

gaagaagcta aagagccaga ggg 23

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

<400> SEQUENCE: 54

atgagaattt caataacctc agg 23

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

<400> SEQUENCE: 55

tcccgctctg ttgcccaggc tgg 23

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

<400> SEQUENCE: 56

cagacaccca ccaccatgcg tgg 23

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

<400> SEQUENCE: 57

tcccaattta ctgggattac agg 23

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

<400> SEQUENCE: 58

tgatttaaaa gttggaaacg tgg 23

<210> SEQ ID NO 59
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Target probe

<400> SEQUENCE: 59

tctagttccc cacctagtct ggg 23

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

<400> SEQUENCE: 60

gattaactga gaattcacaa ggg 23

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

<400> SEQUENCE: 61

cgccaggagg ggtgggtcta agg 23

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

<400> SEQUENCE: 62

gtcccctcca ccccacagtg ggg 23

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

<400> SEQUENCE: 63

ggggccacta gggacaggat tgg 23

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

<400> SEQUENCE: 64

taatactttt atctgtcccc tccaccccac agtggggcca ctagggacag gattggtgac 60

agaaaagccc c 71

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

-continued

<400> SEQUENCE: 65

ggggccacta gggacaggat 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

guuuuagagc uagaaauagc aaguuaaaau aaggcuagcu uguuaucaac uugaaaaagu 60

ggcaccgagu cggugcuuuu 80

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

<400> SEQUENCE: 67

gtcccctcca ccccacagtg cag 23

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

<400> SEQUENCE: 68

gtcccctcca ccccacagtg caa 23

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

<400> SEQUENCE: 69

gtcccctcca ccccacagtg 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

tgtcccctcc accccacagt 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

-continued

```
tgtccccccc accccacagt 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

aaaaccctcc accccacagt 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

tgtcccctcc tttttcagt 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

caccggggtg gtgcccatcc 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 ctggtcgagc 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

cccatcctgg tcgagctgga 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 tcagcgtgtc cgg                                             23


<210> SEQ ID NO 78
```

-continued

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

<400> SEQUENCE: 78

cgcaaataag agctcaccta cgg 23

<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

ctgaagttca tctgcaccac cgg 23

<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

ccggcaagct gcccgtgccc tgg 23

<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

gaccaggatg ggcaccaccc cgg 23

<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

gccgtccagc tcgaccagga tgg 23

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

<400> SEQUENCE: 83

ggccggacac gctgaacttg tgg 23

<210> SEQ ID NO 84
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 84

taacagggta atgtcgaggc cgg 23

<210> SEQ ID NO 85
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 85

aggtgagctc ttatttgcgt agg 23

<210> SEQ ID NO 86
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 86

cttcagggtc agcttgccgt agg 23

<210> SEQ ID NO 87
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 87

gggcacgggc agcttgccgg tgg 23

<210> SEQ ID NO 88
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 88

gagatgatcg ccccttcttc tgg 23

<210> SEQ ID NO 89
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 89

gagatgatcg ccccttcttc 20

<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

-continued

gtgatgaccg gccgttcttc 20

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

<400> SEQUENCE: 91

gtcccctcca ccccacagtg ggg 23

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

<400> SEQUENCE: 92

gagatgatcg cccgttcttc tgg 23

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

<400> SEQUENCE: 93

guccccucca ccccacagug 20

<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 ccccacaguc 20

<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 ccccacagag 20

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

<400> SEQUENCE: 96

guccccucca ccccacacug 20

<210> SEQ ID NO 97

-continued

<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 97

guccccucca ccccacugug 20

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

<400> SEQUENCE: 98

guccccucca ccccagagug 20

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

<400> SEQUENCE: 99

guccccucca ccccucagug 20

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

<400> SEQUENCE: 100

guccccucca cccgacagug 20

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

<400> SEQUENCE: 101

guccccucca ccgcacagug 20

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

<400> SEQUENCE: 102

guccccucca cgccacagug 20

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

-continued

<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 103

guccccucca gcccacagug 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

guccccuccu ccccacagug 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

guccccucga ccccacagug 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

guccccucca ccccacagac 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

guccccucca ccccacucug 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

guccccucca ccccugagug 20

<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

-continued

gucccucca ccggacagug 20

<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

gucccucca ggccacagug 20

<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

guccccucgu ccccacagug 20

<210> SEQ ID NO 112
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 112

ggggccacta gggacaggat ggg 23

<210> SEQ ID NO 113
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 113

gagaugaucg ccccuucuuc 20

<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

gagaugaucg ccccuucuug 20

<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

gagaugaucg ccccuucuac 20

<210> SEQ ID NO 116

-continued

<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 116

gagaugaucg ccccuucauc 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

gagaugaucg ccccuuguuc 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

gagaugaucg ccccuacuuc 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

gagaugaucg ccccaucuuc 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

gagaugaucg cccguucuuc 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

gagaugaucg ccgcuucuuc 20

<210> SEQ ID NO 122
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:

-continued

<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 122

gagaugaucg cgccuucuuc 20

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

<400> SEQUENCE: 123

gagaugaucg gcccuucuuc 20

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

<400> SEQUENCE: 124

gagaugaucc ccccuucuuc 20

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

<400> SEQUENCE: 125

gagaugaugg ccccuucuuc 20

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

<400> SEQUENCE: 126

gagaugaucg ccccuucuag 20

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

<400> SEQUENCE: 127

gagaugaucg ccccuugauc 20

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

<400> SEQUENCE: 128

-continued

gagaugaucg ccccaacuuc 20

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

<400> SEQUENCE: 129

gagaugaucg ccgguucuuc 20

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

<400> SEQUENCE: 130

gagaugaucg ggccuucuuc 20

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

<400> SEQUENCE: 131

gagaugaugc ccccuucuuc 20

<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 ccccttcttc tgg 23

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

<400> SEQUENCE: 133

ggggccacua gggacaggau 20

<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 ggacaggau 19

<210> SEQ ID NO 135

-continued

<211> LENGTH: 18
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 135

ggccacuagg gacaggau 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

gccacuaggg acaggau 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

gagaugaucg ccccuucuuc 20

<210> SEQ ID NO 138
<211> LENGTH: 18
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA target sequence

<400> SEQUENCE: 138

gaugaucgcc ccuucuuc 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

gaucgccccu ucuuc 15

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

<400> SEQUENCE: 140

gccccuucuu c 11

<210> SEQ ID NO 141
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 141

gtcccctcca ccccacagtg c 21

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

<400> SEQUENCE: 142

tgtcnnnnnn accc 14

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

<400> SEQUENCE: 143

tgtcaaaaaa accc 14

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

<400> SEQUENCE: 144

tgtcgggggg accc 14

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

<400> SEQUENCE: 145

tgtcaaaaaa accc 14

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

<400> SEQUENCE: 146

tgtcgggggg accc 14

<210> SEQ ID NO 147
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 147

tgtccccccc accc 14

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

<400> SEQUENCE: 148

tgtctttttt accc 14

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

<400> SEQUENCE: 149

tgtccccccc accc 14

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

<400> SEQUENCE: 150

tgtctttttt accc 14

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

<400> SEQUENCE: 151

ggatcctgtg tccccgagct ggg 23

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

<400> SEQUENCE: 152

gttaatgtgg ctctggttct ggg 23

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

<400> SEQUENCE: 153

-continued

ggggccacta gggacaggat tgg 23

<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

cttcctagtc tcctgatatt ggg 23

<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

tggtcccagc tcggggacac agg 23

<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

agaaccagag ccacattaac cgg 23

<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

gtcaccaatc ctgtccctag tgg 23

<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

agacccaata tcaggagact agg 23

<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

gggatcctgt gtccccgagc tgggaccacc ttatattccc agggccggtt aatgtggctc 60

tggttctggg tactt 75

-continued

<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

gggatcctgt gtccccgagc tgggaccacc ttatattccc agggccggtt aatgtggttc 60

tgggtactt 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

gggatcctgt gtccccgagc tgggaccacc ttatattccc agggcagggc cggttggacc 60

accttatatt cccagggcag ggccggttaa tgtggctctg gttctgggta ctt 113

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

<400> SEQUENCE: 162

gggatcctgt gtccccgtct ggttctgggt actt 34

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

<400> SEQUENCE: 163

gggatcctgt gtccccgagc tgggaccacc ttatattctg 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

gggatcctgt 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 aatgtggctc tggttctggg tacttttatc tgtcccctcc accccacagt 60

-continued

ggggccacta gggacaggat tggtgacaga aaa 93

<210> SEQ ID NO 166
<211> LENGTH: 83
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 166

agggccggtt aatgaatgtg gctctggttc tgggtacttt tatctgtccc ctccacccca 60

cagtggggcc actagacaga aaa 83

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

<400> SEQUENCE: 167

agggccggtt aatgtggctc tggttctggg tacttttatc tgtcccccag tggggccact 60

gattggtgac agaaaa 76

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

<400> SEQUENCE: 168

agggccggtt caggattggt gacagaaaa 29

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

<400> SEQUENCE: 169

agggccggtt aatgtggcga 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 aatgtggctc tggttctggg tacttttatc tgtccccgat 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

-continued

<400> SEQUENCE: 171

agggccggtt aatgtggctc tggttctggg tacttttatc tgtcccctcc accccacagt 60

ggggacagga ttggtgacag aaaa 84

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

<400> SEQUENCE: 172

agggccggtt aatgtggtga 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 aatgtggctc tggttctggg tacttttatc tgtcccctcc accccagggg 60

acagtctgtc ccctccaccc cagggacagg attggtgaca gaaaa 105

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

<400> SEQUENCE: 174

agggccggtt aatgtggctc tggttctggg tacttttatc tgtcccctcc accactaggg 60

acaggattgg tgacagaaaa 80

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

<400> SEQUENCE: 175

cccacagtgg ggccactagg gacaggattg gtgacagaaa 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

cccacagtgg ggccactacc cc 22

<210> SEQ ID NO 177
<211> LENGTH: 96
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 177

cccacagtgg ggccactagt agaaaagccc catccttagg cctcccccat ccttaggcct 60

cctccttcct agtctcctga tattgggtct aacccc 96

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

<400> SEQUENCE: 178

cccacagtgg ggccactagg gacaggattg gtgacagaaa agccccatcc ttaggcctcc 60

tccttcctag tctcctgata ttgggtctaa cccc 94

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

<400> SEQUENCE: 179

cccacagtgg ggccaccctt aggcctcctc cttcctagtc tcctgatatt gggtctaacc 60

cc 62

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

<400> SEQUENCE: 180

cccacagtgg 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

cccacagtgg ggccactagg gacaggattg gtgacaaaaa agccccatcc ttacgcctcc 60

tccttcctag tctcctgata ttgggtctaa cccc 94

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

<400> SEQUENCE: 182

cccacagtgg ggccactagg gacaggcctc ctccttccta gtctcctgat attgggtcta 60

acccc 65

-continued

```
<210> SEQ ID NO 183
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 183

cccacagtgg ggccactagg gacaggggga caggattggt gacagaaaag ccccatcctt     60

aggcctcctc cttcctagtc tcctgatatt gggtctaacc cc                       102


<210> SEQ ID NO 184
<211> LENGTH: 76
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide sequence

<400> SEQUENCE: 184

cccacaggat tggtgacaga aaagccccat ccttaggcct cctccttcct agtctcctga     60

tattgggtct aacccc                                                     76


<210> SEQ ID NO 185
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide

<400> SEQUENCE: 185

ggggccacta gggacaggat ggg                                             23


<210> SEQ ID NO 186
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Target oligonucleotide

<400> SEQUENCE: 186

gagatgatcg ccccttcttc tgg                                             23
```

1. A method of modulating expression of a target nucleic acid in a cell comprising

providing to the cell a guide RNA complementary to the target nucleic acid sequence including a transcriptional activator or repressor domain as a fusion protein for modulating target nucleic acid expression in vivo,

providing to the cell a nuclease null Cas9 protein that interacts with the guide RNA and binds to the target nucleic acid sequence in a site specific manner,

wherein the guide RNA including the transcriptional activator or repressor domain as a fusion protein and the Cas9 protein co-localize to the target nucleic acid sequence and wherein the transcriptional activator or repressor domain modulates expression of the target nucleic acid.

2. The method of claim **2**

wherein the guide RNA including the transcriptional activator or repressor domain as a fusion protein is provided to the cell by introducing to the cell a nucleic acid encoding the guide RNA including the transcriptional activator or repressor domain as a fusion protein,

wherein the Cas9 protein is provided to the cell by introducing to the cell a nucleic acid encoding the Cas9 protein, and

wherein the cell expresses the guide RNA including the transcriptional activator or repressor domain as a fusion protein and the Cas9 protein.

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

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

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

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

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

8. The method of claim **1** wherein the guide RNA is between about 100 to about 250 nucleotides.

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

10. A method of modulating expression of a target nucleic acid in a cell comprising

providing to the cell a guide RNA complementary to viral DNA including the target nucleic acid sequence, wherein the guide RNA includes a transcriptional activator or repressor domain as a fusion protein for modulating target nucleic acid expression in vivo,

providing to the cell a nuclease null Cas9 protein that interacts with the guide RNA and binds to the target nucleic acid sequence in a site specific manner,

wherein the guide RNA including the transcriptional activator or repressor domain as a fusion protein and the Cas9 protein co-localize to the target nucleic acid sequence and wherein the transcriptional activator or repressor domain modulates expression of the target nucleic acid.

11. The method of claim **10**

wherein the guide RNA including the transcriptional activator or repressor domain as a fusion protein is provided to the cell by introducing to the cell a nucleic acid encoding the guide RNA including the transcriptional activator or repressor domain as a fusion protein,

wherein the Cas9 protein is provided to the cell by introducing to the cell a nucleic acid encoding the Cas9 protein, and

wherein the cell expresses the guide RNA including the transcriptional activator or repressor domain as a fusion protein and the Cas9 protein.

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

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

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

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

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

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

* * * * *