US 20140357530A1

(19) **United States**

(12) **Patent Application Publication**

Zhang et al.

(10) **Pub. No.: US 2014/0357530 A1**

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

(54) **FUNCTIONAL GENOMICS USING CRISPR-CAS SYSTEMS, COMPOSITIONS, METHODS, KNOCK OUT LIBRARIES AND APPLICATIONS THEREOF**

(71) Applicants: **THE BROAD INSTITUTE INC.**, CAMBRIDGE, MA (US); **MASSACHUSETTS INSTITUTE OF TECHNOLOGY**, CAMBRIDGE, MA (US)

(72) Inventors: **Feng Zhang**, Cambridge, MA (US); **Neville Espi Sanjana**, Cambridge, MA (US); **Ophir Shalem**, Boston, MA (US)

(21) Appl. No.: **14/463,253**

(22) Filed: **Aug. 19, 2014**

Related U.S. Application Data

(63) Continuation of application No. PCT/US2013/074800, filed on Dec. 12, 2013.

(60) Provisional application No. 61/736,527, filed on Dec. 12, 2012, provisional application No. 61/802,174, filed on Mar. 15, 2013.

Publication Classification

(51) **Int. Cl.**
C12N 15/10 (2006.01)

(52) **U.S. Cl.**
CPC ***C12N 15/1034*** (2013.01)
USPC **506/16**

(57) **ABSTRACT**

The present invention generally relates to compositions, methods applications and screens used in functional genomics that focus on gene function in a cell and that may use vector systems and other aspects related to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas systems and components thereof. Provided are vectors and vector systems, some of which encode one or more components of a CRISPR complex, as well as methods for the design and use of such vectors. Also provided are methods of directing CRISPR complex formation in eukaryotic cells and methods for utilizing the CRISPR-Cas system.

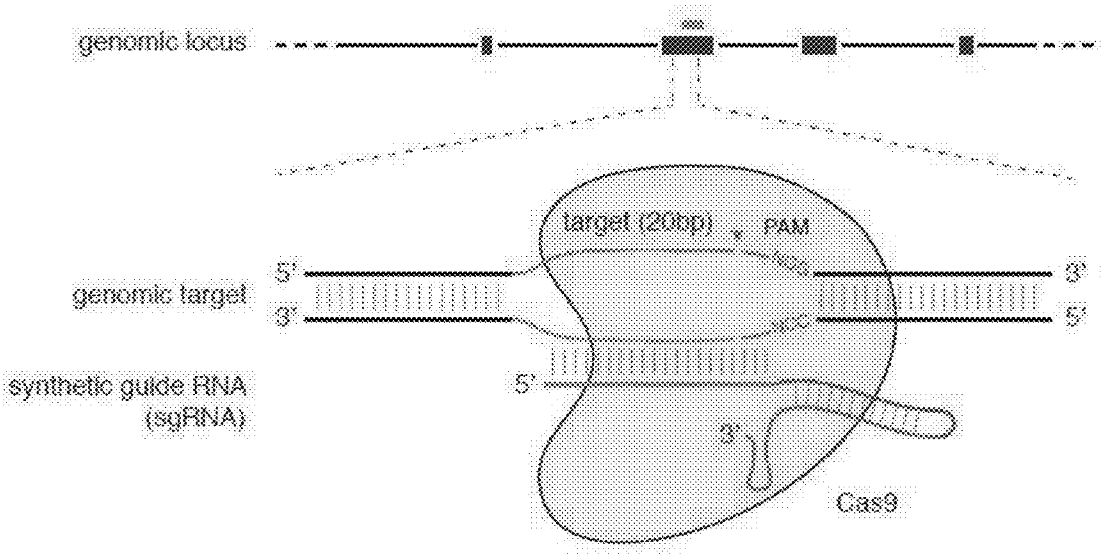

genomic locus
target (20bp)
PAM
5'
3'
genomic target
3'
5'
synthetic guide RNA
(sgRNA)
5'
3'
Cas9

FIG. 1

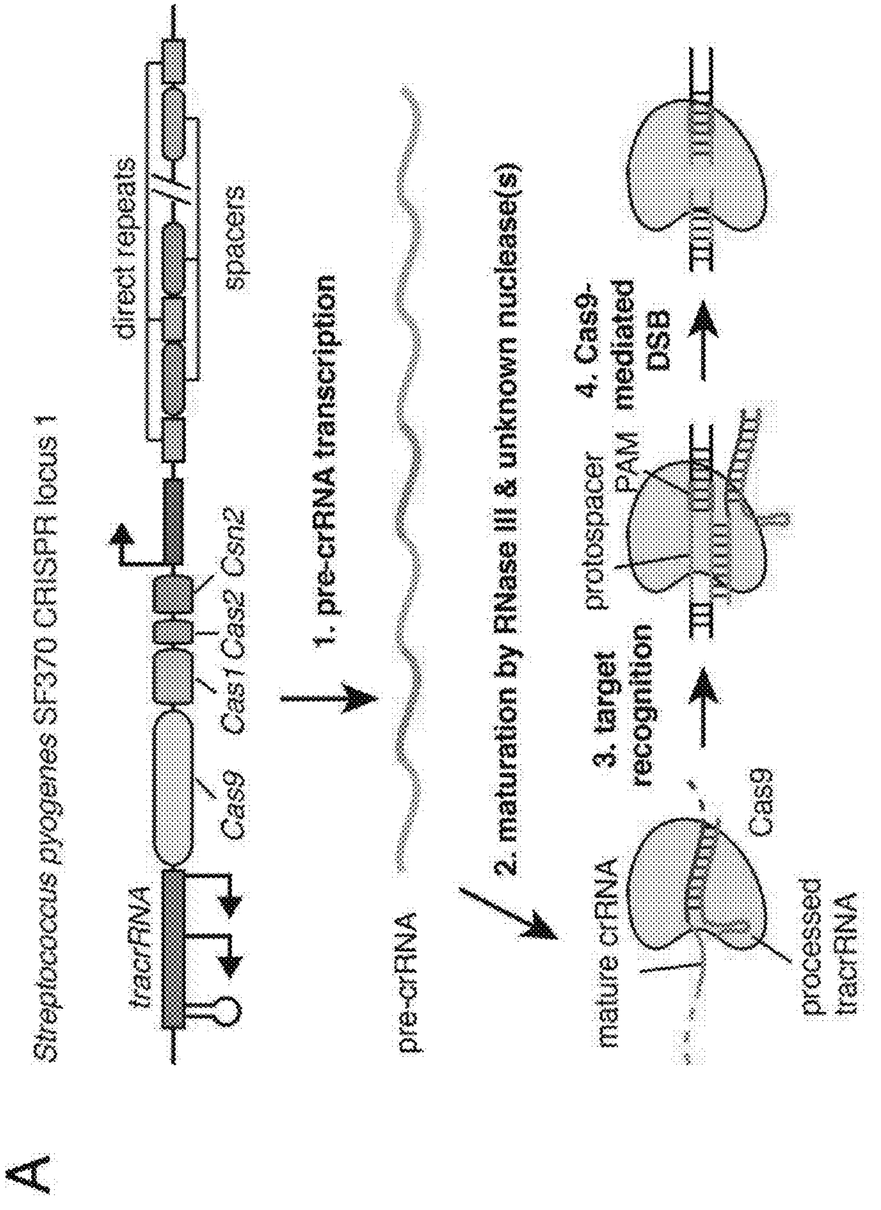

A
Streptococcus pyogenes SF370 CRISPR locus 1
tracrRNA
Cas9
Cas1 Cas2 Csn2
direct repeats
spacers
1. pre-crRNA transcription
pre-crRNA
2. maturation by RNase III & unknown nuclease(s)
mature crRNA
Cas9
processed tracrRNA
3. target recognition
protospacer
PAM
4. Cas9-mediated DSB

FIG. 2A

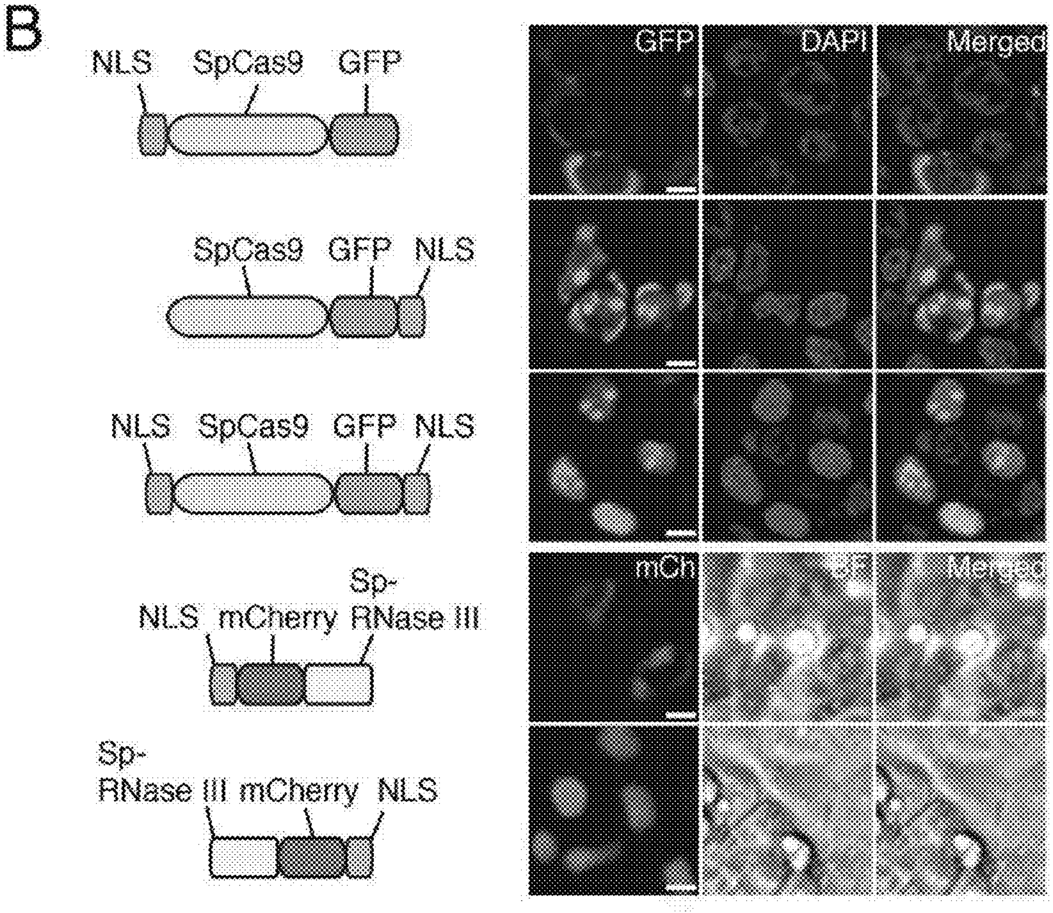

B
NLS
SpCas9
GFP
SpCas9
GFP
NLS
NLS
SpCas9
GFP
NLS
NLS
mCherry
Sp-RNase III
Sp-RNase III
mCherry
NLS
GFP
DAPI
Merged
mCh
BF
Merged

FIG. 2B

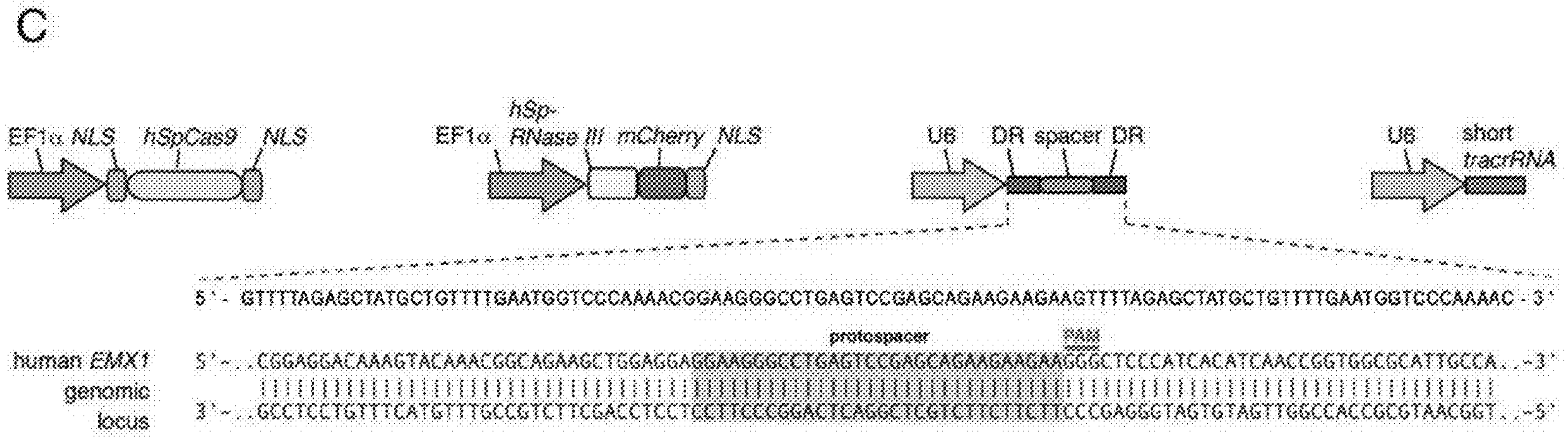

C
EF1α
NLS
hSpCas9
NLS
EF1α
hSp-RNase III
mCherry
NLS
U6
DR
spacer
DR
U6
short tracrRNA
5'- GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC -3'
protospacer
PAM
human EMX1 genomic locus
5'-..CGGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCA..-3'
3'-..GCCTCCTGTTTCATGTTTGCCGTCTTCGACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTAGTGTAGTTGGCCACCGCGTAACGGT..-5'

FIG. 2C

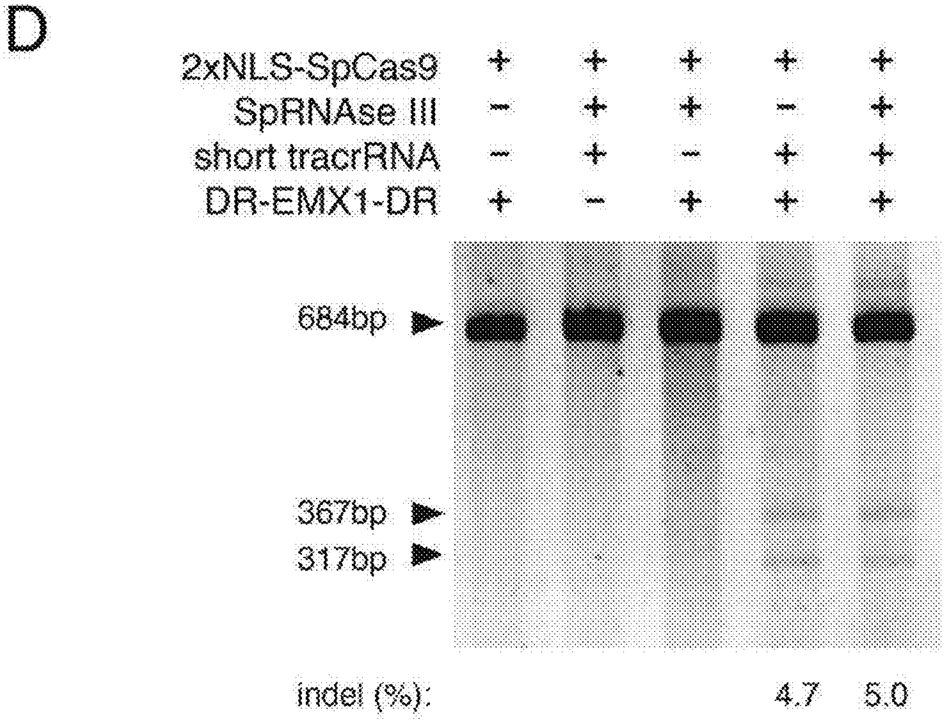

D
2xNLS-SpCas9 + + + + +
SpRNAse III − + + − +
short tracrRNA − + − + +
DR-EMX1-DR + − + + +
684bp
367bp
317bp
indel (%): 4.7 5.0

FIG. 2D

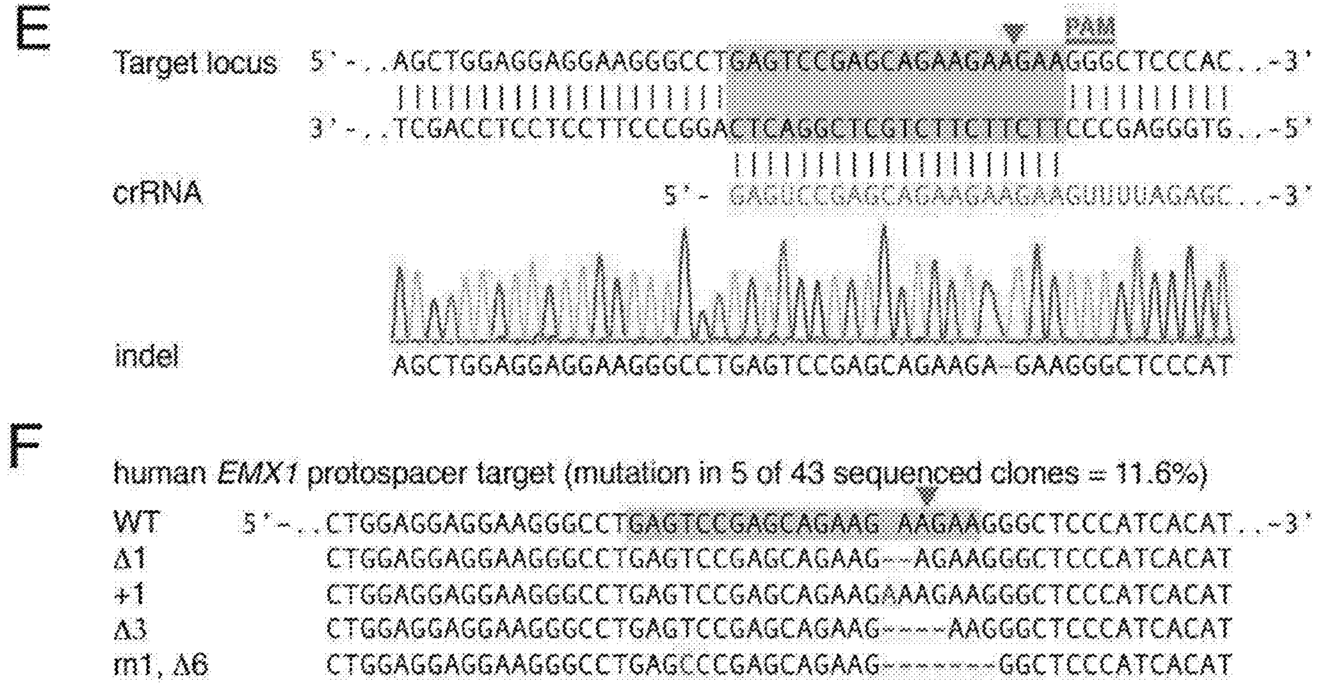

E
Target locus
5'-..AGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAC..-3'
3'-..TCGACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTG..-5'
PAM
crRNA
5'- GAGUCCGAGCAGAAGAAGAAGUUUUAGAGC..-3'
indel
AGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGA-GAAGGGCTCCCAT
F
human EMX1 protospacer target (mutation in 5 of 43 sequenced clones = 11.6%)
WT
5'-..CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG AAGAAGGGCTCCCATCACAT..-3'
Δ1
CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG--AGAAGGGCTCCCATCACAT
+1
CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAAGAAGGGCTCCCATCACAT
Δ3
CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG----AAGGGCTCCCATCACAT
m1, Δ6
CTGGAGGAGGAAGGGCCTGAGCCCGAGCAGAAG-------GGCTCCCATCACAT

FIG. 2E-F

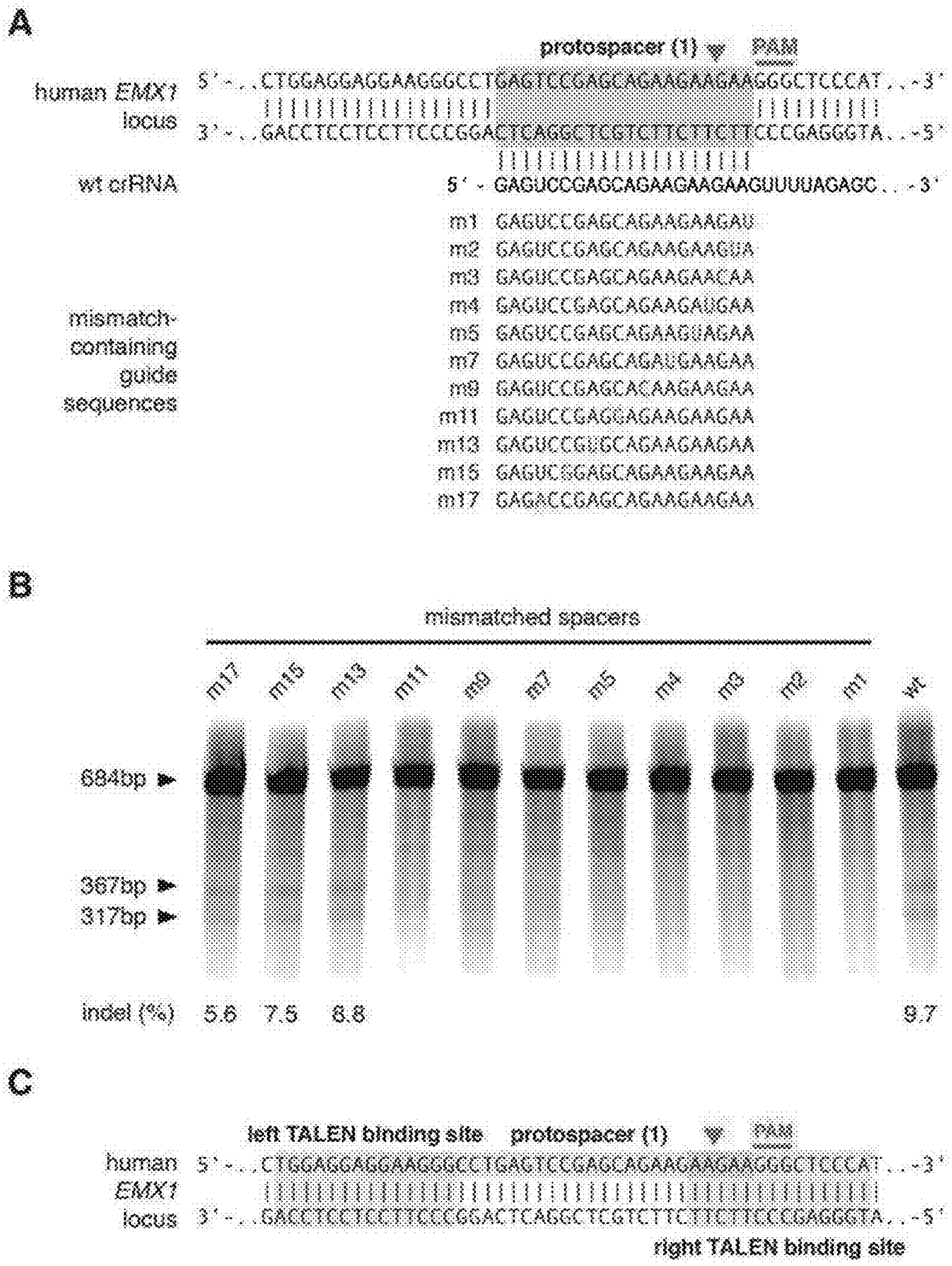

A
protospacer (1)
PAM
human EMX1 locus
5'-..CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAT..-3'
3'-..GACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTA..-5'
wt crRNA
5'-GAGUCCGAGCAGAAGAAGAAGUUUUAGAGC..-3'
mismatch-containing guide sequences
m1 GAGUCCGAGCAGAAGAAGAU
m2 GAGUCCGAGCAGAAGAAGUA
m3 GAGUCCGAGCAGAAGAACAA
m4 GAGUCCGAGCAGAAGAUGAA
m5 GAGUCCGAGCAGAAGUAGAA
m7 GAGUCCGAGCAGAUGAAGAA
m9 GAGUCCGAGCACAAGAAGAA
m11 GAGUCCGAGGAGAAGAAGAA
m13 GAGUCCGUGCAGAAGAAGAA
m15 GAGUCGGAGCAGAAGAAGAA
m17 GAGACCGAGCAGAAGAAGAA
B
mismatched spacers
m17
m15
m13
m11
m9
m7
m5
m4
m3
m2
m1
wt
684bp
367bp
317bp
indel (%)
5.6
7.5
8.8
9.7
C
left TALEN binding site
protospacer (1)
PAM
human EMX1 locus
5'-..CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAT..-3'
3'-..GACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTA..-5'
right TALEN binding site

FIG. 3A-C

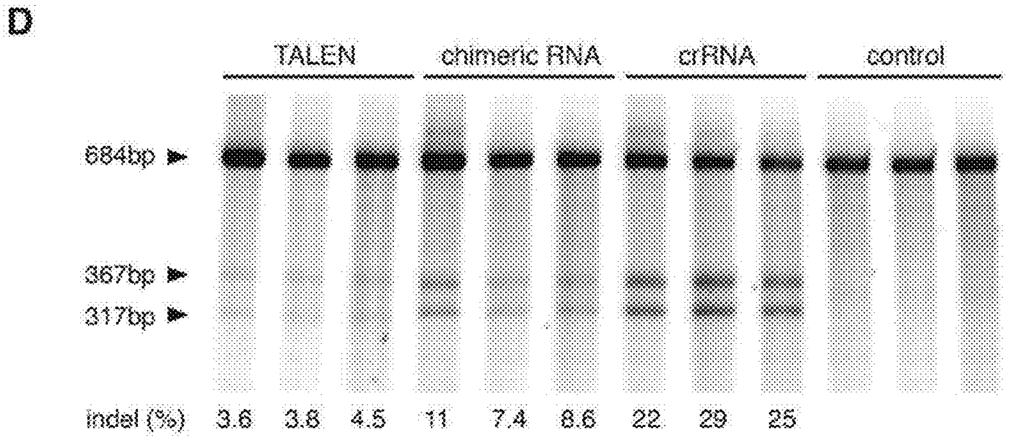

D
TALEN
chimeric RNA
crRNA
control
684bp ►
367bp ►
317bp ►
indel (%) 3.6 3.8 4.5 11 7.4 8.6 22 29 25

FIG. 3D

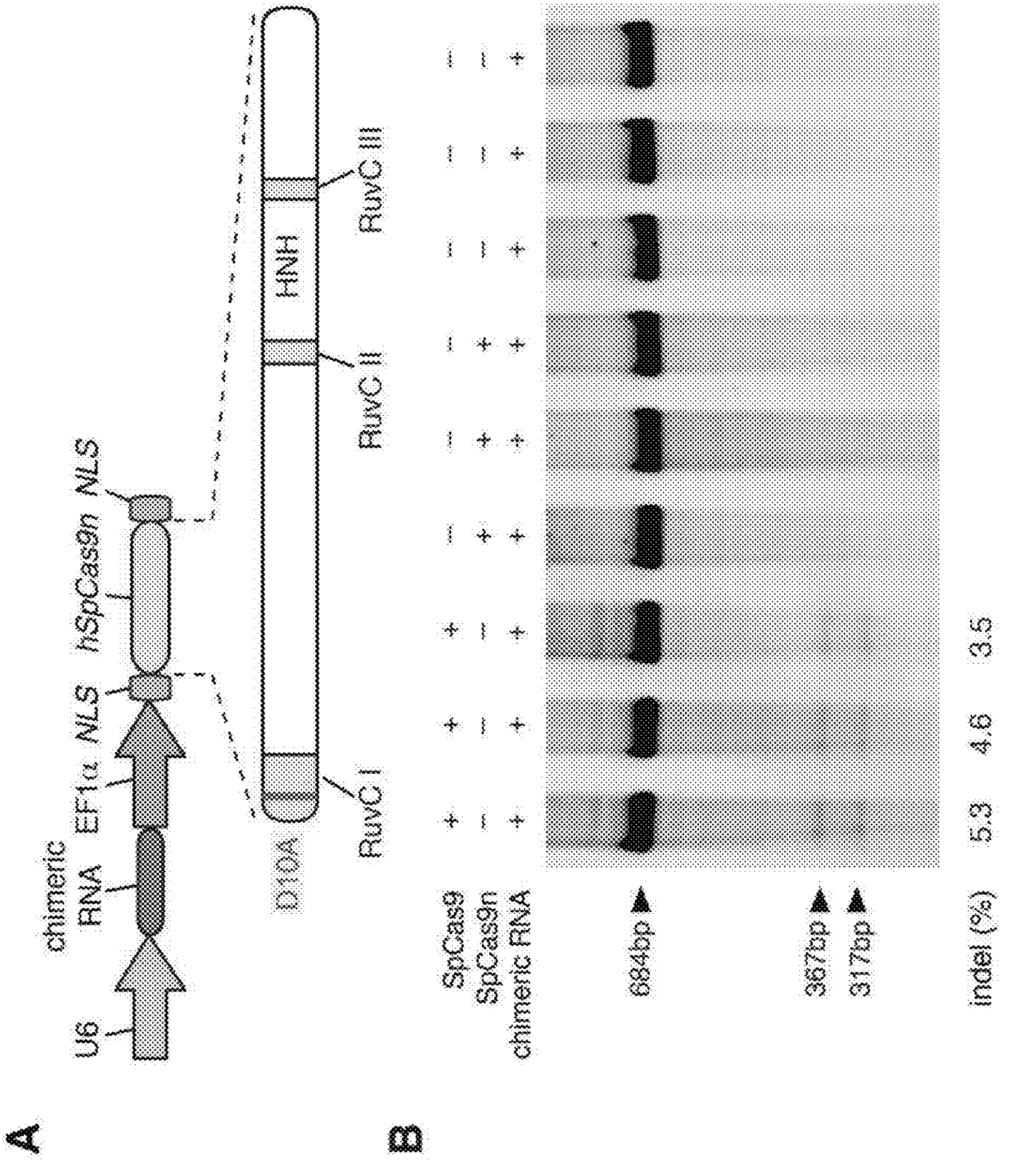

A
U6
chimeric RNA
EF1α
NLS
hSpCas9n
NLS
D10A
RuvC I
RuvC II
HNH
RuvC III
B
SpCas9 + + + − − − − − −
SpCas9n − − − + + + − − −
chimeric RNA + + + + + + + + +
684bp ▶
367bp ▶
317bp ▶
indel (%) 5.3 4.6 3.5

FIG. 4A-B

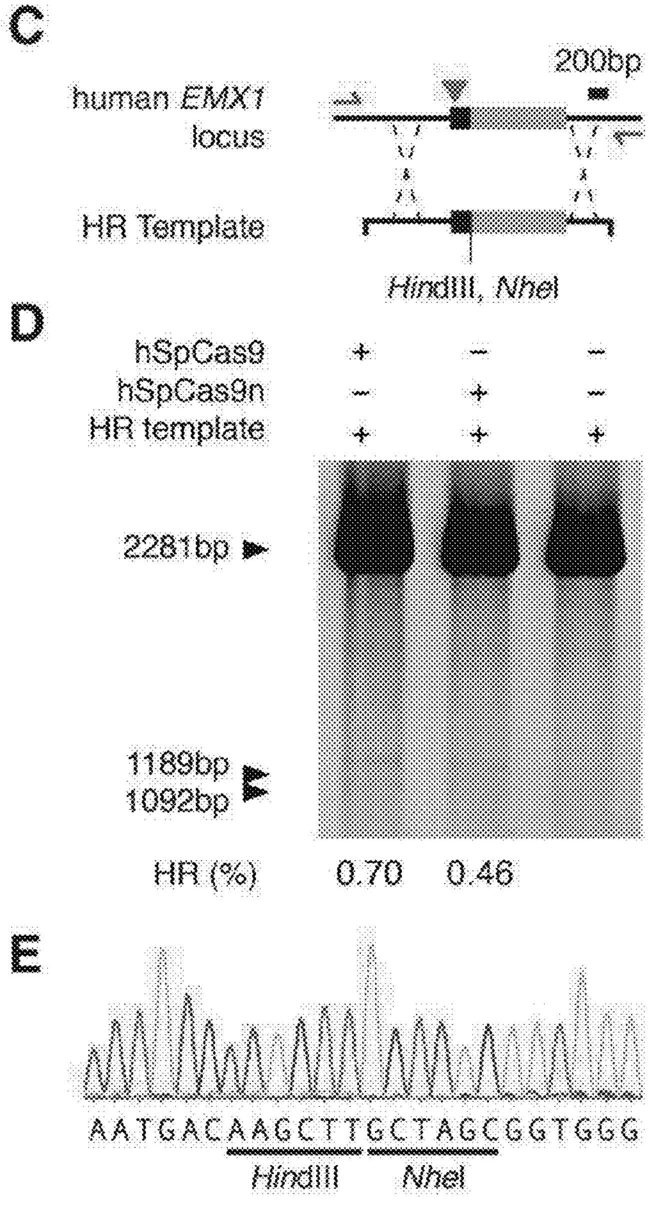

C
200bp
human *EMX1* locus
HR Template
*Hin*dIII, *Nhe*I
D
hSpCas9 + − −
hSpCas9n − + −
HR template + + +
2281bp
1189bp
1092bp
HR (%) 0.70 0.46
E
AATGACAAGCTTGCTAGCGGTGGG
*Hin*dIII
*Nhe*I

FIG. 4C-E

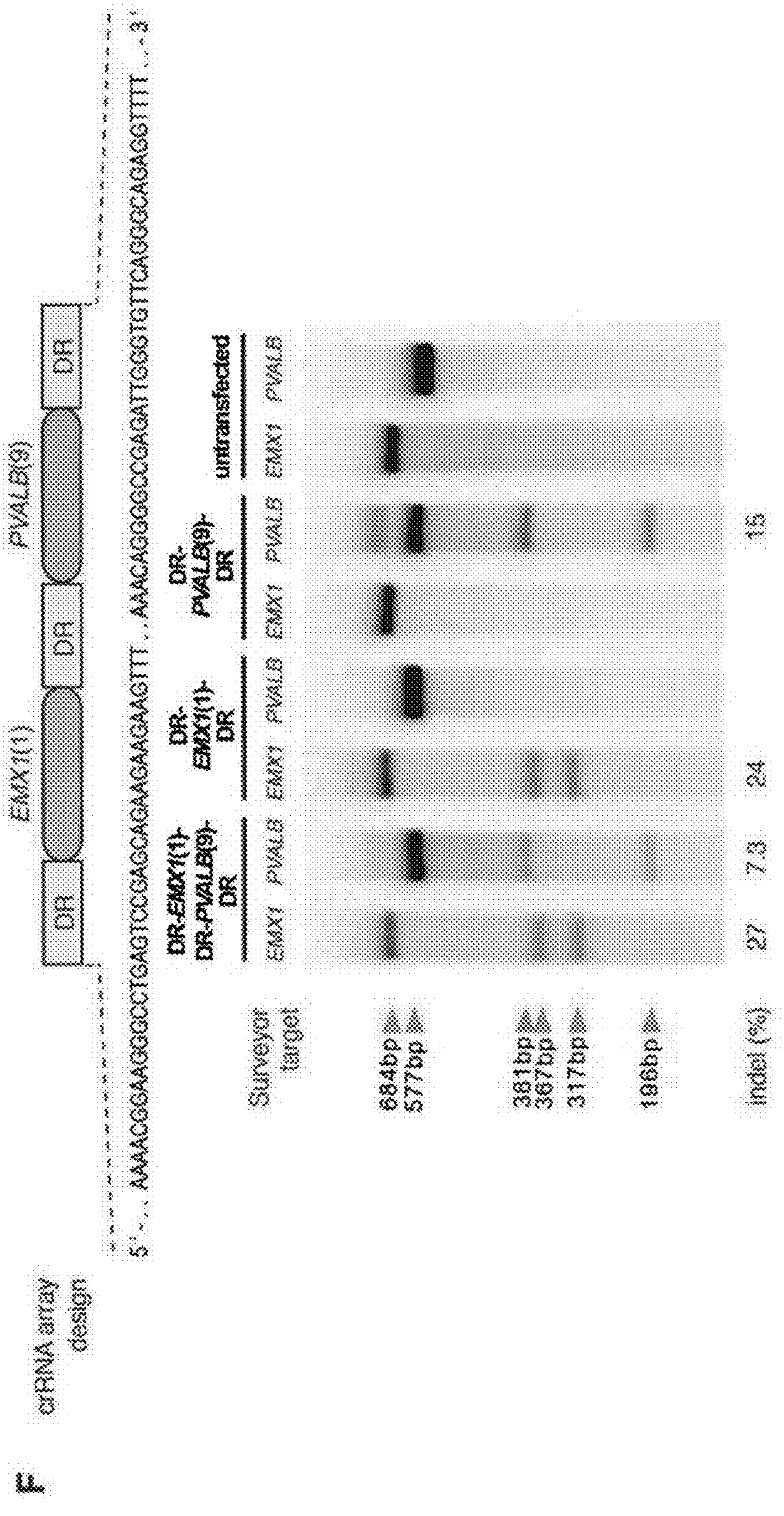

F
crRNA array design
EMX1(1)
PVALB(9)
DR
DR
DR
5'...AAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTT...AAACAGGGGCCGAGATTGGTGTTCAGGGCAGAGGTTTT...3'
Surveyor target
DR-EMX1(1)-DR-PVALB(9)-DR
DR-EMX1(1)-DR
DR-PVALB(9)-DR
untransfected
EMX1
PVALB
EMX1
PVALB
EMX1
PVALB
EMX1
PVALB
684bp
577bp
381bp
367bp
317bp
196bp
indel (%)
27
7.3
24
15

FIG. 4F

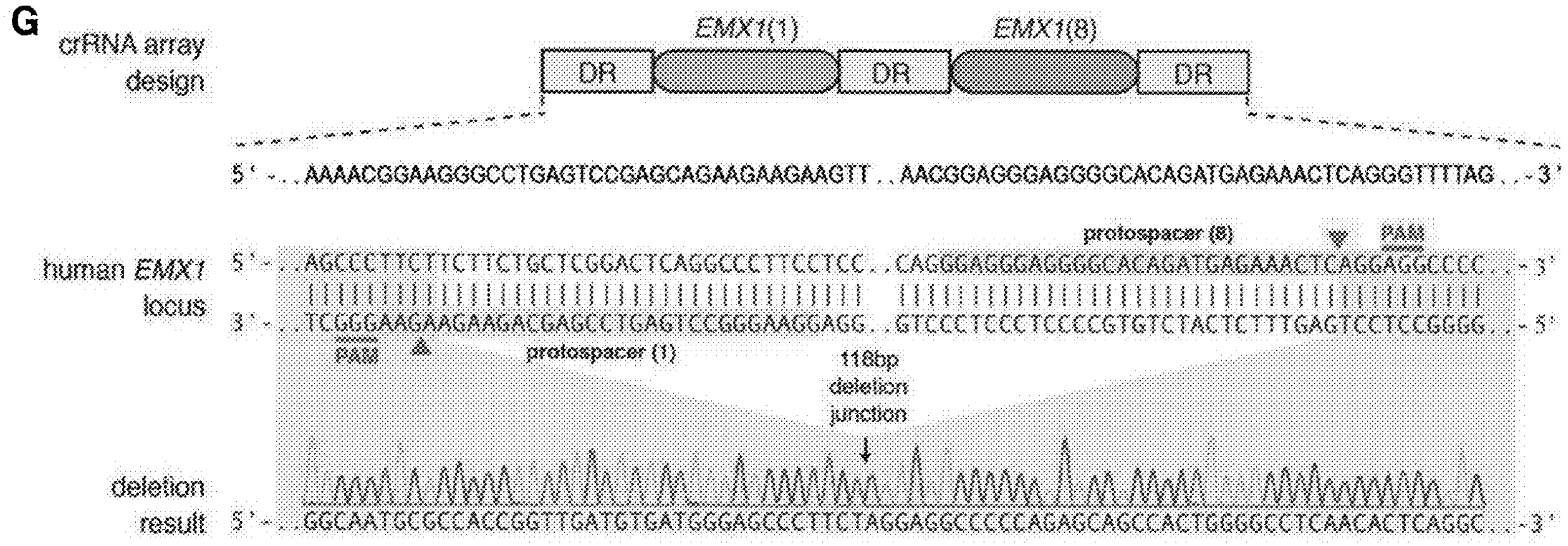

G
crRNA array design
DR
EMX1(1)
DR
EMX1(8)
DR
5'-..AAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTT..AACGGAGGGAGGGGCACAGATGAGAAACTCAGGGTTTTAG..-3'
human EMX1 locus
protospacer (8)
PAM
5'-..AGCCCTTCTTCTTCTGCTCGGACTCAGGCCCTTCCTCC..CAGGGAGGGAGGGGCACAGATGAGAAACTCAGGAGGCCCC..-3'
3'-..TCGGGAAGAAGAAGACGAGCCTGAGTCCGGGAAGGAGG..GTCCCTCCCTCCCCGTGTCTACTCTTTGAGTCCTCCGGGG..-5'
PAM
protospacer (1)
118bp deletion junction
deletion result
5'-..GGCAATGCGCCACCGGTTGATGTGATGGGAGCCCTTCTAGGAGGCCCCCAGAGCAGCCACTGGGGCCTCAACACTCAGGC..-3'

FIG. 4G

Cas9	target species	gene	protospacer ID	protospacer sequence (5' to 3')	PAM	strand	cell line tested	% indel (pre-crRNA + tracrRNA)	% indel (chimeric RNA)
S. pyogenes SF370 type II CRISPR	*Homo sapiens*	*EMX1*	1	GGAAGGGCCTGAGTCCGAGCAGAAGAAGAA	GGG	+	293FT	20 ± 1.8	6.7 ± 0.62
		EMX1	2	CATTGGAGGTGACATCGATGTCCTCCCCAT	TGG	–	293FT	2.1 ± 0.31	N.D.
		EMX1	3	GGACATCGATGTCACCTCCAATGACTAGGG	TGG	+	293FT	14 ± 1.1	N.D.
		EMX1	4	CATCGATGTCCTCCCCATTGGCCTGCTTCG	TGG	–	293FT	11 ± 1.7	N.D.
		EMX1	5	TTCGTGGCAATGCGCCACCGGTTGATGTGA	TGG	–	293FT	4.3 ± 0.46	2.1 ± 0.51
		EMX1	6	TCGTGGCAATGCGCCACCGGTTGATGTGAT	GGG	–	293FT	4.0 ± 0.66	0.41 ± 0.25
		EMX1	7	TCCAGCTTCTGCCGTTTGTACTTTGTCCTC	CGG	–	293FT	1.5 ± 0.12	N.D.
		EMX1	8	GGAGGGAGGGGCACAGATGAGAAACTCAGG	AGG	–	293FT	7.8 ± 0.83	2.3 ± 1.2
	Homo sapiens	*PVALB*	9	AGGGGCCGAGATTGGGTGTTCAGGGCAGAG	AGG	+	293FT	21 ± 2.6	6.5 ± 0.32
		PVALB	10	ATGCAGGAGGGTGGCGAGAGGGGCCGAGAT	TGG	+	293FT	N.D.	N.D.
		PVALB	11	GGTGGCGAGAGGGGCCGAGATTGGGTGTTC	AGG	+	293FT	N.D.	N.D.
	Mus musculus	*Th*	12	CAAGCACTGAGTGCCATTAGCTAAATGCAT	AGG	–	Neuro2A	27 ± 4.3	4.1 ± 2.2
		Th	13	AATGCATAGGGTACCACCCACAGGTGCCAG	GGG	–	Neuro2A	4.8 ± 1.2	N.D.
		Th	14	ACACACATGGGAAAGCCTCTGGGCCAGGAA	AGG	+	Neuro2A	11.3 ± 1.3	N.D.
S. thermophilus LMD-9 CRISPR1	*Homo sapiens*	*EMX1*	15	GGAGGAGGTAGTATACAGAAACACAGAGAA	GTAGAAT	–	293FT	14 ± 0.88	N.T.
		EMX1	16	AGAATGTAGAGGAGTCACAGAAACTCAGCA	CTAGAAA	–	293FT	7.8 ± 0.77	N.T.

FIG. 5

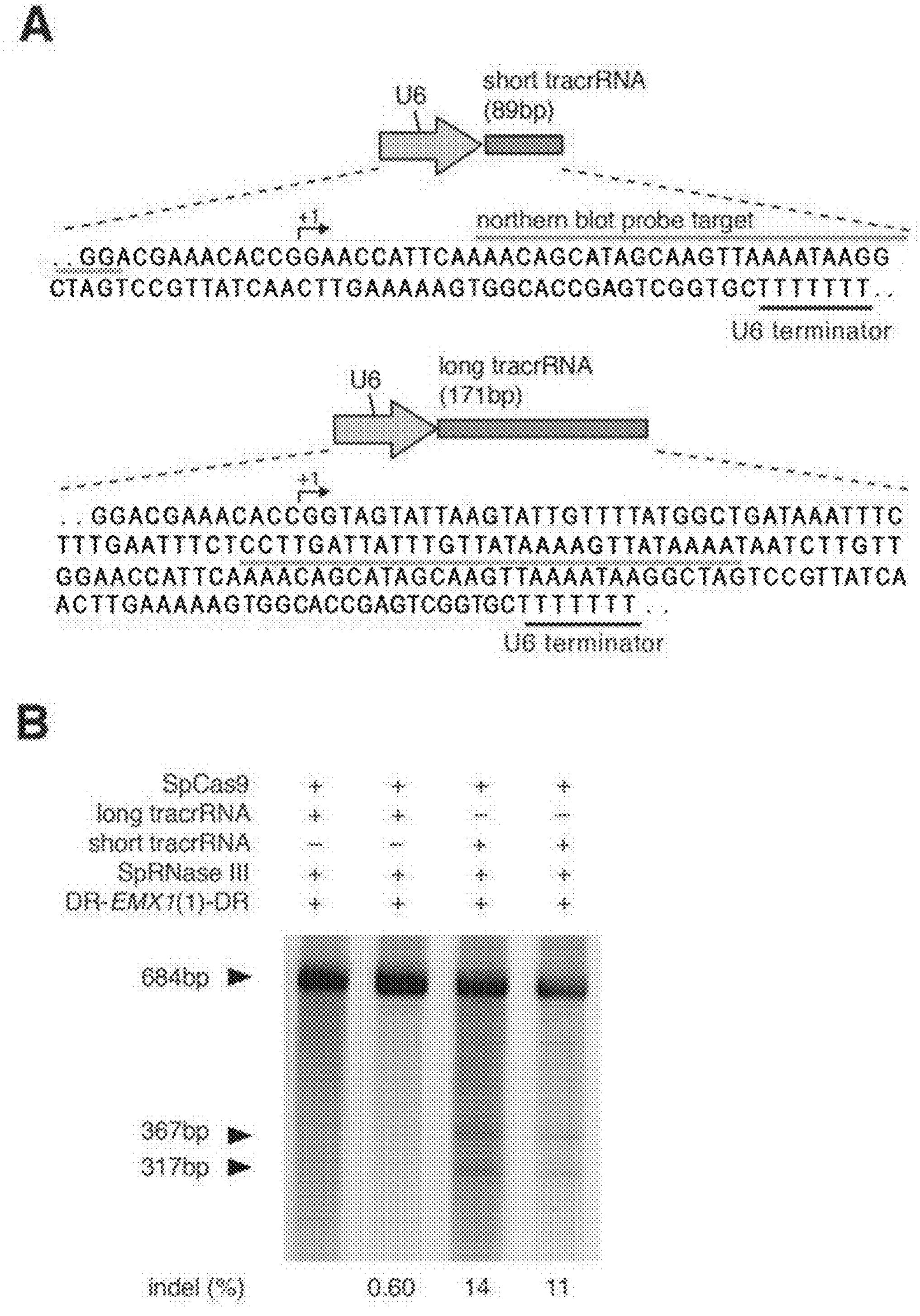

A
U6
short tracrRNA
(89bp)
+1
northern blot probe target
..GGACGAAACACCGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGG
CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT..
U6 terminator
U6
long tracrRNA
(171bp)
+1
..GGACGAAACACCGGTAGTATTAAGTATTGTTTTATGGCTGATAAATTTC
TTTGAATTTCTCCTTGATTATTTGTTATAAAAGTTATAAAATAATCTTGTT
GGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
ACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT..
U6 terminator
B
SpCas9 + + + +
long tracrRNA + + − −
short tracrRNA − − + +
SpRNase III + + + +
DR-*EMX1*(1)-DR + + + +
684bp
367bp
317bp
indel (%) 0.60 14 11

FIG. 6A-B

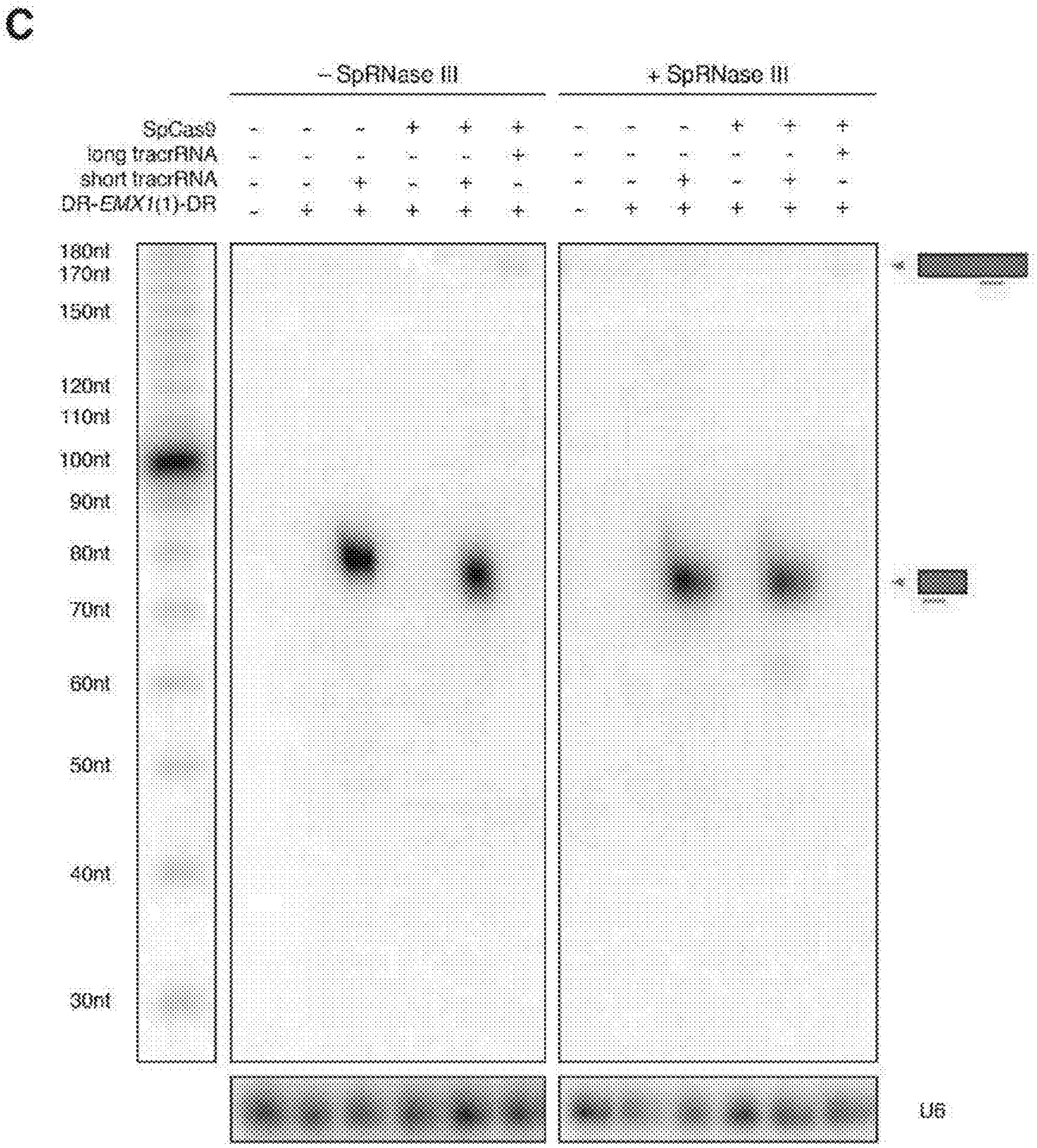

C
– SpRNase III
+ SpRNase III
SpCas9 - - - + + + - - - + + +
long tracrRNA - - - - - + - - - - - +
short tracrRNA - - + - + - - - + - + -
DR-*EMX1*(1)-DR - + + + + + - + + + + +
180nt
170nt
150nt
120nt
110nt
100nt
90nt
80nt
70nt
60nt
50nt
40nt
30nt
U6

FIG. 6C

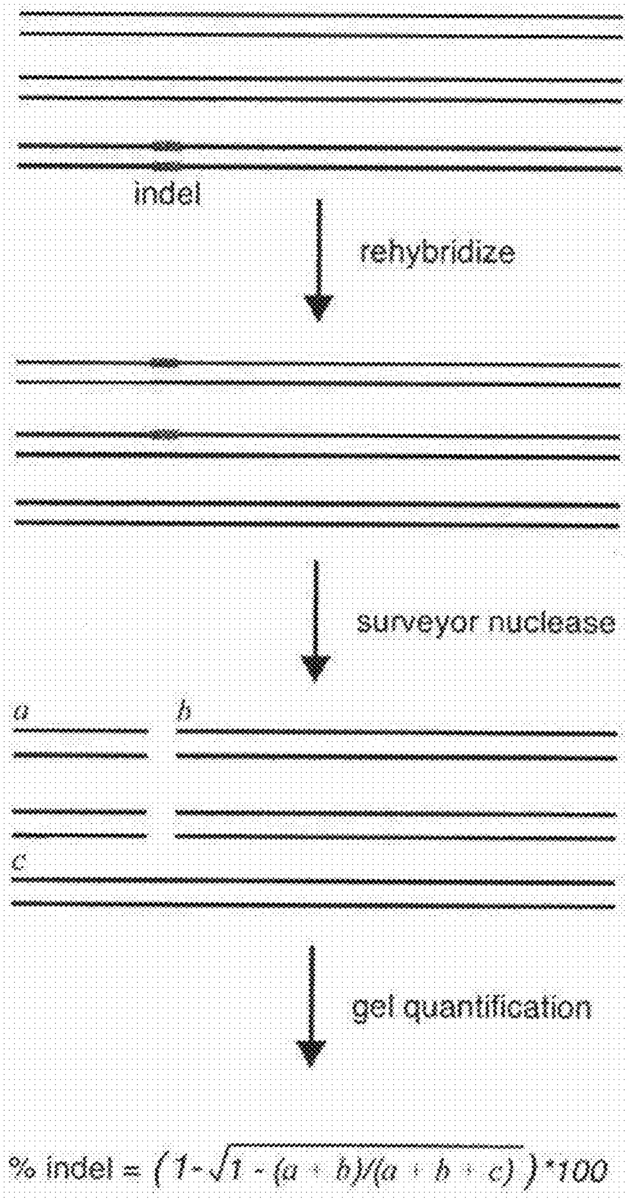

indel
rehybridize
surveyor nuclease
a
b
c
gel quantification
% indel = (1 - √(1 - (a + b)/(a + b + c)))*100

FIG. 7

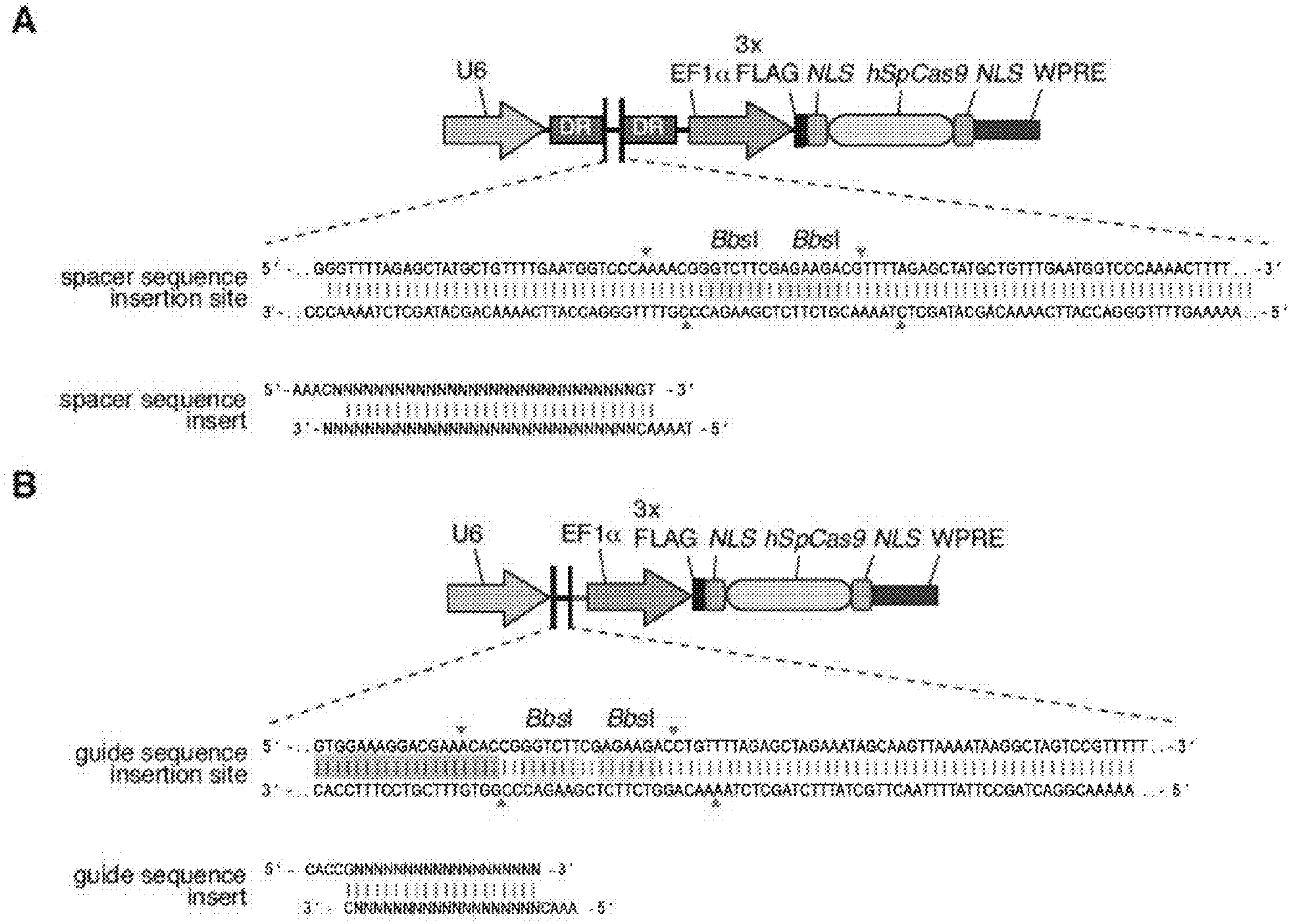
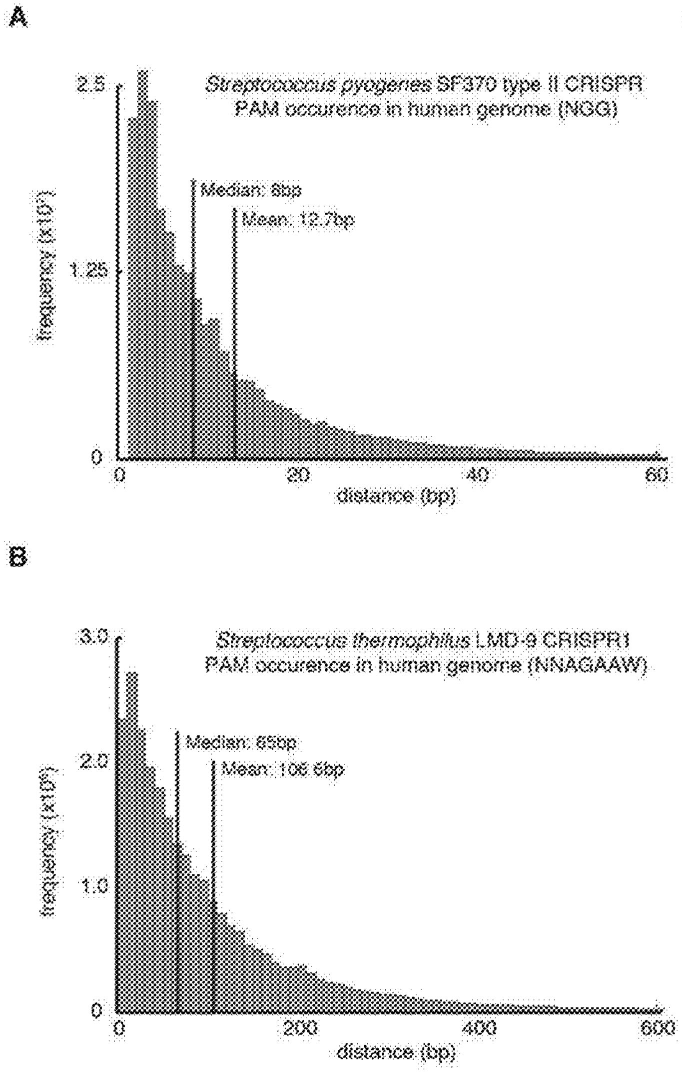

A
U6
3x
EF1α FLAG NLS hSpCas9 NLS WPRE
DR
DR
BbsI BbsI
spacer sequence insertion site
5'-..GGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGGGTCTTCGAGAAGACGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTTT..-3'
3'-..CCCAAAATCTCGATACGACAAAACTTACCAGGGTTTTGCCCAGAAGCTCTTCTGCAAAATCTCGATACGACAAAACTTACCAGGGTTTTGAAAAA..-5'
spacer sequence insert
5'-AAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGT -3'
3'-NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAAAAT -5'
B
U6
3x
EF1α FLAG NLS hSpCas9 NLS WPRE
BbsI BbsI
guide sequence insertion site
5'-..GTGGAAAGGACGAAACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTT..-3'
3'-..CACCTTTCCTGCTTTGTGGCCCAGAAGCTCTTCTGGACAAAATCTCGATCTTTATCGTTCAATTTTATTCCGATCAGGCAAAAA..-5'
guide sequence insert
5'- CACCGNNNNNNNNNNNNNNNNNNN -3'
3'- CNNNNNNNNNNNNNNNNNNNCAAA -5'

FIG. 8A-B



C

	NGG		NNAGAAW	
Chr	median	mean	median	mean
1	7	12.8	67	115.8
2	8	12.7	64	100.8
3	8	13.0	63	98.5
4	9	14.0	61	94.5
5	8	13.1	63	97.9
6	8	13.1	63	98.5
7	8	12.4	64	102.9
8	8	12.8	64	100.9
9	7	13.9	65	120.5
10	7	12.1	66	107.0
11	7	12.0	65	105.8
12	8	12.4	65	103.5
13	8	13.6	62	94.6
14	8	12.0	65	101.5
15	7	11.5	68	107.7
16	7	11.7	74	136.8
17	6	10.3	76	127.9
18	8	13.4	63	101.8
19	6	9.4	82	145.4
20	7	11.1	72	121.8
21	7	13.4	64	111.4
22	6	9.2	85	140.3
X	8	13.2	63	99.0
Y	8	29.2	62	223.7

FIG. 9A-C

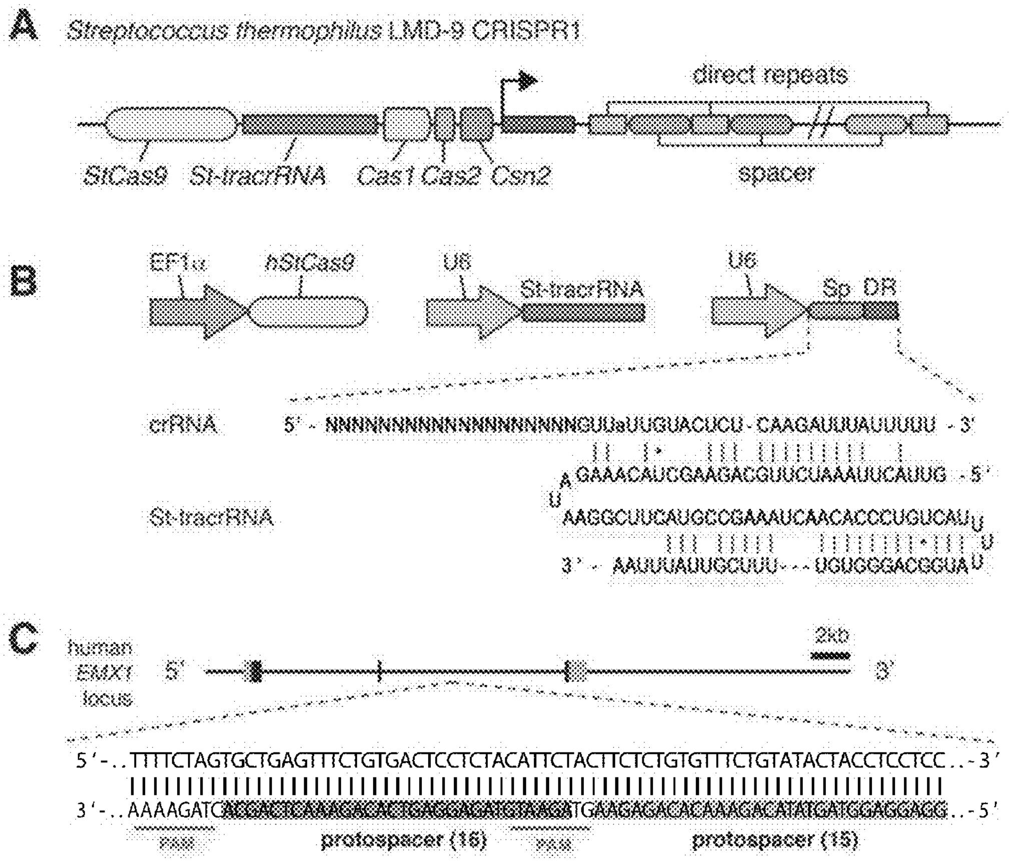

A
Streptococcus thermophilus LMD-9 CRISPR1
direct repeats
StCas9
St-tracrRNA
Cas1
Cas2
Csn2
spacer
B
EF1α
hStCas9
U6
St-tracrRNA
U6
Sp
DR
crRNA
5' - NNNNNNNNNNNNNNNNNNNNGUUaUUGUACUCU - CAAGAUUUAUUUUU - 3'
A GAAACAUCGAAGACGUUCUAAAUUCAUUG - 5'
U
St-tracrRNA
AAGGCUUCAUGCCGAAAUCAACACCCUGUCAU U
U
3' - AAUUUAUUGCUUU - - - UGUGGGACGGUA U
C
human
EMX1
locus
5'
2kb
3'
5'-..TTTTCTAGTGCTGAGTTTCTGTGACTCCTCTACATTCTACTTCTCTGTGTTTCTGTATACTACCTCCTCC..-3'
3'-..AAAAGATCACGACTCAAAGACACTGAGGAGATGTAAGATGAAGAGACACAAAGACATATGATGGAGGAGG..-5'
PAM
protospacer (16)
PAM
protospacer (15)

FIG. 10A-C

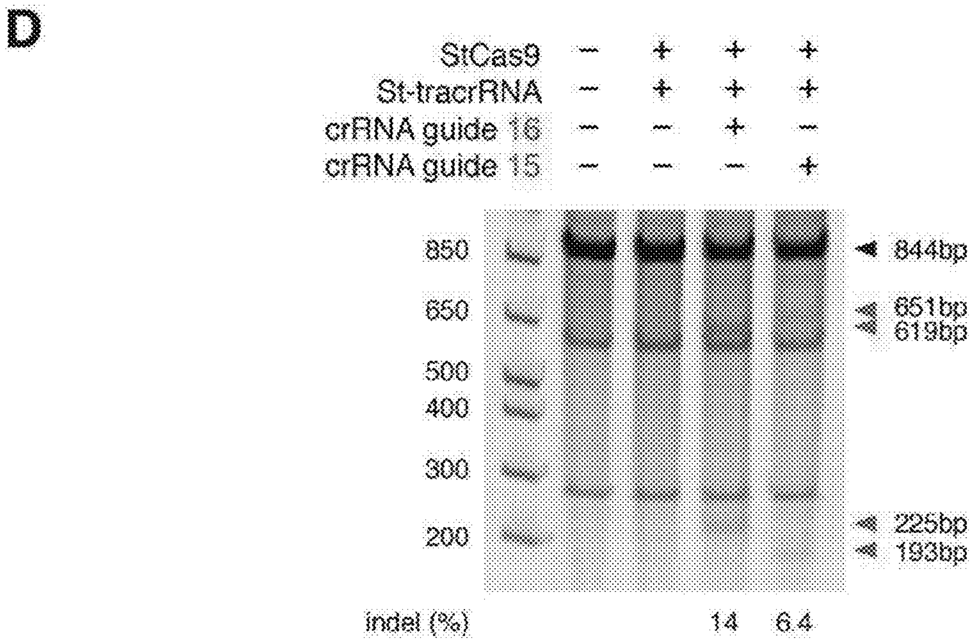

D
StCas9 − + + +
St-tracrRNA − + + +
crRNA guide 16 − − + −
crRNA guide 15 − − − +
850
650
500
400
300
200
844bp
651bp
619bp
225bp
193bp
indel (%) 14 6.4

FIG. 10D

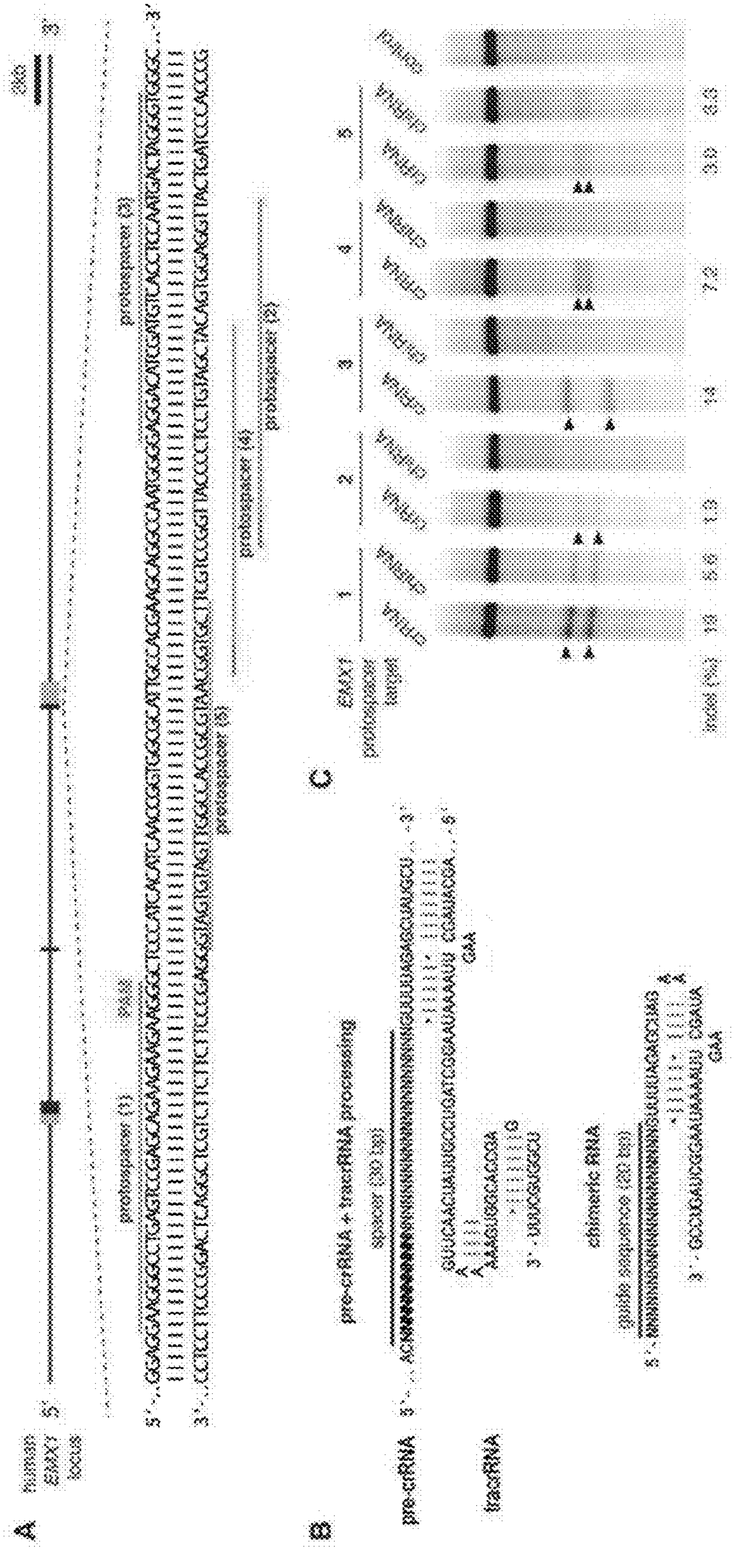

A
human
EMX1
locus
5'
3'
protospacer (1)
PAM
protospacer (2)
protospacer (3)
protospacer (4)
protospacer (5)
B
pre-crRNA
tracrRNA
pre-crRNA + tracrRNA processing
spacer (30 bp)
chimeric RNA
guide sequence (20 bp)
GAA
C
EMX1
protospacer
target
1
2
3
4
5
crRNA
chiRNA
indel (%)
18
5.6
13
14
7.2
3.0
6.0

FIG. 11A-C

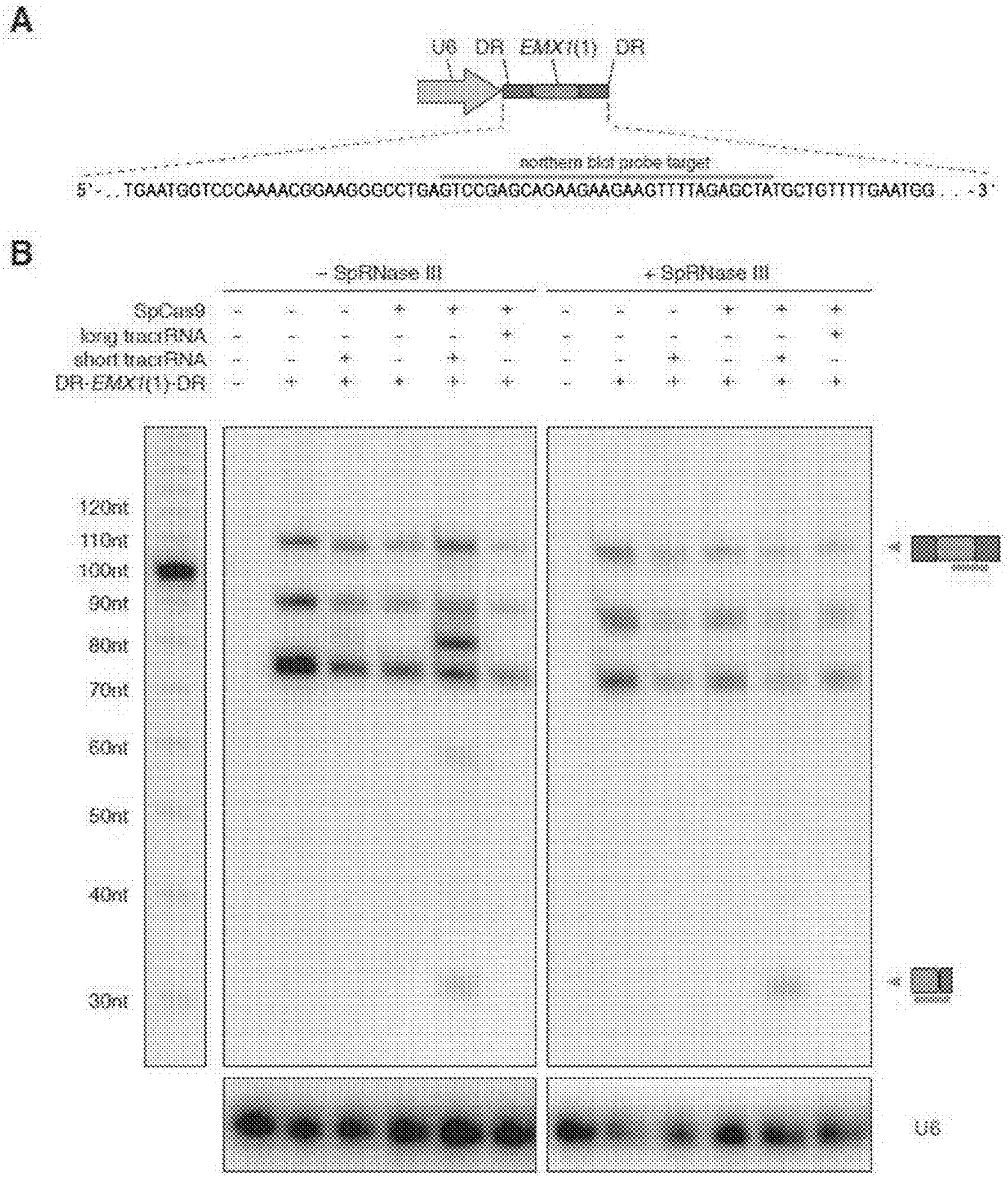

A
U6
DR
EMX1(1)
DR
northern blot probe target
5'-..TGAATGGTCCCAAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTATGCTGTTTTGAATGG...-3'
B
− SpRNase III
+ SpRNase III
SpCas9 - - - + + + - - - + + +
long tracrRNA - - - - - + - - - - - +
short tracrRNA - - + - + - - - + - + -
DR-*EMX1*(1)-DR - + + + + + - + + + + +
120nt
110nt
100nt
90nt
80nt
70nt
60nt
50nt
40nt
30nt
U6

FIG. 12A-B

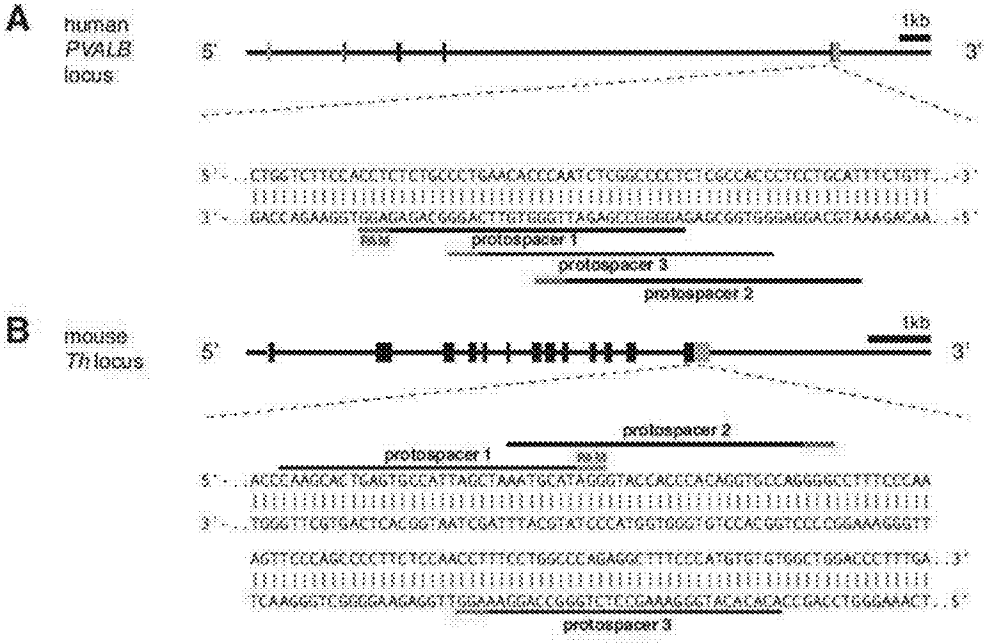

A
human
PVALB
locus
5'
3'
1kb
5'-..CTGGTCTTCCACCTCTCTGCCCTGAACACCCAATCTCGGCCCCTCTCGCCACCCTCCTGCATTTCTGTT..-3'
3'-..GACCAGAAGGTGGAGAGACGGGACTTGTGGGTTAGAGCCGGGGAGAGCGGTGGGAGGACGTAAAGACAA..-5'
PAM
protospacer 1
protospacer 3
protospacer 2
B
mouse
Th locus
5'
3'
1kb
protospacer 2
protospacer 1
PAM
5'-..ACCCAAGCACTGAGTGCCATTAGCTAAATGCATAGGGTACCACCCACAGGTGCCAGGGGCCTTTCCCAA
3'-..TGGGTTCGTGACTCACGGTAATCGATTTACGTATCCCATGGTGGGTGTCCACGGTCCCCGGAAAGGGTT
AGTTCCCAGCCCCTTCTCCAACCTTTCCTGGCCCAGAGGCTTTCCCATGTGTGTGGCTGGACCCTTTGA..3'
TCAAGGGTCGGGGAAGAGGTTGGAAAGGACCGGGTCTCCGAAAGGGTACACACACCGACCTGGGAAACT..5'
protospacer 3

FIG. 13A-B

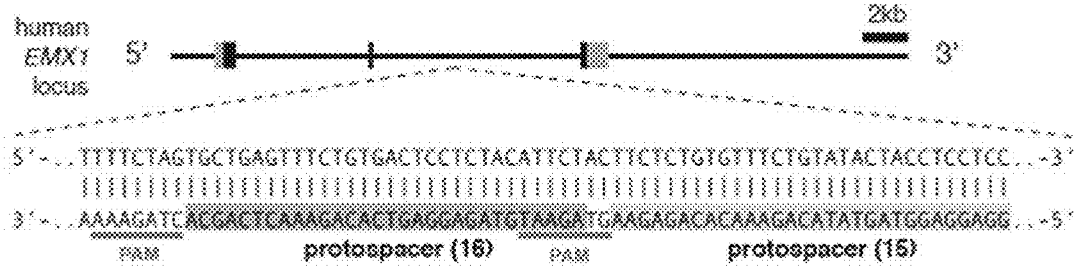

human
EMX1
locus
5'
3'
2kb
5'-..TTTTCTAGTGCTGAGTTTCTGTGACTCCTCTACATTCTACTTCTCTGTGTTTCTGTATACTACCTCCTCC..-3'
3'-..AAAAGATCACGACTCAAAGACACTGAGGAGATGTAAGATGAAGAGACACAAAGACATATGATGGAGGAGG..-5'
PAM
protospacer (16)
PAM
protospacer (15)

FIG. 14

Primer name	**Assay**	**Genomic Target**	**Primer sequence**
Sp-EMX1-F	SURVEYOR assay, sequencing	*EMX1*	AAAACCACCCTTCTCTCTGGC
Sp-EMX1-R	SURVEYOR assay, sequencing	*EMX1*	GGAGATTGGAGACACGGAGAG
Sp-PVALB-F	SURVEYOR assay, sequencing	*PVALB*	CTGGAAAGCCAATGCCTGAC
Sp-PVALB-R	SURVEYOR assay, sequencing	*PVALB*	GGCAGCAAACTCCTTGTCCT
Sp-Th-F	SURVEYOR assay, sequencing	*Th*	GTGCTTTGCAGAGGCCTACC
Sp-Th-R	SURVEYOR assay, sequencing	*Th*	CCTGGAGCGCATGCAGTAGT
St-EMX1-F	SURVEYOR assay, sequencing	*EMX1*	ACCTTCTGTGTTTCCACCATTC
St-EMX1-R	SURVEYOR assay, sequencing	*EMX1*	TTGGGGAGTGCACAGACTTC
Sp-EMX1-RFLP-F	RFLP, sequencing	*EMX1*	GGCTCCCTGGGTTCAAAGTA
Sp-EMX1-RFLP-R	RFLP, sequencing	*EMX1*	AGAGGGGTCTGGATGTCGTAA
Pb_EMX1_sp1	Northern Blot Probe	Not applicable	TAGCTCTAAAACTTCTTCTTCTGCTCGGAC
Pb_tracrRNA	Northern Blot Probe	Not applicable	CTAGCCTTATTTTAACTTGCTATGCTGTTT

FIG. 15

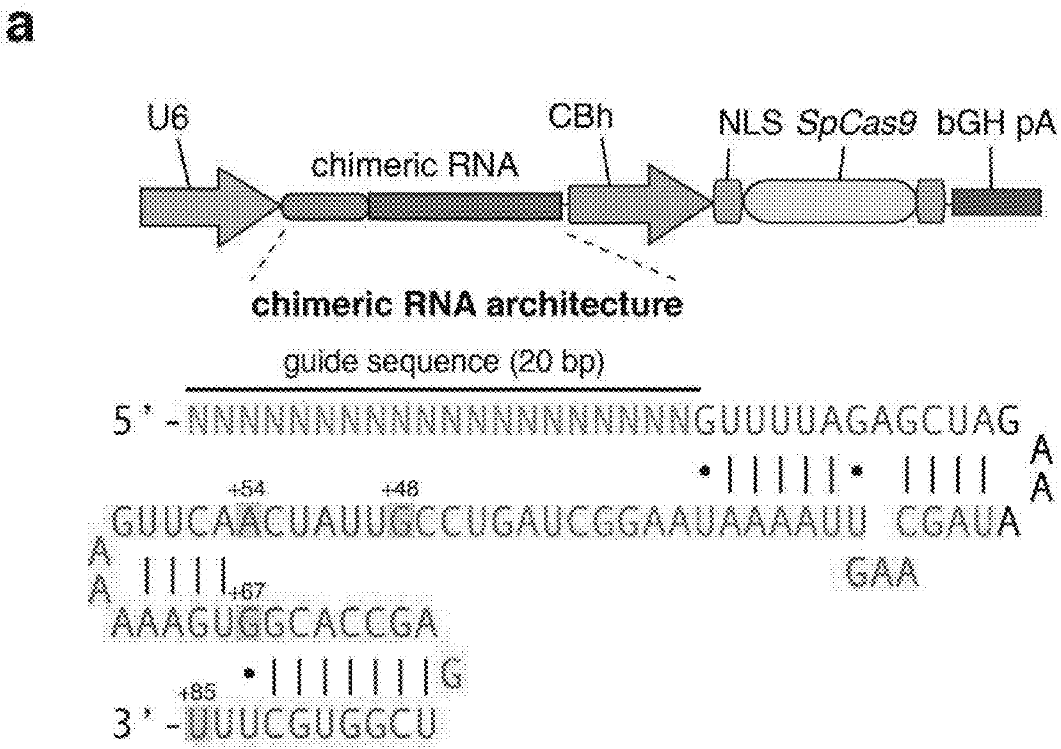

a
U6
chimeric RNA
CBh
NLS SpCas9 bGH pA
chimeric RNA architecture
guide sequence (20 bp)
5'-NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
A
A
+54
+48
GUUCAACUAUUGCCUGAUCGGAAUAAAAUU CGAUA
GAA
A
A
+67
AAAGUGGCACCGA
G
+85
3'-UUUCGUGGCU

FIG. 16A

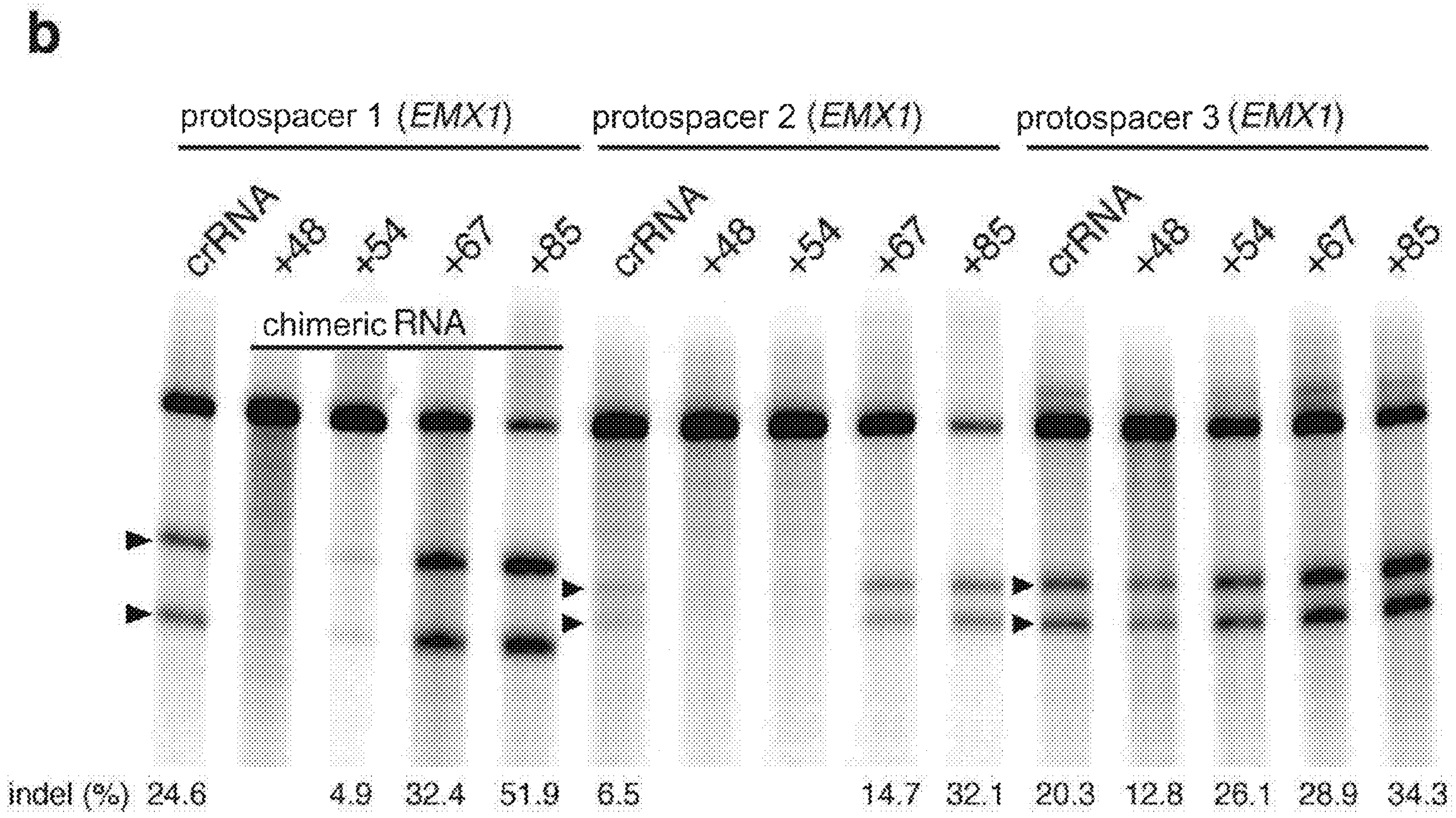

b
protospacer 1 (*EMX1*)
protospacer 2 (*EMX1*)
protospacer 3 (*EMX1*)
crRNA
+48
+54
+67
+85
crRNA
+48
+54
+67
+85
crRNA
+48
+54
+67
+85
chimeric RNA
indel (%)
24.6
4.9
32.4
51.9
6.5
14.7
32.1
20.3
12.8
26.1
28.9
34.3

FIG. 16B

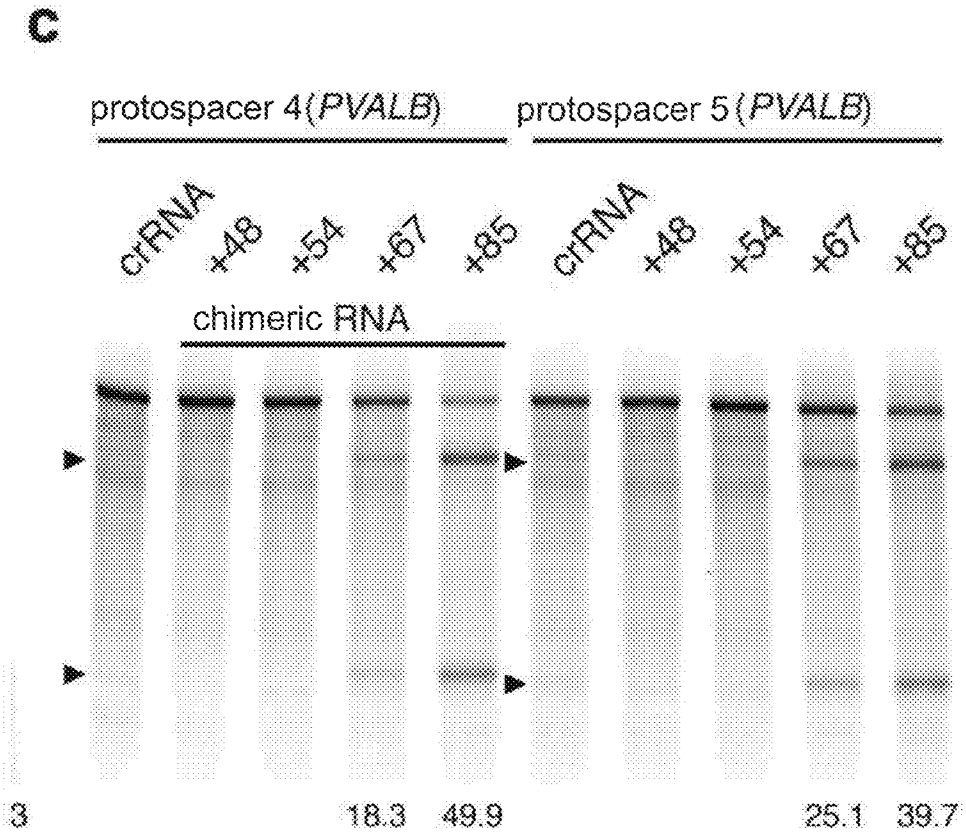

C
protospacer 4 (*PVALB*)
protospacer 5 (*PVALB*)
crRNA
+48
+54
+67
+85
crRNA
+48
+54
+67
+85
chimeric RNA
3
18.3
49.9
25.1
39.7

FIG. 16C

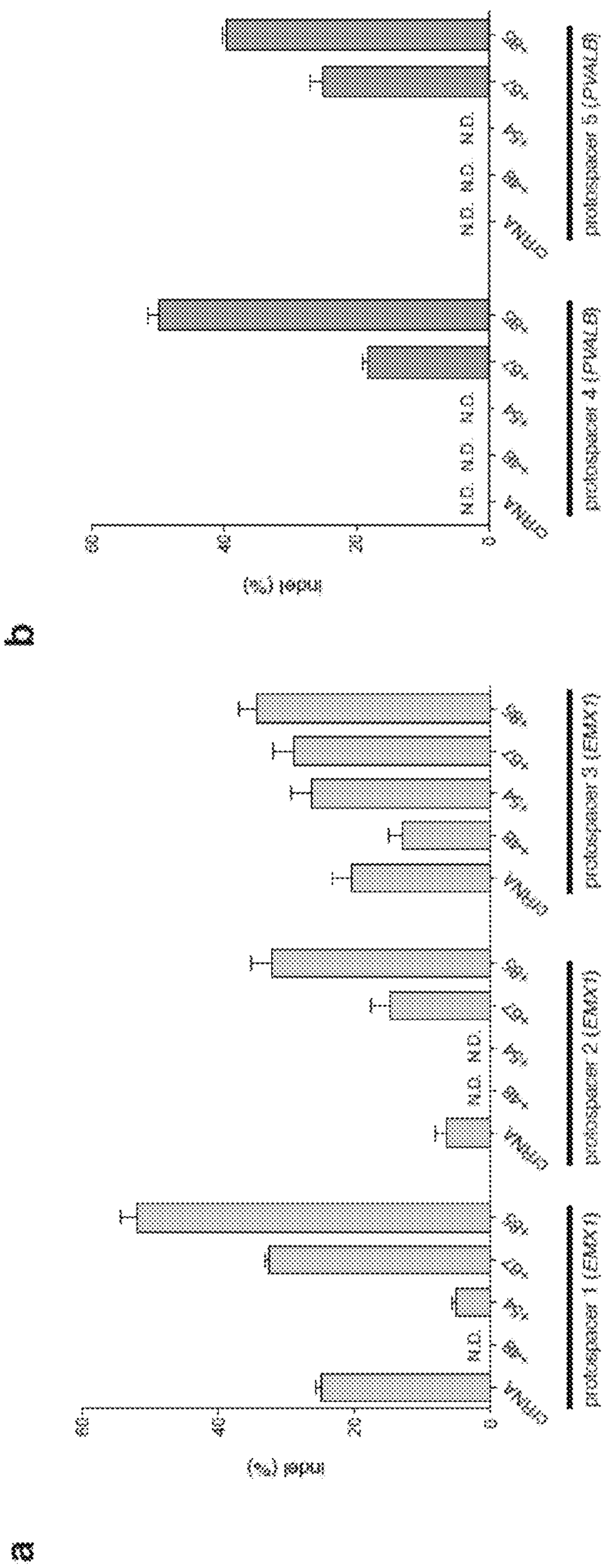

a
indel (%)
0
20
40
60
crRNA
N.D.
protospacer 1 (EMX1)
N.D.
N.D.
protospacer 2 (EMX1)
protospacer 3 (EMX1)
b
indel (%)
0
20
40
60
crRNA
N.D.
N.D.
N.D.
protospacer 4 (PVALB)
N.D.
N.D.
N.D.
protospacer 5 (PVALB)

FIG. 17A-B

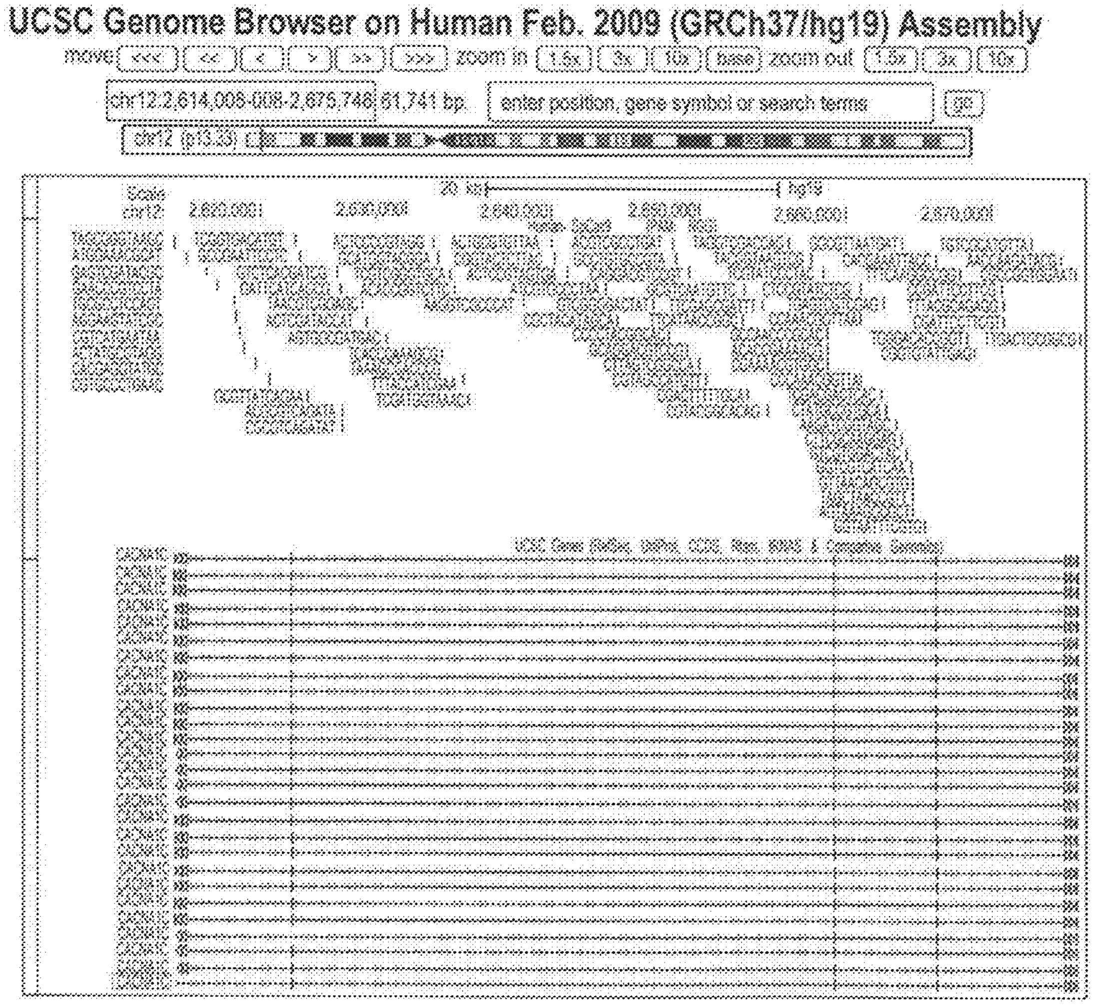

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
chr12:2,614,008-008-2,675,748 61,741 bp.
enter position, gene symbol or search terms
go
chr12 (p13.33)
Scale
20 kb
hg19
chr12: 2,620,000 2,630,000 2,640,000 2,650,000 2,660,000 2,670,000
UCSC Genes (RefSeq, UniProt, CCDS, Rfam, tRNAs & Comparative Genomics)
CACNA1C

FIG. 18

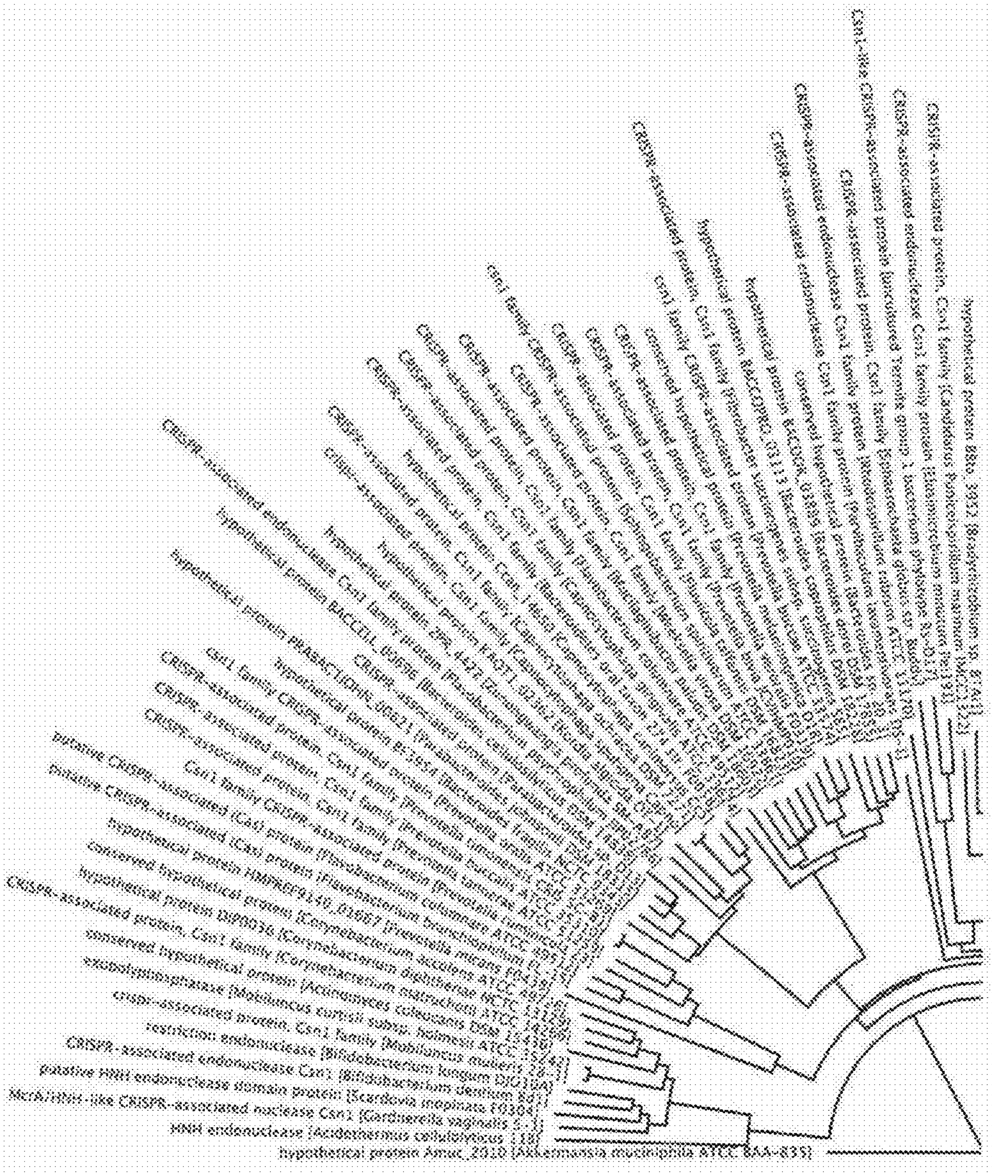

putative CRISPR-associated (Cas) protein [Flavobacterium columnare ATCC 49512]
putative CRISPR-associated (Cas) protein [Flavobacterium branchiophilum FL-15]
hypothetical protein HMPREF9140_01987 [Prevotella micans F0438]
hypothetical protein DIP0036 [Corynebacterium diphtheriae NCTC 13129]
CRISPR-associated protein, Csn1 family [Corynebacterium accolens ATCC 49725]
conserved hypothetical protein [Actinomyces coleocanis DSM 15436]
exopolyphosphatase [Mobiluncus curtisii subsp. holmesii ATCC 35242]
crispr-associated protein, Csn1 family [Mobiluncus mulieris 28-1]
restriction endonuclease [Bifidobacterium longum DJO10A]
CRISPR-associated endonuclease Csn1 [Bifidobacterium dentium Bd1]
putative HNH endonuclease domain protein [Scardovia inopinata F0304]
McrA/HNH-like CRISPR-associated nuclease Csn1 [Gardnerella vaginalis 5-1]
HNH endonuclease [Acidothermus cellulolyticus 11B]
hypothetical protein Amuc_2010 [Akkermansia muciniphila ATCC BAA-835]

FIG. 19A

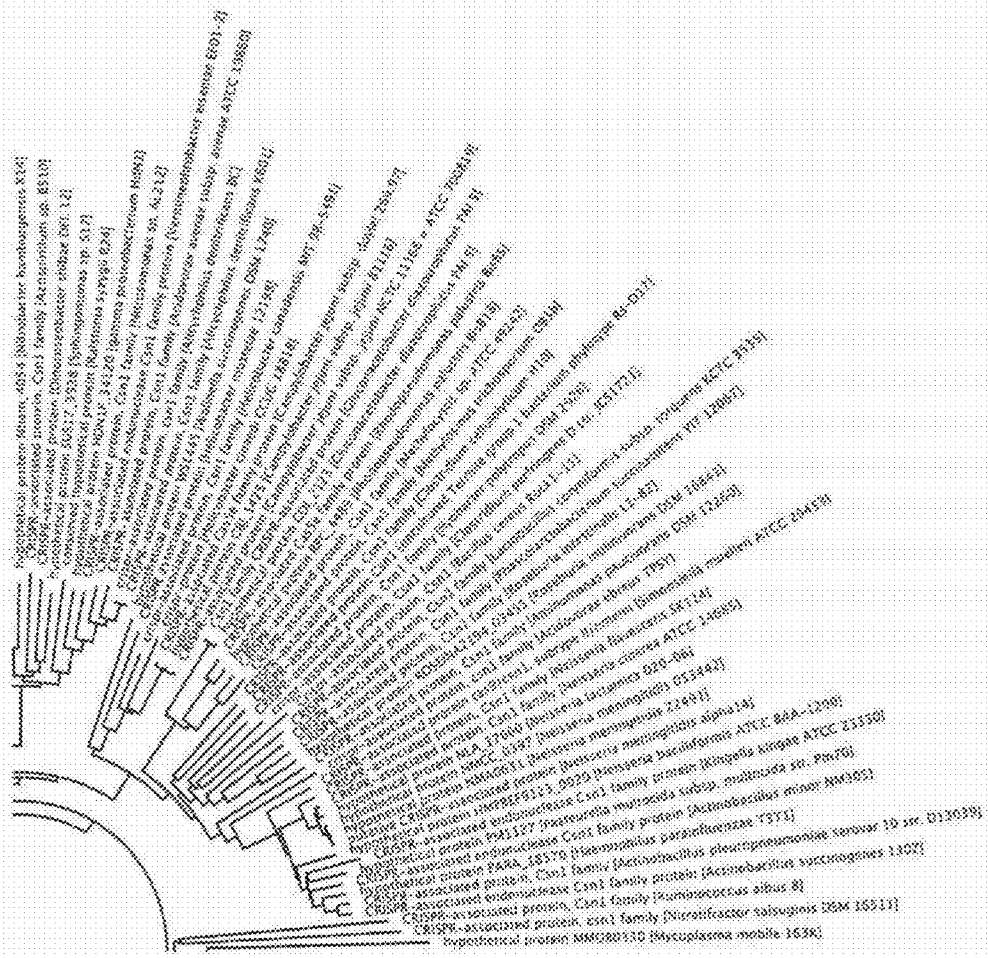

Hypothetical protein MMOB0330 [Mycoplasma mobile 163K]
CRISPR-associated protein, csn1 family [Nitratifractor salsuginis DSM 16511]
CRISPR-associated endonuclease Csn1 family [Ruminococcus albus 8]
Putative CRISPR-associated endonuclease Csn1 family protein [Actinobacillus minor NM305]

FIG. 19B

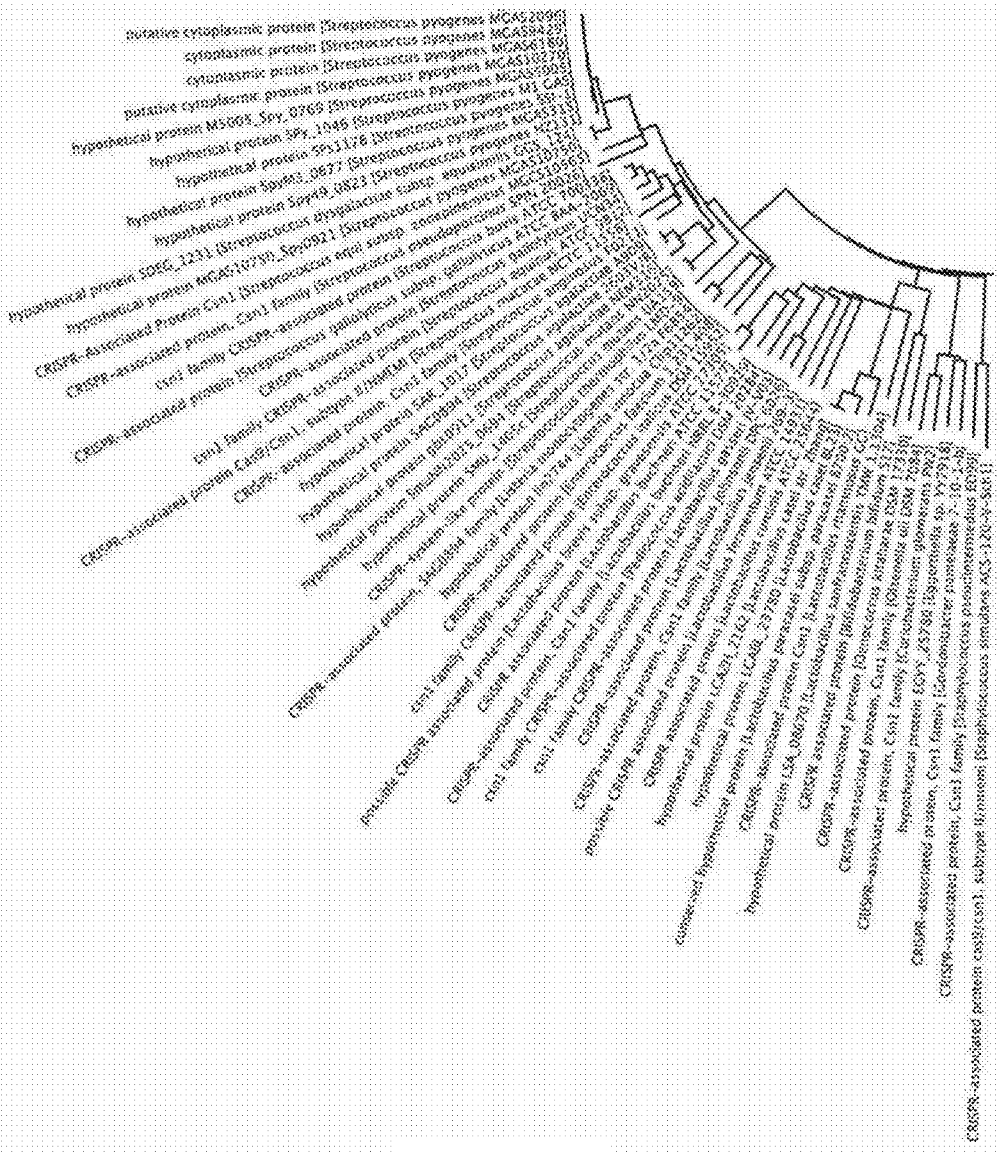

putative cytoplasmic protein [Streptococcus pyogenes MGAS2096]
cytoplasmic protein [Streptococcus pyogenes MGAS9429]
cytoplasmic protein [Streptococcus pyogenes MGAS6180]
cytoplasmic protein [Streptococcus pyogenes MGAS10270]
putative cytoplasmic protein [Streptococcus pyogenes M1 GAS]

FIG. 19C

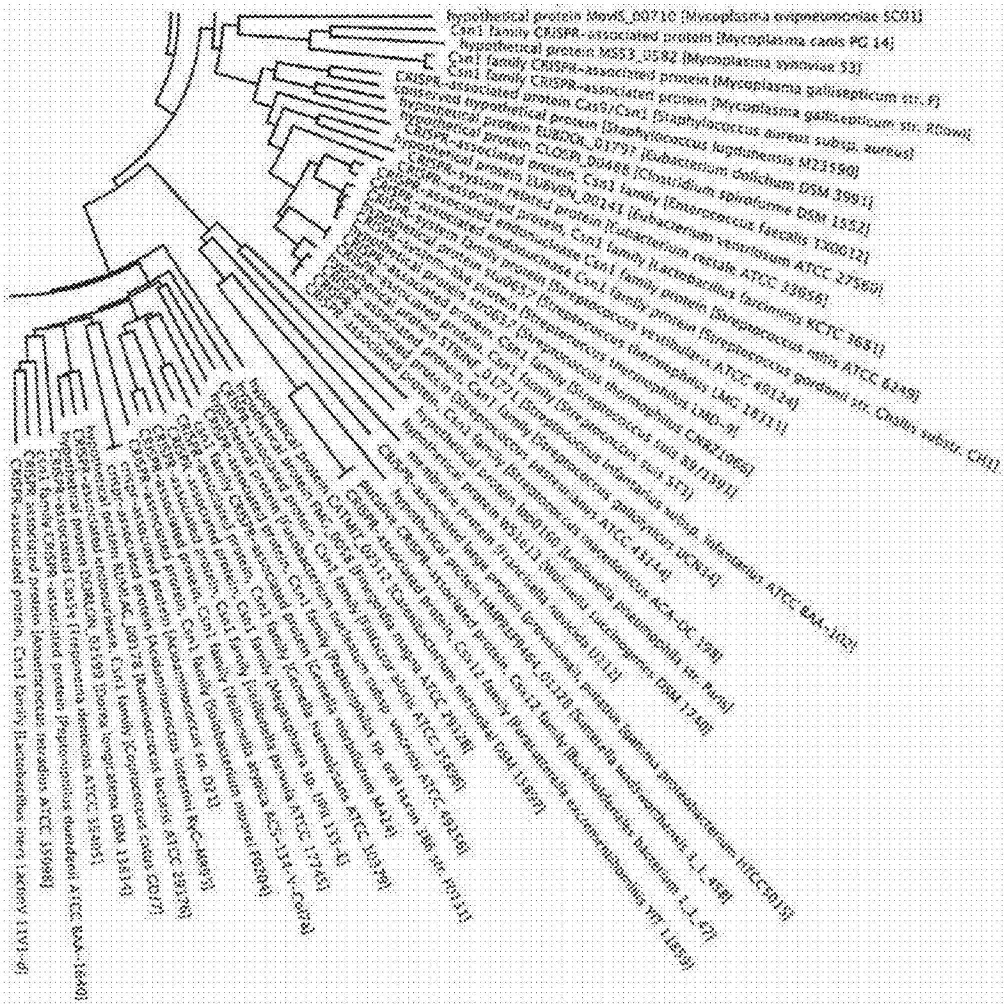

hypothetical protein MovS_00710 [Mycoplasma ovipneumoniae SC01]
Csn1 family CRISPR-associated protein [Mycoplasma canis PG 14]
hypothetical protein MS53_0582 [Mycoplasma synoviae 53]
Csn1 family CRISPR-associated protein [Mycoplasma gallisepticum str. F]
CRISPR-associated protein [Mycoplasma gallisepticum str. R(low)]
conserved hypothetical protein Cas9/Csn1 [Staphylococcus aureus subsp. aureus]
hypothetical protein [Staphylococcus lugdunensis M23590]
hypothetical protein EUBDOL_01797 [Eubacterium dolichum DSM 3991]
hypothetical protein CLOSPI_00468 [Clostridium spiroforme DSM 1552]
CRISPR-associated protein, Csn1 family [Enterococcus faecalis TX0012]
CRISPR-system related protein EUBVEN_00241 [Eubacterium ventriosum ATCC 27560]
CRISPR-associated protein [Eubacterium rectale ATCC 33656]
CRISPR-associated endonuclease, Csn1 family [Lactobacillus farciminis KCTC 3681]

FIG. 19D

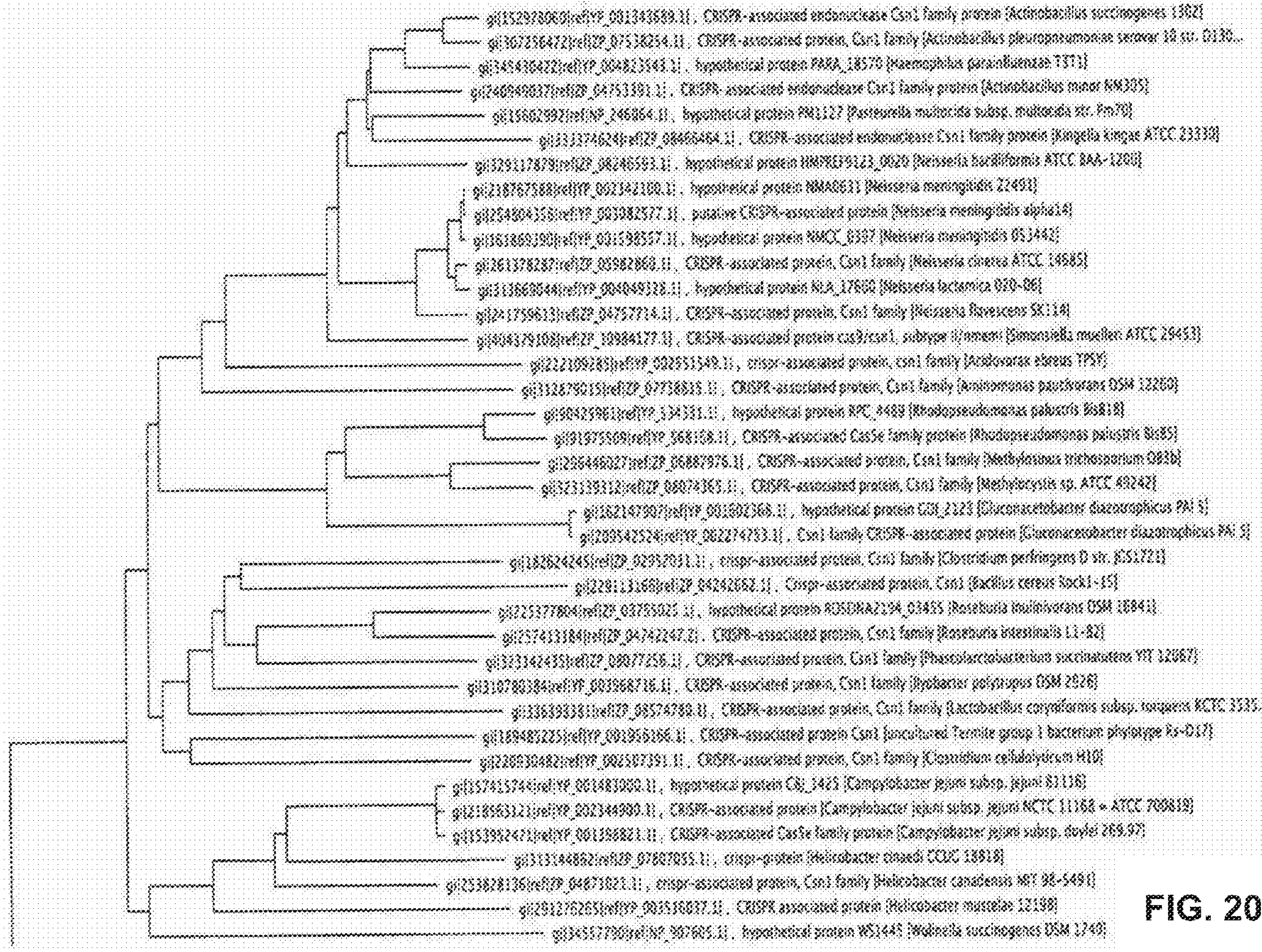

gi|152978060|ref|YP_001343689.1| , CRISPR-associated endonuclease Csn1 family protein [Actinobacillus succinogenes 130Z]
gi|307256472|ref|ZP_07538254.1| , CRISPR-associated protein, Csn1 family [Actinobacillus pleuropneumoniae serovar 10 str. D130...
gi|345430422|ref|YP_004823543.1| , hypothetical protein PARA_18570 [Haemophilus parainfluenzae T3T1]
gi|240949037|ref|ZP_04753381.1| , CRISPR-associated endonuclease Csn1 family protein [Actinobacillus minor NM305]
gi|15602992|ref|NP_246064.1| , hypothetical protein PM1127 [Pasteurella multocida subsp. multocida str. Pm70]
gi|333374624|ref|ZP_08466464.1| , CRISPR-associated endonuclease Csn1 family protein [Kingella kingae ATCC 23330]
gi|329117879|ref|ZP_08246593.1| , hypothetical protein HMPREF9123_0020 [Neisseria bacilliformis ATCC BAA-1200]
gi|218767588|ref|YP_002342100.1| , hypothetical protein NMA0631 [Neisseria meningitidis Z2491]
gi|254804358|ref|YP_003082577.1| , putative CRISPR-associated protein [Neisseria meningitidis alpha14]
gi|161869390|ref|YP_001598557.1| , hypothetical protein NMCC_0397 [Neisseria meningitidis 053442]
gi|261378287|ref|ZP_05982860.1| , CRISPR-associated protein, Csn1 family [Neisseria cinerea ATCC 14685]
gi|313668044|ref|YP_004048328.1| , hypothetical protein NLA_17660 [Neisseria lactamica 020-06]
gi|241759613|ref|ZP_04757714.1| , CRISPR-associated protein, Csn1 family [Neisseria flavescens SK114]
gi|304379108|ref|ZP_10984177.1| , CRISPR-associated protein cas9/csn1, subtype II/nmemi [Simonsiella muelleri ATCC 29453]
gi|222109285|ref|YP_002551549.1| , crispr-associated protein, csn1 family [Acidovorax ebreus TPSY]
gi|312879015|ref|ZP_07738815.1| , CRISPR-associated protein, Csn1 family [Aminomonas paucivorans DSM 12260]
gi|90425961|ref|YP_534331.1| , hypothetical protein RPC_4489 [Rhodopseudomonas palustris BisB18]
gi|91975509|ref|YP_568168.1| , CRISPR-associated Cas5e family protein [Rhodopseudomonas palustris BisB5]
gi|296446027|ref|ZP_06887975.1| , CRISPR-associated protein, Csn1 family [Methylosinus trichosporium OB3b]
gi|323139312|ref|ZP_08074365.1| , CRISPR-associated protein, Csn1 family [Methylocystis sp. ATCC 49242]
gi|162147907|ref|YP_001602368.1| , hypothetical protein GDI_2123 [Gluconacetobacter diazotrophicus PAl 5]
gi|209542524|ref|YP_002274753.1| , Csn1 family CRISPR-associated protein [Gluconacetobacter diazotrophicus PAl 5]
gi|182624245|ref|ZP_02952031.1| , crispr-associated protein, Csn1 family [Clostridium perfringens D str. JGS1721]
gi|229113166|ref|ZP_04242662.1| , Crispr-associated protein, Csn1 [Bacillus cereus Rock1-15]
gi|225377804|ref|ZP_03755025.1| , hypothetical protein ROSEINA2194_03455 [Roseburia inulinivorans DSM 16841]
gi|257413184|ref|ZP_04742247.2| , CRISPR-associated protein, Csn1 family [Roseburia intestinalis L1-82]
gi|323342435|ref|ZP_08077256.1| , CRISPR-associated protein, Csn1 family [Phascolarctobacterium succinatutens YIT 12067]
gi|310780384|ref|YP_003968716.1| , CRISPR-associated protein, Csn1 family [Ilyobacter polytropus DSM 2926]
gi|336393381|ref|ZP_08574780.1| , CRISPR-associated protein, Csn1 family [Lactobacillus coryniformis subsp. torquens KCTC 3535...
gi|189485225|ref|YP_001956166.1| , CRISPR-associated protein Csn1 [uncultured Termite group 1 bacterium phylotype Rs-D17]
gi|220930482|ref|YP_002507391.1| , CRISPR-associated protein, Csn1 family [Clostridium cellulolyticum H10]
gi|157415744|ref|YP_001483000.1| , hypothetical protein C8J_1425 [Campylobacter jejuni subsp. jejuni 81116]
gi|218563121|ref|YP_002344900.1| , CRISPR-associated protein [Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819]
gi|153952471|ref|YP_001398821.1| , CRISPR-associated Cas5e family protein [Campylobacter jejuni subsp. doylei 269.97]
gi|313144862|ref|ZP_07807055.1| , crispr-protein [Helicobacter cinaedi CCUG 18818]
gi|253828136|ref|ZP_04871021.1| , crispr-associated protein, Csn1 family [Helicobacter canadensis MIT 98-5491]
gi|291276265|ref|YP_003516037.1| , CRISPR associated protein [Helicobacter mustelae 12198]
gi|34557790|ref|NP_907605.1| , hypothetical protein WS1445 [Wolinella succinogenes DSM 1740]

FIG. 20A

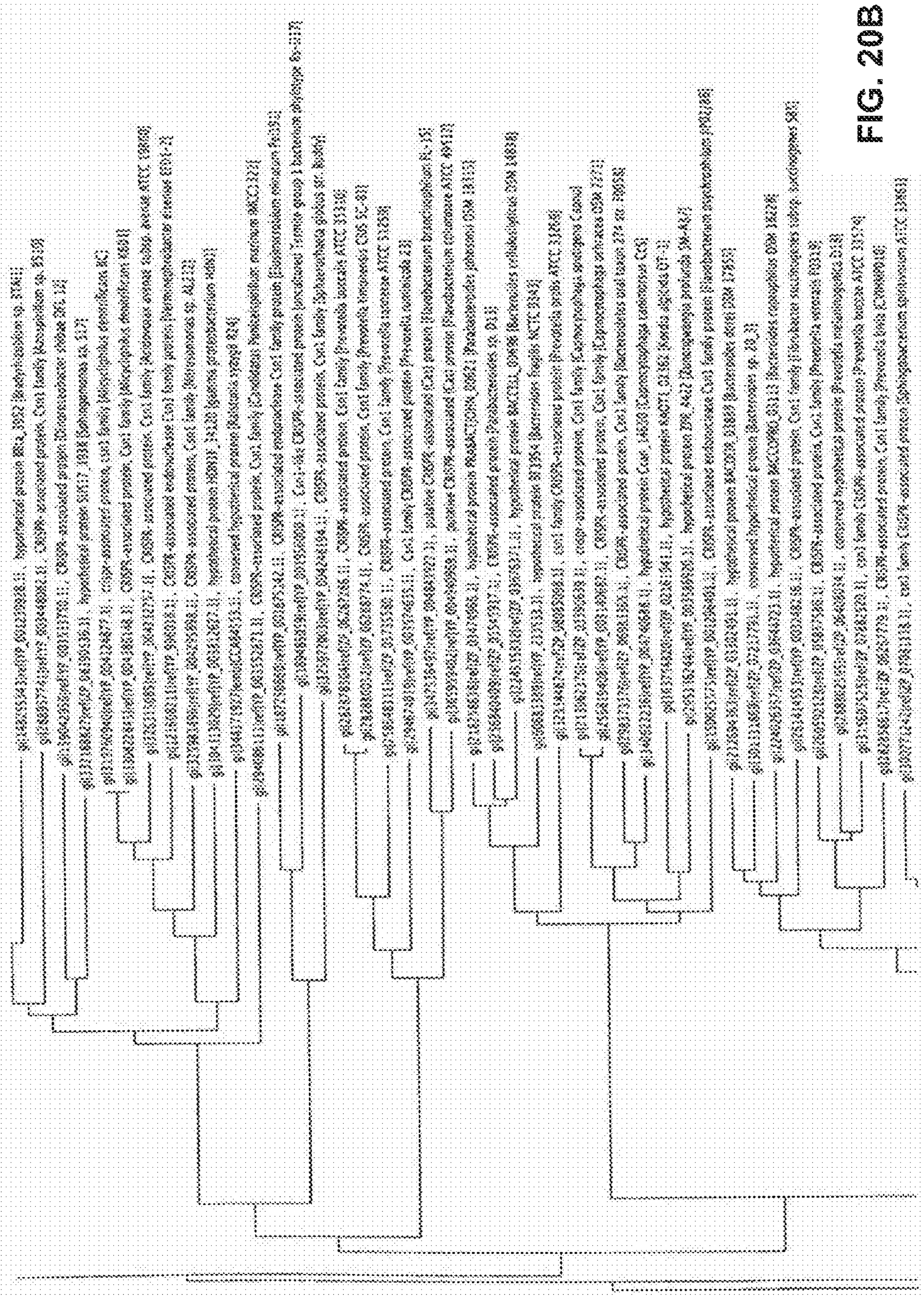


FIG. 20B

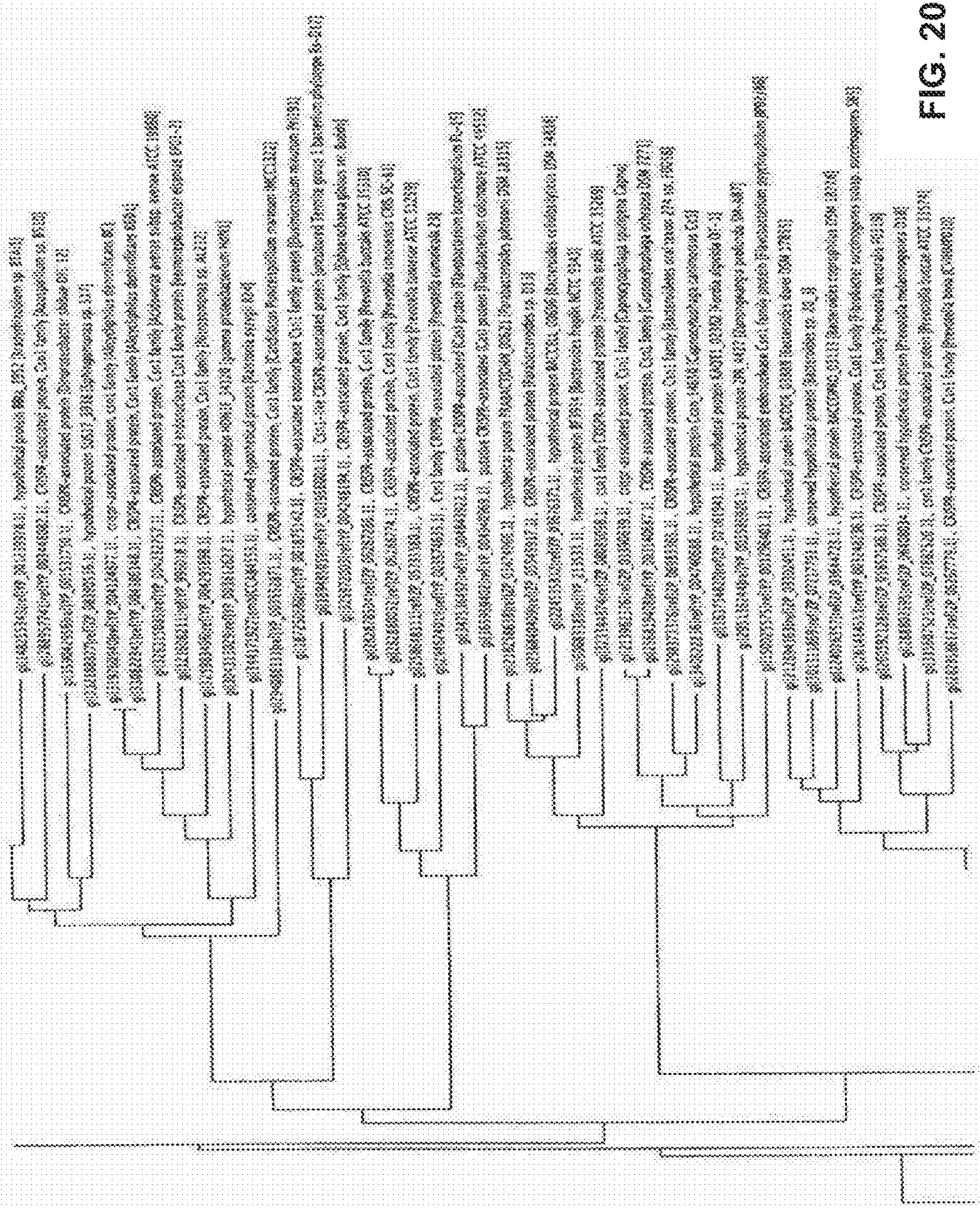


FIG. 20C

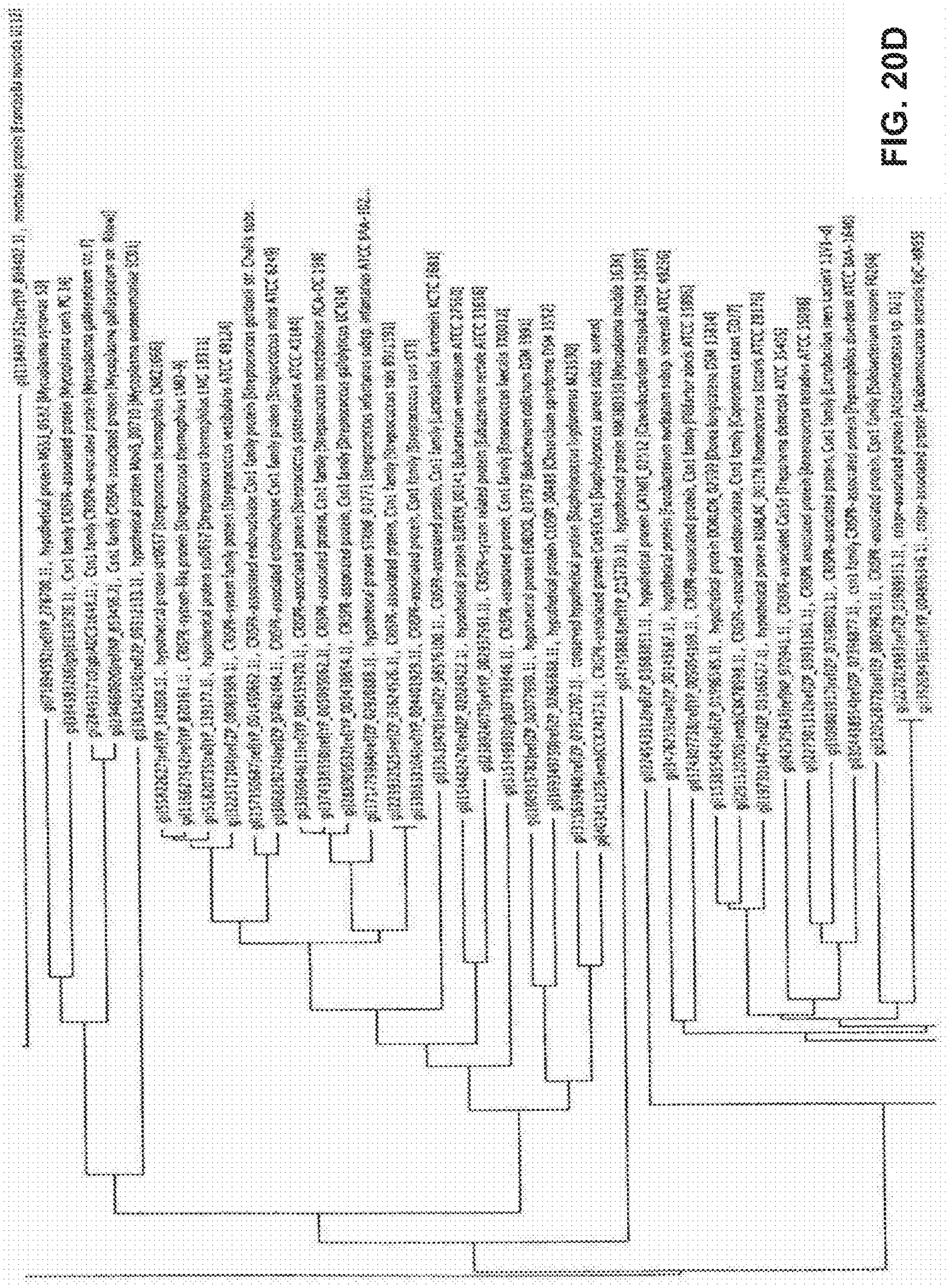


FIG. 20D

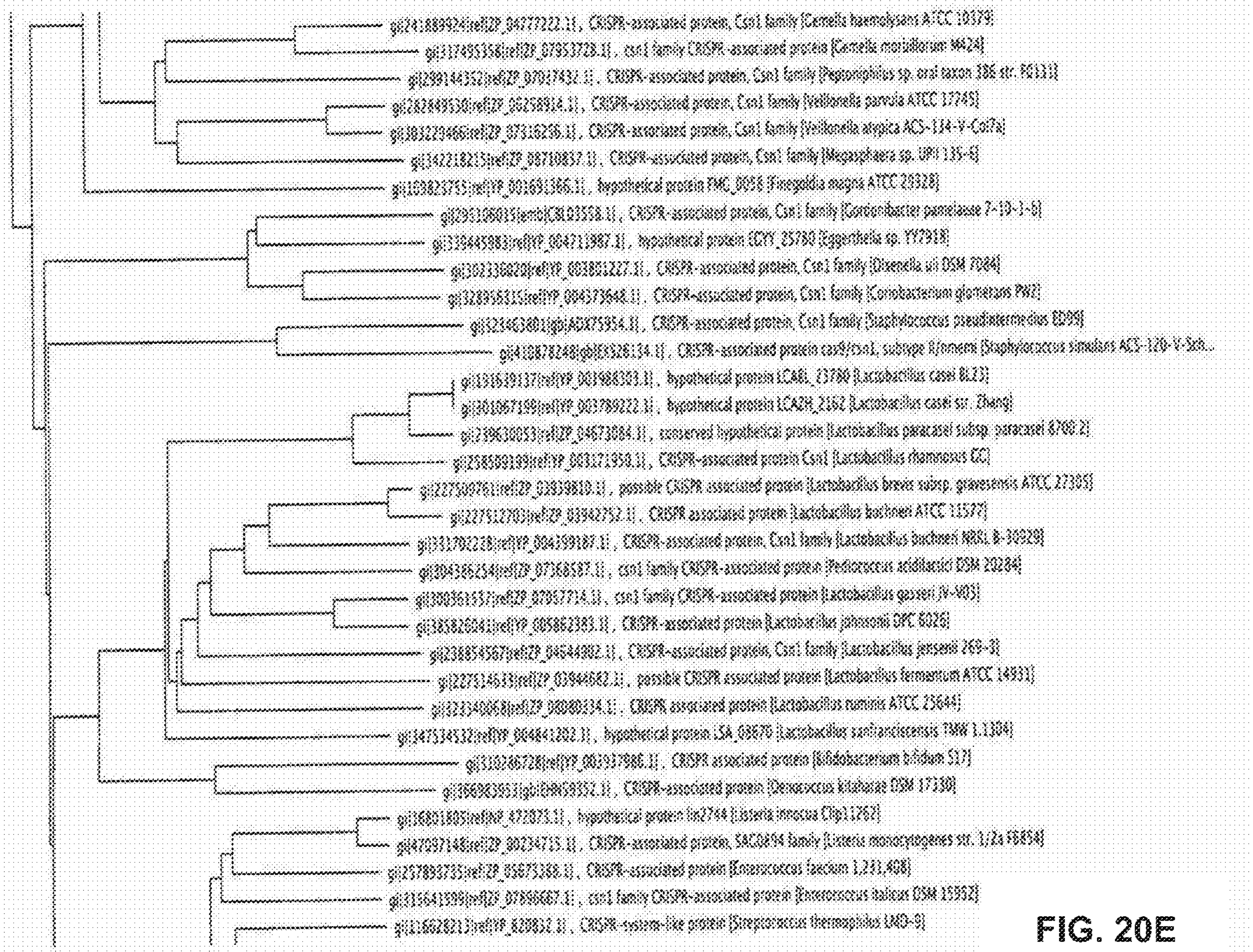

gi|241889924|ref|ZP_04777222.1| , CRISPR-associated protein, Csn1 family [Gemella haemolysans ATCC 10379]
gi|317495358|ref|ZP_07953728.1| , csn1 family CRISPR-associated protein [Gemella morbillorum M424]
gi|299144352|ref|ZP_07037432.1| , CRISPR-associated protein, Csn1 family [Peptoniphilus sp. oral taxon 386 str. F0131]
gi|282849530|ref|ZP_06258914.1| , CRISPR-associated protein, Csn1 family [Veillonella parvula ATCC 17745]
gi|303229466|ref|ZP_07316256.1| , CRISPR-associated protein, Csn1 family [Veillonella atypica ACS-134-V-Col7a]
gi|342218215|ref|ZP_08710837.1| , CRISPR-associated protein, Csn1 family [Megasphaera sp. UPII 135-E]
gi|169823755|ref|YP_001691366.1| , hypothetical protein FMG_0058 [Finegoldia magna ATCC 29328]
gi|295106015|emb|CBL03558.1| , CRISPR-associated protein, Csn1 family [Gordonibacter pamelaeae 7-10-1-b]
gi|319445983|ref|YP_004711987.1| , hypothetical protein EGYY_25780 [Eggerthella sp. YY7918]
gi|302336020|ref|YP_003801227.1| , CRISPR-associated protein, Csn1 family [Olsenella uli DSM 7084]
gi|328956315|ref|YP_004373648.1| , CRISPR-associated protein, Csn1 family [Coriobacterium glomerans PW2]
gi|323463801|gb|ADX75954.1| , CRISPR-associated protein, Csn1 family [Staphylococcus pseudintermedius ED99]
gi|410878248|gb|EKS26134.1| , CRISPR-associated protein cas9/csn1, subtype II/nmemi [Staphylococcus simulans ACS-120-V-Sch...
gi|191639137|ref|YP_001988303.1| , hypothetical protein LCABL_23780 [Lactobacillus casei BL23]
gi|301067199|ref|YP_003789222.1| , hypothetical protein LCAZH_2162 [Lactobacillus casei str. Zhang]
gi|239630053|ref|ZP_04673084.1| , conserved hypothetical protein [Lactobacillus paracasei subsp. paracasei 8700:2]
gi|258509199|ref|YP_003171950.1| , CRISPR-associated protein Csn1 [Lactobacillus rhamnosus GG]
gi|227509076|ref|ZP_03939810.1| , possible CRISPR associated protein [Lactobacillus brevis subsp. gravesensis ATCC 27305]
gi|227512703|ref|ZP_03942752.1| , CRISPR associated protein [Lactobacillus buchneri ATCC 11577]
gi|331702228|ref|YP_004399187.1| , CRISPR-associated protein, Csn1 family [Lactobacillus buchneri NRRL B-30929]
gi|304386254|ref|ZP_07368587.1| , csn1 family CRISPR-associated protein [Pediococcus acidilactici DSM 20284]
gi|300361537|ref|ZP_07057714.1| , csn1 family CRISPR-associated protein [Lactobacillus gasseri JV-V03]
gi|385826041|ref|YP_005862383.1| , CRISPR-associated protein [Lactobacillus johnsonii DPC 6026]
gi|238854567|ref|ZP_04644902.1| , CRISPR-associated protein, Csn1 family [Lactobacillus jensenii 269-3]
gi|227514633|ref|ZP_03944682.1| , possible CRISPR associated protein [Lactobacillus fermentum ATCC 14931]
gi|323340068|ref|ZP_08080334.1| , CRISPR associated protein [Lactobacillus ruminis ATCC 25644]
gi|347534532|ref|YP_004841202.1| , hypothetical protein LSA_08670 [Lactobacillus sanfranciscensis TMW 1.1304]
gi|310286728|ref|YP_003937986.1| , CRISPR associated protein [Bifidobacterium bifidum S17]
gi|366983953|gb|EHN59352.1| , CRISPR-associated protein [Oenococcus kitaharae DSM 17330]
gi|16801805|ref|NP_472073.1| , hypothetical protein lin2744 [Listeria innocua Clip11262]
gi|47097148|ref|ZP_00234715.1| , CRISPR-associated protein, SAG0894 family [Listeria monocytogenes str. 1/2a F6854]
gi|257893735|ref|ZP_05673388.1| , CRISPR-associated protein [Enterococcus faecium 1,231,408]
gi|315641599|ref|ZP_07896687.1| , csn1 family CRISPR-associated protein [Enterococcus italicus DSM 15952]
gi|116628213|ref|YP_820832.1| , CRISPR-system-like protein [Streptococcus thermophilus LMD-9]

FIG. 20E


gi|22537057|ref|NP_687908.1| , hypothetical protein SAG0894 [Streptococcus agalactiae 2603V/R]
gi|76788458|ref|YP_329839.1| , hypothetical protein SAK_1017 [Streptococcus agalactiae A909]
gi|25010965|ref|NP_735360.1| , hypothetical protein gbs0911 [Streptococcus agalactiae NEM316]
gi|24379809|ref|NP_721764.1| , hypothetical protein SMU_1405c [Streptococcus mutans UA159]
gi|290580220|ref|YP_003484612.1| , hypothetical protein SmuNN2025_0694 [Streptococcus mutans NN2025]
gi|306833855|ref|ZP_07466980.1| , csn1 family CRISPR-associated protein [Streptococcus bovis ATCC 700338]
gi|325978669|ref|YP_004288385.1| , CRISPR-associated protein [Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069]
gi|288905639|ref|YP_003430861.1| , CRISPR-associated protein [Streptococcus gallolyticus UCN34]
gi|320547102|ref|ZP_08041398.1| , csn1 family CRISPR-associated protein [Streptococcus equinus ATCC 9812]
gi|315223162|ref|ZP_07865023.1| , CRISPR-associated protein, Csn1 family [Streptococcus anginosus F0211]
gi|357636406|ref|ZP_09134281.1| , CRISPR-associated protein Cas9/Csn1, subtype II/NMEMI [Streptococcus macacae NCTC 11558]
gi|313890160|ref|ZP_07823795.1| , CRISPR-associated protein, Csn1 family [Streptococcus pseudoporcinus SPIN 20026]
gi|195978435|ref|YP_002123679.1| CRISPR-Associated Protein Csn1 [Streptococcus equi subsp. zooepidemicus MGCS10565]
gi|94994317|ref|YP_602415.1| , hypothetical protein MGAS10750_Spy0921 [Streptococcus pyogenes MGAS10750]
gi|251782637|ref|YP_002996940.1| , hypothetical protein SDEG_1231 [Streptococcus dysgalactiae subsp. equisimilis GGS_124]
gi|21910213|ref|NP_664481.1| , hypothetical protein SpyM3_0677 [Streptococcus pyogenes MGAS315]
gi|28896088|ref|NP_802438.1| , hypothetical protein SPs1176 [Streptococcus pyogenes SSI-1]
gi|209559356|ref|YP_002285828.1| , hypothetical protein Spy49_0823 [Streptococcus pyogenes NZ131]
gi|15675041|ref|NP_269215.1| , hypothetical protein SPy_1046 [Streptococcus pyogenes M1 GAS]
gi|71910582|ref|YP_282132.1| , hypothetical protein M5005_Spy_0769 [Streptococcus pyogenes MGAS5005]
gi|94990395|ref|YP_598495.1| , putative cytoplasmic protein [Streptococcus pyogenes MGAS10270]
gi|94988516|ref|YP_596617.1| , cytoplasmic protein [Streptococcus pyogenes MGAS9429]
gi|94992340|ref|YP_600439.1| , putative cytoplasmic protein [Streptococcus pyogenes MGAS2096]
gi|71903413|ref|YP_280216.1| , cytoplasmic protein [Streptococcus pyogenes MGAS6180]

FIG. 20F

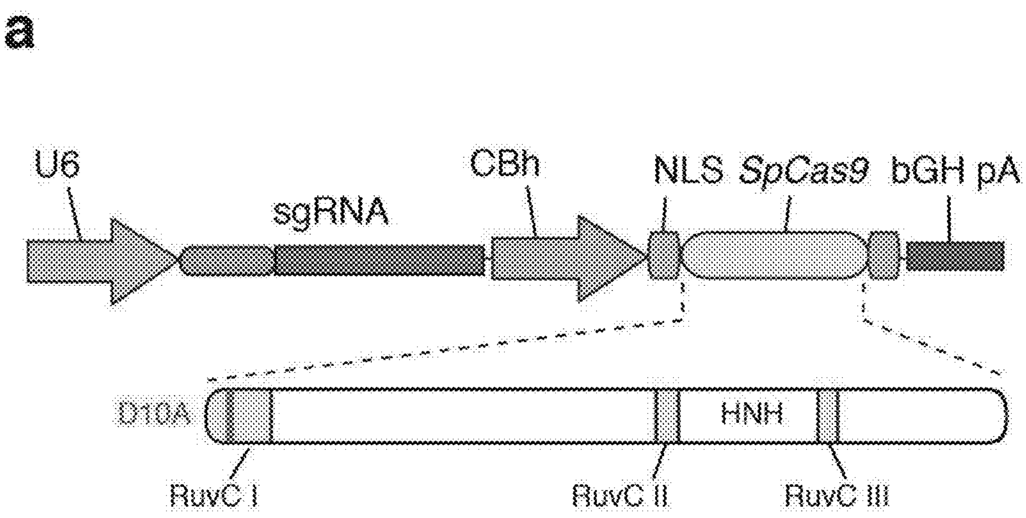

a
U6
sgRNA
CBh
NLS
SpCas9
bGH pA
D10A
HNH
RuvC I
RuvC II
RuvC III

FIG. 21A

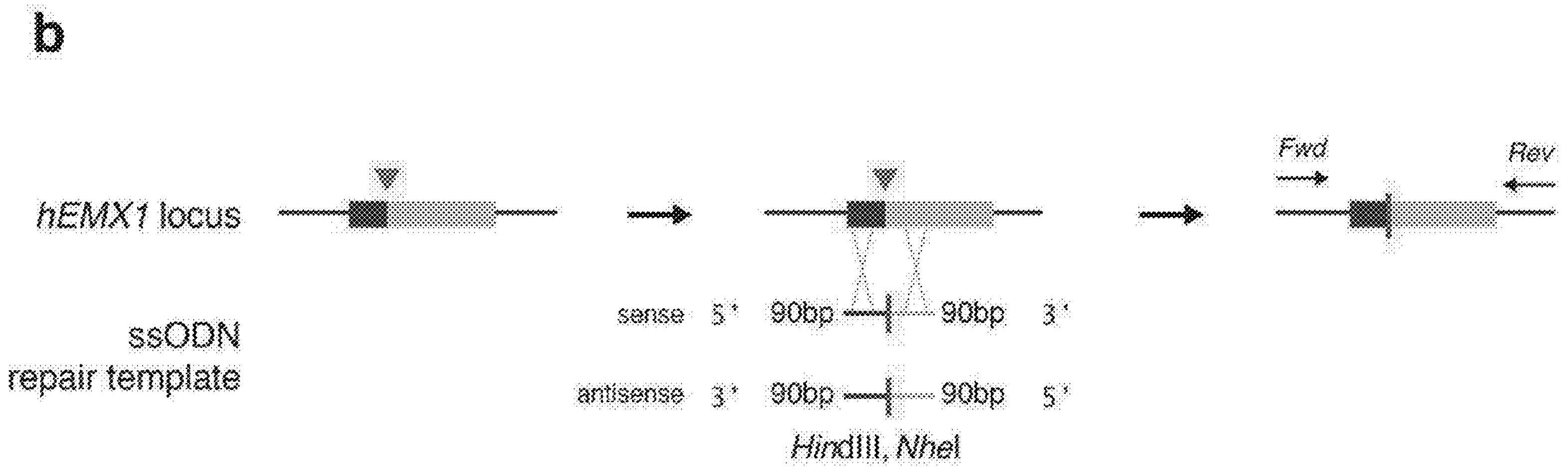

b
hEMX1 locus
ssODN
repair template
sense 5' 90bp 90bp 3'
antisense 3' 90bp 90bp 5'
*Hin*dIII, *Nhe*I
Fwd
Rev

FIG. 21B

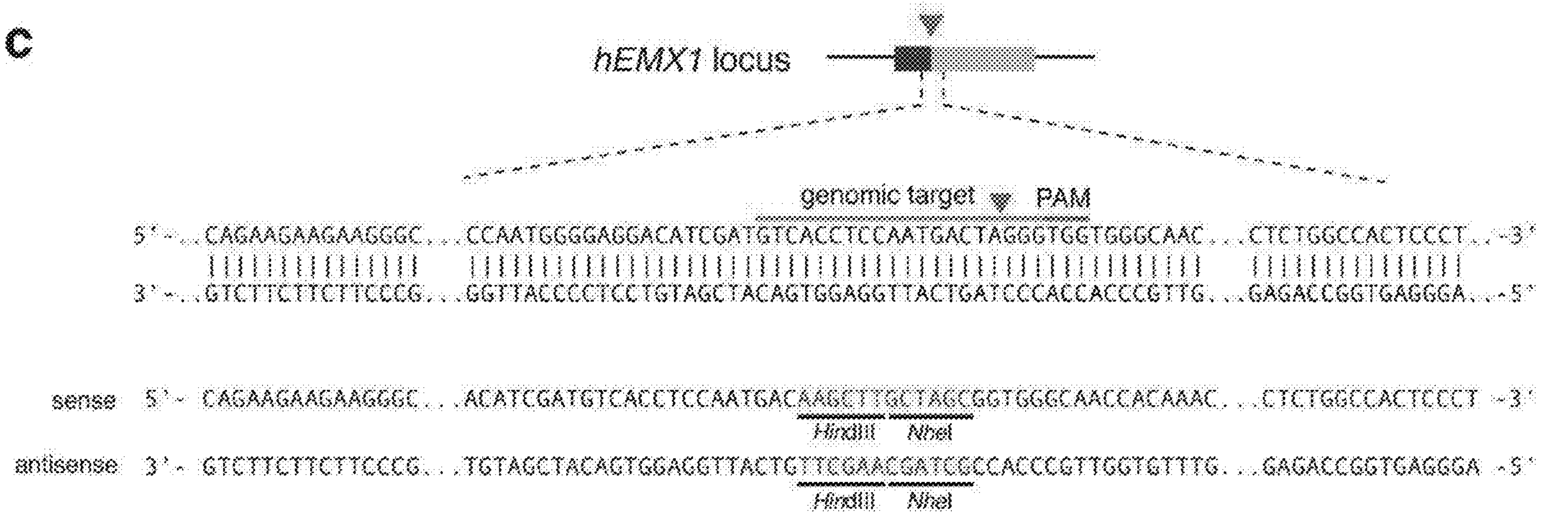

C
hEMX1 locus
genomic target
PAM
5'-...CAGAAGAAGAAGGGC...CCAATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGTGGGCAAC...CTCTGGCCACTCCCT..-3'
3'-...GTCTTCTTCTTCCCG...GGTTACCCCTCCTGTAGCTACAGTGGAGGTTACTGATCCCACCACCCGTTG...GAGACCGGTGAGGGA..-5'
sense 5'- CAGAAGAAGAAGGGC...ACATCGATGTCACCTCCAATGACAAGCTTGCTAGCGGTGGGCAACCACAAAC...CTCTGGCCACTCCCT -3'
HindIII
NheI
antisense 3'- GTCTTCTTCTTCCCG...TGTAGCTACAGTGGAGGTTACTGTTCGAACGATCGCCACCCGTTGGTGTTTG...GAGACCGGTGAGGGA -5'
HindIII
NheI

FIG. 21C

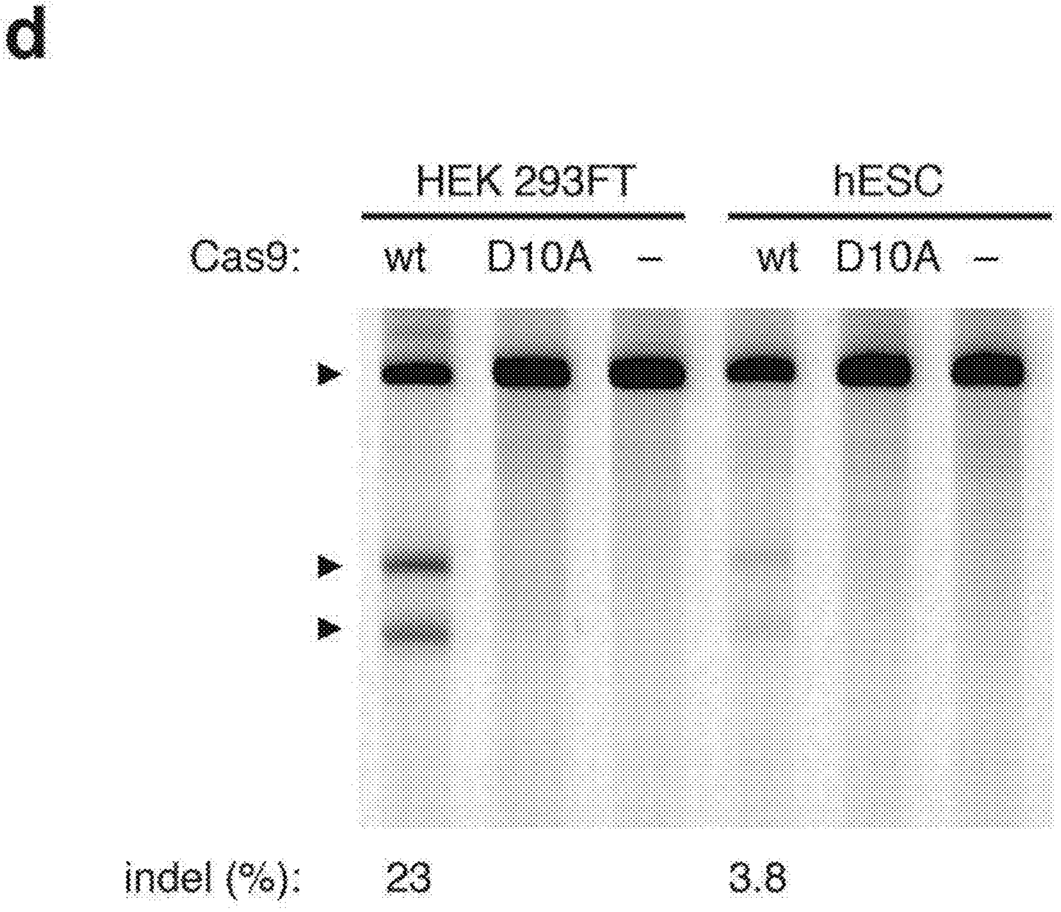


FIG. 21D

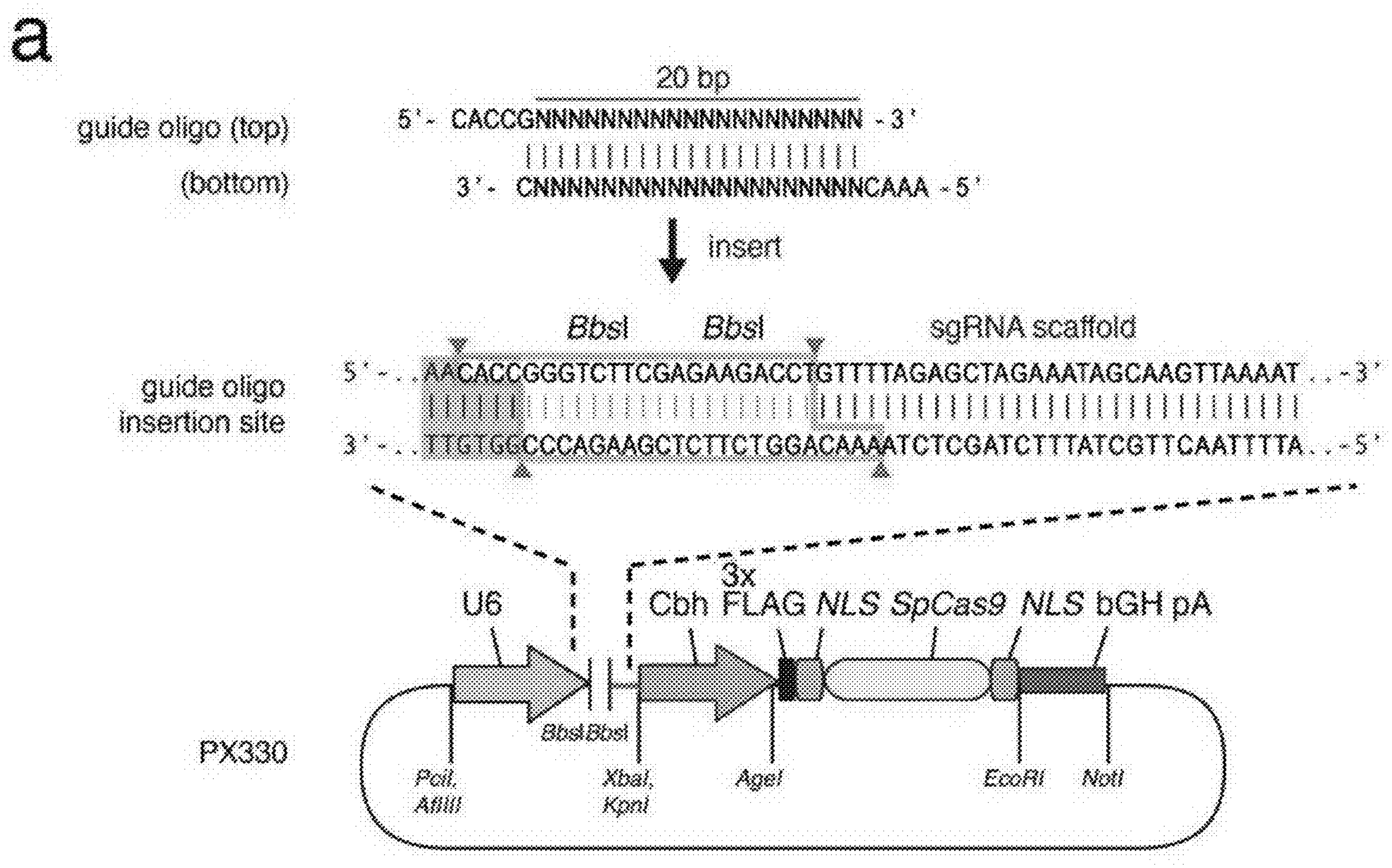

a
20 bp
guide oligo (top)
5'- CACCGNNNNNNNNNNNNNNNNNNNN -3'
(bottom)
3'- CNNNNNNNNNNNNNNNNNNNNCAAA -5'
insert
BbsI
BbsI
sgRNA scaffold
guide oligo insertion site
5'-..AACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAAT...-3'
3'-..TTGTGGCCCAGAAGCTCTTCTGGACAAAATCTCGATCTTTATCGTTCAATTTTA...-5'
U6
Cbh
3x FLAG
NLS
SpCas9
NLS
bGH pA
PX330
BbsI BbsI
PciI, AflIII
XbaI, KpnI
AgeI
EcoRI
NotI

FIG. 22A

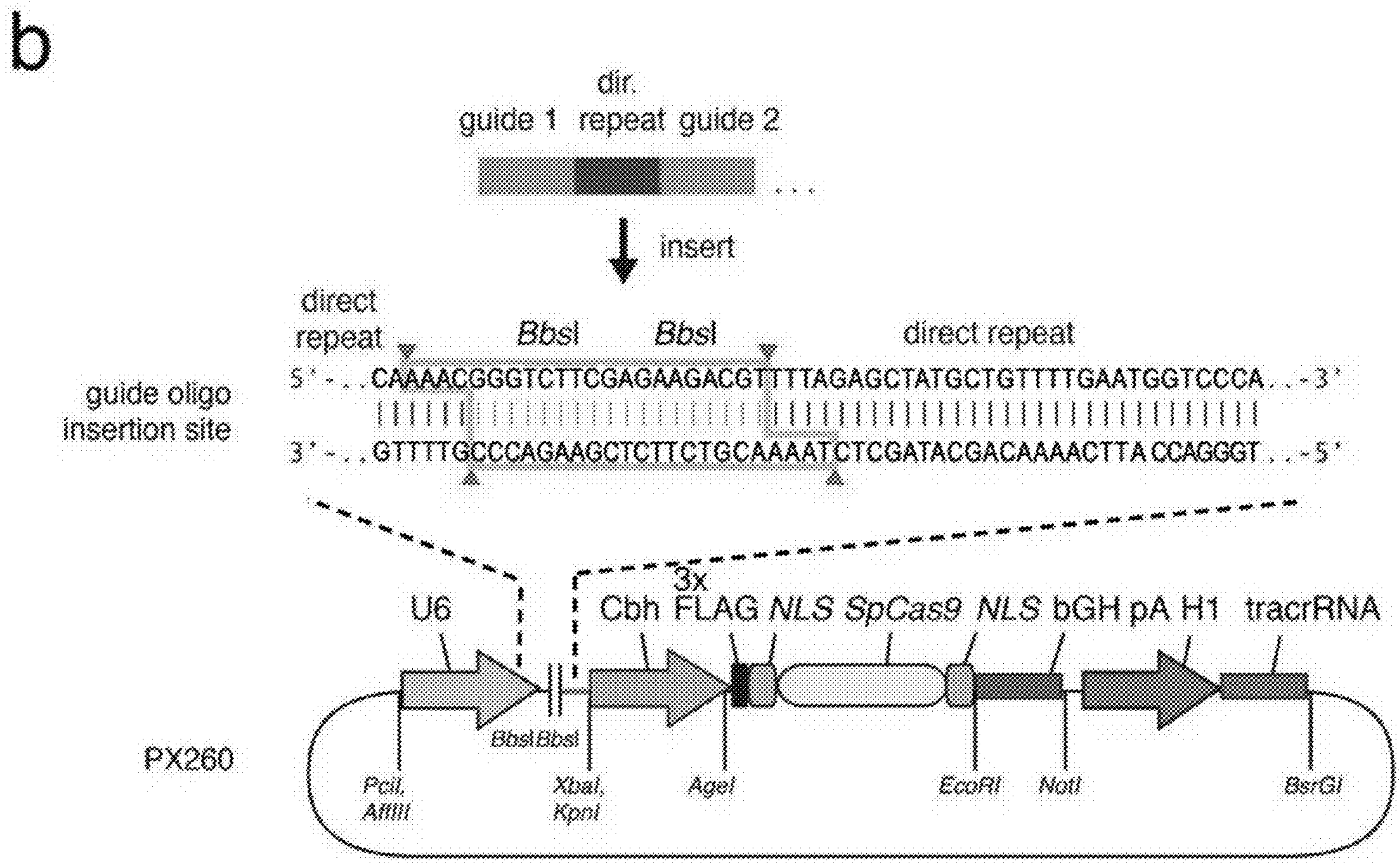

b
guide 1
dir. repeat
guide 2
...
insert
direct repeat
BbsI
BbsI
direct repeat
guide oligo insertion site
5'-..CAAAACGGGTCTTCGAGAAGACGTTTTAGAGCTATGCTGTTTTGAATGGTCCCA..-3'
3'-..GTTTTGCCCAGAAGCTCTTCTGCAAAATCTCGATACGACAAAACTTACCAGGGT..-5'
U6
Cbh
3x FLAG
NLS
SpCas9
NLS
bGH pA
H1
tracrRNA
PX260
PciI, AflIII
BbsI BbsI
XbaI, KpnI
AgeI
EcoRI
NotI
BsrGI

FIG. 22B

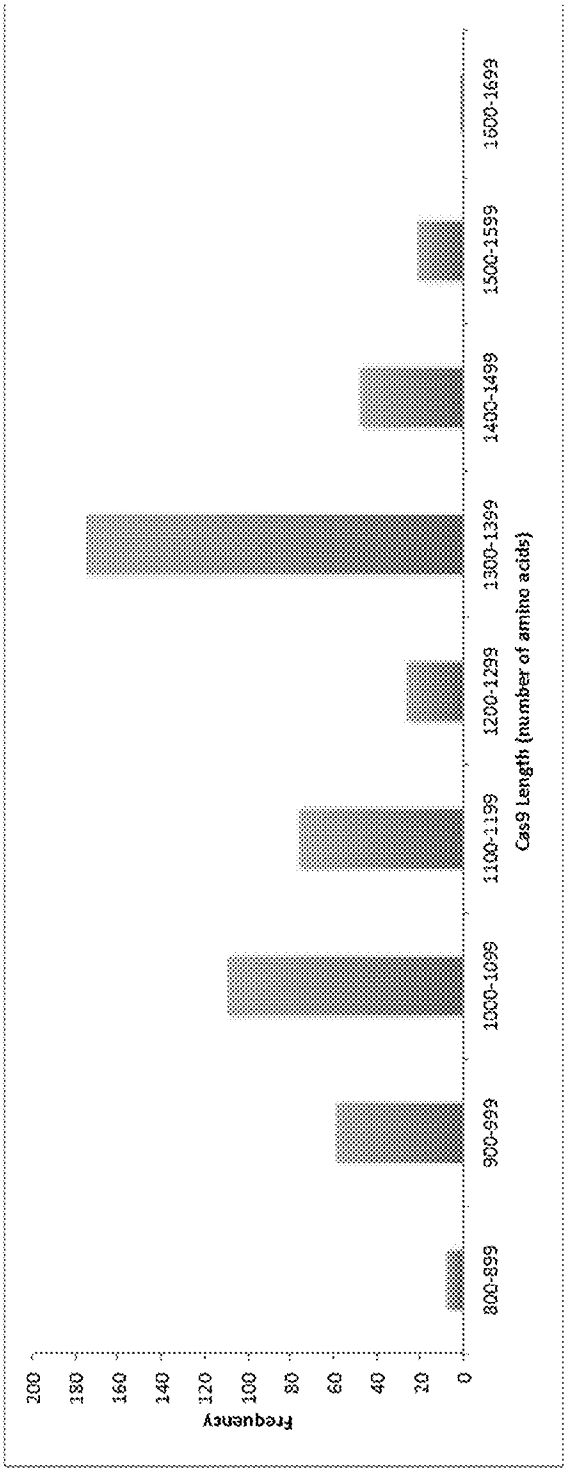

Frequency
200
180
160
140
120
100
80
60
40
20
0
800-899
900-999
1000-1099
1100-1199
1200-1299
1300-1399
1400-1499
1500-1599
1600-1699
Cas9 Length (number of amino acids)

FIG. 23

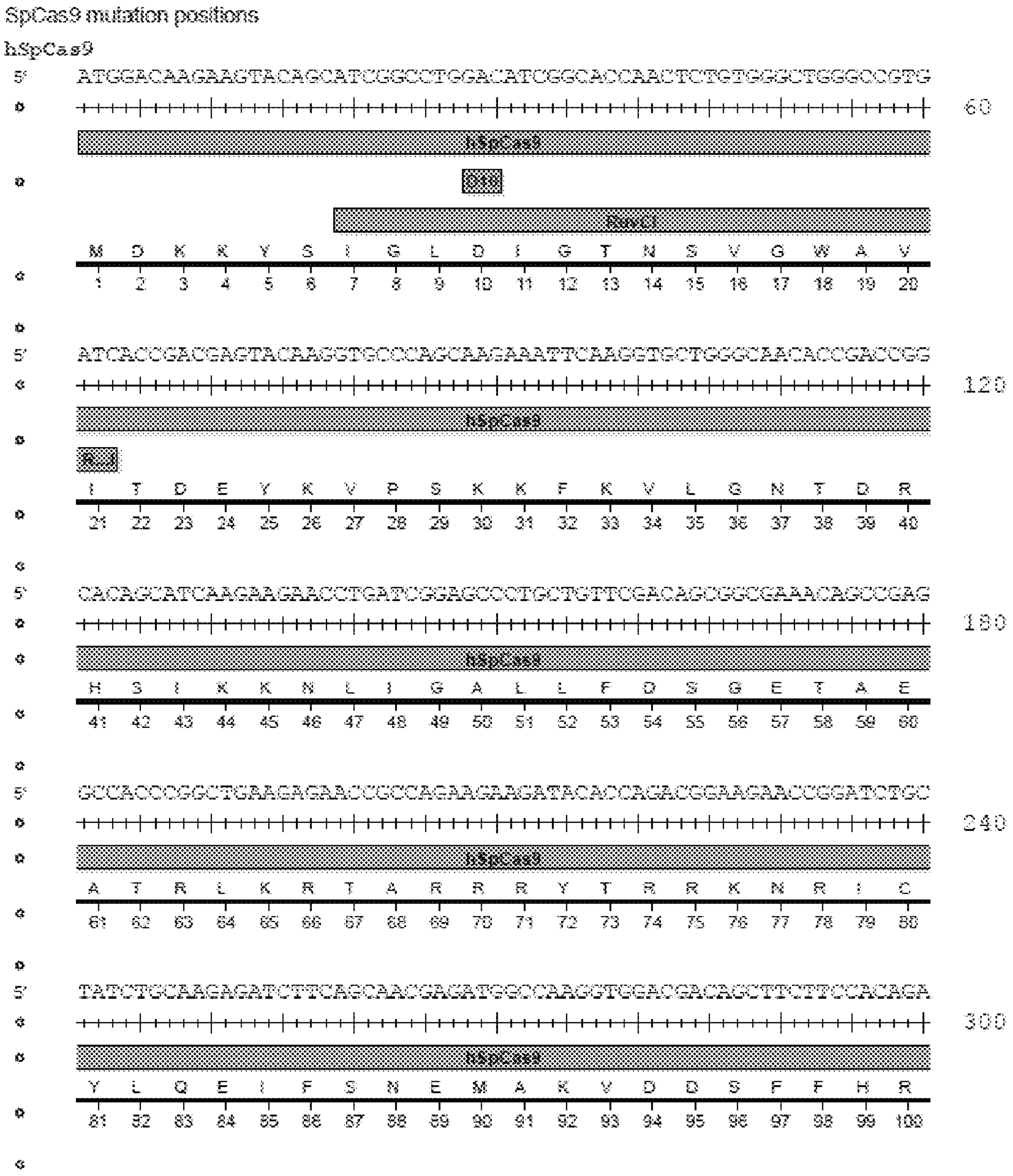

SpCas9 mutation positions
hSpCas9
5'
ATGGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
60
hSpCas9
D10
RuvCI
M D K K Y S I G L D I G T N S V G W A V
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
5'
ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGG
120
hSpCas9
R...I
I T D E Y K V P S K K F K V L G N T D R
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
5'
CACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAG
180
hSpCas9
H S I K K N L I G A L L F D S G E T A E
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
5'
GCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGC
240
hSpCas9
A T R L K R T A R R R Y T R R K N R I C
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
5'
TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
300
hSpCas9
Y L Q E I F S N E M A K V D D S F F H R
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

FIG. 24A

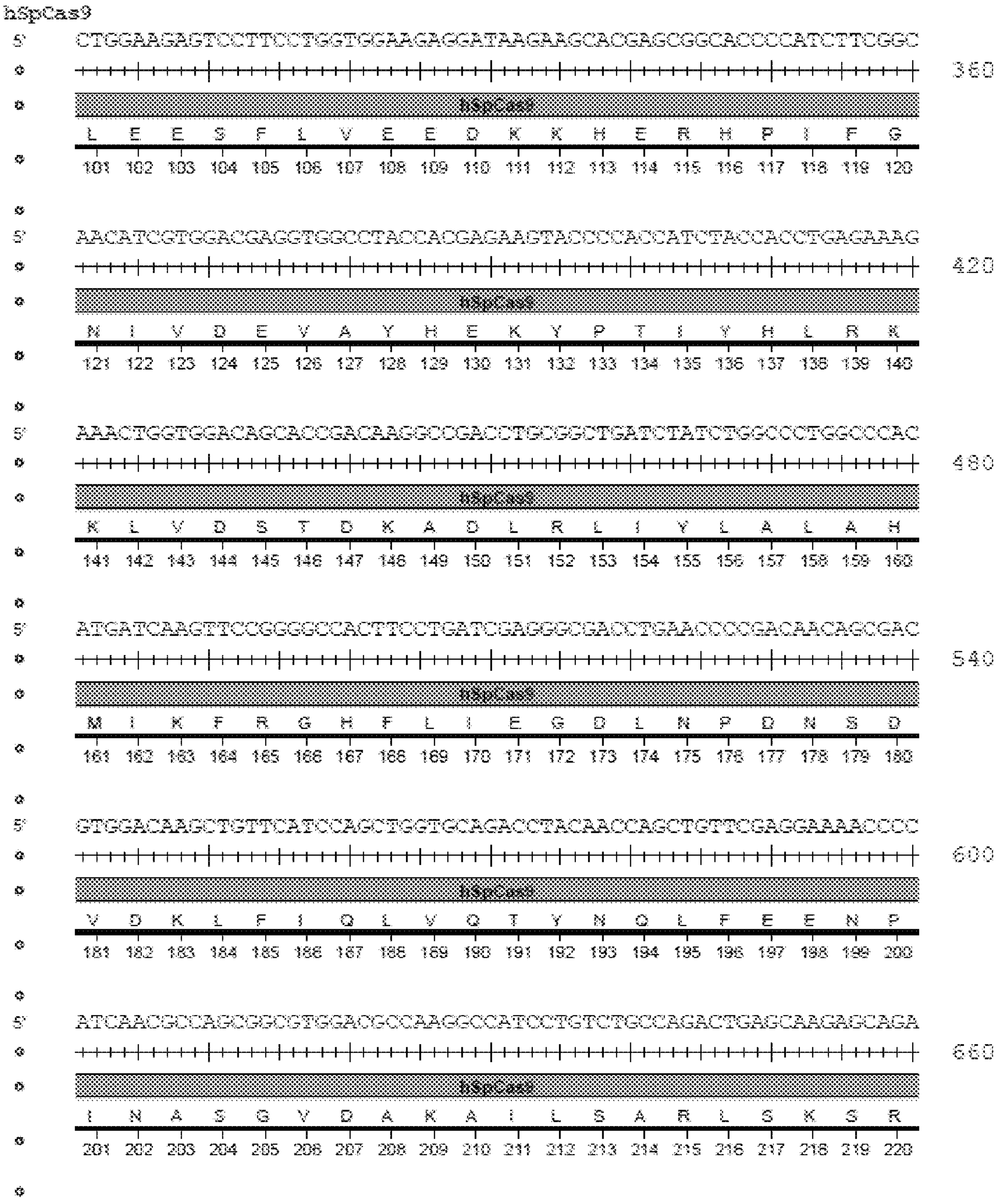

hSpCas9
5' CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGC 360
hSpCas9
L E E S F L V E E D K K H E R H P I F G
101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
5' AACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAG 420
hSpCas9
N I V D E V A Y H E K Y P T I Y H L R K
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
5' AAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCAC 480
hSpCas9
K L V D S T D K A D L R L I Y L A L A H
141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
5' ATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC 540
hSpCas9
M I K F R G H F L I E G D L N P D N S D
161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
5' GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCC 600
hSpCas9
V D K L F I Q L V Q T Y N Q L F E E N P
181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
5' ATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGA 660
hSpCas9
I N A S G V D A K A I L S A R L S K S R
201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220

FIG. 24B

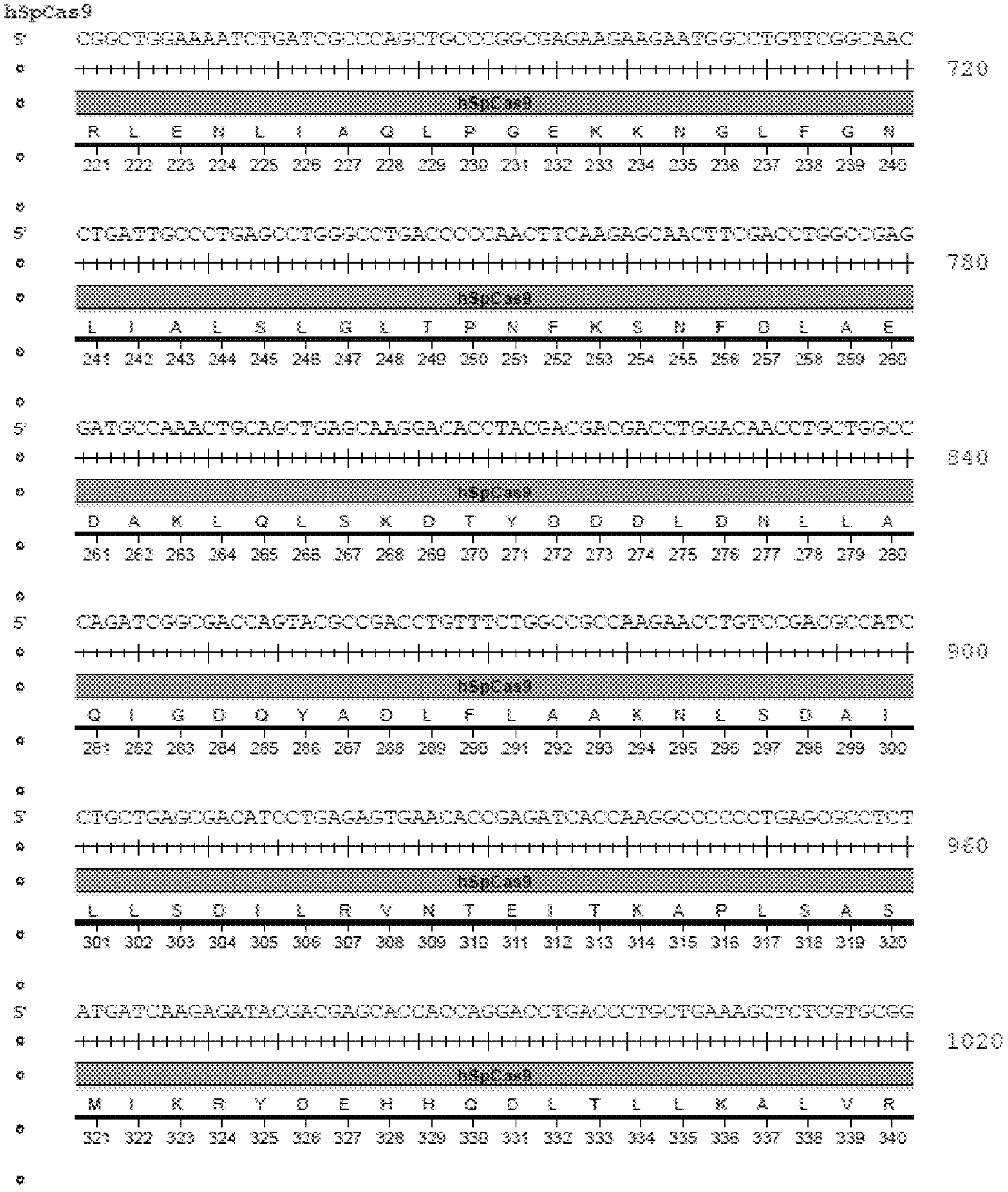

hSpCas9
5'
CGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAAC
720
hSpCas9
R L E N L I A Q L P G E K K N G L F G N
221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
5'
CTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAG
780
hSpCas9
L I A L S L G L T P N F K S N F D L A E
241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
5'
GATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCC
840
hSpCas9
D A K L Q L S K D T Y D D D L D N L L A
261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
5'
CAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATC
900
hSpCas9
Q I G D Q Y A D L F L A A K N L S D A I
281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
5'
CTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCT
960
hSpCas9
L L S D I L R V N T E I T K A P L S A S
301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
5'
ATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGG
1020
hSpCas9
M I K R Y D E H H Q D L T L L K A L V R
321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340

FIG. 24C

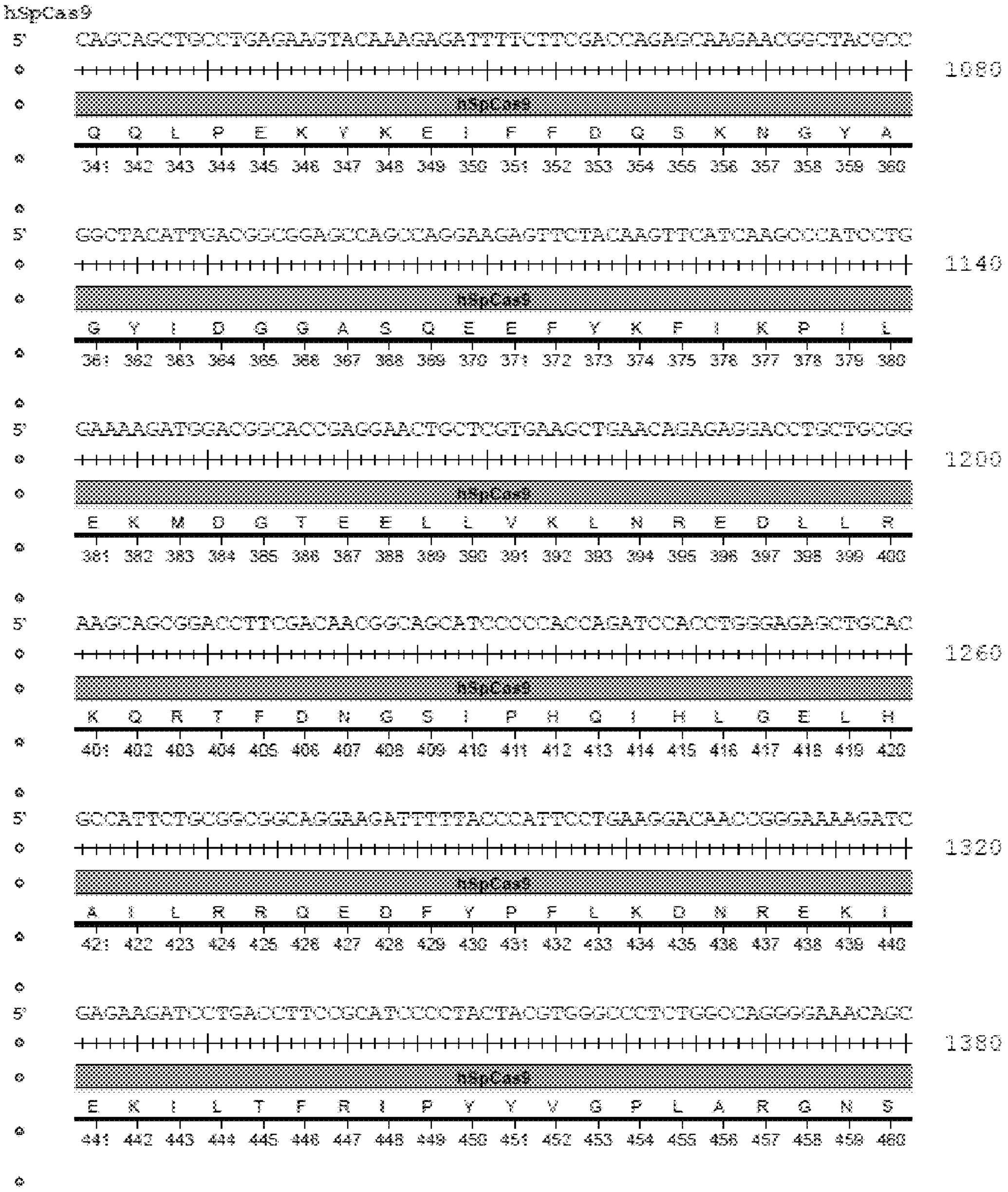

hSpCas9
5' CAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC 1080
hSpCas9
Q Q L P E K Y K E I F F D Q S K N G Y A
341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
5' GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTG 1140
hSpCas9
G Y I D G G A S Q E E F Y K F I K P I L
361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
5' GAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGG 1200
hSpCas9
E K M D G T E E L L V K L N R E D L L R
381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
5' AAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCAC 1260
hSpCas9
K Q R T F D N G S I P H Q I H L G E L H
401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
5' GCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATC 1320
hSpCas9
A I L R R Q E D F Y P F L K D N R E K I
421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
5' GAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGC 1380
hSpCas9
E K I L T F R I P Y Y V G P L A R G N S
441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460

FIG. 24D

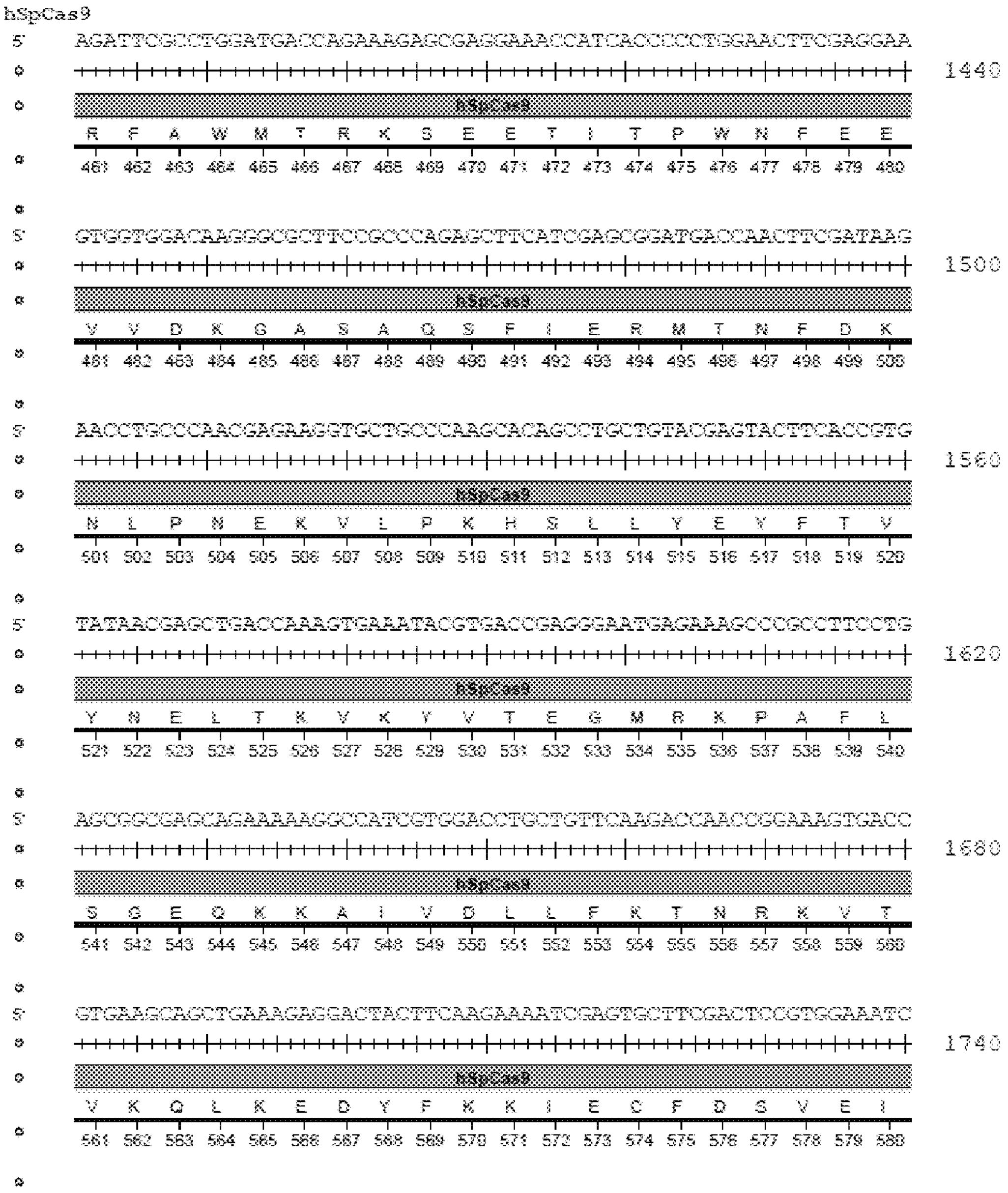

hSpCas9
5'
AGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAA
1440
hSpCas9
R F A W M T R K S E E T I T P W N F E E
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
5'
GTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
1500
hSpCas9
V V D K G A S A Q S F I E R M T N F D K
481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
5'
AACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTG
1560
hSpCas9
N L P N E K V L P K H S L L Y E Y F T V
501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520
5'
TATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTG
1620
hSpCas9
Y N E L T K V K Y V T E G M R K P A F L
521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
5'
AGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACC
1680
hSpCas9
S G E Q K K A I V D L L F K T N R K V T
541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560
5'
GTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATC
1740
hSpCas9
V K Q L K E D Y F K K I E C F D S V E I
561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580

FIG. 24E

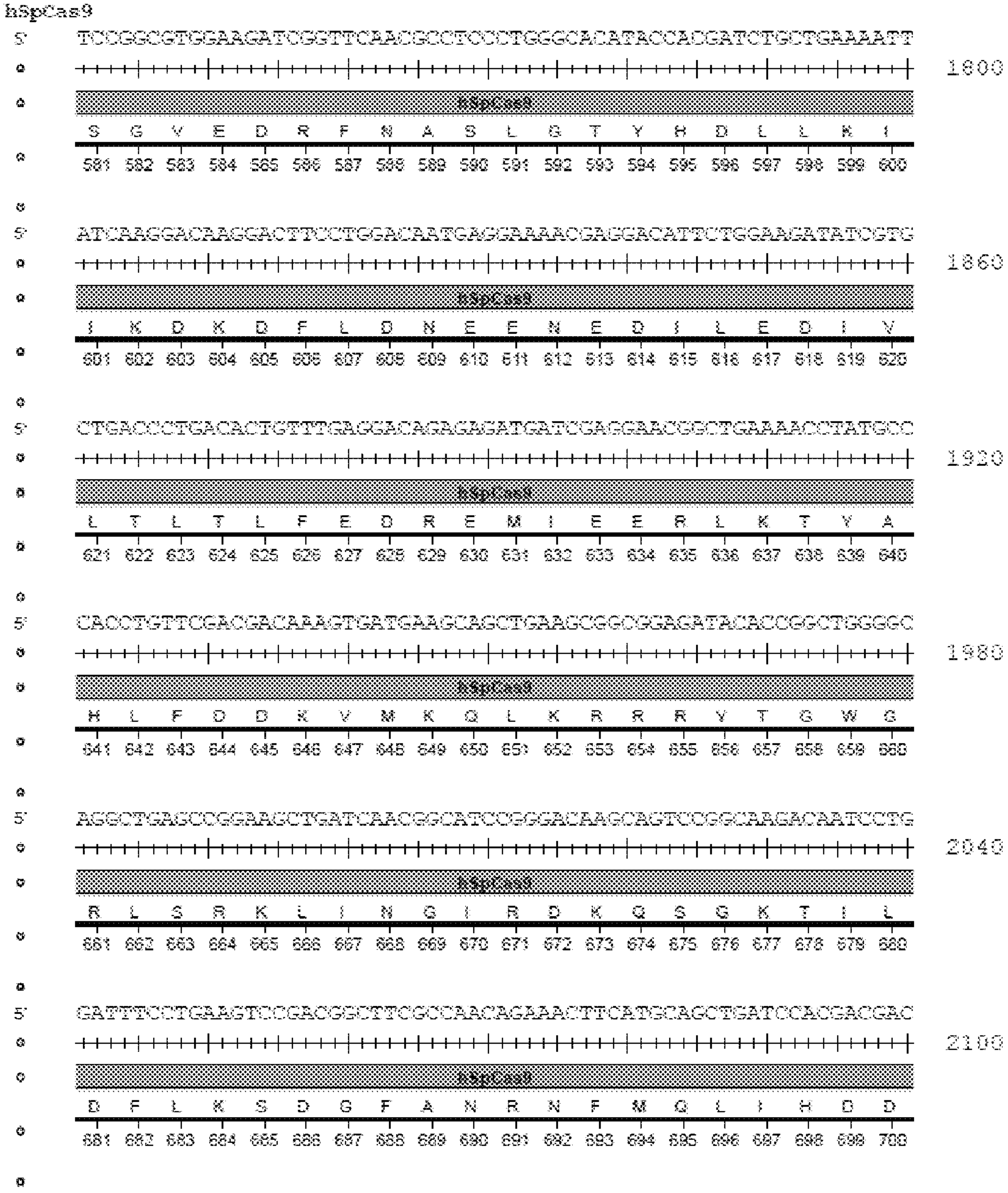

hSpCas9
5' TCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT 1800
hSpCas9
S G V E D R F N A S L G T Y H D L L K I
581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
5' ATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTG 1860
hSpCas9
I K D K D F L D N E E N E D I L E D I V
601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620
5' CTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCC 1920
hSpCas9
L T L T L F E D R E M I E E R L K T Y A
621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640
5' CACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGC 1980
hSpCas9
H L F D D K V M K Q L K R R R Y T G W G
641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
5' AGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTG 2040
hSpCas9
R L S R K L I N G I R D K Q S G K T I L
661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680
5' GATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGAC 2100
hSpCas9
D F L K S D G F A N R N F M Q L I H D D
681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700

FIG. 24F

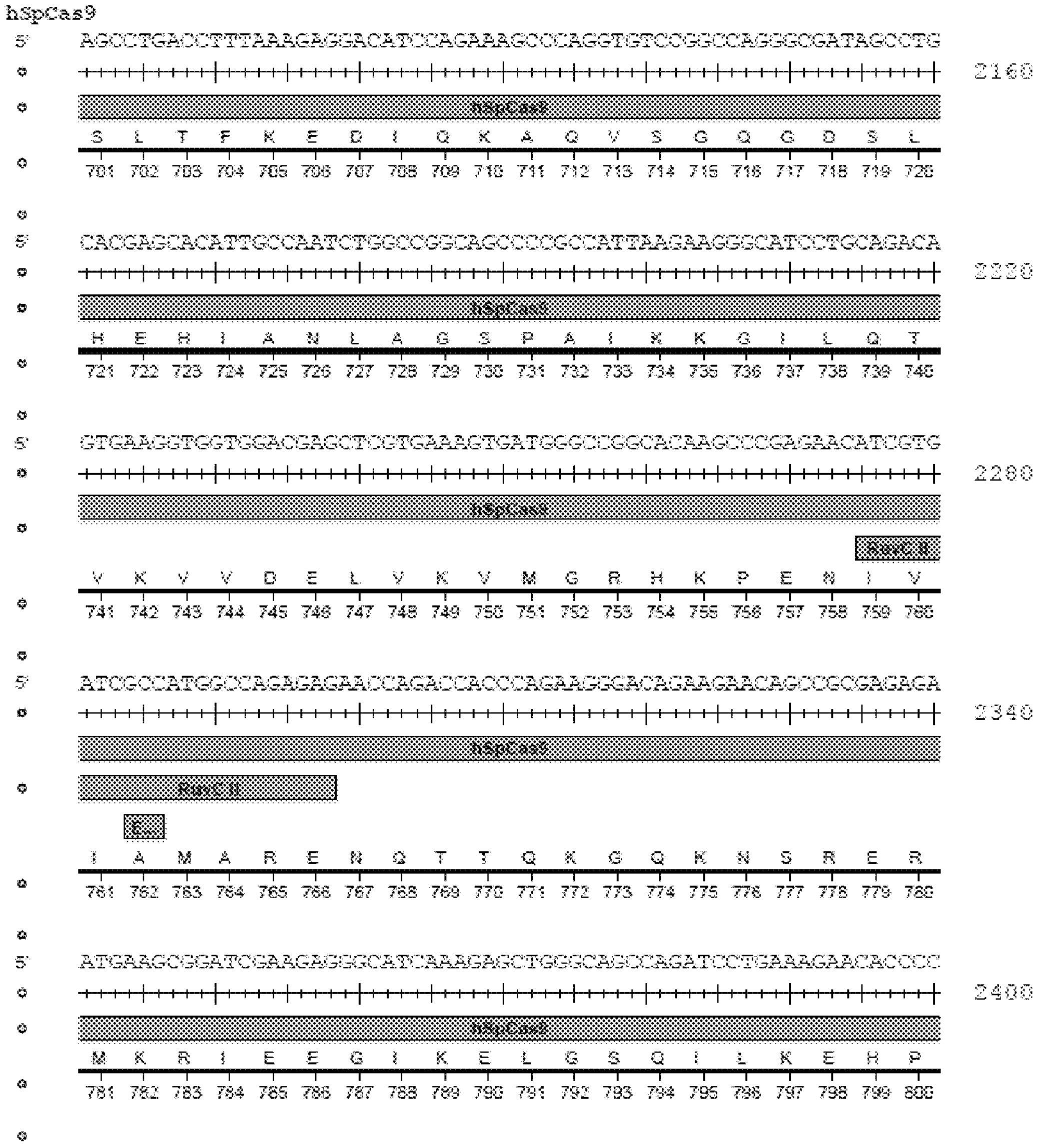

hSpCas9
5' AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
2160
hSpCas9
S L T F K E D I Q K A Q V S G Q G D S L
701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
5' CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACA
2220
hSpCas9
H E H I A N L A G S P A I K K G I L Q T
721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740
5' GTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTG
2280
hSpCas9
RuvC II
V K V V D E L V K V M G R H K P E N I V
741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
5' ATCGCCATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGA
2340
hSpCas9
RuvC II
E...
I A M A R E N Q T T Q K G Q K N S R E R
761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
5' ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCC
2400
hSpCas9
M K R I E E G I K E L G S Q I L K E H P
781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800

FIG. 24G

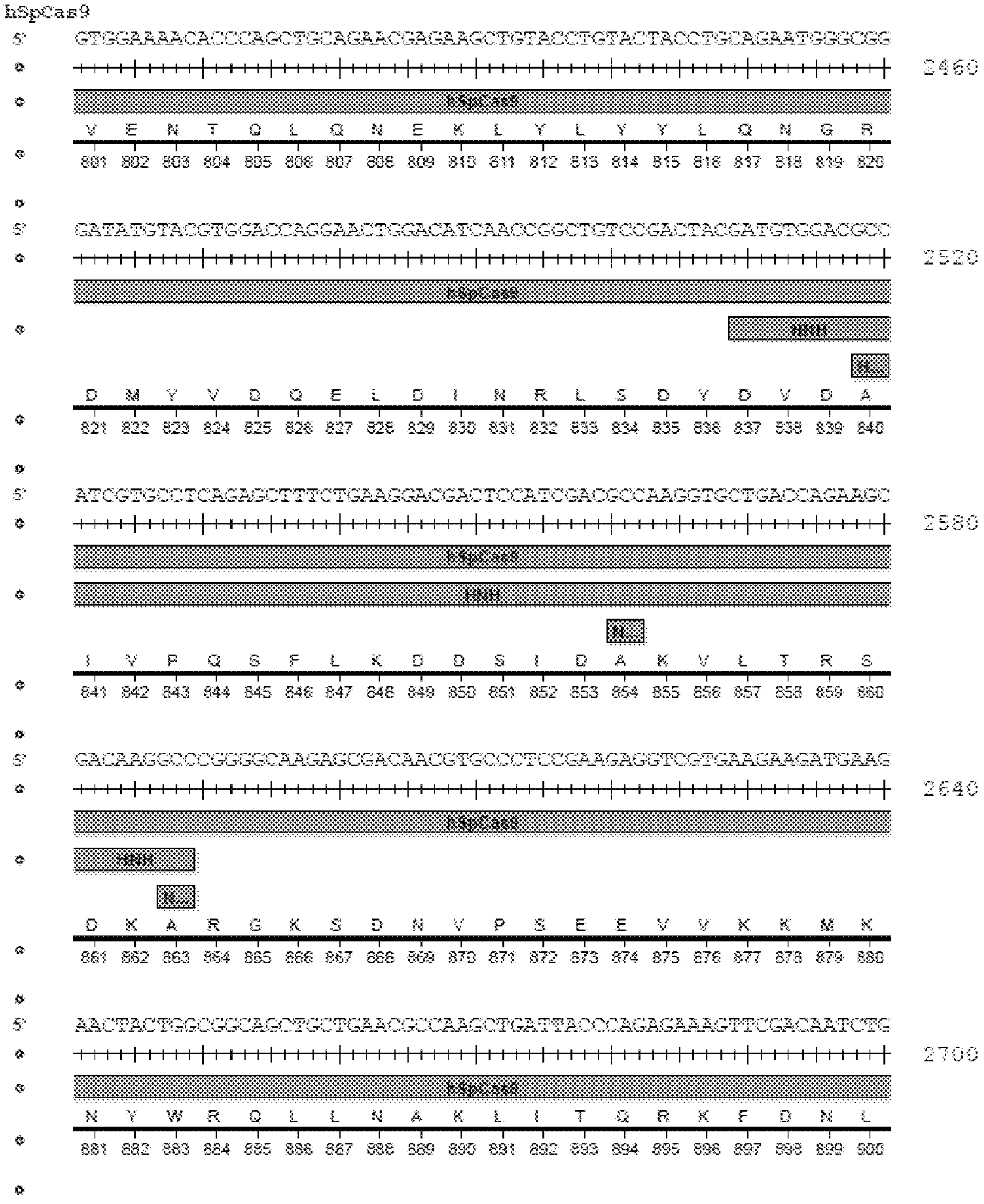

hSpCas9
5' GTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGG
2460
hSpCas9
V E N T Q L Q N E K L Y L Y Y L Q N G R
801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820
5' GATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCC
2520
hSpCas9
HNH
H...
D M Y V D Q E L D I N R L S D Y D V D A
821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
5' ATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACGCCAAGGTGCTGACCAGAAGC
2580
hSpCas9
HNH
N...
I V P Q S F L K D D S I D A K V L T R S
841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860
5' GACAAGGCCCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAG
2640
hSpCas9
HNH
N...
D K A R G K S D N V P S E E V V K K M K
861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880
5' AACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTG
2700
hSpCas9
N Y W R Q L L N A K L I T Q R K F D N L
881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900

FIG. 24H

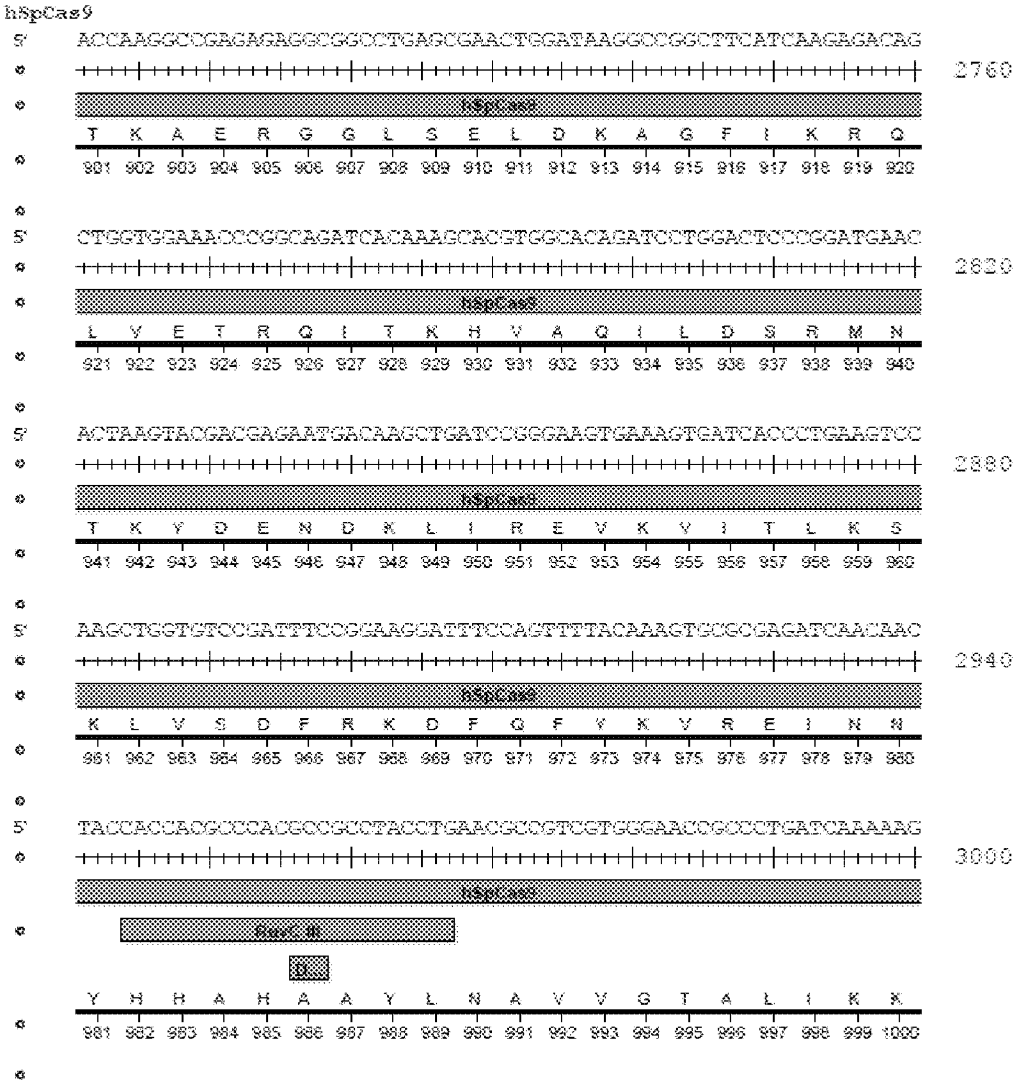

hSpCas9
5' ACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAG
2760
hSpCas9
T K A E R G G L S E L D K A G F I K R Q
901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920
5' CTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAAC
2820
hSpCas9
L V E T R Q I T K H V A Q I L D S R M N
921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940
5' ACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCC
2880
hSpCas9
T K Y D E N D K L I R E V K V I T L K S
941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
5' AAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAAC
2940
hSpCas9
K L V S D F R K D F Q F Y K V R E I N N
961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980
5' TACCACCACGCCCACGCCGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAG
3000
hSpCas9
RuvC III
D...
Y H H A H A A Y L N A V V G T A L I K K
981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

FIG. 24I

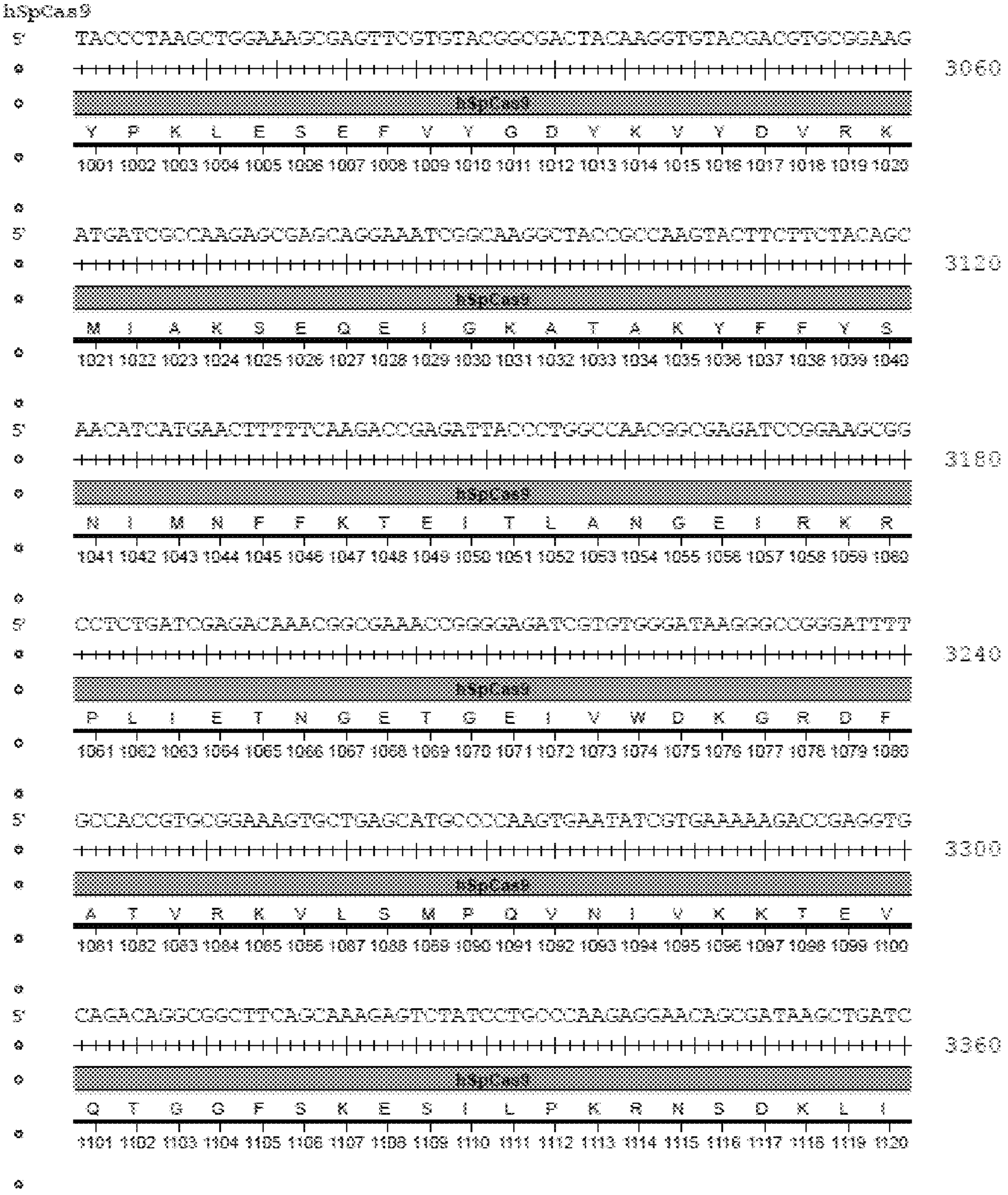

hSpCas9
5' TACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAG 3060
hSpCas9
Y P K L E S E F V Y G D Y K V Y D V R K
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020
5' ATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC 3120
hSpCas9
M I A K S E Q E I G K A T A K Y F F Y S
1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
5' AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGG 3180
hSpCas9
N I M N F F K T E I T L A N G E I R K R
1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060
5' CCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTT 3240
hSpCas9
P L I E T N G E T G E I V W D K G R D F
1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080
5' GCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTG 3300
hSpCas9
A T V R K V L S M P Q V N I V K K T E V
1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100
5' CAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATC 3360
hSpCas9
Q T G G F S K E S I L P K R N S D K L I
1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120

FIG. 24J

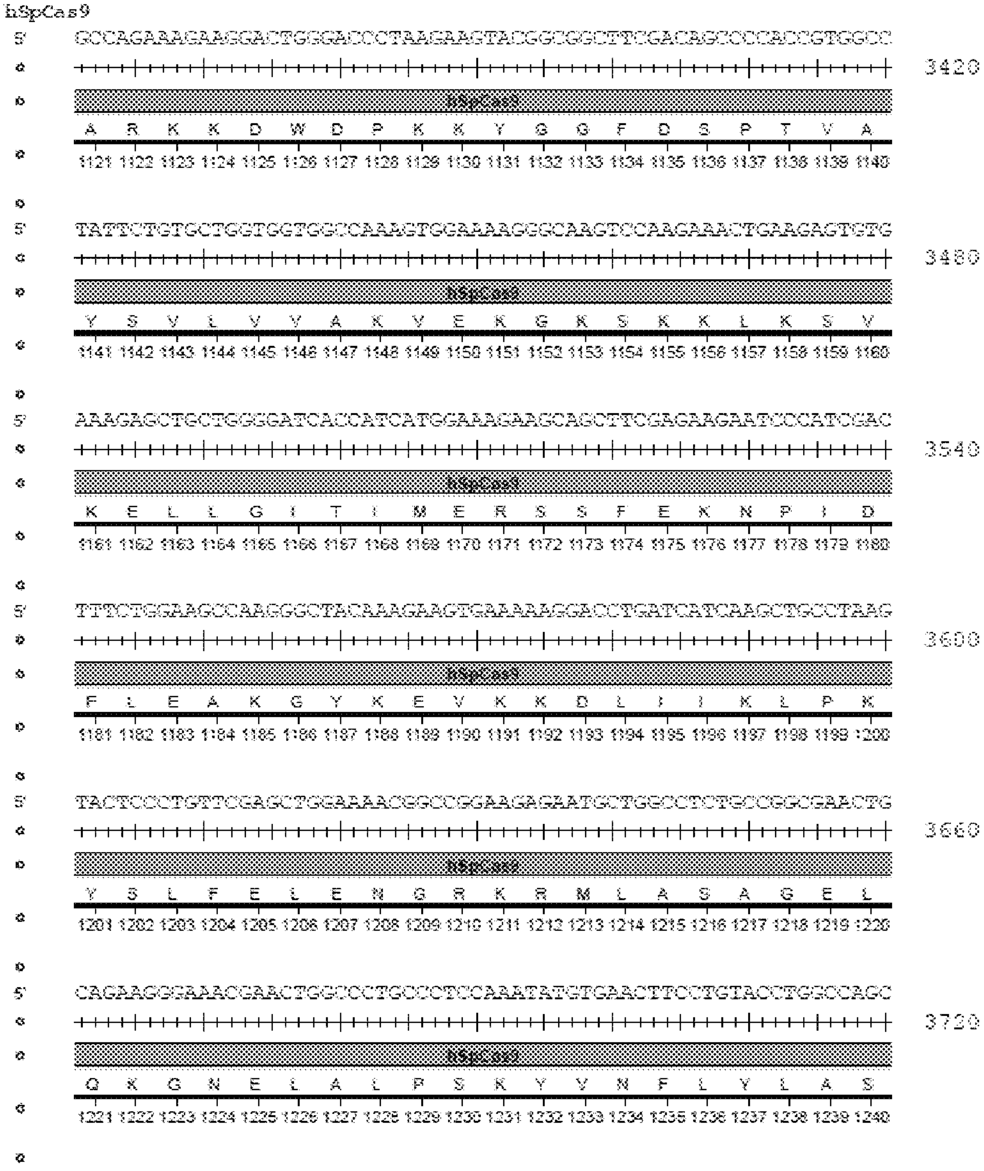

hSpCas9
5' GCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCC
3420
hSpCas9
A R K K D W D P K K Y G G F D S P T V A
1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140
5' TATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTG
3480
hSpCas9
Y S V L V V A K V E K G K S K K L K S V
1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160
5' AAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGAC
3540
hSpCas9
K E L L G I T I M E R S S F E K N P I D
1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180
5' TTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAG
3600
hSpCas9
F L E A K G Y K E V K K D L I I K L P K
1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
5' TACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTG
3660
hSpCas9
Y S L F E L E N G R K R M L A S A G E L
1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220
5' CAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGC
3720
hSpCas9
Q K G N E L A L P S K Y V N F L Y L A S
1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240

FIG. 24K

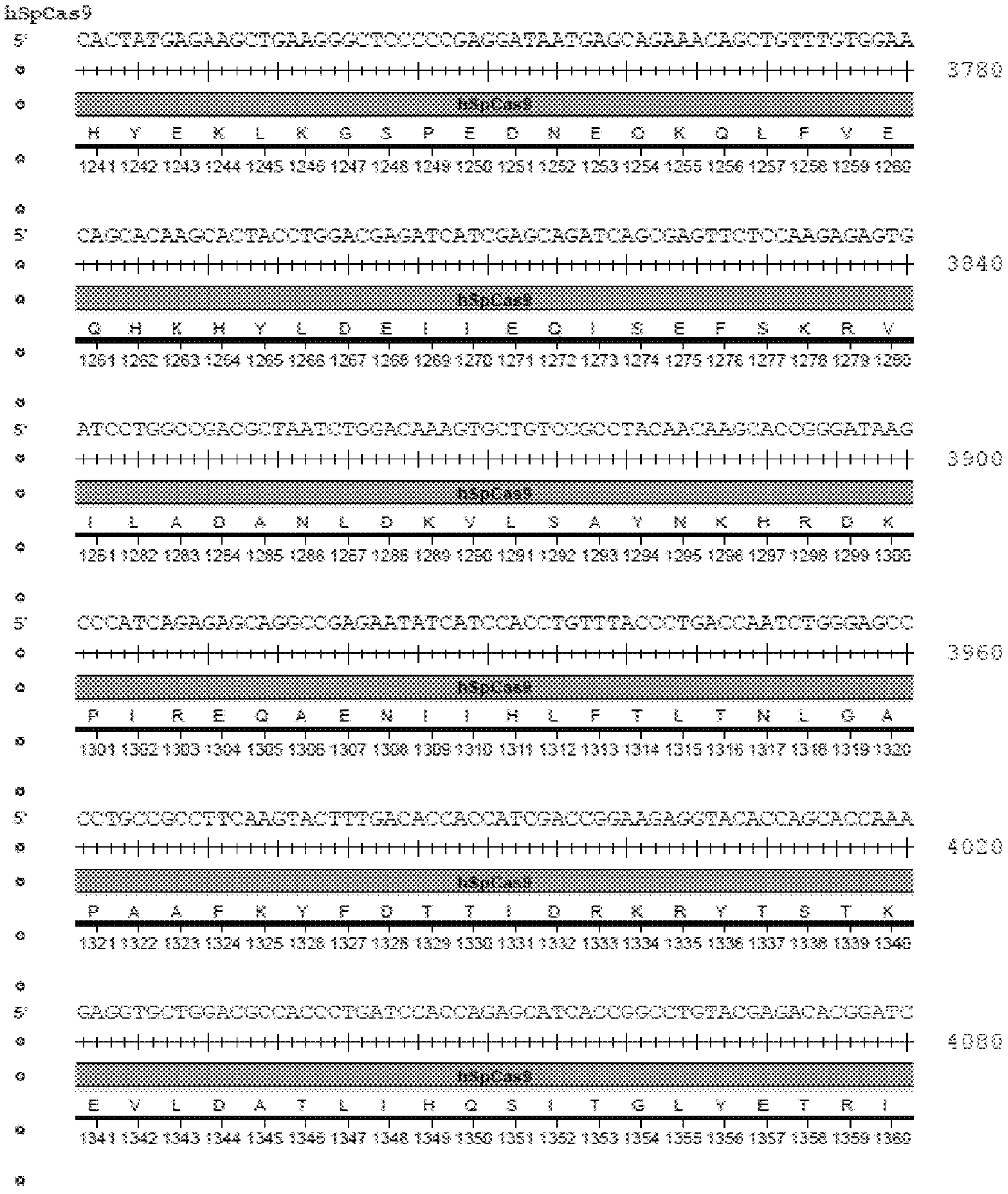

hSpCas9
5' CACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAA 3780
hSpCas9
H Y E K L K G S P E D N E Q K Q L F V E
1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260
5' CAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTG 3840
hSpCas9
Q H K H Y L D E I I E Q I S E F S K R V
1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280
5' ATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG 3900
hSpCas9
I L A D A N L D K V L S A Y N K H R D K
1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300
5' CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCC 3960
hSpCas9
P I R E Q A E N I I H L F T L T N L G A
1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320
5' CCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAA 4020
hSpCas9
P A A F K Y F D T T I D R K R Y T S T K
1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340
5' GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATC 4080
hSpCas9
E V L D A T L I H Q S I T G L Y E T R I
1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360

FIG. 24L

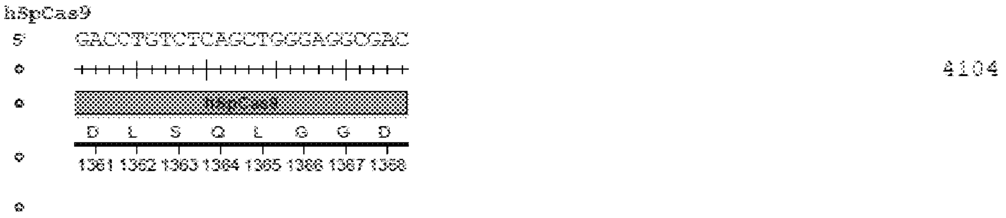

hSpCas9
5'
GACCTGTCTCAGCTGGGAGGCGAC
4104
hSpCas9
D L S Q L G G D
1361 1362 1363 1364 1365 1366 1367 1368

FIG. 24M

Conditional Cas9, Rosa26 targeting vector map

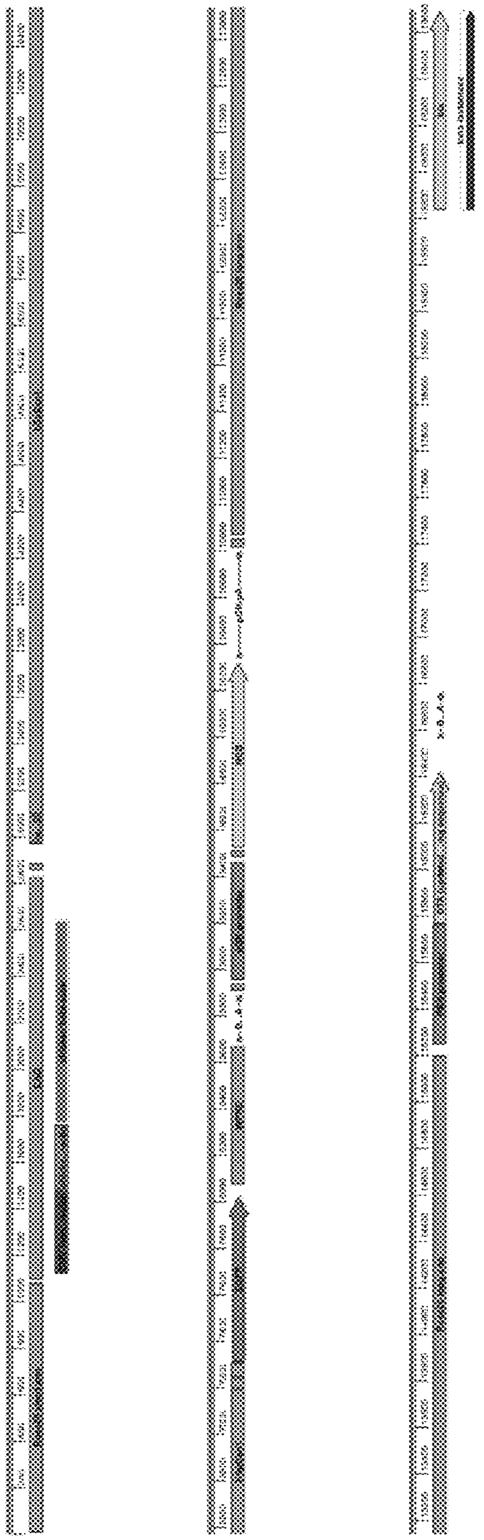
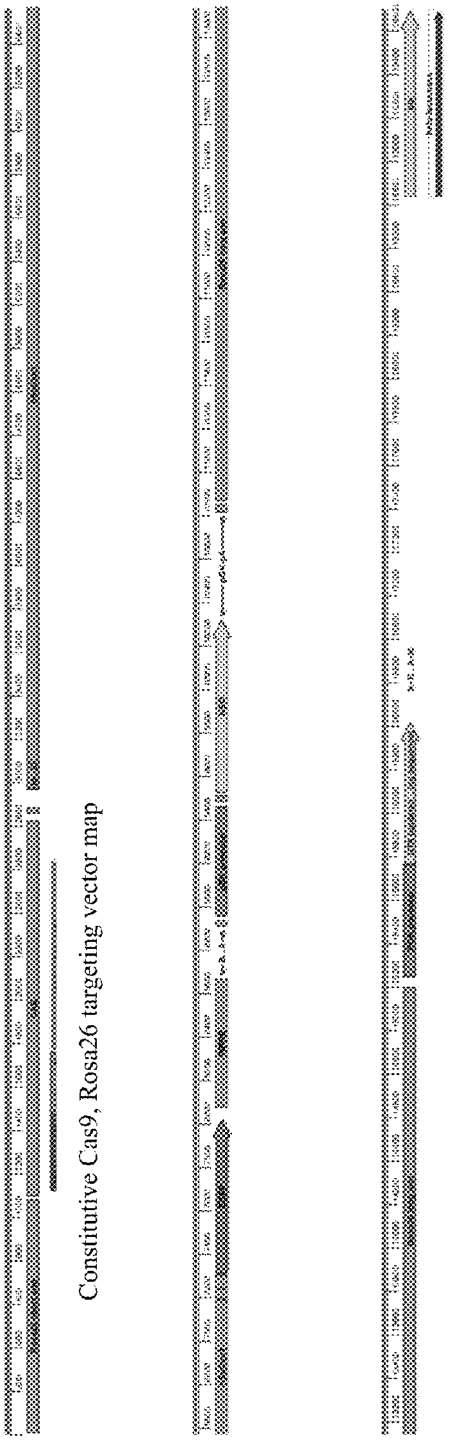


FIG. 25A



Constitutive Cas9, Rosa26 targeting vector map

FIG. 25B

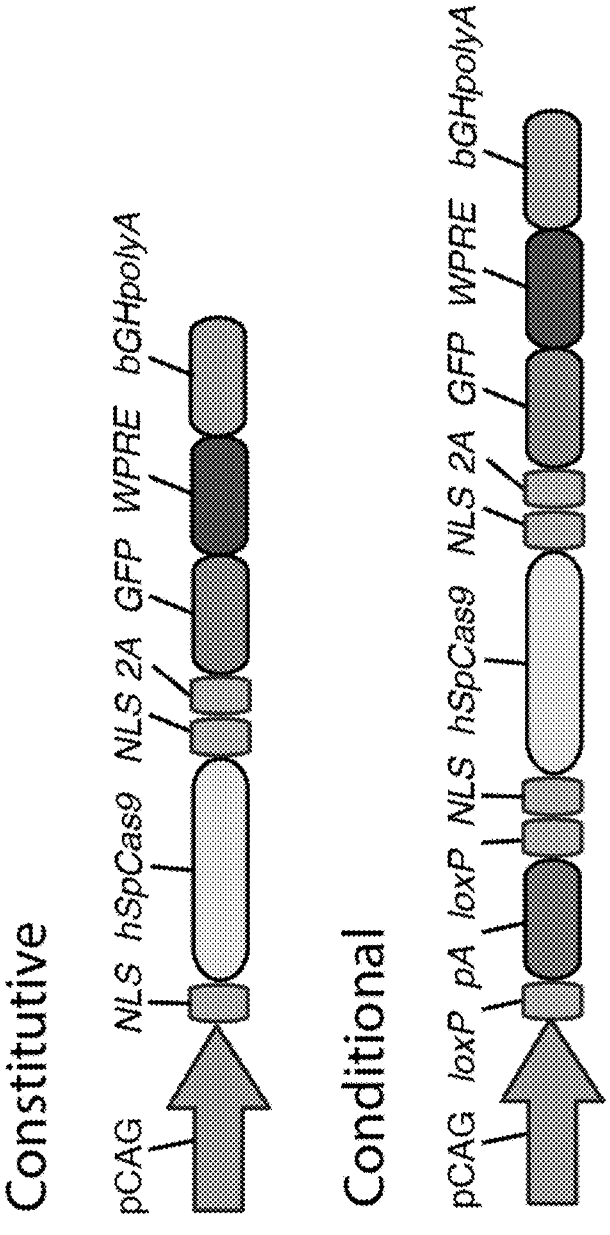
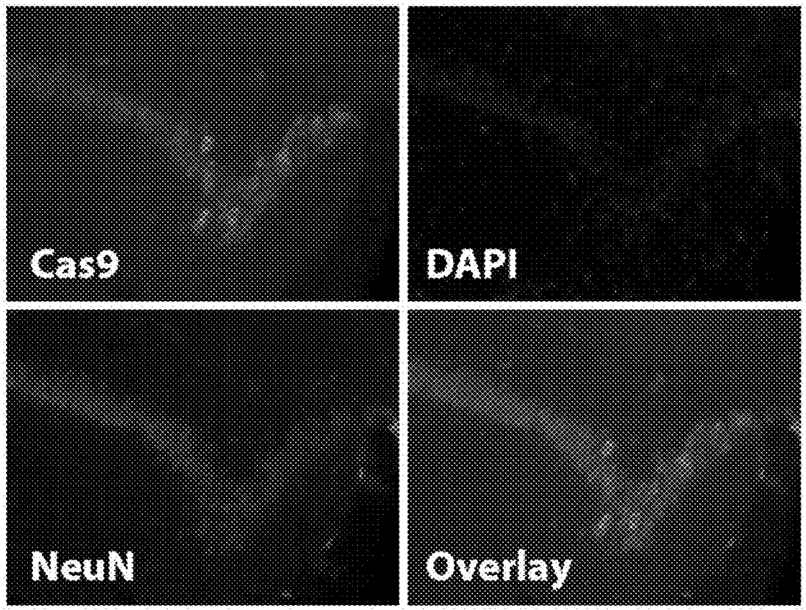

Constitutive
pCAG
NLS
hSpCas9
NLS
2A
GFP
WPRE
bGHpolyA
Conditional
pCAG
loxP
pA
loxP
NLS
NLS
hSpCas9
NLS
2A
GFP
WPRE
bGHpolyA

FIG. 26

Cas9 Expression in Mouse Hippocampus (AAV)



Cas9 Expression in Mouse Cortex (AAV)

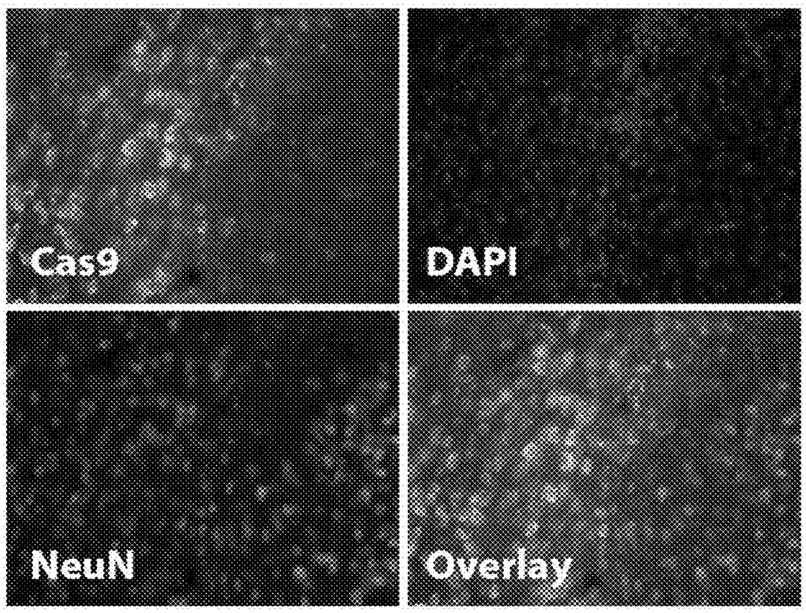


FIG. 27

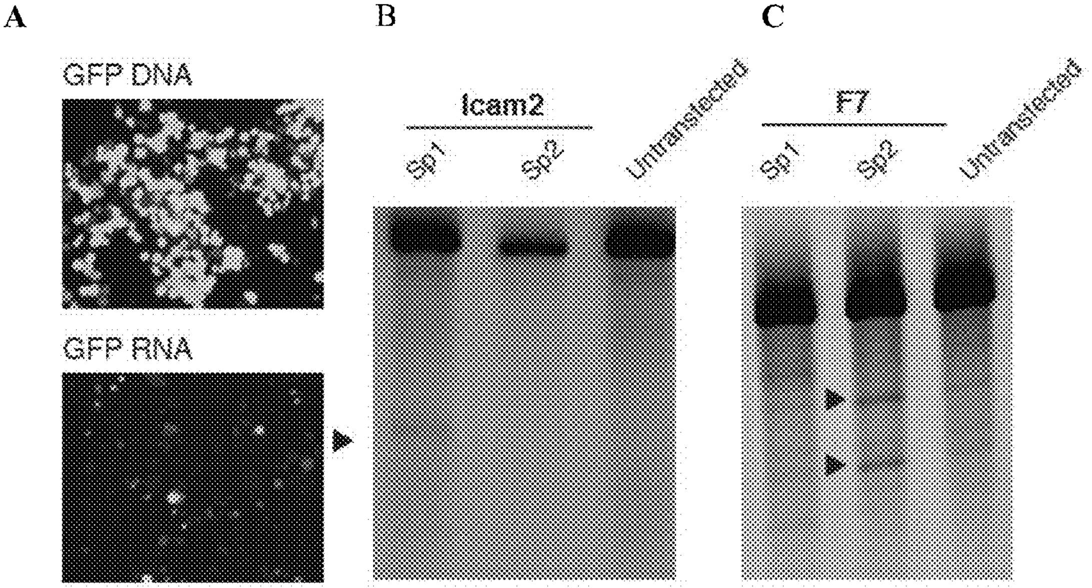

A
B
C
GFP DNA
GFP RNA
Icam2
Sp1
Sp2
Untransfected
F7
Sp1
Sp2
Untransfected

FIG. 28

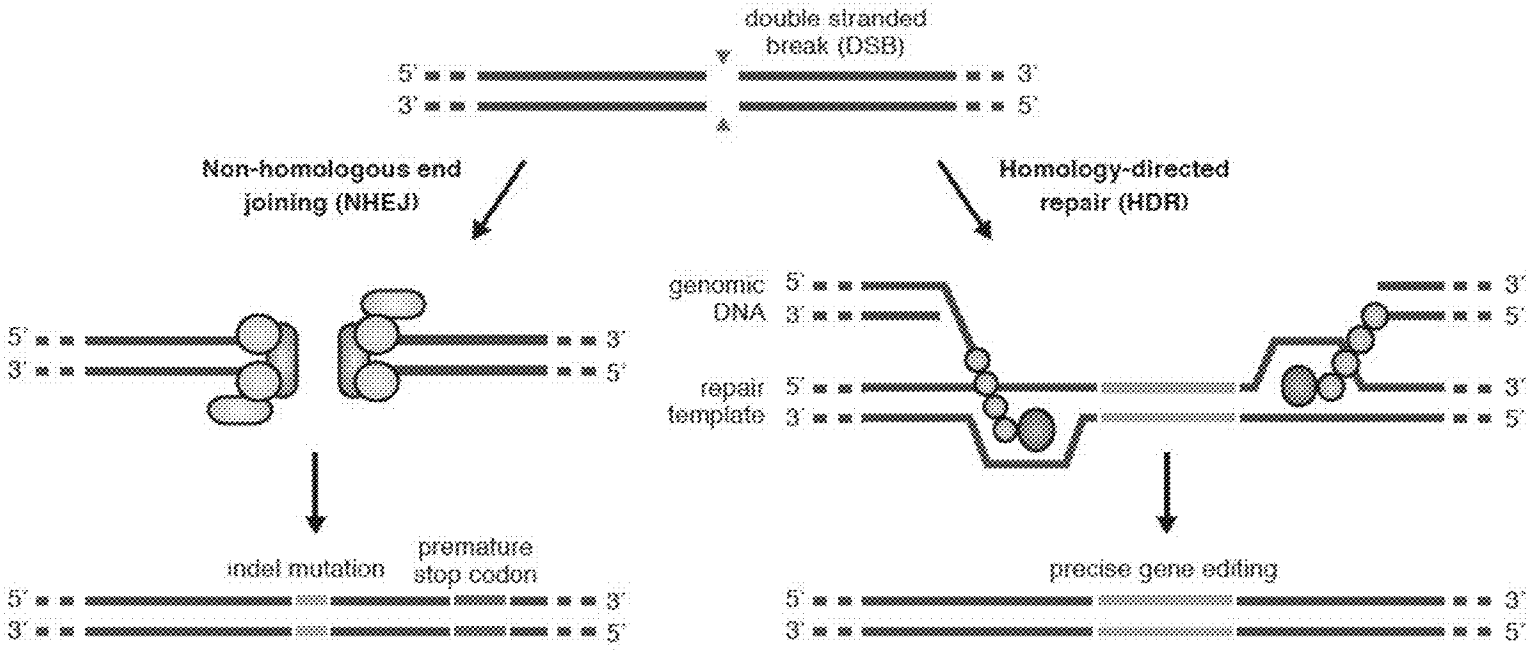

double stranded
break (DSB)
5'
3'
3'
5'
Non-homologous end
joining (NHEJ)
Homology-directed
repair (HDR)
genomic
DNA
repair
template
indel mutation
premature
stop codon
precise gene editing

FIG. 29

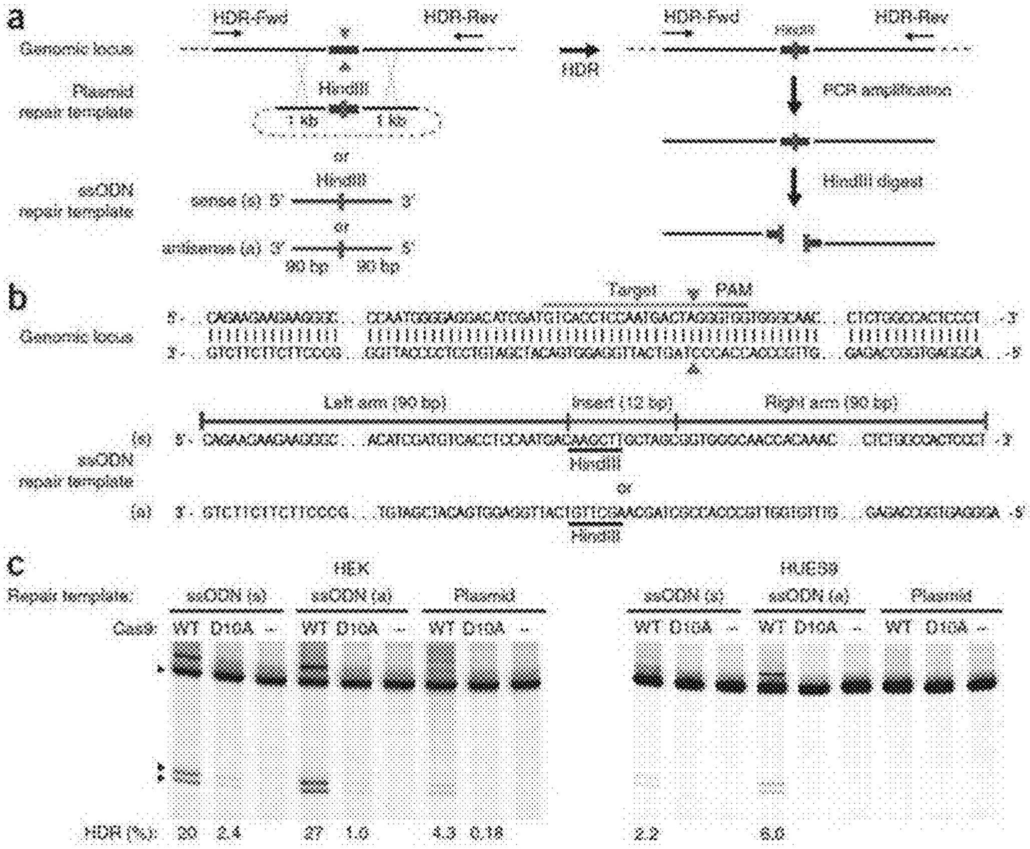

a
HDR-Fwd
HDR-Rev
Genomic locus
Plasmid repair template
HindIII
1 kb
1 kb
or
ssODN repair template
HindIII
sense (s) 5' 3'
or
antisense (a) 3' 5'
90 bp
90 bp
HDR
HDR-Fwd
HindIII
HDR-Rev
PCR amplification
HindIII digest
b
Target
PAM
Genomic locus
5' - CAGAAGAAGAAGGGC ... CCAATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGTGGGCAAC ... CTCTGGCCACTCCCT - 3'
3' - GTCTTCTTCTTCCCG ... GGTTACCCCTCCTGTAGCTACAGTGGAGGTTACTGATCCCACCACCCGTTG ... GAGACCGGTGAGGGA - 5'
Left arm (90 bp)
Insert (12 bp)
Right arm (90 bp)
(s) 5' - CAGAAGAAGAAGGGC ... ACATCGATGTCACCTCCAATGACAAGCTTGCTAGCGGTGGGCAACCACAAAC ... CTCTGGCCACTCCCT - 3'
ssODN repair template
HindIII
or
(a) 3' - GTCTTCTTCTTCCCG ... TGTAGCTACAGTGGAGGTTACTGTTCGAACGATCGCCACCCGTTGGTGTTTG ... GAGACCGGTGAGGGA - 5'
HindIII
c
HEK
HUES9
Repair template:
ssODN (s)
ssODN (a)
Plasmid
ssODN (s)
ssODN (a)
Plasmid
Cas9:
WT D10A –
WT D10A –
WT D10A –
WT D10A –
WT D10A –
WT D10A –
HDR (%):
20 2.4
27 1.0
4.3 0.18
2.2
6.0

FIG. 30A-C

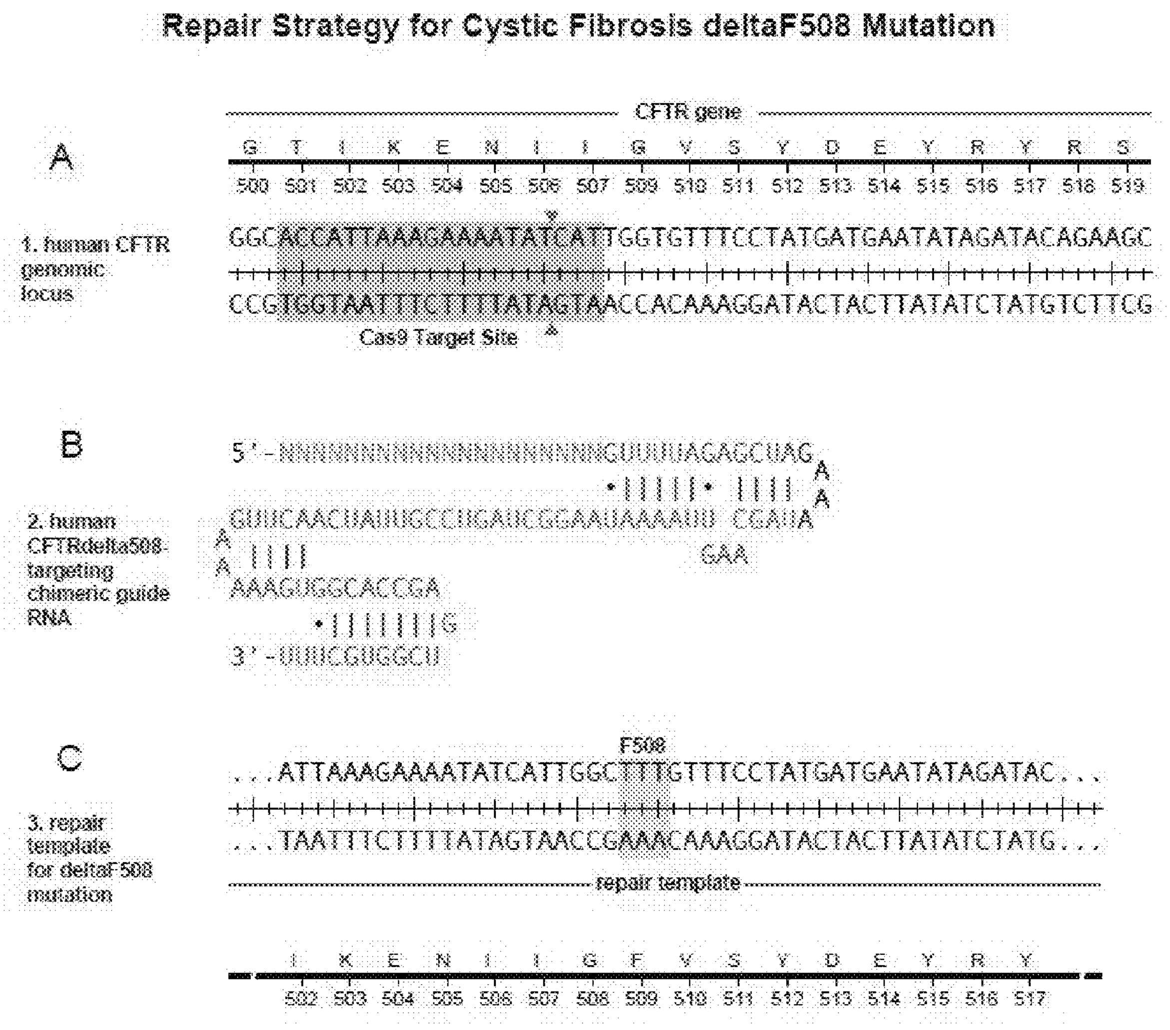
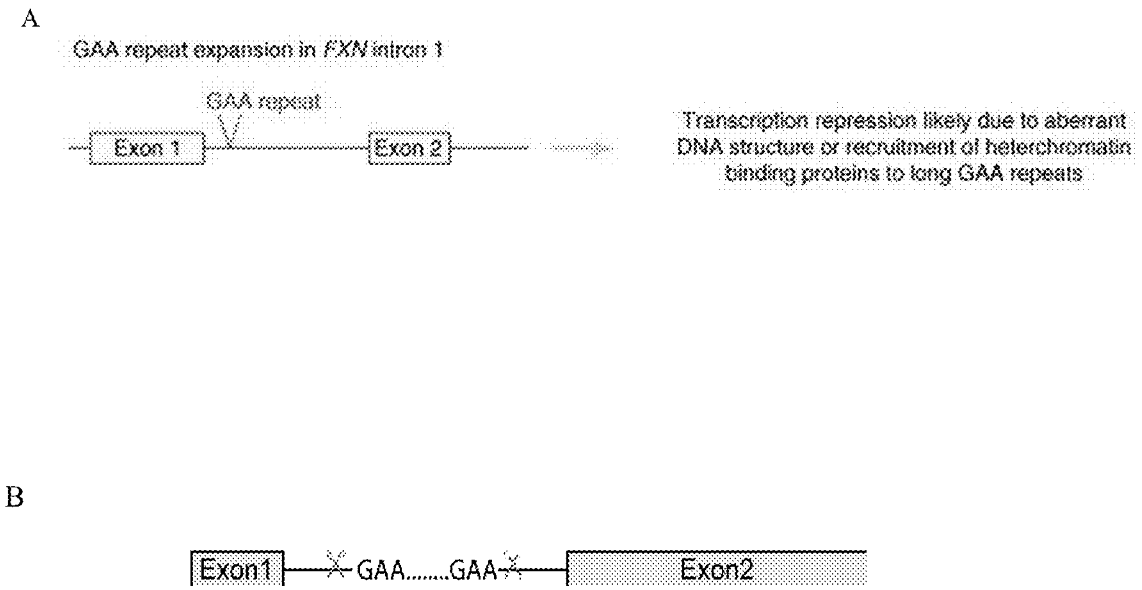

Repair Strategy for Cystic Fibrosis deltaF508 Mutation
A
CFTR gene
G T I K E N I I G V S Y D E Y R Y R S
500 501 502 503 504 505 506 507 509 510 511 512 513 514 515 516 517 518 519
1. human CFTR genomic locus
GGCACCATTAAAGAAAATATCATTGGTGTTTCCTATGATGAATATAGATACAGAAGC
CCGTGGTAATTTCTTTTATAGTAACCACAAAGGATACTACTTATATCTATGTCTTCG
Cas9 Target Site
B
2. human CFTRdelta508-targeting chimeric guide RNA
5'-NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
A
A
GUUCAACUAUUGCCUGAUCGGAAUAAAAUU CGAUA
GAA
A
A
AAAGUGGCACCGA
G
3'-UUUCGUGGCU
C
3. repair template for deltaF508 mutation
F508
...ATTAAAGAAAATATCATTGGCTTTGTTTCCTATGATGAATATAGATAC...
...TAATTTCTTTTATAGTAACCGAAACAAAGGATACTACTTATATCTATG...
repair template
I K E N I I G F V S Y D E Y R Y
502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517

FIG. 31



GAA repeat expansion in *FXN* intron 1

Transcription repression likely due to aberrant DNA structure or recruitment of heterchromatin binding proteins to long GAA repeats

FIG. 32

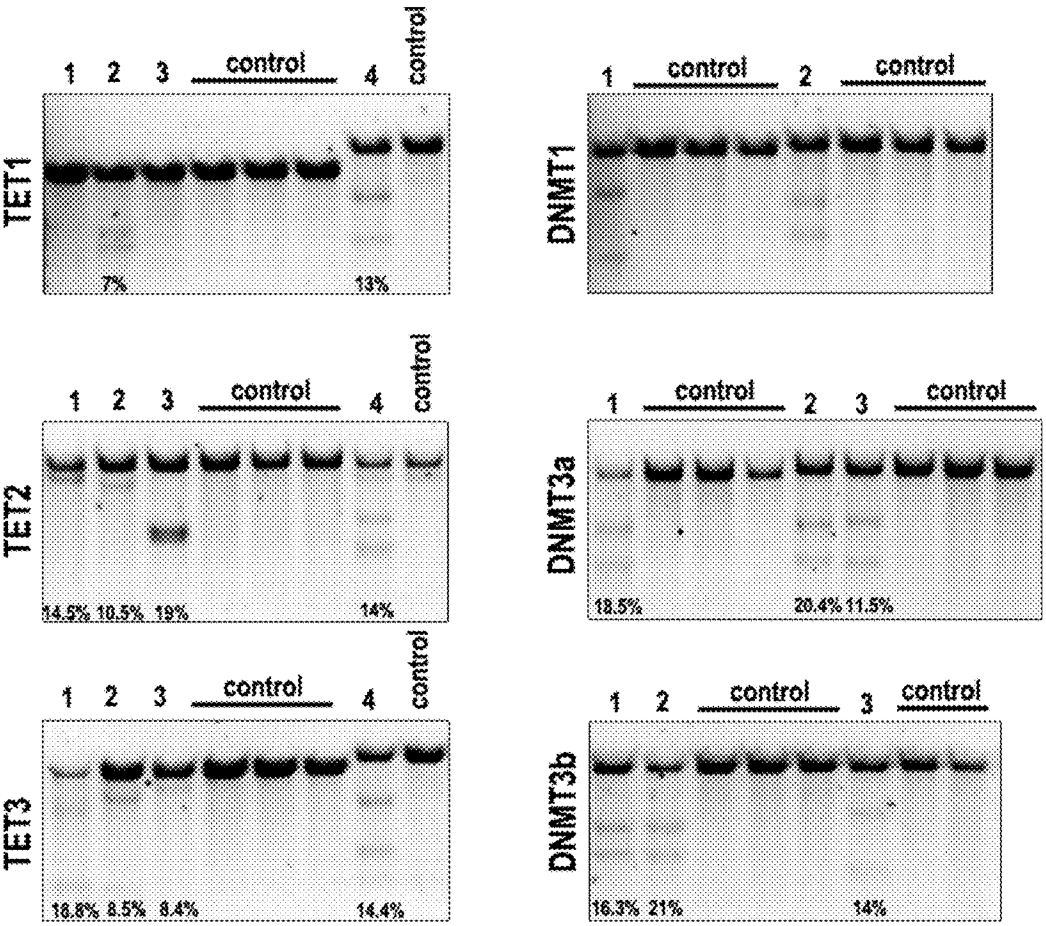

TET1
1
2
3
control
4
control
7%
13%
DNMT1
1
control
2
control
TET2
1
2
3
control
4
control
14.5%
10.5%
19%
14%
DNMT3a
1
control
2
3
control
18.5%
20.4%
11.5%
TET3
1
2
3
control
4
control
18.8%
8.5%
8.4%
14.4%
DNMT3b
1
2
control
3
control
16.3%
21%
14%

FIG. 33

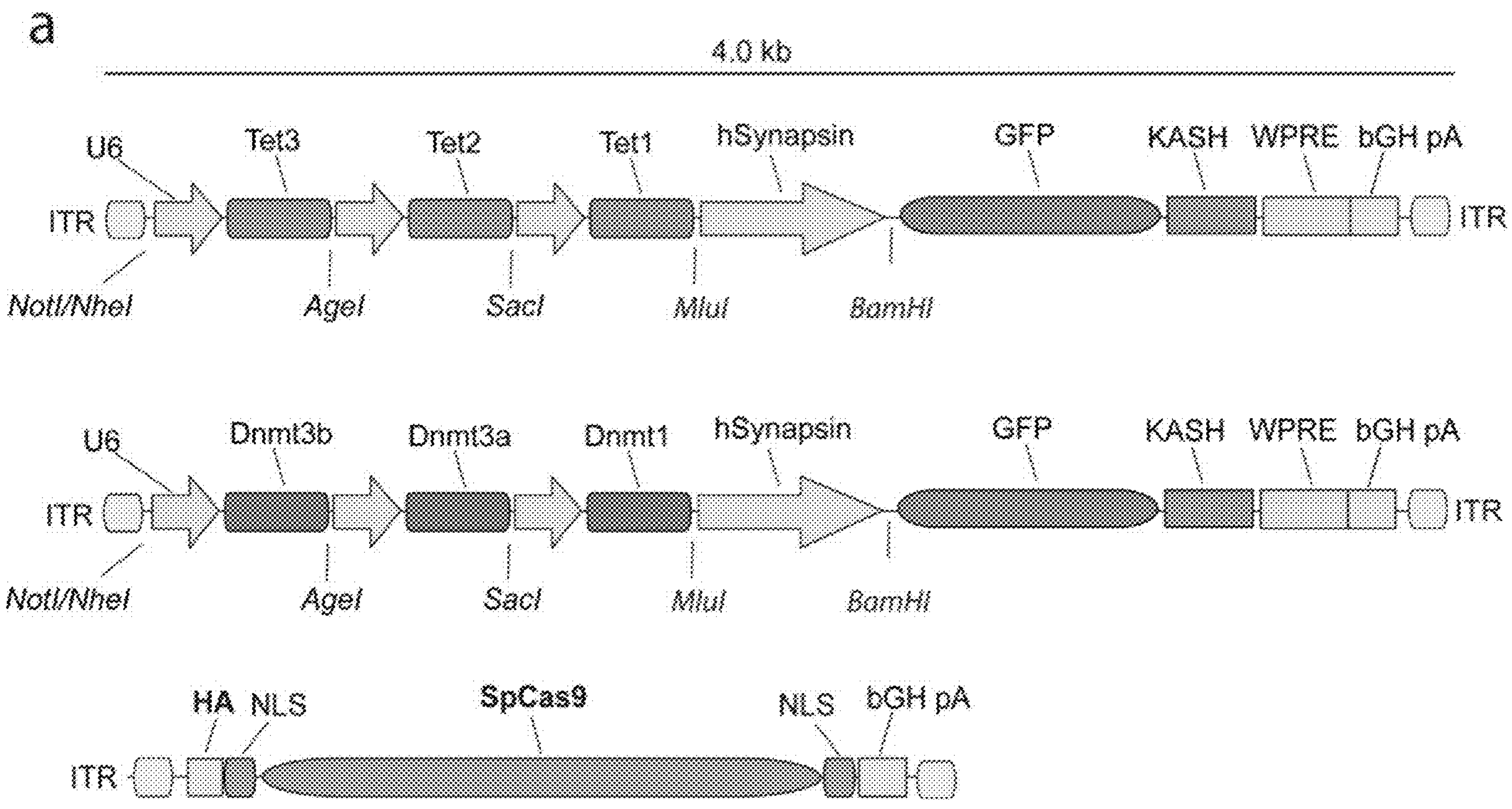

a
4.0 kb
U6
Tet3
Tet2
Tet1
hSynapsin
GFP
KASH
WPRE
bGH pA
ITR
ITR
NotI/NheI
AgeI
SacI
MluI
BamHI
U6
Dnmt3b
Dnmt3a
Dnmt1
hSynapsin
GFP
KASH
WPRE
bGH pA
ITR
ITR
NotI/NheI
AgeI
SacI
MluI
BamHI
HA
NLS
SpCas9
NLS
bGH pA
ITR

FIG. 34

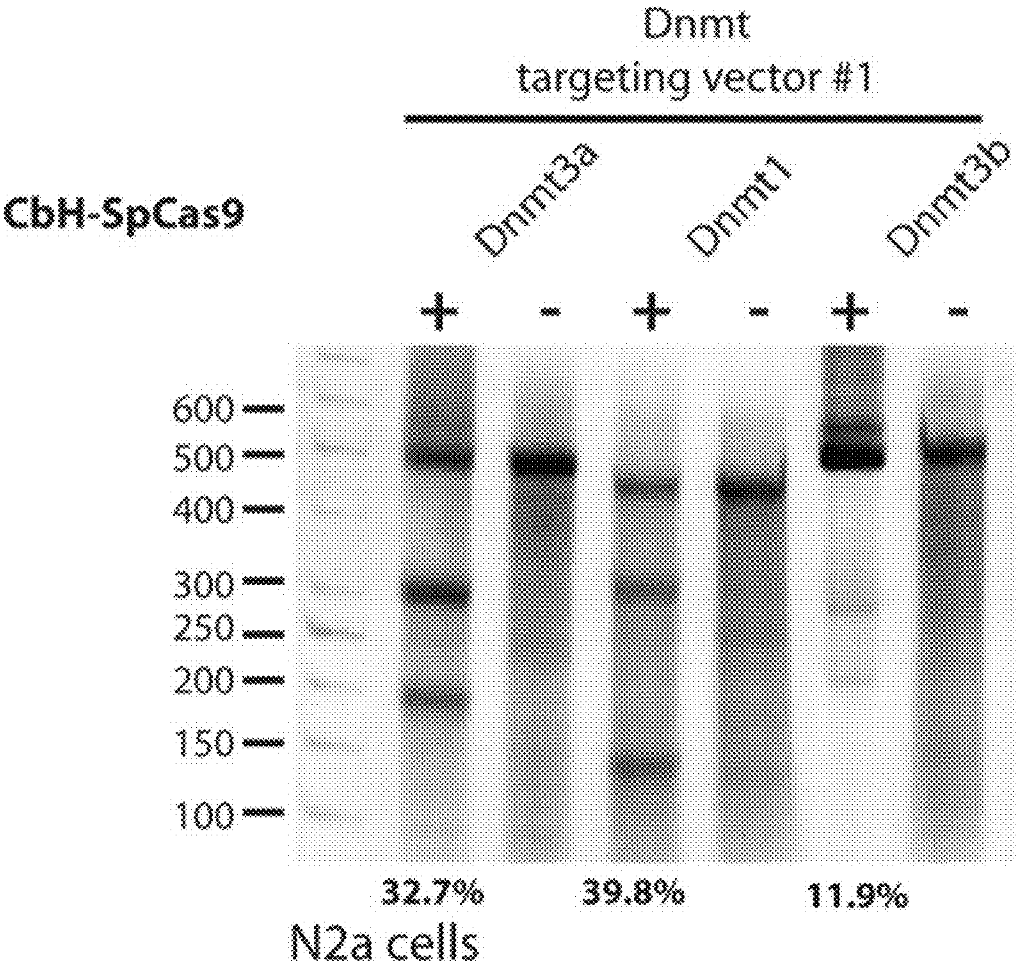

Dnmt
targeting vector #1
CbH-SpCas9
Dnmt3a
Dnmt1
Dnmt3b
+
-
+
-
+
-
600
500
400
300
250
200
150
100
32.7%
39.8%
11.9%
N2a cells

FIG. 35

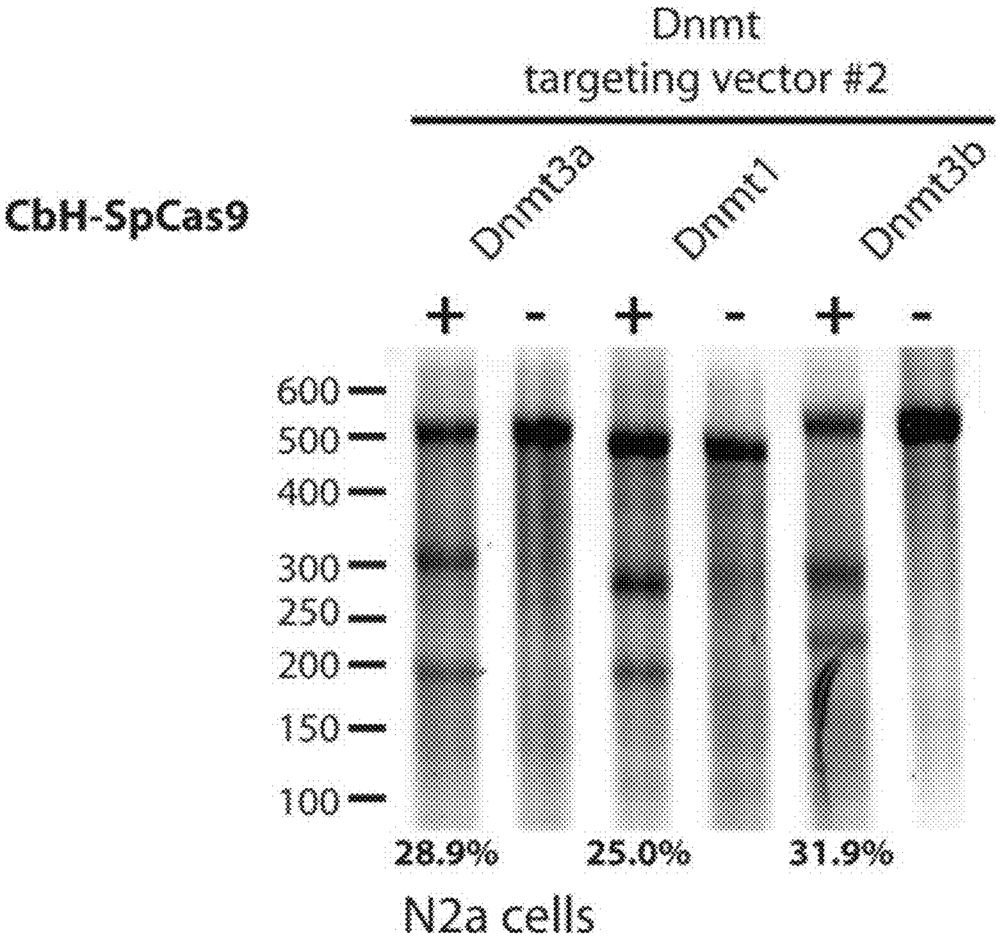

Dnmt
targeting vector #2
CbH-SpCas9
Dnmt3a
Dnmt1
Dnmt3b
+
-
+
-
+
-
600
500
400
300
250
200
150
100
28.9%
25.0%
31.9%
N2a cells

FIG. 36

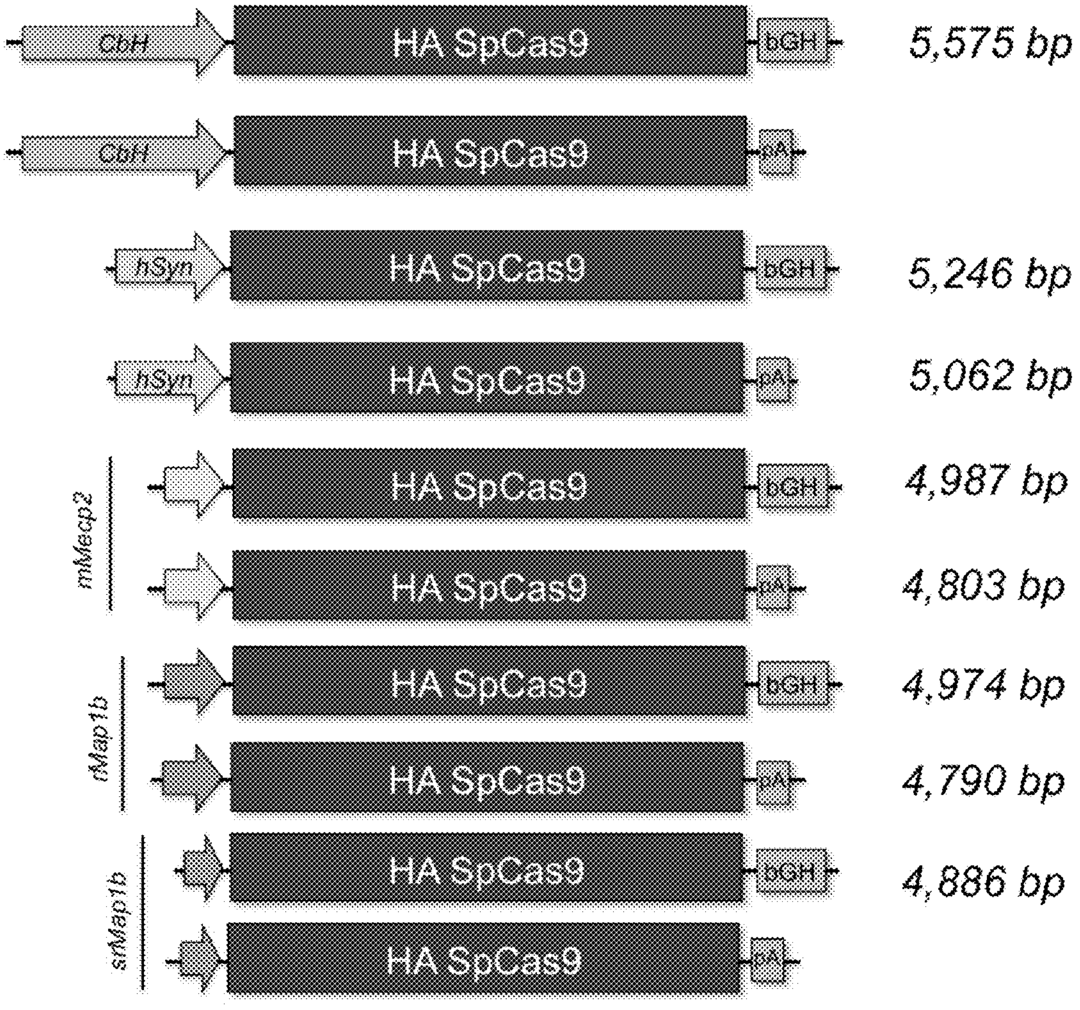

CbH
HA SpCas9
bGH
5,575 bp
CbH
HA SpCas9
pA
hSyn
HA SpCas9
bGH
5,246 bp
hSyn
HA SpCas9
pA
5,062 bp
mMecp2
HA SpCas9
bGH
4,987 bp
HA SpCas9
pA
4,803 bp
rMap1b
HA SpCas9
bGH
4,974 bp
HA SpCas9
pA
4,790 bp
srMap1b
HA SpCas9
bGH
4,886 bp
HA SpCas9
pA

FIG. 37

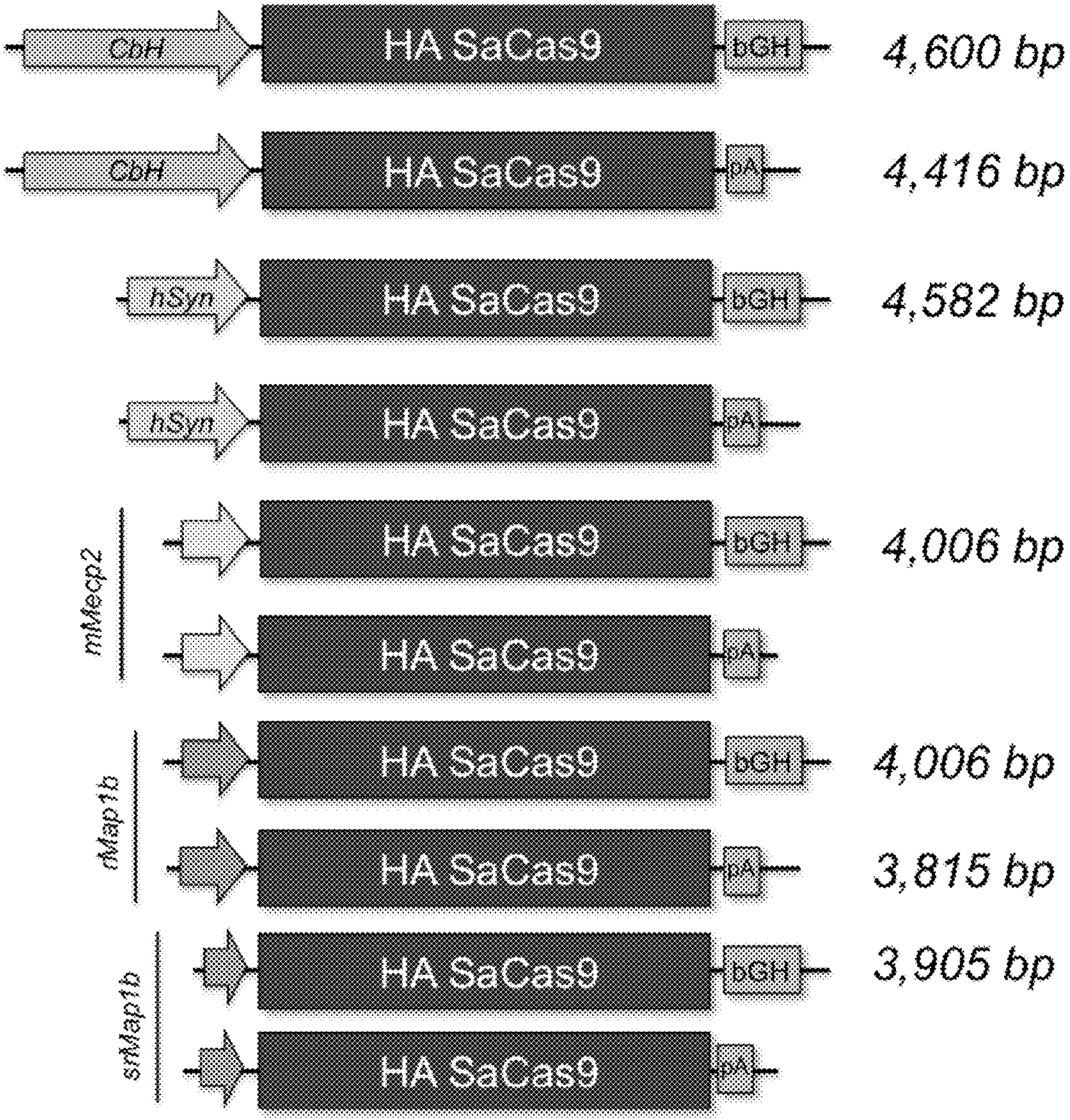

CbH
HA SaCas9
bGH
4,600 bp
CbH
HA SaCas9
pA
4,416 bp
hSyn
HA SaCas9
bGH
4,582 bp
hSyn
HA SaCas9
pA
mMecp2
HA SaCas9
bGH
4,006 bp
HA SaCas9
pA
rMap1b
HA SaCas9
bGH
4,006 bp
HA SaCas9
pA
3,815 bp
srMap1b
HA SaCas9
bGH
3,905 bp
HA SaCas9
pA

FIG. 38

Expression of SpCas9 & SaCas9 in N2a cells

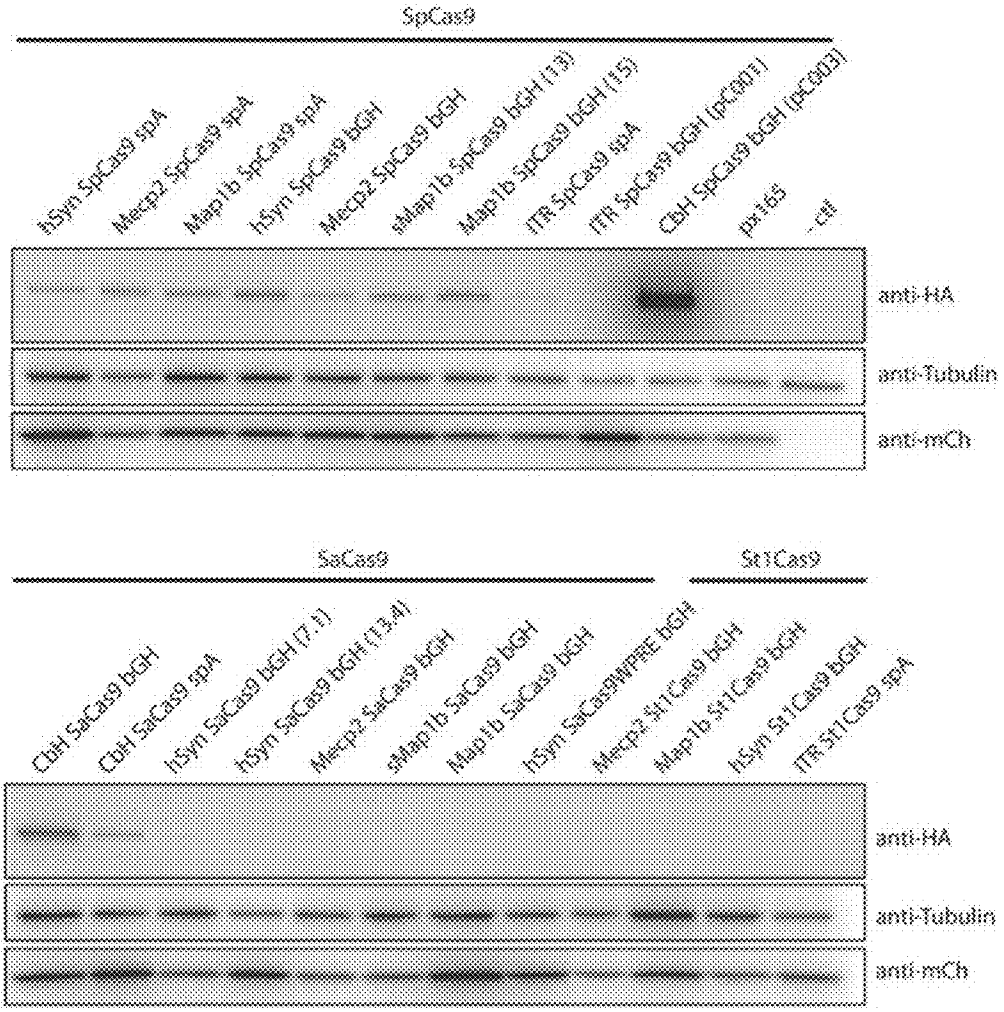


FIG. 39

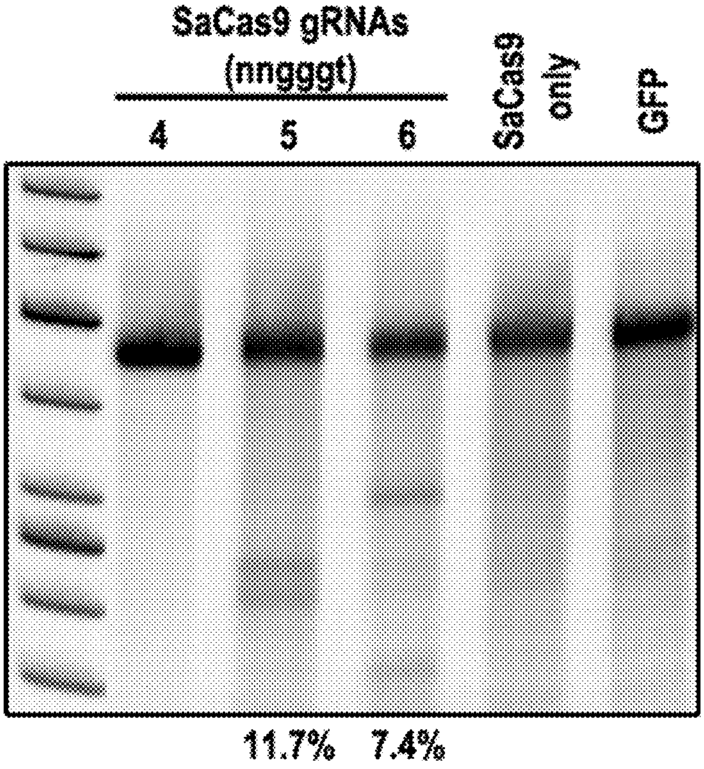


FIG. 40

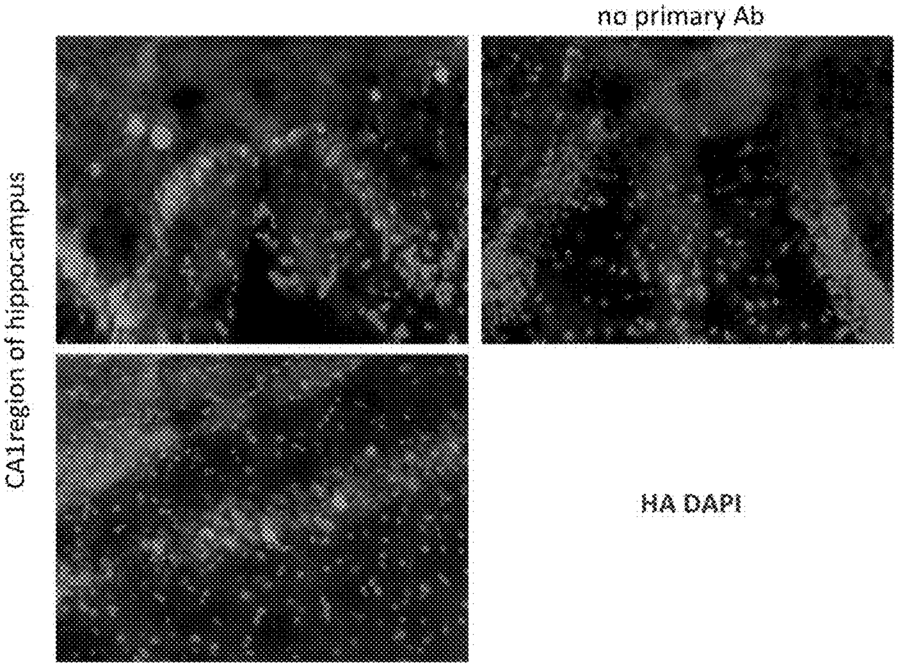


FIG. 41

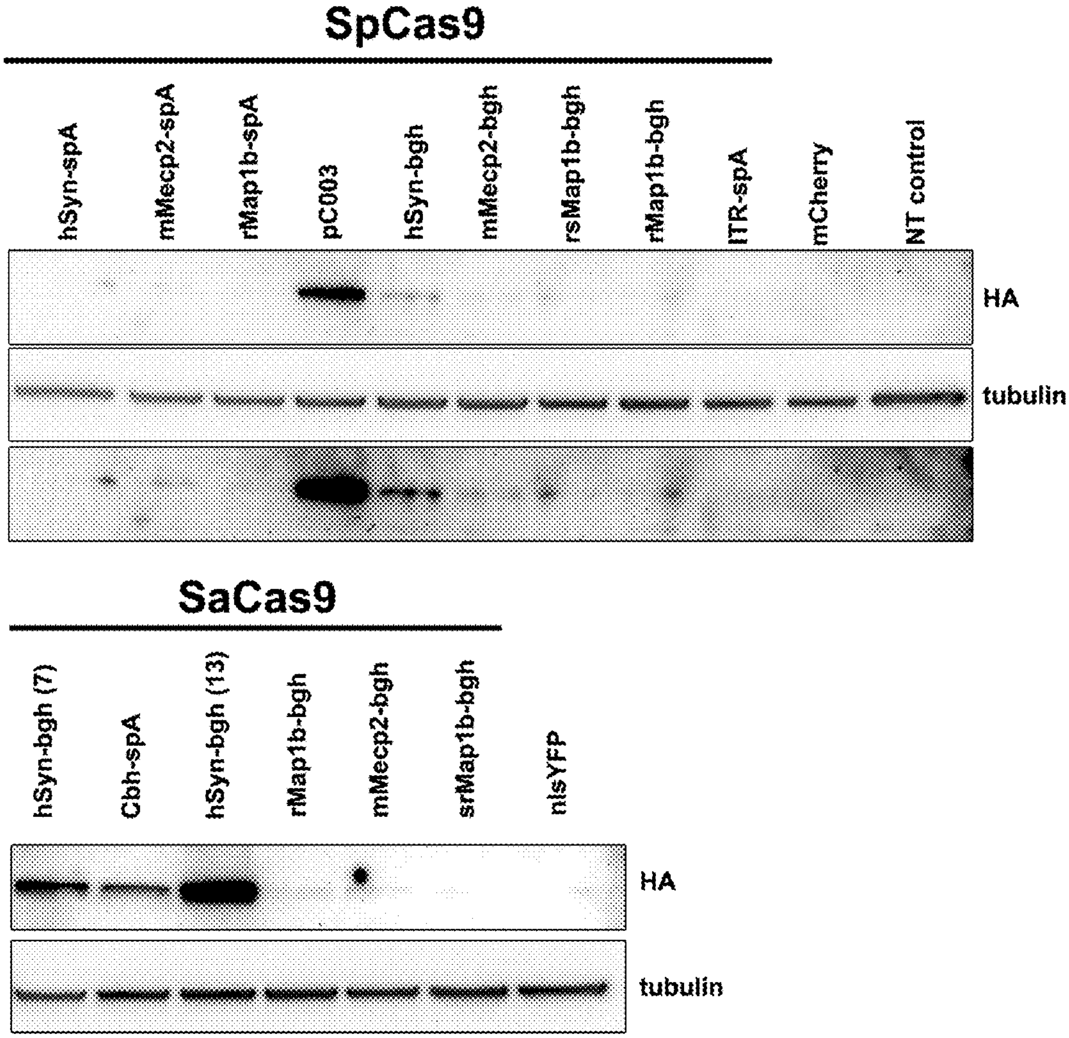

SpCas9
hSyn-spA
mMecp2-spA
rMap1b-spA
pC003
hSyn-bgh
mMecp2-bgh
rsMap1b-bgh
rMap1b-bgh
ITR-spA
mCherry
NT control
HA
tubulin
SaCas9
hSyn-bgh (7)
Cbh-spA
hSyn-bgh (13)
rMap1b-bgh
mMecp2-bgh
srMap1b-bgh
nlsYFP
HA
tubulin

FIG. 42

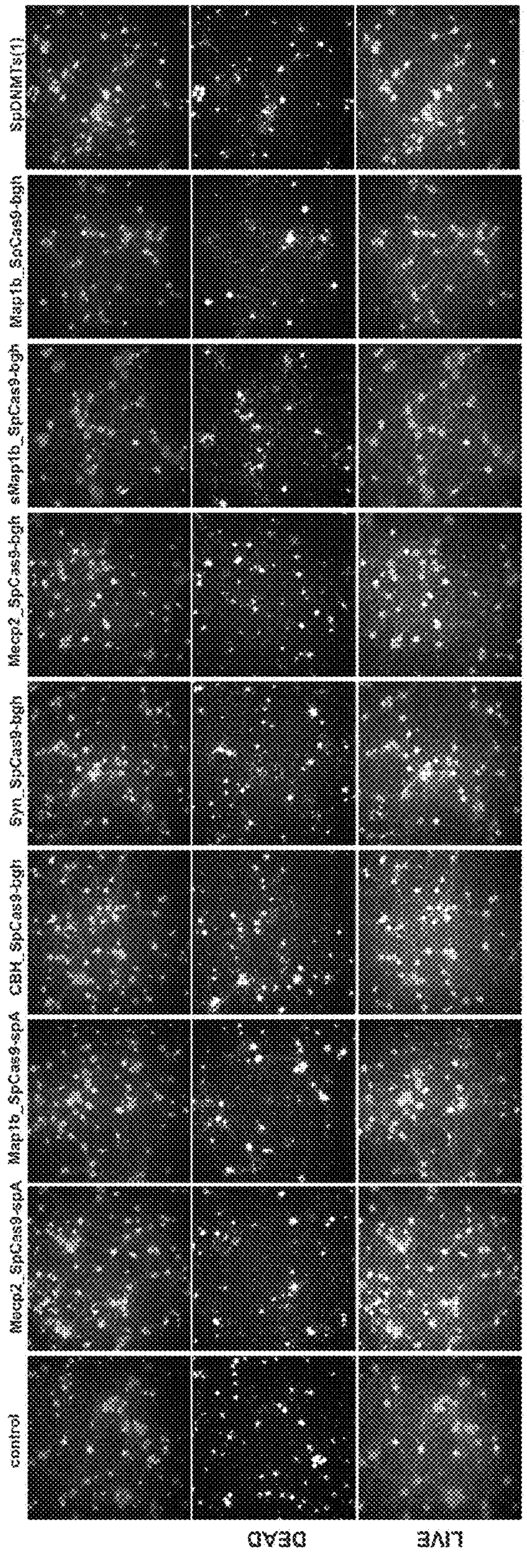

control
Mecp2_SpCas9-spA
Map1b_SpCas9-spA
CBH_SpCas9-bgh
Syn_SpCas9-bgh
Mecp2_SpCas9-bgh
sMap1b_SpCas9-bgh
Map1b_SpCas9-bgh
SpDNMTs(1)
DEAD
LIVE

FIG. 43

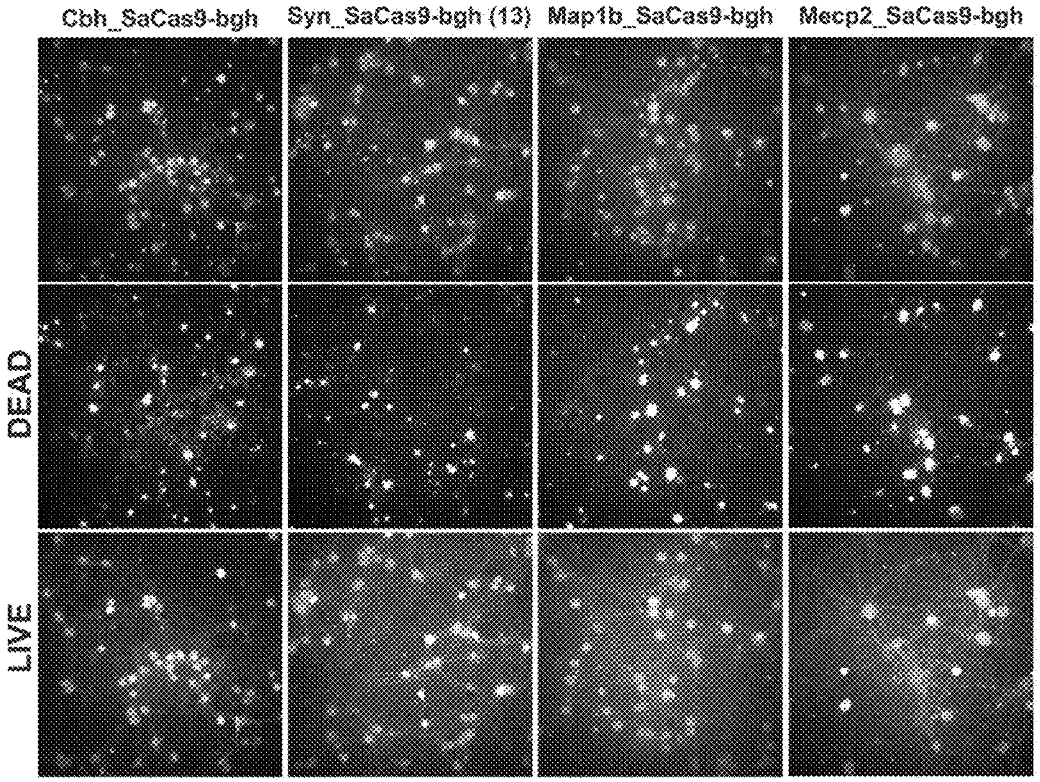

Cbh_SaCas9-bgh
Syn_SaCas9-bgh (13)
Map1b_SaCas9-bgh
Mecp2_SaCas9-bgh
DEAD
LIVE

FIG. 44

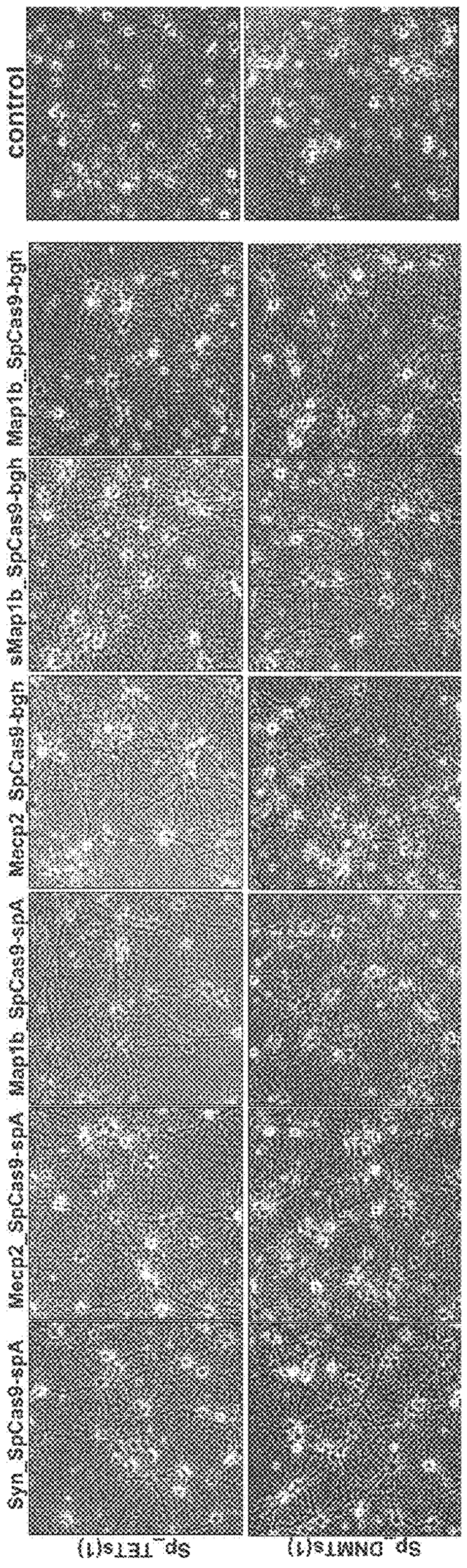

Syn_SpCas9-spA
Mecp2_SpCas9-spA
Map1b_SpCas9-spA
Mecp2_SpCas9-bgh
sMap1b_SpCas9-bgh
Map1b_SpCas9-bgh
control
Sp_TETs(1)
Sp_DNMTs(1)

FIG. 45

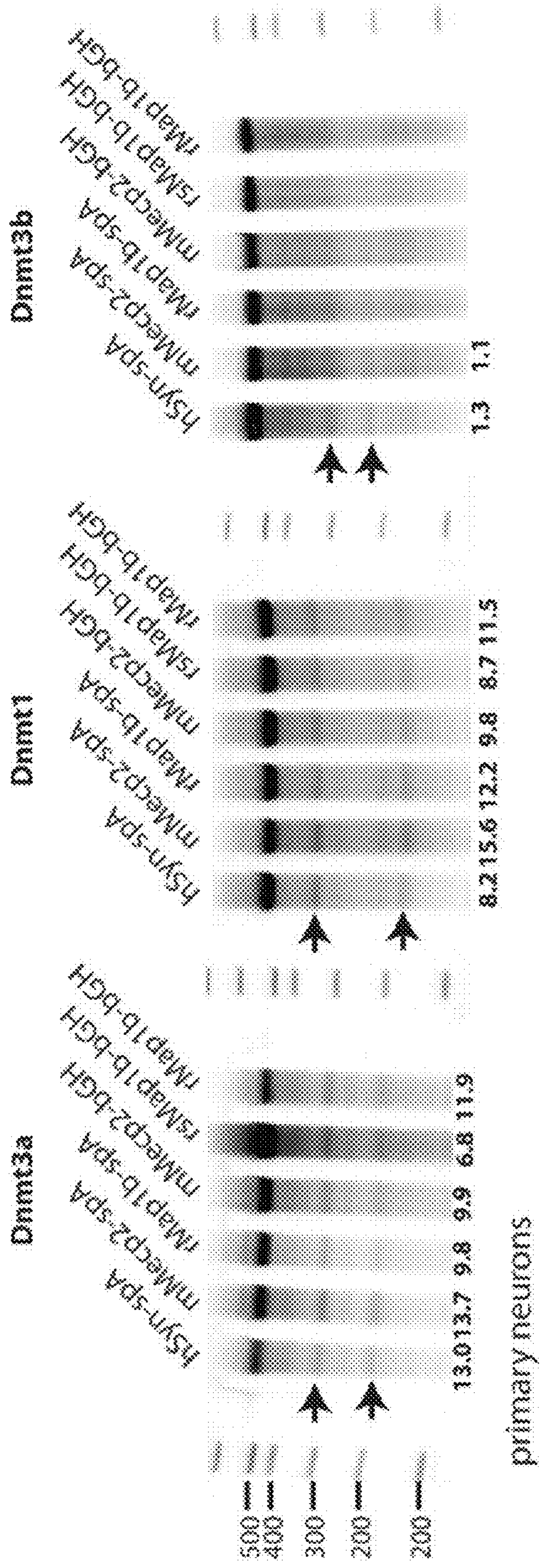

Dnmt3a
hSyn-spA
rMecp2-spA
rMap1b-spA
mMecp2-bGH
rsMap1b-bGH
rMap1b-bGH
13.0
13.7
9.8
9.9
6.8
11.9
Dnmt1
hSyn-spA
rMecp2-spA
rMap1b-spA
mMecp2-bGH
rsMap1b-bGH
rMap1b-bGH
8.2
15.6
12.2
9.8
8.7
11.5
Dnmt3b
hSyn-spA
rMecp2-spA
rMap1b-spA
mMecp2-bGH
rsMap1b-bGH
rMap1b-bGH
1.3
1.1
500
400
300
200
200
primary neurons

FIG. 46

SpCas9 *in vivo: Mecp2*

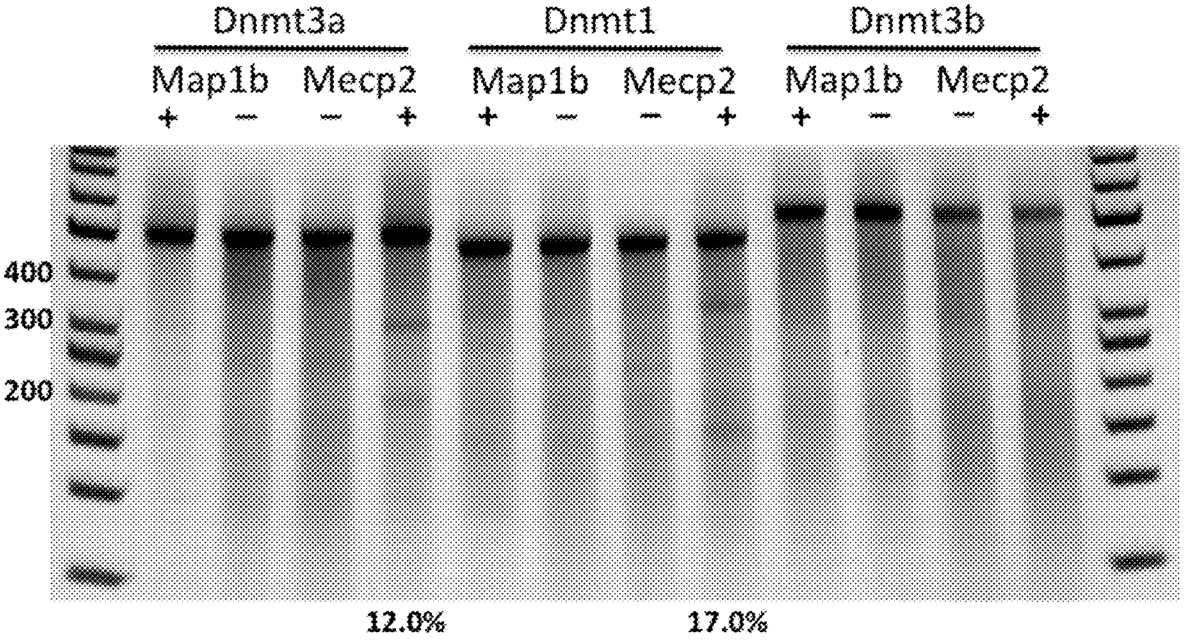


FIG. 47

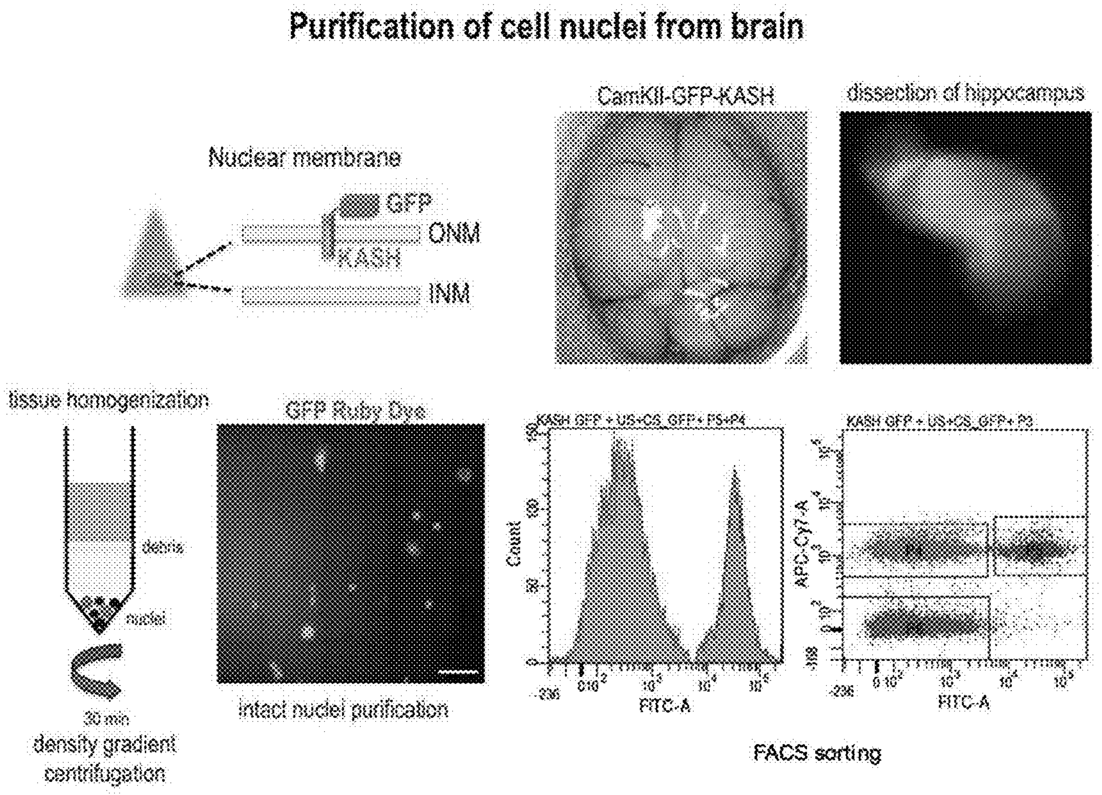

Purification of cell nuclei from brain
Nuclear membrane
GFP
ONM
KASH
INM
CamKII-GFP-KASH
dissection of hippocampus
tissue homogenization
debris
nuclei
30 min
density gradient centrifugation
GFP Ruby Dye
intact nuclei purification
KASH GFP + US+CS_GFP+ P5+P4
Count
0
50
100
150
FITC-A
-236
0
10^2
10^3
10^4
10^5
KASH GFP + US+CS_GFP+ P3
APC-Cy7-A
-169
0
10^2
10^3
10^4
10^5
FITC-A
-236
0
10^2
10^3
10^4
10^5
FACS sorting

FIG. 48

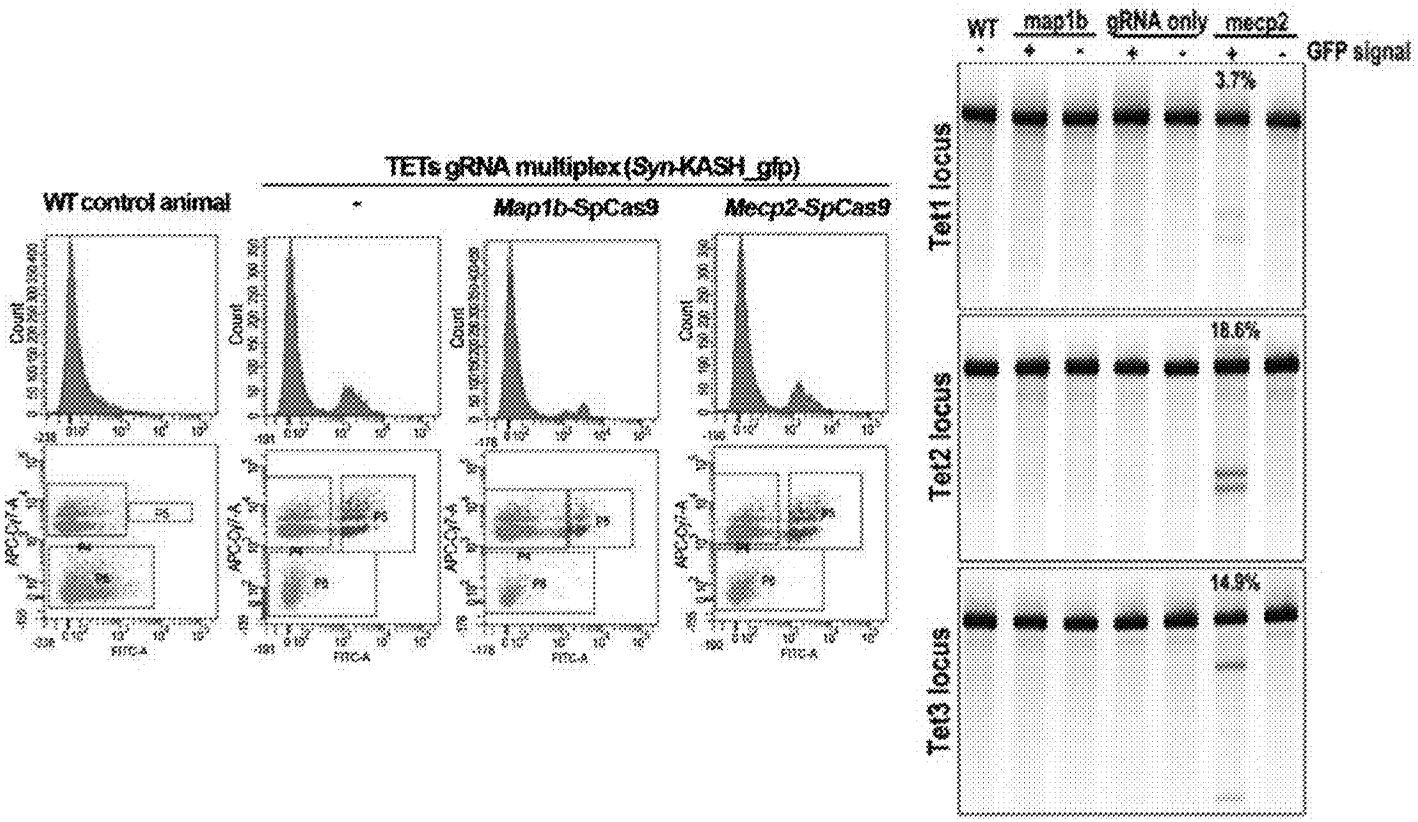

WT control animal
TETs gRNA multiplex (Syn-KASH_gfp)
-
Map1b-SpCas9
Mecp2-SpCas9
Count
APC-Cy7-A
FITC-A
P5
P6
WT
map1b
gRNA only
mecp2
GFP signal
-
+
-
+
-
+
-
3.7%
18.6%
14.9%
Tet1 locus
Tet2 locus
Tet3 locus

FIG. 49

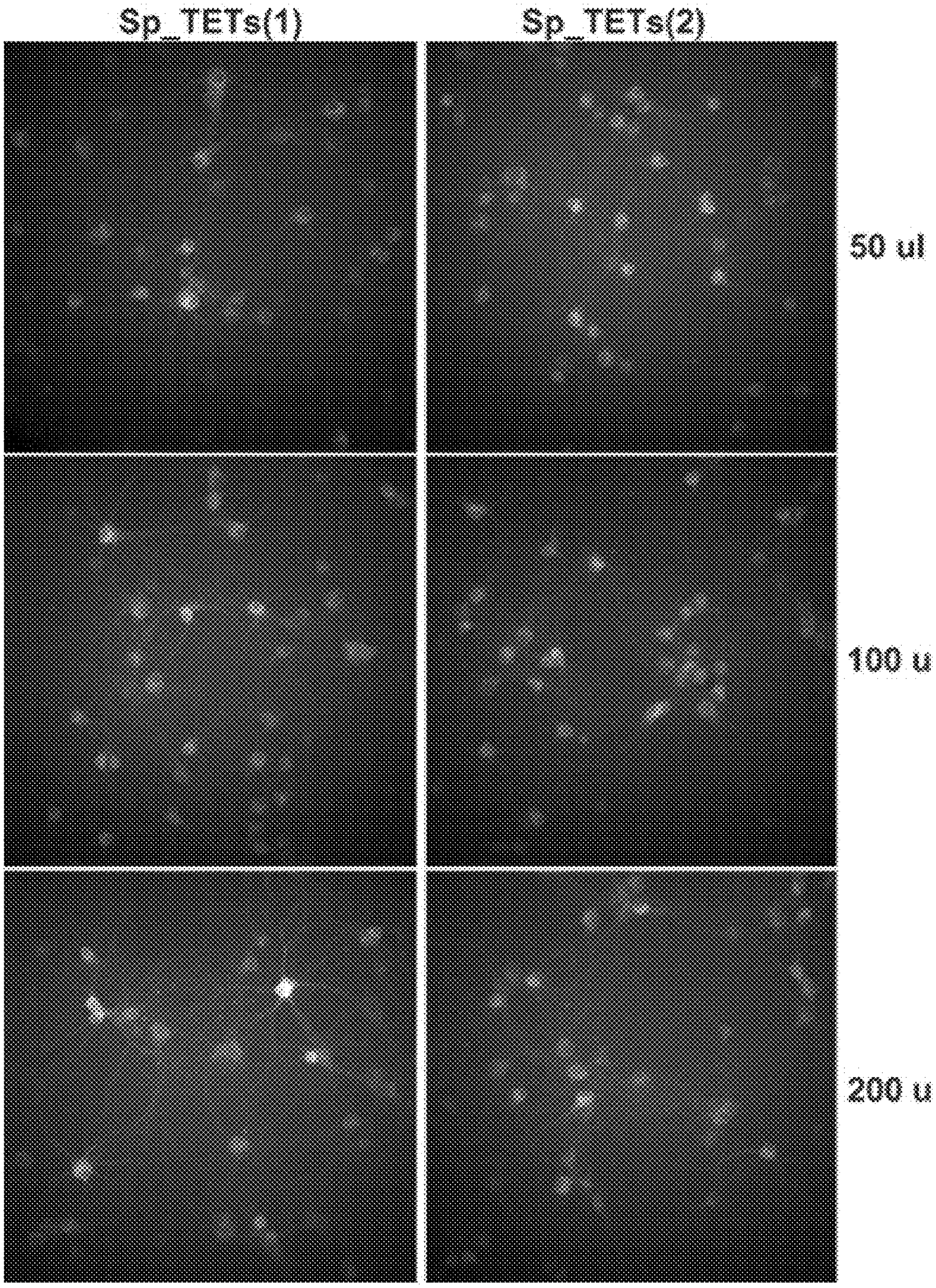
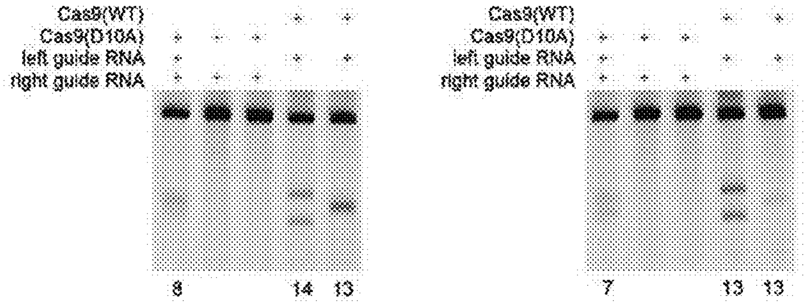
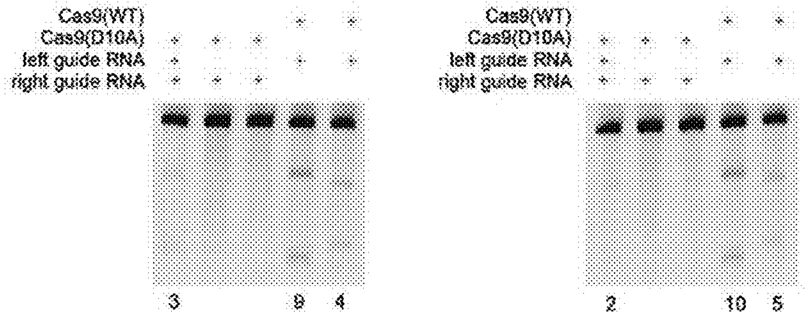


FIG. 50

	pattern	overhang
1	CCnNNNNNNNNNNNNNNNNnnnnnnnnnnnnNNNNNNNNNNNNNNNNnGG	38
2	CCnNNNNNNNNNNNNNNNNnnnnnnnnNNNNNNNNNNNNNNNNnGG	34
3	CCnNNNNNNNNNNNNNNNNnnnnNNNNNNNNNNNNNNNNnGG	30
4	CCnNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNnGG	26
5	CCnNNNNNNNNNNNNNNNNNNNNNNNNNNNNnGG	22
6	CCnNNNNNNNNNNNNNNNNNNNNNNNNnGG	18
7	CCnNNNNNNNNNNNNNNNNNNNNnGG	14
8	CCnNNNNNNNNNNNNNNNNnGG	10
9	CCnNNNNNNNNNNNNnGG	6
10	CCnNNNNNNNNNNnGG	4
11	CCnNNNNNNNNNnGG	3
12	CCnNNNNNNNnGG	2
13	CCnNNNNNNNnGG	1
14	CCnNNNNNNnGG	0
15	CCnNNNNNnGG	1
16	CCnNNNNnGG	2
17	CCnNNNnGG	3
18	CCnnGG	4
19	CCGG	8
20	GGCC	10
21	GGNNCC	12
22	GGNNNCC	14
23	GGNNNNCC	16
24	GGNNNNNNNNCC	20



Pattern 1 Pattern 2



Pattern 3 Pattern 4

FIG. 51

Reverse PCR primers for generating U6-guide RNA expression Cassette (pair with U6 forward primer)

FZ0260	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCACATCAACCGGTGGCGCAGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 1
FZ0261	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCACATCAACCGGTGGCGCAGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 2
FZ0262	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGGACAAAGTACAAACGGCGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 3
FZ0263	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGGACAAAGTACAAACGGCGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 4
FZ0264	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGTGGCGCATTGCCACGAAGCGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 5
FZ0265	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGAGGGCAGAGTGCTGCTTGGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 6
FZ0266	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGTCCGAGCAGAAGAAGAAGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 7
FZ0267	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGGACAAAGTACAAACGGCGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 8
FZ0268	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGCAGAAGAAGAAGGCTCCGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 9
FZ0269	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCACATCAACCGGTGGCGCAGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 10
FZ0270	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCCTGGCCCAGGTGAAGGTGGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 11
FZ0271	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCGGTCCCTGGCCCAGGTGAGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 12
FZ0272	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAACCGGAGGACAAAGTACAGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 13
FZ0273	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGGTGAAGGTGTGGTTCCAGGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 14
FZ0274	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGTGAAGGTGTGGTTCCAGAGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 15
FZ0275	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAACCGGAGGACAAAGTACAGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 16
FZ0276	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCCTGGCCCAGGTGAAGGTGGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 17
FZ0277	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGGTGAAGGTGTGGTTCCAGGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 18
FZ0278	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGGACAAAGTACAAACGGCGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 19
FZ0279	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGAGGGAGGGGCACAGATGGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 20
FZ0280	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCACCTTCAGCTGGGCCAGGGAGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 21
FZ0281	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACACCGTAGTCATTGGAGGTGAGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 22
FZ0282	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCAGAGCAGCCACTGGGGCCTGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 23
FZ0283	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGACCTTCAGCTGGGCCAGGGAGGTGTTTCGTCCTTTCCACAAG	Left guide RNA 24
FZ0284	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCCCATTGGCCTGCTTCGTGGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 1
FZ0285	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACATTGGCCTGCTTCGTGGCAAGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 2
FZ0286	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCCTCCTGCAGCTTCTGCCGGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 3
FZ0287	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCTCCAGCTTCTGCCGTTTGGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 4
FZ0288	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACATTGGCCTGCTTCGTGGCAAGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 5
FZ0289	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGCAGCAAGCAGCACTCTGCCGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 6
FZ0290	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTTCTTCTTCTGCTCGGACTCGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 7
FZ0291	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACACCGGAGGACAAAGTACAAAGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 8
FZ0292	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCTTCTTCTGCTCGGACTCAGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 9
FZ0293	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGTTGATGTGATGGGAGGCGCTGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 10
FZ0294	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGCCAGGGAGGGAGGGGCAGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 11
FZ0295	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGGGAGGGAGGGGCACAGATGGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 12
FZ0296	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCGGTTCTGGAACCACACCTGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 13
FZ0297	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCACCTGGGCCAGGGAGGCGACGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 14
FZ0298	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCACCTGGGCCAGGGAGGGGAGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 15
FZ0299	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGTTCTGGAACCACACCTTCAGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 16
FZ0300	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGGAGGGAGGGGCACAGATGGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 17
FZ0301	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGCCAGGGAGGGAGGGGCAGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 18
FZ0302	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGTTCTGGAACCACACCTTCAGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 19
FZ0303	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGGTGAAGGTGTGGTTCCAGGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 20
FZ0304	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAACCGGAGGACAAAGTACAGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 21
FZ0305	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCAAACCCACGAGGGCAGAGTGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 22
FZ0306	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGTTTCTCATCTGTGCCCCGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 23
FZ0307	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGGACAAAGTACAAACGGCGGTGTTTCGTCCTTTCCACAAG	Right guide RNA 24

FIG. 52

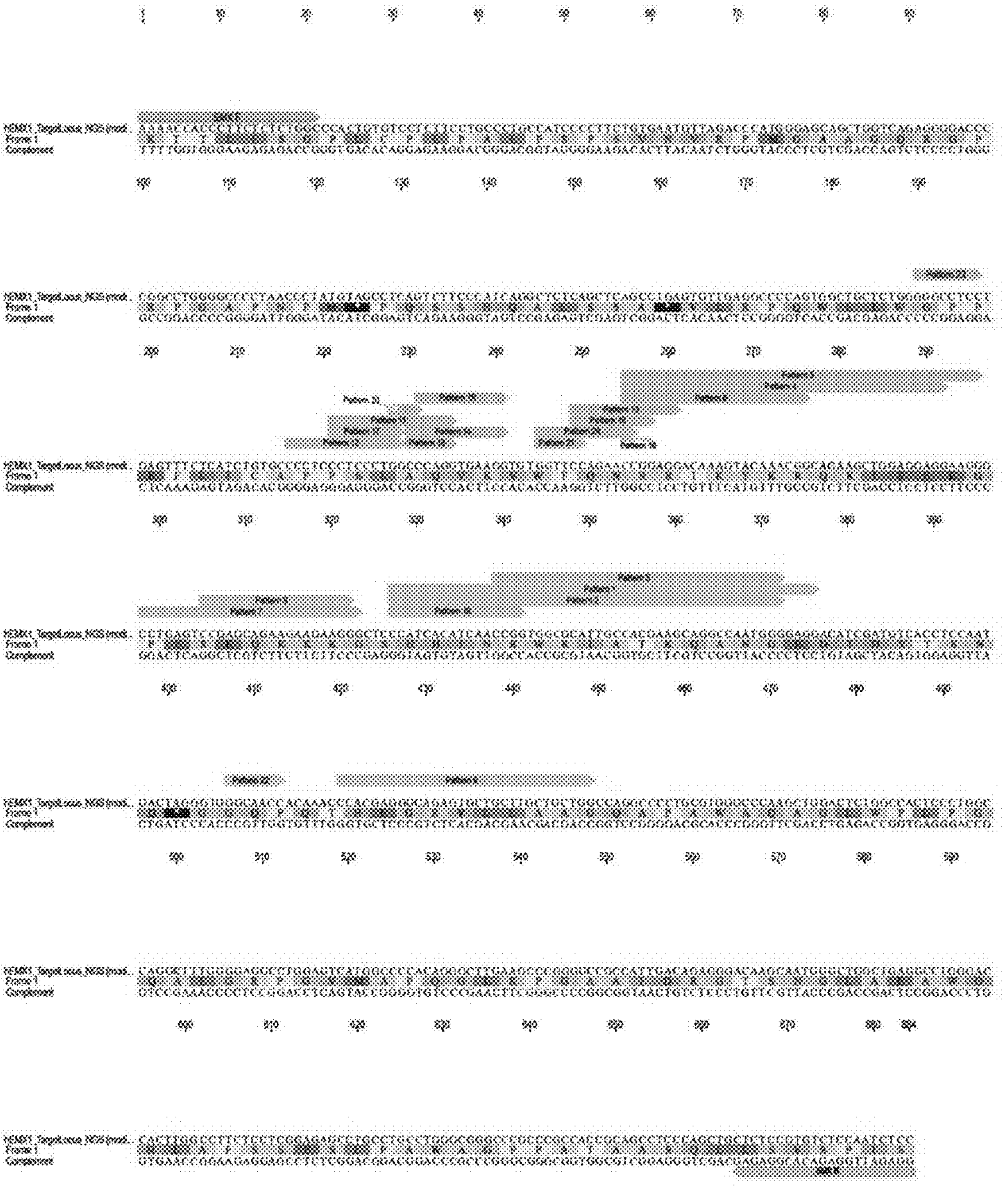


FIG. 53

lane #	Cas9 to use	Left gRNA	Right gRNA	Overhang (D10A) - number of bases protruding from 5'
1	D10A	left 23	left 12	-36
2	D10A	right 4	left 9	-25
3	D10A	left 23	right 23	-16
4	D10A	right 7	left 10	-15
5	D10A	right 16	left 3	-8
6	D10A	right 22	right 6	26
7	D10A	left 12	right 16	31
8	D10A	left 12	right 13	34
9	D10A	left 10	right 1	38
10	D10A	right 23	right 16	51
11	D10A	right 23	right 13	54
12	D10A	left 3	right 7	57
13	D10A	left 12	right 4	65
14	D10A	left 12	right 3	69
15	D10A	left 3	right 10	76
16	D10A	right 23	right 4	85
17	D10A	left 12	right 9	95
18	D10A	left 12	right 10	115
19	D10A	right 23	right 10	135
20	D10A	left 12	right 2	145
21	D10A	left 12	left 22	181
22	D10A	right 23	left 22	201
23	D10A	left 12	right 6	222
24	D10A	right 23	right 6	242

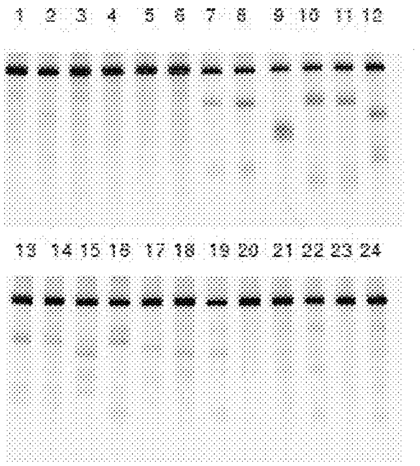


FIG. 54

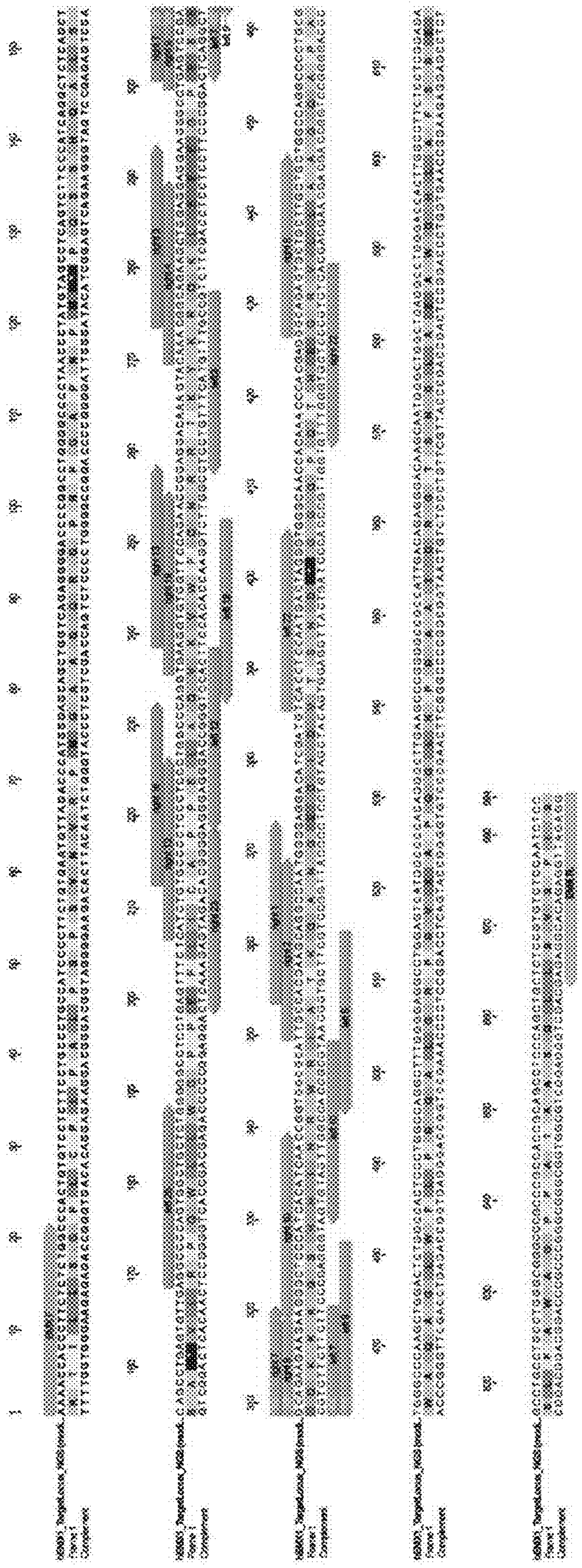


FIG. 55

FUNCTIONAL GENOMICS USING CRISPR-CAS SYSTEMS, COMPOSITIONS, METHODS, KNOCK OUT LIBRARIES AND APPLICATIONS THEREOF

RELATED APPLICATIONS AND INCORPORATION BY REFERENCE

[0001] This application is a continuation of US international application PCT/US2013/074800 filed Dec. 12, 2013, which claims benefit of and priority to U.S. provisional patent application Nos. 61/736,527 filed Dec. 12, 2012 and 61/802,174 filed Mar. 15, 2013.

[0002] Reference is also made to U.S. provisional patent application Nos. 61/960,777 filed on Sep. 25, 2013 and 61/961,980 filed on Oct. 28, 2013. Reference is made to U.S. provisional patent applications 61/758,468; 61/769,046; 61/802,174; 61/806,375; 61/814,263; 61/819,803 and 61/828,130 filed on Jan. 30, 2013; Feb. 25, 2013; Mar. 15, 2013; Mar. 28, 2013; Apr. 20, 2013; May 6, 2013 and May 28, 2013 respectively. Reference is also made to U.S. provisional patent applications 61/836,123, 61/847,537, 61/862,355 and 61/871,301 filed on Jun. 17, 2013; Jul. 17, 2013, Aug. 5, 2013 and Aug. 28, 2013 respectively. Reference is also made to U.S. provisional patent applications 61/736,527 and 61/748,427 on Dec. 12, 2012 and Jan. 2, 2013, respectively. Reference is also made to U.S. provisional patent application 61/791,409 filed on Mar. 15, 2013. Reference is also made to U.S. provisional patent application 61/799,800 filed Mar. 15, 2013. Reference is also made to U.S. provisional patent applications 61/835,931, 61/835,936, 61/836,127, 61/836,101, 61/836,080, and 61/835,973 each filed Jun. 17, 2013.

[0003] The foregoing applications, and all documents cited therein or during their prosecution ("appln cited documents") and all documents cited or referenced in the appln cited documents, and all documents cited or referenced herein ("herein cited documents"), and all documents cited or referenced in herein cited documents, together with any manufacturer's instructions, descriptions, product specifications, and product sheets for any products mentioned herein or in any document incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention. More specifically, all referenced documents are incorporated by reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0004] This invention was made with government support under the NIH Pioneer Award (1DP1MH100706) awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD OF THE INVENTION

[0005] The present invention generally relates to compositions, methods, applications and screens used in functional genomics that focus on gene function in a cell and that may use vector systems and other aspects related to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas systems and components thereof.

BACKGROUND OF THE INVENTION

[0006] Recent advances in genome sequencing techniques and analysis methods have significantly accelerated the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. Functional genomics is a field of molecular biology that may be considered to utilize the vast wealth of data produced by genomic projects (such as genome sequencing projects) to describe gene (and protein) functions and interactions. Contrary to classical genomics, functional genomics focuses on the dynamic aspects such as gene transcription, translation, and protein-protein interactions, as opposed to the static aspects of the genomic information such as DNA sequence or structures, though these static aspects are very important and supplement one's understanding of cellular and molecular mechanisms. Functional genomics attempts to answer questions about the function of DNA at the levels of genes, RNA transcripts, and protein products. A key characteristic of functional genomics studies is a genome-wide approach to these questions, generally involving high-throughput methods rather than a more traditional "gene-by-gene" approach. Given the vast inventory of genes and genetic information it is advantageous to use genetic screens to provide information of what these genes do, what cellular pathways they are involved in and how any alteration in gene expression can result in a particular biological process.

[0007] Functional genomic screens and libraries attempt to characterize gene function in the context of living cells and hence are likely to generate biologically significant data. There are three key elements for a functional genomics screen: a good reagent to perturb the gene, a good tissue culture model and a good readout of cell state. Good reagents that allow for precise genome targeting technologies are needed to enable systematic reverse engineering of causal genetic variations by allowing selective perturbation of individual genetic elements, as well as to advance synthetic biology, biotechnological, and medical applications. Although genome-editing techniques such as designer zinc fingers, transcription activator-like effectors (TALEs), or homing meganucleases are available for producing targeted genome perturbations, there remains a need for new genome engineering technologies that are affordable, easy to set up, scalable, and amenable to targeting multiple positions within the eukaryotic genome.

SUMMARY OF THE INVENTION

[0008] The CRISPR-Cas system does not require the generation of customized proteins to target specific sequences but rather a single Cas enzyme can be programmed by a short RNA molecule to recognize a specific DNA target. Adding the CRISPR-Cas system to the repertoire of genome sequencing techniques and analysis methods may significantly simplify the methodology and accelerate the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. To utilize the CRISPR-Cas system effectively for genome editing without deleterious effects, it is critical to understand aspects of engineering, optimization and tissue/organ specific delivery of these genome engineering tools, which are aspects of the claimed invention.

[0009] There exists a pressing need for alternative and robust systems and techniques for sequence targeting with a wide array of applications. Aspects of this invention address this need and provide related advantages. An exemplary CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within the target polynucleotide. The guide sequence is linked to a tracr mate sequence, which in turn hybridizes to a tracr sequence.

[0010] One aspect of the invention comprehends a genome wide library that may comprise a plurality of CRISPR-Cas system guide RNAs that may comprise guide sequences that are capable of targeting a plurality of target sequences in a plurality of genomic loci, wherein said targeting results in a knockout of gene function. This library may potentially comprise guide RNAs that target each and every gene in the genome of an organism. In some embodiments of the invention the organism or subject is a eukaryote (including mammal including human) or a non-human eukaryote or a non-human animal or a non-human mammal. In some embodiments, the organism or subject is a non-human animal, and may be an arthropod, for example, an insect, or may be a nematode. In some methods of the invention the organism or subject is a plant. In some methods of the invention the organism or subject is a mammal or a non-human mammal. A non-human mammal may be for example a rodent (preferably a mouse or a rat), an ungulate, or a primate. In some methods of the invention the organism or subject is algae, including microalgae, or is a fungus.

[0011] In another aspect, the invention provides a method of generating a gene knockout cell library comprising introducing into each cell in a population of cells a vector system of one or more vectors that may comprise an engineered, non-naturally occurring CRISPR-Cas system comprising I. a Cas protein, and II. one or more guide RNAs of the library of the invention, wherein components I and II may be on the same or on different vectors of the system, integrating components I and II into each cell, wherein the guide sequence targets a unique gene in each cell,

wherein the Cas protein is operably linked to a regulatory element, wherein when transcribed, the guide RNA comprising the guide sequence directs sequence-specific binding of a CRISPR-Cas system to a target sequence in the genomic loci of the unique gene, inducing cleavage of the genomic loci by the Cas protein, and confirming different knockout mutations in a plurality of unique genes in each cell of the population of cells thereby generating a gene knockout cell library. In an embodiment of the invention, the Cas protein is a Cas9 protein. In another embodiment, the one or more vectors are plasmid vectors. In a further embodiment, the regulatory element operably linked to the Cas protein is an inducible promoter, e.g. a doxycycline inducible promoter. The invention comprehends that the population of cells is a population of eukaryotic cells, and in a preferred embodiment, the population of cells is a population of embryonic stem (ES) cells. In another embodiment the confirming of different knockout mutations is by whole exome sequencing. The invention also provides kits that comprise the genome wide libraries mentioned herein. The kit may comprise a single container comprising vectors or plasmids comprising the library of the invention. The kit may also comprise a panel comprising a selection of unique CRISPR-Cas system guide RNAs comprising guide sequences from the library of the invention, wherein the selection is indicative of a particular physiological condition. The invention comprehends that the targeting is of about 100 or more sequences, about 1000 or more sequences or about 20,000 or more sequences or the entire genome. Furthermore, a panel of target sequences may be focused on a relevant or desirable pathway, such as an immune pathway or cell division.

[0012] In another aspect the invention provides for use of genome wide libraries for functional genomic studies. Such studies focus on the dynamic aspects such as gene transcription, translation, and protein-protein interactions, as opposed to the static aspects of the genomic information such as DNA sequence or structures, though these static aspects are very important and supplement one's understanding of cellular and molecular mechanisms. Functional genomics attempts to answer questions about the function of DNA at the levels of genes, RNA transcripts, and protein products. A key characteristic of functional genomics studies is a genome-wide approach to these questions, generally involving high-throughput methods rather than a more traditional "gene-by-gene" approach. Given the vast inventory of genes and genetic information it is advantageous to use genetic screens to provide information of what these genes do, what cellular pathways they are involved in and how any alteration in gene expression can result in particular biological process.

[0013] In one aspect, the invention provides methods for using one or more elements of a CRISPR-Cas system. The CRISPR complex of the invention provides an effective means for modifying a target polynucleotide. The CRISPR complex of the invention has a wide variety of utilities including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target polynucleotide in a multiplicity of cell types in various tissues and organs. As such the CRISPR complex of the invention has a broad spectrum of applications in, e.g., gene or genome editing, gene therapy, drug discovery, drug screening, disease diagnosis, and prognosis.

[0014] Aspects of the invention relate to Cas9 enzymes having improved target specificity in a CRISPR-Cas9 system having guide RNAs having optimal activity, smaller in length than wild-type Cas9 enzymes and nucleic acid molecules coding therefor, and chimeric Cas9 enzymes, as well as methods of improving the target specificity of a Cas9 enzyme or of designing a CRISPR-Cas9 system comprising designing or preparing guide RNAs having optimal activity and/or selecting or preparing a Cas9 enzyme having a smaller size or length than wild-type Cas9 whereby packaging a nucleic acid coding therefor into a delivery vector is advanced as there is less coding therefor in the delivery vector than for wild-type Cas9, and/or generating chimeric Cas9 enzymes.

[0015] Also provided are uses of the present sequences, vectors, enzymes or systems, in medicine or in therapy. Also provided are uses of the same in gene or genome editing. Also provided are the present sequences, vectors, enzymes, or systems for use in medicine or in therapy; or for use in gene or genome editing. Still further provided are uses of the present sequences, vectors, enzymes, or systems in the manufacture of a medicament.

[0016] In an additional aspect of the invention, a Cas9 enzyme may comprise one or more mutations and may be used as a generic DNA binding protein with or without fusion to a functional domain. The mutations may be artificially introduced mutations or gain- or loss-of-function mutations. The mutations may include but are not limited to mutations in one of the catalytic domains (D10 and H840) in the RuvC and HNH catalytic domains, respectively. Further mutations have been characterized. In one aspect of the invention, the functional domain may be a transcriptional activation domain, which may be VP64. In other aspects of the invention, the functional domain may be a transcriptional repressor domain, which may be KRAB or SID4X. Other aspects of the invention relate to the mutated Cas 9 enzyme being fused to domains which include but are not limited to a transcriptional activator, repressor, a recombinase, a transposase, a histone remodeler, a demethylase, a DNA methyltransferase, a cryptochrome, a light inducible/controllable domain or a chemically inducible/controllable domain.

[0017] In a further embodiment, the invention provides for methods to generate mutant tracrRNA and direct repeat sequences or mutant chimeric guide sequences that allow for enhancing performance of these RNAs in cells. Aspects of the invention also provide for selection of said sequences.

[0018] Aspects of the invention also provide for methods of simplifying the cloning and delivery of components of the CRISPR complex. In a preferred embodiment of the invention, a suitable promoter, such as the U6 promoter, is amplified with a DNA oligo and added onto the guide RNA. The resulting PCR product can then be transfected into cells to drive expression of the guide RNA. Aspects of the invention also relate to the guide RNA being transcribed in vitro or ordered from a synthesis company and directly transfected.

[0019] In one aspect, the invention provides for methods to improve activity by using a more active polymerase. In a preferred embodiment, the expression of guide RNAs under the control of the T7 promoter is driven by the expression of the T7 polymerase in the cell. In an advantageous embodiment, the cell is a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a human cell. In a more preferred embodiment the human cell is a patient specific cell.

[0020] In one aspect, the invention provides for methods of reducing the toxicity of Cas enzymes. In certain aspects, the Cas enzyme is any Cas9 as described herein, for instance any naturally-occurring bacterial Cas9 as well as any chimaeras, mutants, homologs or orthologs. In a preferred embodiment, the Cas9 is delivered into the cell in the form of mRNA. This allows for the transient expression of the enzyme thereby reducing toxicity. In another preferred embodiment, the invention also provides for methods of expressing Cas9 under the control of an inducible promoter, and the constructs used therein.

[0021] In another aspect, the invention provides for methods of improving the in vivo applications of the CRISPR-Cas system. In the preferred embodiment, the Cas enzyme is wildtype Cas9 or any of the modified versions described herein, including any naturally-occurring bacterial Cas9 as well as any chimaeras, mutants, homologs or orthologs. An advantageous aspect of the invention provides for the selection of Cas9 homologs that are easily packaged into viral vectors for delivery. Cas9 orthologs typically share the general organization of 3-4 RuvC domains and a HNH domain. The 5' most RuvC domain cleaves the non-complementary strand, and the HNH domain cleaves the complementary strand. All notations are in reference to the guide sequence.

[0022] The catalytic residue in the 5' RuvC domain is identified through homology comparison of the Cas9 of interest with other Cas9 orthologs (from *S. pyogenes* type II CRISPR locus, *S. thermophilus* CRISPR locus 1, *S. thermophilus* CRISPR locus 3, and *Franciscilla novicida* type II CRISPR locus), and the conserved Asp residue (D10) is mutated to alanine to convert Cas9 into a complementary-strand nicking enzyme. Similarly, the conserved His and Asn residues in the HNH domains are mutated to Alanine to convert Cas9 into a non-complementary-strand nicking enzyme. In some embodiments, both sets of mutations may be made, to convert Cas9 into a non-cutting enzyme.

[0023] In some embodiments, the CRISPR enzyme is a type I or III CRISPR enzyme, preferably a type II CRISPR enzyme. This type II CRISPR enzyme may be any Cas enzyme. A preferred Cas enzyme may be identified as Cas9 as this can refer to the general class of enzymes that share homology to the biggest nuclease with multiple nuclease domains from the type II CRISPR system. Most preferably, the Cas9 enzyme is from, or is derived from, spCas9 or saCas9. By derived, Applicants mean that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as described herein

[0024] It will be appreciated that the terms Cas and CRISPR enzyme are generally used herein interchangeably, unless otherwise apparent. As mentioned above, many of the residue numberings used herein refer to the Cas9 enzyme from the type II CRISPR locus in *Streptococcus pyogenes*. However, it will be appreciated that this invention includes many more Cas9s from other species of microbes, such as SpCas9, SaCas9, St1Cas9 and so forth. Further examples are provided herein. The skilled person will be able to determine appropriate corresponding residues in Cas9 enzymes other than SpCas9 by comparison of the relevant amino acid sequences. Thus, where a specific amino acid replacement is referred to using the SpCas9 numbering, then, unless the context makes it apparent this is not intended to refer to other Cas9 enzymes, the disclosure is intended to encompass corresponding modifications in other Cas9 enzymes.

[0025] An example of a codon optimized sequence, in this instance optimized for humans (i.e. being optimized for expression in humans) is provided herein, see the SaCas9 human codon optimized sequence. Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species is known.

[0026] In further embodiments, the invention provides for methods of enhancing the function of Cas9 by generating chimeric Cas9 proteins. Chimeric Cas9 proteins may be new Cas9 containing fragments from more than one naturally occurring Cas9. These methods may comprise fusing N-terminal fragments of one Cas9 homolog with C-terminal fragments of another Cas9 homolog. These methods also allow for the selection of new properties displayed by the chimeric Cas9 proteins.

[0027] It will be appreciated that in the present methods, where the organism is an animal or a plant, the modification may occur ex vivo or in vitro, for instance in a cell culture and in some instances not in vivo. In other embodiments, it may occur in vivo. Where the modification occurs ex vivo or in vitro, a modified cell may be used to generate a complete organism, or a modified cell may be introduced or reintroduced into a host organism.

[0028] In one aspect, the invention provides a method of modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest comprising: delivering a non-naturally occurring or engineered composition comprising:

A)—I. a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises:

(a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell,

(b) a tracr mate sequence, and

(c) a tracr sequence, and

II. a polynucleotide sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences,

wherein (a), (b) and (c) are arranged in a 5' to 3' orientation, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and

wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence and the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA,

or

(B) I. polynucleotides comprising:

(a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, and

(b) at least one or more tracr mate sequences,

II. a polynucleotide sequence encoding a CRISPR enzyme, and

III. a polynucleotide sequence comprising a tracr sequence,

wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and

wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence, and the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA.

[0029] Any or all of the polynucleotide sequence encoding a CRISPR enzyme, guide sequence, tracr mate sequence or tracr sequence, may be RNA. The polynucleotides encoding the sequence encoding a CRISPR enzyme, the guide sequence, tracr mate sequence or tracr sequence may be RNA and may be delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun.

[0030] It will be appreciated that where reference is made to a polynucleotide, which is RNA and is said to 'comprise' a feature such as a tracr mate sequence, the RNA sequence includes the feature. Where the polynucleotide is DNA and is said to comprise a feature such as a tracr mate sequence, the DNA sequence is or can be transcribed into the RNA including the feature at issue. Where the feature is a protein, such as the CRISPR enzyme, the DNA or RNA sequence referred to is, or can be, translated (and in the case of DNA transcribed first).

[0031] Accordingly, in certain embodiments the invention provides a method of modifying an organism, e.g., mammal including human or a non-human mammal or organism by manipulation of a target sequence in a genomic locus of interest comprising delivering a non-naturally occurring or engineered composition comprising a viral or plasmid vector system comprising one or more viral or plasmid vectors operably encoding a composition for expression thereof, wherein the composition comprises: (A) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, (b) a tracr mate sequence, and (c) a tracr sequence, and II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences (or optionally at least one or more nuclear localization sequences as some embodiments can involve no NLS), wherein (a), (b) and (c) are arranged in a 5' to 3' orientation, wherein components I and II are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence, or (B) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, and (b) at least one or more tracr mate sequences, II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and III. a third regulatory element operably linked to a tracr sequence, wherein components I, II and III are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence. In some embodiments, components I, II and III are located on the same vector. In other embodiments, components I and II are located on the same vector, while component III is located on another vector. In other embodiments, components I and III are located on the same vector, while component II is located on another vector. In other embodiments, components II and III are located on the same vector, while component I is located on another vector. In other embodiments, each of components I, II and III is located on different vectors. The invention also provides a viral or plasmid vector system as described herein.

[0032] Preferably, the vector is a viral vector, such as a lenti- or baculo- or preferably adeno-viral/adeno-associated viral vectors, but other means of delivery are known (such as yeast systems, microvesicles, gene guns/means of attaching vectors to gold nanoparticles) and are provided. In some embodiments, one or more of the viral or plasmid vectors may be delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun.

[0033] By manipulation of a target sequence, Applicants also mean the epigenetic manipulation of a target sequence. This may be of the chromatin state of a target sequence, such as by modification of the methylation state of the target sequence (i.e. addition or removal of methylation or methylation patterns or CpG islands), histone modification, increasing or reducing accessibility to the target sequence, or by promoting 3D folding.

[0034] It will be appreciated that where reference is made to a method of modifying an organism or mammal including human or a non-human mammal or organism by manipulation of a target sequence in a genomic locus of interest, this may apply to the organism (or mammal) as a whole or just a single cell or population of cells from that organism (if the organism is multicellular). In the case of humans, for instance, Applicants envisage, inter alia, a single cell or a population of cells and these may preferably be modified ex vivo and then re-introduced. In this case, a biopsy or other tissue or biological fluid sample may be necessary. Stem cells are also particularly preferred in this regard. But, of course, in vivo embodiments are also envisaged.

[0035] In certain embodiments the invention provides a method of treating or inhibiting a condition caused by a defect in a target sequence in a genomic locus of interest in a subject (e.g., mammal or human) or a non-human subject (e.g., mammal) in need thereof comprising modifying the subject or a non-human subject by manipulation of the target sequence and wherein the condition is susceptible to treatment or inhibition by manipulation of the target sequence comprising providing treatment comprising: delivering a non-naturally occurring or engineered composition comprising an AAV or lentivirus vector system comprising one or more AAV or lentivirus vectors operably encoding a composition for expression thereof, wherein the target sequence is manipulated by the composition when expressed, wherein the composition comprises: (A) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, (b) a tracr mate sequence, and (c) a tracr sequence, and II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences (or optionally at least one or more nuclear localization sequences as some embodiments can involve no NLS) wherein (a), (b) and (c) are arranged in a 5' to 3' orientation, wherein components I and II are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence, or (B) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, and (b) at least one or more tracr mate sequences, II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and III. a third regulatory element operably linked to a tracr sequence, wherein components I, II and III are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence. In some embodiments, components I, II and III are located on the same vector. In other embodiments, components I and II are located on the same vector, while component III is located on another vector. In other embodiments, components I and III are located on the same vector, while component II is located on another vector. In other embodiments, components II and III are located on the same vector, while component I is located on another vector. In other embodiments, each of components I, II and III is located on different vectors. The invention also provides a viral (e.g. AAV or lentivirus) vector system as described herein.

[0036] Some methods of the invention can include inducing expression. In some methods of the invention the organism or subject is a eukaryote (including mammal including human) or a non-human eukaryote or a non-human animal or a non-human mammal. In some embodiments, the organism or subject is a non-human animal, and may be an arthropod, for example, an insect, or may be a nematode. In some methods of the invention the organism or subject is a plant. In some methods of the invention the organism or subject is a mammal or a non-human mammal. A non-human mammal may be for example a rodent (preferably a mouse or a rat), an ungulate, or a primate. In some methods of the invention the organism or subject is algae, including microalgae, or is a fungus. In some methods of the invention the viral vector is an AAV or a lentivirus, and can be part of a vector system as described herein. In some methods of the invention the CRISPR enzyme is a Cas9. In some methods of the invention the expression of the guide sequence is under the control of the T7 promoter and is driven by the expression of T7 polymerase.

[0037] The invention in some embodiments comprehends a method of delivering a CRISPR enzyme comprising delivering to a cell mRNA encoding the CRISPR enzyme. In some of these methods the CRISPR enzyme is a Cas9.

[0038] The invention also provides methods of preparing the vector systems of the invention, in particular the viral vector systems as described herein. The invention in some embodiments comprehends a method of preparing the AAV of the invention comprising transfecting plasmid(s) containing or consisting essentially of nucleic acid molecule(s) coding for the AAV into AAV-infected cells, and supplying AAV rep and/or cap obligatory for replication and packaging of the AAV. In some embodiments the AAV rep and/or cap obligatory for replication and packaging of the AAV are supplied by transfecting the cells with helper plasmid(s) or helper virus(es). In some embodiments the helper virus is a poxvirus, adenovirus, herpesvirus or baculovirus. In some embodiments the poxvirus is a vaccinia virus. In some embodiments the cells are mammalian cells. And in some embodiments the cells are insect cells and the helper virus is baculovirus. In other embodiments, the virus is a lentivirus.

[0039] In plants, pathogens are often host-specific. For example, *Fusarium oxyvsporum* f. sp. *lycopersici* causes tomato wilt but attacks only tomato, and *F. oxysporum f. dianthii Puccinia graminis* f. sp. *tritici* attacks only wheat. Plants have existing and induced defenses to resist most pathogens. Mutations and recombination events across plant generations lead to genetic variability that gives rise to susceptibility, especially as pathogens reproduce with more frequency than plants. In plants there can be non-host resistance, e.g., the host and pathogen are incompatible. There can also be Horizontal Resistance, e.g., partial resistance against all races of a pathogen, typically controlled by many genes and Vertical Resistance, e.g., complete resistance to some races of a pathogen but not to other races, typically controlled by a few genes. In a Gene-for-Gene level, plants and pathogens evolve together, and the genetic changes in one balance changes in other. Accordingly, using Natural Variability, breeders combine most useful genes for Yield, Quality, Uniformity, Hardiness, Resistance. The sources of resistance genes include native or foreign Varieties, Heirloom Varieties, Wild Plant Relatives, and Induced Mutations, e.g., treating plant material with mutagenic agents. Using the present invention, plant breeders are provided with a new tool to induce mutations. Accordingly, one skilled in the art can analyze the genome of sources of resistance genes, and in Varieties having desired characteristics or traits employ the present invention to induce the rise of resistance genes, with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs.

[0040] The invention further comprehends a composition of the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme) for use in medicine or in therapy. In some embodiments the invention comprehends a composition according to the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme) for use in a method according to the invention. In some embodiments the invention provides for the use of a composition of the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme) in ex vivo gene or genome editing. In certain embodiments the invention comprehends use of a composition of the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme) in the manufacture of a medicament for ex vivo gene or genome editing or for use in a method according of the invention. The invention comprehends in some embodiments a composition of the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme), wherein the target sequence is flanked at its 3' end by a PAM (protospacer adjacent motif) sequence comprising 5'-motif, especially where the Cas9 is (or is derived from) *S. pyogenes* or *S. aureus* Cas9. For example, a suitable PAM is 5'-NRG or 5'-NNGRR (where N is any Nucleotide) for SpCas9 or SaCas9 enzymes (or derived enzymes), respectively, as mentioned below.

[0041] It will be appreciated that SpCas9 or SaCas9 are those from or derived from *S. pyogenes* or *S. aureus* Cas9.

[0042] Aspects of the invention comprehend improving the specificity of a CRISPR enzyme, e.g. Cas9, mediated gene targeting and reducing the likelihood of off-target modification by the CRISPR enzyme, e.g. Cas9. The invention in some embodiments comprehends a method of modifying an organism or a non-human organism with a reduction in likelihood of off-target modifications by manipulation of a first and a second target sequence on opposite strands of a DNA duplex in a genomic locus of interest in a cell comprising delivering a non-naturally occurring or engineered composition comprising:

[0043] I. a first CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence,

wherein the first polynucleotide sequence comprises:

(a) a first guide sequence capable of hybridizing to the first target sequence,

(b) a first tracr mate sequence, and

(c) a first tracr sequence,

[0044] II. a second CRISPR-Cas system chiRNA polynucleotide sequence, wherein the second polynucleotide sequence comprises:

(a) a second guide sequence capable of hybridizing to the second target sequence,

(b) a second tracr mate sequence, and

(c) a second tracr sequence, and

[0045] III. a polynucleotide sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences and comprising one or more mutations, wherein (a), (b) and (c) are arranged in a 5' to 3' orientation, wherein when transcribed, the first and the second tracr mate sequence hybridize to the first and second tracr sequence respectively and the first and the second guide sequence directs sequence-specific binding of a first and a second CRISPR complex to the first and second target sequences respectively, wherein the first CRISPR complex comprises the CRISPR enzyme complexed with (1) the first guide sequence that is hybridized to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tract mate sequence that is hybridized to the second tracr sequence, wherein the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA, and wherein the first guide sequence directs cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directs cleavage of the other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism with a reduction in likelihood of off-target modifications.

[0046] In some methods of the invention any or all of the polynucleotide sequence encoding the CRISPR enzyme, the first and the second guide sequence, the first and the second tracr mate sequence or the first and the second tracr sequence, is/are RNA. In further embodiments of the invention the polynucleotides encoding the sequence encoding the CRISPR enzyme, the first and the second guide sequence, the first and the second tracr mate sequence or the first and the second tracr sequence, is/are RNA and are delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun. In certain embodiments of the invention, the first and second tracr mate sequence share 100% identity and/or the first and second tracr sequence share 100% identity. In some embodiments, the polynucleotides may be comprised within a vector system comprising one or more vectors. In preferred embodiments of the invention the CRISPR enzyme is a Cas9 enzyme, e.g. SpCas9. In an aspect of the invention the CRISPR enzyme comprises one or more mutations in a catalytic domain, wherein the one or more mutations are selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the CRISPR enzyme has the D10A mutation. In preferred embodiments, the first CRISPR enzyme has one or more mutations such that the enzyme is a complementary strand nicking enzyme, and the second CRISPR enzyme has one or more mutations such that the enzyme is a non-complementary strand nicking enzyme. Alternatively the first enzyme may be a non-complementary strand nicking enzyme, and the second enzyme may be a complementary strand nicking enzyme.

[0047] In preferred methods of the invention the first guide sequence directing cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directing cleavage of the other strand near the second target sequence results in a 5' overhang. In embodiments of the invention the 5' overhang is at most 200 base pairs, preferably at most 100 base pairs, or more preferably at most 50 base pairs. In embodiments of the invention the 5' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs.

[0048] The invention in some embodiments comprehends a method of modifying an organism or a non-human organism with a reduction in likelihood of off-target modifications by manipulation of a first and a second target sequence on opposite strands of a DNA duplex in a genomic locus of interest in a cell comprising delivering a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising

[0049] I. a first regulatory element operably linked to

(a) a first guide sequence capable of hybridizing to the first target sequence, and

(b) at least one or more tracr mate sequences,

[0050] II. a second regulatory element operably linked to

(a) a second guide sequence capable of hybridizing to the second target sequence, and

(b) at least one or more tracr mate sequences,

[0051] III. a third regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and

[0052] IV. a fourth regulatory element operably linked to a tracr sequence,

[0053] wherein components I, II, III and IV are located on the same or different vectors of the system, when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the first and the second guide sequence direct sequence-specific binding of a first and a second CRISPR complex to the first and second target sequences respectively, wherein the first CRISPR complex comprises the CRISPR enzyme complexed with (1) the first guide sequence that is hybridized to the first target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence, wherein the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA, and wherein the first guide sequence directs cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directs cleavage of the other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism with a reduction in likelihood of off-target modifications.

[0054] The invention also provides a vector system as described herein. The system may comprise one, two, three or four different vectors. Components I, II, III and IV may thus be located on one, two, three or four different vectors, and all combinations for possible locations of the components are herein envisaged, for example: components I, II, III and IV can be located on the same vector; components I, II, III and IV can each be located on different vectors; components I, II, III and IV may be located on a total of two or three different vectors, with all combinations of locations envisaged, etc.

[0055] In some methods of the invention any or all of the polynucleotide sequence encoding the CRISPR enzyme, the first and the second guide sequence, the first and the second tracr mate sequence or the first and the second tracr sequence, is/are RNA. In further embodiments of the invention the first and second tracr mate sequence share 100% identity and/or the first and second tracr sequence share 100% identity. In preferred embodiments of the invention the CRISPR enzyme is a Cas9 enzyme, e.g. SpCas9. In an aspect of the invention the CRISPR enzyme comprises one or more mutations in a catalytic domain, wherein the one or more mutations are selected from the group consisting of D10A, E762A, H840A. N854A, N863A and D986A. In a highly preferred embodiment the CRISPR enzyme has the D10A mutation. In preferred embodiments, the first CRISPR enzyme has one or more mutations such that the enzyme is a complementary strand nicking enzyme, and the second CRISPR enzyme has one or more mutations such that the enzyme is a non-complementary strand nicking enzyme. Alternatively the first enzyme may be a non-complementary strand nicking enzyme, and the second enzyme may be a complementary strand nicking enzyme. In a further embodiment of the invention, one or more of the viral vectors are delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun.

[0056] In preferred methods of the invention the first guide sequence directing cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directing cleavage of other strand near the second target sequence results in a 5' overhang. In embodiments of the invention the 5' overhang is at most 200 base pairs, preferably at most 100 base pairs, or more preferably at most 50 base pairs. In embodiments of the invention the 5' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs.

[0057] The invention in some embodiments comprehends a method of modifying a genomic locus of interest with a reduction in likelihood of off-target modifications by introducing into a cell containing and expressing a double stranded DNA molecule encoding a gene product of interest an engineered, non-naturally occurring CRISPR-Cas system comprising a Cas protein having one or more mutations and two guide RNAs that target a first strand and a second strand of the DNA molecule respectively, whereby the guide RNAs target the DNA molecule encoding the gene product and the Cas protein nicks each of the first strand and the second strand of the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas protein and the two guide RNAs do not naturally occur together.

[0058] In preferred methods of the invention the Cas protein nicking each of the first strand and the second strand of the DNA molecule encoding the gene product results in a 5' overhang. In embodiments of the invention the 5' overhang is at most 200 base pairs, preferably at most 100 base pairs, or more preferably at most 50 base pairs. In embodiments of the invention the 5' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs.

[0059] Embodiments of the invention also comprehend the guide RNAs comprising a guide sequence fused to a tracr mate sequence and a tracr sequence. In an aspect of the invention the Cas protein is codon optimized for expression in a eukaryotic cell, preferably a mammalian cell or a human cell. In further embodiments of the invention the Cas protein is a type II CRISPR-Cas protein, e.g. a Cas 9 protein. In a highly preferred embodiment the Cas protein is a Cas9 protein, e.g. SpCas9. In aspects of the invention the Cas protein has one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the Cas protein has the D10A mutation.

[0060] Aspects of the invention relate to the expression of the gene product being decreased or a template polynucleotide being further introduced into the DNA molecule encoding the gene product or an intervening sequence being excised precisely by allowing the two 5' overhangs to reanneal and ligate or the activity or function of the gene product being altered or the expression of the gene product being increased. In an embodiment of the invention, the gene product is a protein.

[0061] The invention also comprehends an engineered, non-naturally occurring CRISPR-Cas system comprising a Cas protein having one or more mutations and two guide RNAs that target a first strand and a second strand respectively of a double stranded DNA molecule encoding a gene product in a cell, whereby the guide RNAs target the DNA molecule encoding the gene product and the Cas protein nicks each of the first strand and the second strand of the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas protein and the two guide RNAs do not naturally occur together.

[0062] In aspects of the invention the guide RNAs may comprise a guide sequence fused to a tracr mate sequence and a tracr sequence. In an embodiment of the invention the Cas protein is a type II CRISPR-Cas protein. In an aspect of the invention the Cas protein is codon optimized for expression in a eukaryotic cell, preferably a mammalian cell or a human cell. In further embodiments of the invention the Cas protein is a type II CRISPR-Cas protein, e.g. a Cas 9 protein. In a highly preferred embodiment the Cas protein is a Cas9 protein, e.g. SpCas9. In aspects of the invention the Cas protein has one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the Cas protein has the D10A mutation.

[0063] Aspects of the invention relate to the expression of the gene product being decreased or a template polynucleotide being further introduced into the DNA molecule encoding the gene product or an intervening sequence being excised precisely by allowing the two 5' overhangs to reanneal and ligate or the activity or function of the gene product being altered or the expression of the gene product being increased. In an embodiment of the invention, the gene product is a protein.

[0064] The invention also comprehends an engineered, non-naturally occurring vector system comprising one or more vectors comprising:

a) a first regulatory element operably linked to each of two CRISPR-Cas system guide RNAs that target a first strand and a second strand respectively of a double stranded DNA molecule encoding a gene product,

b) a second regulatory element operably linked to a Cas protein,

wherein components (a) and (b) are located on same or different vectors of the system, whereby the guide RNAs target the DNA molecule encoding the gene product and the Cas protein nicks each of the first strand and the second strand of the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas protein and the two guide RNAs do not naturally occur together.

[0065] In aspects of the invention the guide RNAs may comprise a guide sequence fused to a tracr mate sequence and a tracr sequence. In an embodiment of the invention the Cas protein is a type II CRISPR-Cas protein. In an aspect of the invention the Cas protein is codon optimized for expression in a eukaryotic cell, preferably a mammalian cell or a human cell. In further embodiments of the invention the Cas protein is a type II CRISPR-Cas protein, e.g. a Cas 9 protein. In a highly preferred embodiment the Cas protein is a Cas9 protein, e.g. SpCas9. In aspects of the invention the Cas protein has one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the Cas protein has the D10A mutation.

[0066] Aspects of the invention relate to the expression of the gene product being decreased or a template polynucleotide being further introduced into the DNA molecule encoding the gene product or an intervening sequence being excised precisely by allowing the two 5' overhangs to reanneal and ligate or the activity or function of the gene product being altered or the expression of the gene product being increased. In an embodiment of the invention, the gene product is a protein. In preferred embodiments of the invention the vectors of the system are viral vectors. In a further embodiment, the vectors of the system are delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun.

[0067] In one aspect, the invention provides a method of modifying a target polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by homologous recombination with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid changes in a protein expressed from a gene comprising the target sequence. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cell, wherein the one or more vectors drive expression of one or more of: the CRISPR enzyme, the guide sequence linked to the tracr mate sequence, and the tracr sequence. In some embodiments, said vectors are delivered to the eukaryotic cell in a subject. In some embodiments, said modifying takes place in said eukaryotic cell in a cell culture. In some embodiments, the method further comprises isolating said eukaryotic cell from a subject prior to said modifying. In some embodiments, the method further comprises returning said eukaryotic cell and/or cells derived therefrom to said subject.

[0068] In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cells, wherein the one or more vectors drive expression of one or more of: the CRISPR enzyme, the guide sequence linked to the tracr mate sequence, and the tracr sequence.

[0069] In one aspect, the invention provides a method of generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, a disease gene is any gene associated with an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) introducing one or more vectors into a eukaryotic cell, wherein the one or more vectors drive expression of one or more of: a CRISPR enzyme, a guide sequence linked to a tracr mate sequence, and a tracr sequence; and (b) allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said disease gene, wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence within the target polynucleotide, and (2) the tracr mate sequence that is hybridized to the tracr sequence, thereby generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by homologous recombination with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid changes in a protein expression from a gene comprising the target sequence.

[0070] In one aspect the invention provides for a method of selecting one or more prokaryotic cell(s) by introducing one or more mutations in a gene in the one or more prokaryotic cell (s), the method comprising: introducing one or more vectors into the prokaryotic cell (s), wherein the one or more vectors drive expression of one or more of: a CRISPR enzyme, a guide sequence linked to a tracr mate sequence, a tracr sequence, and an editing template; wherein the editing template comprises the one or more mutations that abolish CRISPR enzyme cleavage; allowing homologous recombination of the editing template with the target polynucleotide in the cell(s) to be selected; allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said gene, wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence within the target polynucleotide, and (2) the tracr mate sequence that is hybridized to the tracr sequence, wherein binding of the CRISPR complex to the target polynucleotide induces cell death, thereby allowing one or more prokaryotic cell(s) in which one or more mutations have been introduced to be selected. In a preferred embodiment, the CRISPR enzyme is Cas9. In another aspect of the invention the cell to be selected may be a eukaryotic cell. Aspects of the invention allow for selection of specific cells without requiring a selection marker or a two-step process that may include a counter-selection system.

[0071] In one aspect, the invention provides for methods of modifying a target polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence.

[0072] In other embodiments, this invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. The method comprises increasing or decreasing expression of a target polynucleotide by using a CRISPR complex that binds to the polynucleotide.

[0073] Where desired, to effect the modification of the expression in a cell, one or more vectors comprising a tracr sequence, a guide sequence linked to the tracr mate sequence, a sequence encoding a CRISPR enzyme is delivered to a cell. In some methods, the one or more vectors comprises a regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence; and a regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting a guide sequence upstream of the tracr mate sequence. When expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a cell. Typically, the CRISPR complex comprises a CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence.

[0074] In some methods, a target polynucleotide can be inactivated to effect the modification of the expression in a cell. For example, upon the binding of a CRISPR complex to a target sequence in a cell, the target polynucleotide is inactivated such that the sequence is not transcribed, the coded protein is not produced, or the sequence does not function as the wild-type sequence does. For example, a protein or microRNA coding sequence may be inactivated such that the protein is not produced.

[0075] In certain embodiments, the CRISPR enzyme comprises one or more mutations selected from the group consisting of D10A, E762A. H840A, N854A, N863A or D986A and/or the one or more mutations is in a RuvC1 or HNH domain of the CRISPR enzyme or is a mutation as otherwise as discussed herein. In some embodiments, the CRISPR enzyme has one or more mutations in a catalytic domain, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the enzyme further comprises a functional domain. In some embodiments, the functional domain is a transcriptional activation domain, preferably VP64. In some embodiments, the functional domain is a transcription repression domain, preferably KRAB. In some embodiments, the transcription repression domain is SID, or concatemers of SID (eg SID4X). In some embodiments, the functional domain is an epigenetic modifying domain, such that an epigenetic modifying enzyme is provided. In some embodiments, the functional domain is an activation domain, which may be the P65 activation domain.

[0076] In some embodiments, the CRISPR enzyme is a type I or III CRISPR enzyme, but is preferably a type II CRISPR enzyme. This type II CRISPR enzyme may be any Cas enzyme. A Cas enzyme may be identified as Cas9 as this can refer to the general class of enzymes that share homology to the biggest nuclease with multiple nuclease domains from the type II CRISPR system. Most preferably, the Cas9 enzyme is from, or is derived from, spCas9 or saCas9. By derived, Applicants mean that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as described herein.

[0077] It will be appreciated that the terms Cas and CRISPR enzyme are generally used herein interchangeably, unless otherwise apparent. As mentioned above, many of the residue numberings used herein refer to the Cas9 enzyme from the type II CRISPR locus in *Streptococcus pyogenes*. However, it will be appreciated that this invention includes many more Cas9s from other species of microbes, such as SpCas9, SaCa9, St1Cas9 and so forth.

[0078] An example of a codon optimized sequence, in this instance optimized for humans (i.e. being optimized for expression in humans) is provided herein, see the SaCas9 human codon optimized sequence. Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species is known.

[0079] Preferably, delivery is in the form of a vector which may be a viral vector, such as a lenti- or baculo- or preferably adeno-viral/adeno-associated viral vectors, but other means of delivery are known (such as yeast systems, microvesicles, gene guns/means of attaching vectors to gold nanoparticles) and are provided. A vector may mean not only a viral or yeast system (for instance, where the nucleic acids of interest may be operably linked to and under the control of (in terms of expression, such as to ultimately provide a processed RNA) a promoter), but also direct delivery of nucleic acids into a host cell. While in herein methods the vector may be a viral vector and this is advantageously an AAV, other viral vectors as herein discussed can be employed, such as lentivirus. For example, baculoviruses may be used for expression in insect cells. These insect cells may, in turn be useful for producing large quantities of further vectors, such as AAV or lentivirus vectors adapted for delivery of the present invention. Also envisaged is a method of delivering the present CRISPR enzyme comprising delivering to a cell mRNA encoding the CRISPR enzyme. It will be appreciated that in certain embodiments the CRISPR enzyme is truncated, and/or comprised of less than one thousand amino acids or less than four thousand amino acids, and/or is a nuclease or nickase, and/or is codon-optimized, and/or comprises one or more mutations, and/or comprises a chimeric CRISPR enzyme, and/or the other options as herein discussed. AAV and lentiviral vectors are preferred.

[0080] In certain embodiments, the target sequence is flanked or followed, at its 3' end, by a PAM suitable for the CRISPR enzyme, typically a Cas and in particular a Cas9.

[0081] For example, a suitable PAM is 5'-NRG or 5'-NNGRR for SpCas9 or SaCas9 enzymes (or derived enzymes), respectively.

[0082] It will be appreciated that SpCas9 or SaCas9 are those from or derived from *S. pyogenes* or *S. aureus* Cas9.

[0083] Accordingly, it is an object of the invention to not encompass within the invention any previously known product, process of making the product, or method of using the product such that Applicants reserve the right and hereby disclose a disclaimer of any previously known product, process, or method. It is further noted that the invention does not intend to encompass within the scope of the invention any product, process, or making of the product or method of using the product, which does not meet the written description and enablement requirements of the USPTO (35 U.S.C. §112, first paragraph) or the EPO (Article 83 of the EPC), such that Applicants reserve the right and hereby disclose a disclaimer of any previously described product, process of making the product, or method of using the product.

[0084] It is noted that in this disclosure and particularly in the claims and/or paragraphs, terms such as "comprises", "comprised", "comprising" and the like can have the meaning attributed to it in U.S. Patent law; e.g., they can mean "includes", "included", "including", and the like; and that terms such as "consisting essentially of" and "consists essentially of" have the meaning ascribed to them in U.S. Patent law, e.g., they allow for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention.

[0085] These and other embodiments are disclosed or are obvious from and encompassed by, the following Detailed Description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0086] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0087] FIG. **1** shows a schematic model of the CRISPR system. The Cas9 nuclease from *Streptococcus pyogenes* (yellow) is targeted to genomic DNA by a synthetic guide RNA (sgRNA) consisting of a 20-nt guide sequence (blue) and a scaffold (red). The guide sequence base-pairs with the DNA target (blue), directly upstream of a requisite 5'-NGG protospacer adjacent motif (PAM; magenta), and Cas9 mediates a double-stranded break (DSB)-3 bp upstream of the PAM (red triangle).

[0088] FIG. **2**A-F shows an exemplary CRISPR system, a possible mechanism of action, an example adaptation for expression in eukaryotic cells, and results of tests assessing nuclear localization and CRISPR activity.

[0089] FIG. **3**A-D shows results of an evaluation of SpCas9 specificity for an example target.

[0090] FIG. **4**A-G show an exemplary vector system and results for its use in directing homologous recombination in eukaryotic cells.

[0091] FIG. **5** provides a table of protospacer sequences and summarizes modification efficiency results for protospacer targets designed based on exemplary *S. pyogenes* and *S. thermophilus* CRISPR systems with corresponding PAMs against loci in human and mouse genomes. Cells were transfected with Cas9 and either pre-crRNA/tracrRNA or chimeric RNA, and analyzed 72 hours after transfection. Percent indels are calculated based on Surveyor assay results from indicated cell lines (N=3 for all protospacer targets, errors are S.E.M., N.D. indicates not detectable using the Surveyor assay, and N.T. indicates not tested in this study).

[0092] FIG. **6**A-C shows a comparison of different tracrRNA transcripts for Cas9-mediated gene targeting.

[0093] FIG. **7** shows a schematic of a surveyor nuclease assay for detection of double strand break-induced micro-insertions and -deletions.

[0094] FIG. **8**A-B shows exemplary bicistronic expression vectors for expression of CRISPR system elements in eukaryotic cells.

[0095] FIG. **9**A-C shows histograms of distances between adjacent *S. pyogenes* SF370 locus 1 PAM (NGG) (FIG. **9**A) and *S. thermophilus* LMD9 locus 2 PAM (NNAGAAW) (FIG. **9**B) in the human genome; and distances for each PAM by chromosome (Chr) (FIG. **9**C).

[0096] FIG. **10**A-D shows an exemplary CRISPR system, an example adaptation for expression in eukaryotic cells, and results of tests assessing CRISPR activity.

[0097] FIG. **11**A-C shows exemplary manipulations of a CRISPR system for targeting of genomic loci in mammalian cells.

[0098] FIG. **12**A-B shows the results of a Northern blot analysis of crRNA processing in mammalian cells.

[0099] FIG. **13**A-B shows an exemplary selection of protospacers in the human PVALB and mouse Th loci.

[0100] FIG. **14** shows example protospacer and corresponding PAM sequence targets of the *S. thermophilus* CRISPR system in the human EMX1 locus.

[0101] FIG. **15** provides a table of sequences for primers and probes used for Surveyor, RFLP, genomic sequencing, and Northern blot assays.

[0102] FIG. **16**A-C shows exemplary manipulation of a CRISPR system with chimeric RNAs and results of SURVEYOR assays for system activity in eukaryotic cells.

[0103] FIG. **17**A-B shows a graphical representation of the results of SURVEYOR assays for CRISPR system activity in eukaryotic cells.

[0104] FIG. **18** shows an exemplary visualization of some *S. pyogenes* Cas9 target sites in the human genome using the UCSC genome browser.

[0105] FIG. **19**A-D shows a circular depiction of the phylogenetic analysis revealing five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids).

[0106] FIG. **20**A-F shows the linear depiction of the phylogenetic analysis revealing five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids).

[0107] FIG. **21**A-D shows genome editing via homologous recombination. (a) Schematic of SpCas9 nickase, with D10A mutation in the RuvC I catalytic domain. (b) Schematic representing homologous recombination (HR) at the human EMX1 locus using either sense or antisense single stranded oligonucleotides as repair templates. Red arrow above indicates sgRNA cleavage site; PCR primers for genotyping (Tables J and K) are indicated as arrows in right panel. (c) Sequence of region modified by HR. d, SURVEYOR assay for wildtype (wt) and nickase (D10A) SpCas9-mediated indels at the EMX1 target 1 locus (n=3). Arrows indicate positions of expected fragment sizes.

[0108] FIG. **22**A-B shows single vector designs for SpCas9.

[0109] FIG. **23** shows a graph representing the length distribution of Cas9 orthologs.

[0110] FIG. **24**A-M shows sequences where the mutation points are located within the SpCas9 gene.

[0111] FIG. **25**A shows the Conditional Cas9, Rosa26 targeting vector map.

[0112] FIG. **25**B shows the Constitutive Cas9, Rosa26 targeting vector map.

[0113] FIG. **26** shows a schematic of the important elements in the Constitutive and Conditional Cas9 constructs.

[0114] FIG. **27** shows delivery and in vivo mouse brain Cas9 expression data.

[0115] FIG. **28** shows RNA delivery of Cas9 and chimeric RNA into cells (A) Delivery of a GFP reporter as either DNA or mRNA into Neuro-2A cells. (B) Delivery of Cas9 and chimeric RNA against the Icam2 gene as RNA results in cutting for one of two spacers tested. (C) Delivery of Cas9 and chimeric RNA against the F7 gene as RNA results in cutting for one of two spacers tested.

[0116] FIG. **29** shows how DNA double-strand break (DSB) repair promotes gene editing. In the error-prone non-homologous end joining (NHEJ) pathway, the ends of a DSB are processed by endogenous DNA repair machineries and rejoined together, which can result in random insertion/deletion (indel) mutations at the site of junction. Indel mutations occurring within the coding region of a gene can result in frame-shift and a premature stop codon, leading to gene knockout. Alternatively, a repair template in the form of a plasmid or single-stranded oligodeoxynucleotides (ssODN) can be supplied to leverage the homology-directed repair (HDR) pathway, which allows high fidelity and precise editing.

[0117] FIG. **30**A-C shows anticipated results for HDR in HEK and HUES9 cells. (a) Either a targeting plasmid or an ssODN (sense or antisense) with homology arms can be used to edit the sequence at a target genomic locus cleaved by Cas9 (red triangle). To assay the efficiency of HDR, we introduced a HindIII site (red bar) into the target locus, which was PCR-amplified with primers that anneal outside of the region of homology. Digestion of the PCR product with HindIII reveals the occurrence of HDR events. (b) ssODNs, oriented in either the sense or the antisense (s or a) direction relative to the locus of interest, can be used in combination with Cas9 to achieve efficient HDR-mediated editing at the target locus. A minimal homology region of 40 bp, and preferably 90 bp, is recommended on either side of the modification (red bar). (c) Example of the effect of ssODNs on HDR in the EMX1 locus is shown using both wild-type Cas9 and Cas9 nickase (D10A). Each ssODN contains homology arms of 90 bp flanking a 12-bp insertion of two restriction sites.

[0118] FIG. **31**A-C shows the repair strategy for Cystic Fibrosis delta F508 mutation.

[0119] FIG. **32**A-B (a) shows a schematic of the GAA repeat expansion in FXN intron 1 and (b) shows a schematic of the strategy adopted to excise the GAA expansion region using the CRISPR/Cas system.

[0120] FIG. **33** shows a screen for efficient SpCas9 mediated targeting of Tet1-3 and Dnmt1, 3a and 3b gene loci. Surveyor assay on DNA from transfected N2A cells demonstrates efficient DNA cleavage by using different gRNAs.

[0121] FIG. **34** shows a strategy of multiplex genome targeting using a 2-vector system in an AAV1/2 delivery system. Tet1-3 and Dnmt1, 3a and 3b gRNA under the control of the U6 promoter. GFP-KASH under the control of the human synapsin promoter. Restriction sides shows simple gRNA replacement strategy by subcloning. HA-tagged SpCas9 flanked by two nuclear localization signals (NLS) is shown. Both vectors are delivered into the brain by AAV1/2 virus in a 1:1 ratio.

[0122] FIG. **35** shows verification of multiplex DNMT targeting vector #1 functionality using Surveyor assay. N2A cells were co-transfected with the DNMT targeting vector #1

(+) and the SpCas9 encoding vector for testing SpCas9 mediated cleavage of DNMTs genes family loci. gRNA only (−) is negative control. Cells were harvested for DNA purification and downstream processing 48 h after transfection.

[0123] FIG. **36** shows verification of multiplex DNMT targeting vector #2 functionality using Surveyor assay. N2A cells were co-transfected with the DNMT targeting vector #1 (+) and the SpCas9 encoding vector for testing SpCas9 mediated cleavage of DNMTs genes family loci. gRNA only (−) is negative control. Cells were harvested for DNA purification and downstream processing 48 h after transfection.

[0124] FIG. **37** shows schematic overview of short promoters and short polyA versions used for HA-SpCas9 expression in vivo. Sizes of the encoding region from L-ITR to R-ITR are shown on the right.

[0125] FIG. **38** shows schematic overview of short promoters and short polyA versions used for HA-SaCas9 expression in vivo. Sizes of the encoding region from L-ITR to R-ITR are shown on the right.

[0126] FIG. **39** shows expression of SpCas9 and SaCas9 in N2A cells. Representative Western blot of HA-tagged SpCas9 and SaCas9 versions under the control of different short promoters and with or short polyA (spA) sequences. Tubulin is loading control. mCherry (mCh) is a transfection control. Cells were harvested and further processed for Western blotting 48 h after transfection.

[0127] FIG. **40** shows screen for efficient SaCas9 mediated targeting of Tet3 gene locus. Surveyor assay on DNA from transfected N2A cells demonstrates efficient DNA cleavage by using different gRNAs with NNGGGT PUM sequence. GFP transfected cells and cells expressing only SaCas9 are controls.

[0128] FIG. **41** shows expression of HA-SaCas9 in the mouse brain. Animals were injected into dentate gyri with virus driving expression of HA-SaCas9 under the control of human Synapsin promoter. Animals were sacrificed 2 weeks after surgery. HA tag was detected using rabbit monoclonal antibody C29F4 (Cell Signaling). Cell nuclei stained in blue with DAPI stain.

[0129] FIG. **42** shows expression of SpCas9 and SaCas9 in cortical primary neurons in culture 7 days after transduction. Representative Western blot of HA-tagged SpCas9 and SaCas9 versions under the control of different promoters and with bgh or short polyA (spA) sequences. Tubulin is loading control.

[0130] FIG. **43** shows LIVE/DEAD stain of primary cortical neurons 7 days after transduction with AAV1 particles carrying SpCas9 with different promoters and multiplex gRNAs constructs (example shown on the last panel for DNMTs). Neurons after AAV transduction were compared with control untransduced neurons. Red nuclei indicate permeabilized, dead cells (second line of panels). Live cells are marked in green color (third line of panels).

[0131] FIG. **44** shows LIVE/DEAD stain of primary cortical neurons 7 days after transduction with AAV1 particles carrying SaCas9 with different promoters. Red nuclei indicate permeabilized, dead cells (second line of panels). Live cells are marked in green color (third line of panels).

[0132] FIG. **45** shows comparison of morphology of neurons after transduction with AAV1 virus carrying SpCas9 and gRNA multiplexes for TETs and DNMTs genes loci. Neurons without transduction are shown as a control.

[0133] FIG. **46** shows verification of multiplex DNMT targeting vector #1 functionality using Surveyor assay in primary cortical neurons. Cells were co-transduced with the DNMT targeting vector #1 and the SpCas9 viruses with different promoters for testing SpCas9 mediated cleavage of DNMTs genes family loci.

[0134] FIG. **47** shows in vivo efficiency of SpCas9 cleavage in the brain. Mice were injected with AAV1/2 virus carrying gRNA multiplex targeting DNMT family genes loci together with SpCas9 viruses under control of 2 different promoters: mouse Mecp2 and rat Map1b. Two weeks after injection brain tissue was extracted and nuclei were prepped and sorted using FACS, based on the GFP expression driven by Synapsin promoter from gRNA multiplex construct. After gDNA extraction Surveyor assay was run. + indicates GFP positive nuclei and − control, GFP-negative nuclei from the same animal. Numbers on the gel indicate assessed SpCas9 efficiency.

[0135] FIG. **48** shows purification of GFP-KASH labeled cell nuclei from hippocampal neurons. The outer nuclear membrane (ONM) of the cell nuclear membrane is tagged with a fusion of GFP and the KASH protein transmembrane domain. Strong GFP expression in the brain after one week of stereotactic surgery and AAV1/2 injection. Density gradient centrifugation step to purify cell nuclei from intact brain. Purified nuclei are shown. Chromatin stain by Vybrant® DyeCycle™ Ruby Stain is shown in red, GFP labeled nuclei are green. Representative FACS profile of GFP+ and GFP− cell nuclei (Magenta: Vybrant® DyeCycle™ Ruby Stain, Green: GFP).

[0136] FIG. **49** shows efficiency of SpCas9 cleavage in the mouse brain. Mice were injected with AAV1/2 virus carrying gRNA multiplex targeting TET family genes loci together with SpCas9 viruses under control of 2 different promoters: mouse Mecp2 and rat Map1b. Three weeks after injection brain tissue was extracted, nuclei were prepped and sorted using FACS, based on the GFP expression driven by Synapsin promoter from gRNA multiplex construct. After gDNA extraction Surveyor assay was run. + indicates GFP positive nuclei and − control, GFP-negative nuclei from the same animal. Numbers on the gel indicate assessed SpCas9 efficiency.

[0137] FIG. **50** shows GFP-KASH expression in cortical neurons in culture. Neurons were transduced with AAV1 virus carrying gRNA multiplex constructs targeting TET genes loci. The strongest signal localize around cells nuclei due to KASH domain localization.

[0138] FIG. **51** shows (top) a list of spacing (as indicated by the pattern of arrangement for two PAM sequences) between pairs of guide RNAs. Only guide RNA pairs satisfying patterns 1, 2, 3, 4 exhibited indels when used with SpCas9 (D10A) nickase. (bottom) Gel images showing that combination of SpCas9(D10A) with pairs of guide RNA satisfying patterns 1, 2, 3, 4 led to the formation of indels in the target site.

[0139] FIG. **52** shows a list of U6 reverse primer sequences used to generate U6-guide RNA expression cassettes. Each primer needs to be paired with the U6 forward primer "gcactgagggcctatttcccatgattc" to generate amplicons containing U6 and the desired guide RNA.

[0140] FIG. **53** shows a Genomic sequence map from the human Emx1 locus showing the locations of the 24 patterns listed in FIG. **33**.

[0141] FIG. **54** shows on (right) a gel image indicating the formation of indels at the target site when variable 5' overhangs are present after cleavage by the Cas9 nickase targeted by different pairs of guide RNAs. on (left) a table indicating the lane numbers of the gel on the right and various parameters including identifying the guide RNA pairs used and the length of the 5' overhang present following cleavage by the Cas9 nickase.

[0142] FIG. **55** shows a Genomic sequence map from the human Emx1 locus showing the locations of the different pairs of guide RNAs that result in the gel patterns of FIG. **54** (right) and which are further described in Example 35.

[0143] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE INVENTION

[0144] The invention relates to the engineering and optimization of systems, methods and compositions used for the control of gene expression involving sequence targeting, such as genome perturbation or gene-editing, that relate to the CRISPR-Cas system and components thereof. In advantageous embodiments, the Cas enzyme is Cas9.

[0145] An advantage of the present methods is that the CRISPR system avoids off-target binding and its resulting side effects. This is achieved using systems arranged to have a high degree of sequence specificity for the target DNA.

[0146] Cas9 optimization may be used to enhance function or to develop new functions, one can generate chimeric Cas9 proteins. Examples that the Applicants have generated are provided in Example 12. Chimeric Cas9 proteins can be made by combining fragments from different Cas9 homologs. For example, two example chimeric Cas9 proteins from the Cas9s described herein. For example, Applicants fused the N-term of St1Cas9 (fragment from this protein is in bold) with C-term of SpCas9. The benefit of making chimeric Cas9s include any or all of: reduced toxicity, improved expression in eukaryotic cells; enhanced specificity; reduced molecular weight of protein, for example, making the protein smaller by combining the smallest domains from different Cas9 homologs; and/or altering the PAM sequence requirement.

[0147] The Cas9 may be used as a generic DNA binding protein. For example, and as shown in Example 13, Applicants used Cas9 as a generic DNA binding protein by mutating the two catalytic domains (D10 and H840) responsible for cleaving both strands of the DNA target. In order to upregulate gene transcription at a target locus Applicants fused a transcriptional activation domain (VP64) to Cas9. Other transcriptional activation domains are known. As shown in Example 17, transcriptional activation is possible. As also shown in Example 17, gene repression (in this case of the beta-catenin gene) is possible using a Cas9 repressor (DNA-binding domain) that binds to the target gene sequence, thus repressing its activity.

[0148] Cas9 and one or more guide RNA can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other plasmid or viral vector types, in particular, using formulations and doses from, for example, U.S. Pat. No. 8,454,972 (formulations, doses for adenovirus), U.S. Pat. No. 8,404,658 (formulations, doses for AAV) and U.S. Pat. No. 5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For examples, for AAV, the route of administration, formulation and dose can be as in U.S. Pat. No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in U.S. Pat. No. 8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as in U.S. Pat. No. 5,846,946 and as in clinical studies involving plasmids. Doses may be based on or extrapolated to an average 70 kg individual, and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed.

[0149] The viral vectors can be injected into the tissue of interest. For cell-type specific genome modification, the expression of Cas9 can be driven by a cell-type specific promoter. For example, liver-specific expression might use the Albumin promoter and neuron-specific expression might use the Synapsin I promoter.

[0150] Transgenic animals are also provided. Preferred examples include animals comprising Cas9, in terms of polynucleotides encoding Cas9 or the protein itself. Mice, rats and rabbits are preferred. To generate transgenic mice with the constructs, as exemplified herein one may inject pure, linear DNA into the pronucleus of a zygote from a pseudo pregnant female, e.g. a CB56 female. Founders may then be identified, genotyped, and backcrossed to CB57 mice. The constructs may then be cloned and optionally verified, for instance by Sanger sequencing. Knock outs are envisaged where for instance one or more genes are knocked out in a model. However, are knockins are also envisaged (alone or in combination). An example knockin Cas9 mouse was generated and this is exemplified, but Cas9 knockins are preferred. To generate a Cas9 knock in mice one may target the same constitutive and conditional constructs to the Rosa26 locus, as described herein (FIGS. **25**A-B and 26). Methods of US Patent Publication Nos. 20120017290 and 20110265198 assigned to Sangamo BioSciences, Inc. directed to targeting the *Rosa* locus may be modified to utilize the CRISPR Cas system of the present invention. In another embodiment, the methods of US Patent Publication No. 20130236946 assigned to Cellectis directed to targeting the *Rosa* locus may also be modified to utilize the CRISPR Cas system of the present invention.

[0151] Utility of the conditional Cas9 mouse: Applicants have shown in 293 cells that the Cas9 conditional expression construct can be activated by co-expression with Cre. Applicants also show that the correctly targeted R1 mESCs can have active Cas9 when Cre is expressed. Because Cas9 is followed by the P2A peptide cleavage sequence and then EGFP Applicants identify successful expression by observing EGFP. Applicants have shown Cas9 activation in mESCs. This same concept is what makes the conditional Cas9 mouse so useful. Applicants may cross their conditional Cas9 mouse with a mouse that ubiquitously expresses Cre (ACTB-Cre line) and may arrive at a mouse that expresses Cas9 in every cell. It should only take the delivery of chimeric RNA to induce genome editing in embryonic or adult mice. Interestingly, if the conditional Cas9 mouse is crossed with a mouse expressing Cre under a tissue specific promoter, there should only be Cas9 in the tissues that also express Cre. This approach may be used to edit the genome in only precise tissues by delivering chimeric RNA to the same tissue.

[0152] As mentioned above, transgenic animals are also provided, as are transgenic plants, especially crops and algae. The transgenic plants may be useful in applications outside of providing a disease model. These may include food or feed production through expression of, for instance, higher protein, carbohydrate, nutrient or vitamin levels than would normally be seen in the wildtype. In this regard, transgenic plants, especially pulses and tubers, and animals, especially mammals such as livestock (cows, sheep, goats and pigs), but also poultry and edible insects, are preferred.

[0153] Transgenic algae or other plants such as rape may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol), for instance. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[0154] In terms of in vivo delivery, AAV is advantageous over other viral vectors for a couple of reasons:

[0155] Low toxicity (this may be due to the purification method not requiring ultra centrifugation of cell particles that can activate the immune response)

[0156] Low probability of causing insertional mutagenesis because it doesn't integrate into the host genome.

[0157] AAV has a packaging limit of 4.5 or 4.75 Kb. This means that Cas9 as well as a promoter and transcription terminator have to be all fit into the same viral vector. Constructs larger than 4.5 or 4.75 Kb will lead to significantly reduced virus production. SpCas9 is quite large, the gene itself is over 4.1 Kb, which makes it difficult for packing into AAV. Therefore embodiments of the invention include utilizing homologs of Cas9 that are shorter. For example:

Species	Cas9 Size
Corynebacter diphtheriae	3252
Eubacterium ventriosum	3321
Streptococcus pasteurianus	3390
Lactobacillus farciminis	3378
Sphaerochaeta globus	3537
Azospirillum B510	3504
Gluconacetobacter diazotrophicus	3150
Neisseria cinerea	3246
Roseburia intestinalis	3420
Parvibaculum lavamentivorans	3111
Staphylococcus aureus	3159
Nitratifractor salsuginis DSM 16511	3396
Campylobacter lari CF89-12	3009
Streptococcus thermophilus LMD-9	3396

[0158] These species are therefore, in general, preferred Cas9 species. Applicants have shown delivery and in vivo mouse brain Cas9 expression data.

[0159] Two ways to package Cas9 coding nucleic acid molecules, e.g., DNA, into viral vectors to mediate genome modification in vivo are preferred:

[0160] To achieve NHEJ-mediated gene knockout:

[0161] Single virus vector:

[0162] Vector containing two or more expression cassettes:

[0163] Promoter-Cas9 coding nucleic acid molecule-terminator

[0164] Promoter-gRNA1-terminator

[0165] Promoter-gRNA2-terminator

[0166] Promoter-gRNA(N)-terminator (up to size limit of vector)

[0167] Double virus vector:

[0168] Vector 1 containing one expression cassette for driving the expression of Cas9

[0169] Promoter-Cas9 coding nucleic acid molecule-terminator

[0170] Vector 2 containing one more expression cassettes for driving the expression of one or more guideRNAs

[0171] Promoter-gRNA 1-terminator

[0172] Promoter-gRNA(N)-terminator (up to size limit of vector)

[0173] To mediate homology-directed repair. In addition to the single and double virus vector approaches described above, an additional vector is used to deliver a homology-direct repair template.

[0174] Promoter used to drive Cas9 coding nucleic acid molecule expression can include:

[0175] AAV ITR can serve as a promoter: this is advantageous for eliminating the need for an additional promoter element (which can take up space in the vector). The additional space freed up can be used to drive the expression of additional elements (gRNA, etc.). Also, ITR activity is relatively weaker, so can be used to reduce toxicity due to over expression of Cas9.

[0176] For ubiquitous expression, can use promoters: CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains, etc.

[0177] For brain expression, can use promoters: SynapsinI for all neurons, CaMKIIalpha for excitatory neurons, GAD67 or GAD65 or VGAT for GABAergic neurons, etc.

[0178] For liver expression, can use Albumin promoter.

[0179] For lung expression, can use SP-B.

[0180] For endothelial cells, can use ICAM.

[0181] For hematopoietic cells can use IFNbeta or CD45.

[0182] For Osteoblasts can use OG-2.

[0183] Promoter used to drive guide RNA can include:

[0184] Pol III promoters such as U6 or H1

[0185] Use of Pol II promoter and intronic cassettes to express gRNA

[0186] As to AAV, the AAV can be AAV1, AAV2, AAV5 or any combination thereof. One can select the AAV of the AAV with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid or capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. The above promoters and vectors are preferred individually.

[0187] RNA delivery is also a useful method of in vivo delivery. FIG. **27** shows delivery and in vivo mouse brain Cas9 expression data. It is possible to deliver Cas9 and gRNA (and, for instance, HR repair template) into cells using liposomes or nanoparticles. Thus delivery of the CRISPR enzyme, such as a Cas9 and/or delivery of the RNAs of the invention may be in RNA form and via microvesicles, liposomes or nanoparticles. For example, Cas9 mRNA and gRNA can be packaged into liposomal particles for delivery in vivo. Liposomal transfection reagents such as lipofectamine from Life Technologies and other reagents on the market can effectively deliver RNA molecules into the liver.

[0188] Enhancing NHEJ or HR efficiency is also helpful for delivery. It is preferred that NHEJ efficiency is enhanced by co-expressing end-processing enzymes such as Trex2 (Dumitrache et al. Genetics. 2011 August; 188(4): 787-797). It is preferred that HR efficiency is increased by transiently inhibiting NHEJ machineries such as Ku70 and Ku86. HR efficiency can also be increased by co-expressing prokaryotic or eukaryotic homologous recombination enzymes such as RecBCD, RecA.

[0189] Various means of delivery are described herein, and further discussed in this section.

[0190] Viral delivery: The CRISPR enzyme, for instance a Cas9, and/or any of the present RNAs, for instance a guide RNA, can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other viral vector types, or combinations thereof. Cas9 and one or more guide RNAs can be packaged into one or more viral vectors. In some embodiments, the viral vector is delivered to the tissue of interest by, for example, an intramuscular injection, while other times the viral delivery is via intravenous, transdermal, intranasal, oral, mucosal, or other delivery methods. Such delivery may be either via a single dose, or multiple doses. One skilled in the art understands that the actual dosage to be delivered herein may vary greatly depending upon a variety of factors, such as the vector chose, the target cell, organism, or tissue, the general condition of the subject to be treated, the degree of transformation/modification sought, the administration route, the administration mode, the type of transformation/modification sought, etc.

[0191] Such a dosage may further contain, for example, a carrier (water, saline, ethanol, glycerol, lactose, sucrose, calcium phosphate, gelatin, dextran, agar, pectin, peanut oil, sesame oil, etc.), a diluent, a pharmaceutically-acceptable carrier (e.g., phosphate-buffered saline), a pharmaceutically-acceptable excipient, an adjuvant to enhance antigenicity, an immunostimulatory compound or molecule, and/or other compounds known in the art. The adjuvant herein may contain a suspension of minerals (alum, aluminum hydroxide, aluminum phosphate) on which antigen is adsorbed; or water-in-oil emulsion in which antigen solution is emulsified in oil (MF-59, Freund's incomplete adjuvant), sometimes with the inclusion of killed mycobacteria (Freund's complete adjuvant) to further enhance antigenicity (inhibits degradation of antigen and/or causes influx of macrophages). Adjuvants also include immunostimulatory molecules, such as cytokines, costimulatory molecules, and for example, immunostimulatory DNA or RNA molecules, such as CpG oligonucleotides. Such a dosage formulation is readily ascertainable by one skilled in the art. The dosage may further contain one or more pharmaceutically acceptable salts such as, for example, a mineral acid salt such as a hydrochloride, a hydrobromide, a phosphate, a sulfate, etc.; and the salts of organic acids such as acetates, propionates, malonates, benzoates, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, gels or gelling materials, flavorings, colorants, microspheres, polymers, suspension agents, etc. may also be present herein. In addition, one or more other conventional pharmaceutical ingredients, such as preservatives, humectants, suspending agents, surfactants, antioxidants, anticaking agents, fillers, chelating agents, coating agents, chemical stabilizers, etc. may also be present, especially if the dosage form is a reconstitutable form. Suitable exemplary ingredients include microcrystalline cellulose, carboxymethylcellulose sodium, polysorbate 80, phenylethyl alcohol, chlorobutanol, potassium sorbate, sorbic acid, sulfur dioxide, propyl gallate, the parabens, ethyl vanillin, glycerin, phenol, parachlorophenol, gelatin, albumin and a combination thereof. A thorough discussion of pharmaceutically acceptable excipients is available in REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991) which is incorporated by reference herein.

[0192] In an embodiment herein the delivery is via an adenovirus, which may be at a single booster dose containing at least 1×10^5 particles (also referred to as particle units, pu) of adenoviral vector. In an embodiment herein, the dose preferably is at least about 1×10^6 particles (for example, about 1×10^6-1×10^{12} particles), more preferably at least about 1×10^7 particles, more preferably at least about 1×10^8 particles (e.g., about 1×10^8-1×10^{11} particles or about 1×10^8-1×10^{12} particles), and most preferably at least about 1×10^0 particles (e.g., about 1×10^9-1×10^{10} particles or about 1×10^9-$\times10^{12}$ particles), or even at least about 1×10^{10} particles (e.g., about 1×10^{10}-1×10^{12} particles) of the adenoviral vector. Alternatively, the dose comprises no more than about 1×10^{14} particles, preferably no more than about 1×10^{13} particles, even more preferably no more than about 1×10^{12} particles, even more preferably no more than about 1×10^{11} particles, and most preferably no more than about 1×10^{10} particles (e.g., no more than about 1×10^9 articles). Thus, the dose may contain a single dose of adenoviral vector with, for example, about 1×10^6 particle units (pu), about 2×10^6 pu, about 4×10^6 pu, about 1×10^7 pu, about 2×10^7 pu, about 4×10^7 pu, about 1×10^8 pu, about 2×10^8 pu, about 4×10^8 pu, about 1×10^9 pu, about 2×10^9 pu, about 4×10^9 pu, about 1×10^{10} pu, about 2×10^{10} pu, about 4×10^{10} pu, about 1×10^{11} pu, about 2×10^{11} pu, about 4×10^{11} pu, about 1×10^{12} pu, about 2×10^{12} pu, or about 4×10^{12} pu of adenoviral vector. See, for example, the adenoviral vectors in U.S. Pat. No. 8,454,972 B2 to Nabel, et. al., granted on Jun. 4, 2013; incorporated by reference herein, and the dosages at col 29, lines 36-58 thereof. In an embodiment herein, the adenovirus is delivered via multiple doses.

[0193] In an embodiment herein, the delivery is via an AAV. A therapeutically effective dosage for in vivo delivery of the AAV to a human is believed to be in the range of from about 20 to about 50 ml of saline solution containing from about 1×10^{10} to about 1×10^{10} functional AAV/ml solution. The dosage may be adjusted to balance the therapeutic benefit against any side effects. In an embodiment herein, the AAV dose is generally in the range of concentrations of from about 1×10^5 to 1×10^{50} genomes AAV, from about 1×10^8 to 1×10^{20} genomes AAV, from about 1×10^{10} to about 1×10^{16} genomes, or about 1×10^{11} to about 1×10^{16} genomes AAV. A human dosage may be about 1×10^{13} genomes AAV. Such concentrations may be delivered in from about 0.001 ml to about 100 ml, about 0.05 to about 50 ml, or about 10 to about 25 ml of a carrier solution. Other effective dosages can be readily established by one of ordinary skill in the art through routine trials establishing dose response curves. See, for example, U.S. Pat. No. 8,404,658 B2 to Hajjar, et al., granted on Mar. 26, 2013, at col. 27, lines 45-60.

[0194] In an embodiment herein the delivery is via a plasmid. In such plasmid compositions, the dosage should be a sufficient amount of plasmid to elicit a response. For instance, suitable quantities of plasmid DNA in plasmid compositions can be from about 0.1 to about 2 mg, or from about 1 μg to about 10 μg.

[0195] The doses herein are based on an average 70 kg individual. The frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), or scientist skilled in the art.

[0196] Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells. The most commonly known lentivirus is the human immunodeficiency virus (HIV), which uses the envelope glycoproteins of other viruses to target a broad range of cell types.

[0197] Lentiviruses may be prepared as follows. After cloning pCasES10 (which contains a lentiviral transfer plasmid backbone), HEK293FT at low passage (p=5) were seeded in a T-75 flask to 50% confluence the day before transfection in DMEM with 10% fetal bovine serum and without antibiotics. After 20 hours, media was changed to OptiMEM (serum-free) media and transfection was done 4 hours later. Cells were transfected with 10 μg of lentiviral transfer plasmid (pCasES10) and the following packaging plasmids: 5 μg of pMD2.G (VSV-g pseudotype), and 7.5 μg of psPAX2 (gag/pol/rev/tat). Transfection was done in 4 mL OptiMEM with a cationic lipid delivery agent (50 uL Lipofectamine 2000 and 100 ul Plus reagent). After 6 hours, the media was changed to antibiotic-free DMEM with 10% fetal bovine serum.

[0198] Lentivirus may be purified as follows. Viral supernatants were harvested after 48 hours. Supernatants were first cleared of debris and filtered through a 0.45 um low protein binding (PVDF) filter. They were then spun in a ultracentrifuge for 2 hours at 24,000 rpm. Viral pellets were resuspended in 50 ul of DMEM overnight at 4 C. They were then aliquotted and immediately frozen at −80 C.

[0199] In another embodiment, minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV) are also contemplated, especially for ocular gene therapy (see, e.g., Balagaan, J Gene Med 2006; 8: 275-285, Published online 21 Nov. 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jgm.845). In another embodiment, RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostain and angiostatin that is delivered via a subretinal injection for the treatment of the web form of age-related macular degeneration is also contemplated (see, e.g., Binley et al., HUMAN GENE THERAPY 23:980-991 (September 2012)) may be modified for the CRISPR-Cas system of the present invention.

[0200] In another embodiment, self-inactivating lentiviral vectors with an siRNA targeting a common exon shared by HIV tat/rev, a nucleolar-localizing TAR decoy, and an anti-CCR5-specific hammerhead ribozyme (see, e.g., DiGiusto et al. (2010) Sci Transl Med 2:36ra43) may be used/and or adapted to the CRISPR-Cas system of the present invention. A minimum of 2.5×10^6 CD34+ cells per kilogram patient weight may be collected and prestimulated for 16 to 20 hours in X-VIVO 15 medium (Lonza) containing 2 mML-glutamine, stem cell factor (100 ng/ml), Flt-3 ligand (Flt-3L) (100 ng/ml), and thrombopoietin (10 ng/ml) (CellGenix) at a density of 2×10^6 cells/ml. Prestimulated cells may be transduced with lentiviral at a multiplicity of infection of 5 for 16 to 24 hours in 75-cm^2 tissue culture flasks coated with fibronectin (25 mg/cm^2) (RetroNectin, Takara Bio Inc.).

[0201] Lentiviral vectors have been disclosed as in the treatment for Parkinson's Disease, see, e.g., US Patent Publication No. 20120295960 and U.S. Pat. Nos. 7,303,910 and 7,351,585. Lentiviral vectors have also been disclosed for the treatment of ocular diseases, see e.g., US Patent Publication Nos. 20060281180, 20090007284, US20110117189; US20090017543; US20070054961, US20100317109. Lentiviral vectors have also been disclosed for delivery to the train, see, e.g., US Patent Publication Nos. US20110293571; US20110293571, US20040013648, US20070025970, US20090111106 and U.S. Pat. No. 7,259,015.

[0202] RNA delivery: The CRISPR enzyme, for instance a Cas9, and/or any of the present RNAs, for instance a guide RNA, can also be delivered in the form of RNA. Cas9 mRNA can be generated using in vitro transcription. For example, Cas9 mRNA can be synthesized using a PCR cassette containing the following elements: T7_promoter-kozak sequence (GCCACC)-Cas9-3' UTR from beta globin-polyA tail (a string of 120 or more adenines). The cassette can be used for transcription by T7 polymerase. Guide RNAs can also be transcribed using in vitro transcription from a cassette containing T7_promoter-GG-guide RNA sequence.

[0203] To enhance expression and reduce toxicity, the CRISPR enzyme and/or guide RNA can be modified using pseudo-U or 5-Methyl-C.

[0204] mRNA delivery methods are especially promising for liver delivery currently. In particular, for AAV8 is particularly preferred for delivery to the liver.

[0205] CRISPR enzyme mRNA and guide RNA might also be delivered separately. CRISPR enzyme mRNA can be delivered prior to the guide RNA to give time for CRISPR enzyme to be expressed. CRISPR enzyme mRNA might be administered 1-12 hours (preferably around 2-6 hours) prior to the administration of guide RNA.

[0206] Alternatively, CRISPR enzyme mRNA and guide RNA can be administered together. Advantageously, a second booster dose of guide RNA can be administered 1-12 hours (preferably around 2-6 hours) after the initial administration of CRISPR enzyme mRNA+guide RNA.

[0207] Additional administrations of CRISPR enzyme mRNA and/or guide RNA might be useful to achieve the most efficient levels of genome modification.

[0208] For minimization of toxicity and off-target effect, it will be important to control the concentration of CRISPR enzyme mRNA and guide RNA delivered. Optimal concentrations of CRISPR enzyme mRNA and guide RNA can be determined by testing different concentrations in a cellular or animal model and using deep sequencing the analyze the extent of modification at potential off-target genomic loci. For example, for the guide sequence targeting 5'-GAGTCCGAGCAGAAGAAGAA-3' in the EMX1 gene of the human genome, deep sequencing can be used to assess the level of modification at the following two off-target loci, 1: 5'-GAGTCCTAGCAGGAGAAGAA-3' and 2: 5'-GAGTCTAAGCAGAAGAAGAA-3'. The concentration that gives the highest level of on-target modification while minimizing the level of off-target modification should be chosen for in vivo delivery.

[0209] Alternatively, to minimize the level of toxicity and off-target effect, CRISPR enzyme nickase mRNA (for example *S. pyogenes* Cas9 with the D10A mutation) can be delivered with a pair of guide RNAs targeting a site of interest. The two guide RNAs need to be spaced as follows. Guide sequences in red (single underline) and blue (double underline) respectively (these examples are based on the PAM requirement for *Streptococcus pyogenes* Cas9).

Overhang length (bp)	Guide RNA design (guide sequence and PAM color coded)
14	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNN-5'

-continued

Overhang length (bp)	Guide RNA design (guide sequence and PAM color coded)
13	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNN-5'
12	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
11	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNN-5'
10	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNN-5'
9	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNN-5'
8	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNN-5'
7	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNN-5'
6	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNN-5'
5	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNN-5'
4	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNN-5'
3	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
2	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
1	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
blunt	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
1	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
2	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
3	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
4	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
5	5'-NNNNNNNNNNNNNNNNNNNNCCNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
6	5'-NNNNNNNNNNNNNNNNNNNNCCNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
7	5'-NNNNNNNNNNNNNNNNNNNNCCNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNGGNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
8	5'-NNNNNNNNNNNNNNNNNNNNNCCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNNGGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
12	5'-NNNNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNNNNCCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
13	5'-NNNNNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNNNNCCNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'

-continued

Overhang length (bp)	Guide RNA design (guide sequence and PAM color coded)
14	5'-NNNNNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNNNNCCNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
15	5'-NNNNNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNNNNCCNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
16	5'-NNNNNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNNNCCNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
17	5'-NNNNNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3' 3'-NNNNNNNNNNNNNNNNNNNNNNNCCNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'

[0210] Further interrogation of the system have given Applicants evidence of the 5'overhang (see, e.g., Ran et al., Cell. 2013 Sep. 12; 154(6):1380-9 and U.S. Provisional Patent Application Ser. No. 61/871,301 filed Aug. 28, 2013). Applicants have further identified parameters that relate to efficient cleavage by the Cas9 nickase mutant when combined with two guide RNAs and these parameters include but are not limited to the length of the 5' overhang. In embodiments of the invention the 5' overhang is at most 200 base pairs, preferably at most 100 base pairs, or more preferably at most 50 base pairs. In embodiments of the invention the 5' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs or 1-34 base pairs. In other preferred methods of the invention the first guide sequence directing cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directing cleavage of other strand near the second target sequence results in a blunt cut or a 3' overhang. In embodiments of the invention the 3' overhang is at most 150, 100 or 25 base pairs or at least 15, 10 or 1 base pairs. In preferred embodiments the 3' overhang is 1-100 basepairs.

[0211] Aspects of the invention relate to the expression of the gene product being decreased or a template polynucleotide being further introduced into the DNA molecule encoding the gene product or an intervening sequence being excised precisely by allowing the two 5' overhangs to reanneal and ligate or the activity or function of the gene product being altered or the expression of the gene product being increased. In an embodiment of the invention, the gene product is a protein.

[0212] Only sgRNA pairs creating 5' overhangs with less than 8 bp overlap between the guide sequences (offset greater than −8 bp) were able to mediate detectable indel formation. Importantly, each guide used in these assays is able to efficiently induce indels when paired with wildtype Cas9, indicating that the relative positions of the guide pairs are the most important parameters in predicting double nicking activity.

[0213] Since Cas9n and Cas9H840A nick opposite strands of DNA, substitution of Cas9n with Cas9H840A with a given sgRNA pair should result in the inversion of the overhang type. For example, a pair of sgRNAs that will generate a 5' overhang with Cas9n should in principle generate the corresponding 3' overhang instead. Therefore, sgRNA pairs that lead to the generation of a 3' overhang with Cas9n might be used with Cas9H840A to generate a 5' overhang. Unexpectedly, Applicants tested Cas9H840A with a set of sgRNA pairs designed to generate both 5' and 3' overhangs (offset range from −278 to +58 bp), but were unable to observe indel formation. Further work may be needed to identify the necessary design rules for sgRNA pairing to allow double nicking by Cas9H840A.

[0214] Targeted deletion of genes is preferred. Examples are exemplified in Example 18. Preferred are, therefore, genes involved in cholesterol biosynthesis, fatty acid biosynthesis, and other metabolic disorders, genes encoding misfolded proteins involved in amyloid and other diseases, oncogenes leading to cellular transformation, latent viral genes, and genes leading to dominant-negative disorders, amongst other disorders. As exemplified here, Applicants prefer gene delivery of a CRISPR-Cas system to the liver, brain, ocular, epithelial, hematopoetic, or another tissue of a subject or a patient in need thereof, suffering from metabolic disorders, amyloidosis and protein-aggregation related diseases, cellular transformation arising from genetic mutations and translocations, dominant negative effects of gene mutations, latent viral infections, and other related symptoms, using either viral or nanoparticle delivery system.

[0215] Therapeutic applications of the CRISPR-Cas system include Glaucoma, Amyloidosis, and Huntington's disease. These are exemplified in Example 20 and the features described therein are preferred alone or in combination.

[0216] Another example of a polyglutamine expansion disease that may be treated by the present invention includes spinocerebellar ataxia type 1 (SCA1). Upon intracerebellar injection, recombinant adenoassociated virus (AAV) vectors expressing short hairpin RNAs profoundly improve motor coordination, restored cerebellar morphology and resolved characteristic ataxin-inclusions in Purkinje cells of SCA1 mice (see, e.g., Xia et al., Nature Medicine. Vol. 10, No. 8, August 2004). In particular, AAV1 and AAV5 vectors are preferred and AAV titers of about 1×10^{12} vector genomes/ml are desirable.

[0217] As an example, chronic infection by HIV-1 may be treated or prevented. In order to accomplish this, one may generate CRISPR-Cas guide RNAs that target the vast majority of the HIV-1 genome while taking into account HIV-1 strain variants for maximal coverage and effectiveness. One may accomplish delivery of the CRISPR-Cas system by conventional adenoviral or lentiviral-mediated infection of the host immune system. Depending on approach, host immune cells could be a) isolated, transduced with CRISPR-Cas, selected, and re-introduced in to the host or b) transduced in vivo by systemic delivery of the CRISPR-Cas system. The first approach allows for generation of a resistant immune population whereas the second is more likely to target latent viral reservoirs within the host. This is discussed in more detail in the Examples section.

[0218] In another example, US Patent Publication No. 20130171732 assigned to Sangamo BioSciences, Inc. relates to insertion of an anti-HIV transgene into the genome, methods of which may be applied to the CRISPR Cas system of the present invention. In another embodiment, the CXCR4 gene may be targeted and the TALE system of US Patent Publication No. 20100291048 assigned to Sangamo BioSciences, Inc. may be modified to the CRISPR Cas system of the present invention. The method of US Patent Publication Nos. 20130137104 and 20130122591 assigned to Sangamo BioSciences, Inc. and US Patent Publication No. 20100146651 assigned to Cellectis may be more generally applicable for transgene expression as it involves modifying a hypoxanthine-guanine phosphoribosyltransferase (HPRT) locus for increasing the frequency of gene modification.

[0219] It is also envisaged that the present invention generates a gene knockout cell library. Each cell may have a single gene knocked out. This is exemplified in Example 23.

[0220] One may make a library of ES cells where each cell has a single gene knocked out, and the entire library of ES cells will have every single gene knocked out. This library is useful for the screening of gene function in cellular processes as well as diseases. To make this cell library, one may integrate Cas9 driven by an inducible promoter (e.g. doxycycline inducible promoter) into the ES cell. In addition, one may integrate a single guide RNA targeting a specific gene in the ES cell. To make the ES cell library, one may simply mix ES cells with a library of genes encoding guide RNAs targeting each gene in the human genome. One may first introduce a single BxB1 attB site into the AAVS1 locus of the human ES cell. Then one may use the BxB1 integrase to facilitate the integration of individual guide RNA genes into the BxB1 attB site in AAVS1 locus. To facilitate integration, each guide RNA gene may be contained on a plasmid that carries of a single attP site. This way BxB1 will recombine the attB site in the genome with the attP site on the guide RNA containing plasmid. To generate the cell library, one may take the library of cells that have single guide RNAs integrated and induce Cas9 expression. After induction, Cas9 mediates double strand break at sites specified by the guide RNA.

[0221] Chronic administration of protein therapeutics may elicit unacceptable immune responses to the specific protein. The immunogenicity of protein drugs can be ascribed to a few immunodominant helper T lymphocyte (HTL) epitopes. Reducing the MHC binding affinity of these HTL epitopes contained within these proteins can generate drugs with lower immunogenicity (Tangri S, et al. ("Rationally engineered therapeutic proteins with reduced immunogenicity" J Immunol. 2005 Mar. 15; 174(6):3187-96.) In the present invention, the immunogenicity of the CRISPR enzyme in particular may be reduced following the approach first set out in Tangri et al with respect to erythropoietin and subsequently developed. Accordingly, directed evolution or rational design may be used to reduce the immunogenicity of the CRISPR enzyme (for instance a Cas9) in the host species (human or other species).

[0222] In Example 28, Applicants used 3 guideRNAs of interest and able to visualize efficient DNA cleavage in vivo occurring only in a small subset of cells. Essentially, what Applicants have shown here is targeted in vivo cleavage. In particular, this provides proof of concept that specific targeting in higher organisms such as mammals can also be achieved. It also highlights multiplex aspect in that multiple guide sequences (i.e. separate targets) can be used simultaneously (in the sense of co-delivery). In other words, Applicants used a multiple approach, with several different sequences targeted at the same time, but independently.

[0223] A suitable example of a protocol for producing AAV, a preferred vector of the invention is provided in Example 34.

[0224] Trinucleotide repeat disorders are preferred conditions to be treated. These are also exemplified herein.

[0225] For example, US Patent Publication No. 20110016540, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with trinucleotide repeat expansion disorders. Trinucleotide repeat expansion disorders are complex, progressive disorders that involve developmental neurobiology and often affect cognition as well as sensori-motor functions.

[0226] Trinucleotide repeat expansion proteins are a diverse set of proteins associated with susceptibility for developing a trinucleotide repeat expansion disorder, the presence of a trinucleotide repeat expansion disorder, the severity of a trinucleotide repeat expansion disorder or any combination thereof. Trinucleotide repeat expansion disorders are divided into two categories determined by the type of repeat. The most common repeat is the triplet CAG, which, when present in the coding region of a gene, codes for the amino acid glutamine (Q). Therefore, these disorders are referred to as the polyglutamine (polyQ) disorders and comprise the following diseases: Huntington Disease (HD); Spinobulbar Muscular Atrophy (SBMA); Spinocerebellar Ataxias (SCA types 1, 2, 3, 6, 7, and 17); and Dentatorubro-Pallidoluysian Atrophy (DRPLA). The remaining trinucleotide repeat expansion disorders either do not involve the CAG triplet or the CAG triplet is not in the coding region of the gene and are, therefore, referred to as the non-polyglutamine disorders. The non-polyglutamine disorders comprise Fragile X Syndrome (FRAXA); Fragile XE Mental Retardation (FRAXE); Friedreich Ataxia (FRDA); Myotonic Dystrophy (DM); and Spinocerebellar Ataxias (SCA types 8, and 12).

[0227] The proteins associated with trinucleotide repeat expansion disorders are typically selected based on an experimental association of the protein associated with a trinucleotide repeat expansion disorder to a trinucleotide repeat expansion disorder. For example, the production rate or circulating concentration of a protein associated with a trinucleotide repeat expansion disorder may be elevated or depressed in a population having a trinucleotide repeat expansion disorder relative to a population lacking the trinucleotide repeat expansion disorder. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the proteins associated with trinucleotide repeat expansion disorders may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[0228] Non-limiting examples of proteins associated with trinucleotide repeat expansion disorders include AR (androgen receptor), FMR1 (fragile X mental retardation 1), HTT (huntingtin), DMPK (dystrophia myotonica-protein kinase), FXN (frataxin), ATXN2 (ataxin 2), ATN1 (atrophin 1), FEN1 (flap structure-specific endonuclease 1), TNRC6A (trinucleotide repeat containing 6A), PABPN1 (poly(A) binding protein, nuclear 1), JPH3 (junctophilin 3), MED15 (mediator complex subunit 15), ATXN1 (ataxin 1), ATXN3 (ataxin 3), TBP (TATA box binding protein), CACNA1A (calcium channel, voltage-dependent, P/Q type, alpha 1A subunit), ATXN80S (ATXN8 opposite strand (non-protein coding)), PPP2R2B (protein phosphatase 2, regulatory subunit B, beta), ATXN7 (ataxin 7), TNRC6B (trinucleotide repeat containing 6B), TNRC6C (trinucleotide repeat containing 6C), CELF3 (CUGBP, Elav-like family member 3), MAB21L1 (mab-21-like 1 (*C. elegans*)), MSH2 (mutS homolog 2, colon cancer, nonpolyposis type I (*E. coli*)). TMEM185A (transmembrane protein 185A), SIX5 (SIX homeobox 5). CNPY3 (canopy 3 homolog (zebrafish)), FRAXE (fragile site, folic acid type, rare, fra(X)q28) E), GNB2 (guanine nucleotide binding protein (G protein), beta polypeptide 2), RPL14 (ribosomal protein L14), ATXN8 (ataxin 8), INSR (insulin receptor), TTR (transthyretin), EP400 (E1A binding protein p400), GIGYF2 (GRB10 interacting GYF protein 2), OGG1 (8-oxoguanine DNA glycosylase). STC1 (stanniocalcin 1), CNDP1 (carnosine dipeptidase 1 (metallopeptidase M20 family)), C10orf2 (chromosome 10 open reading frame 2), MAML3 mastermind-like 3 (*Drosophila*), DKC1 (dyskeratosis congenita 1, dyskerin), PAXIP1 (PAX interacting (with transcription-activation domain) protein 1), CASK (calcium/calmodulin-dependent serine protein kinase (MAGUK family)), MAPT (microtubule-associated protein tau), SP1 (Sp1 transcription factor), POLG (polymerase (DNA directed), gamma), AFF2 (AF4/FMR2 family, member 2), THBS1 (thrombospondin 1), TP53 (tumor protein p53), ESR1 (estrogen receptor 1), CGGBP1 (CGG triplet repeat binding protein 1), ABT1 (activator of basal transcription 1), KLK3 (kallikrein-related peptidase 3), PRNP (prion protein), JUN (jun oncogene), KCNN3 (potassium intermediate/small conductance calcium-activated channel, subfamily N, member 3), BAX (BCL2-associated X protein), FRAXA (fragile site, folic acid type, rare, fra(X)(q27.3) A (macroorchidism, mental retardation)). KBTBD10 (kelch repeat and BTB (POZ) domain containing 10), MBNL1 (muscleblind-like (*Drosophila*)), RAD51 (RAD51 homolog (RecA homolog, *E. coli*) (*S. cerevisiae*)), NCOA3 (nuclear receptor coactivator 3), ERDA1 (expanded repeat domain, CAG/CTG 1), TSC1 (tuberous sclerosis 1), COMP (cartilage oligomeric matrix protein), GCLC (glutamate-cysteine ligase, catalytic subunit), RRAD (Ras-related associated with diabetes). MSH3 (mutS homolog 3 (*E. coli*)), DRD2 (dopamine receptor D2), CD44 (CD44 molecule (Indian blood group)), CTCF (CCCTC-binding factor (zinc finger protein)), CCND1 (cyclin D1), CLSPN (claspin homolog (*Xenopus laevis*)), MEF2A (myocyte enhancer factor 2A), PTPRU (protein tyrosine phosphatase, receptor type, U), GAPDH (glyceraldehyde-3-phosphate dehydrogenase), TRIM22 (tripartite motif-containing 22), WT1 (Wilms tumor 1), AHR (aryl hydrocarbon receptor), GPX1 (glutathione peroxidase 1), TPMT (thiopurine S-methyltransferase), NDP (Norrie disease (pseudoglioma)), ARX (aristaless related homeobox), MUS81 (MUS81 endonuclease homolog (*S. cerevisiae*)), TYR (tyrosinase (oculocutaneous albinism IA)), EGR1 (early growth response 1), UNG (uracil-DNA glycosylase). NUMBL (numb homolog (*Drosophila*)-like), FABP2 (fatty acid binding protein 2, intestinal), EN2 (engrailed homeobox 2), CRYGC (crystallin, gamma C). SRP14 (signal recognition particle 14 kDa (homologous Alu RNA binding protein)), CRYGB (crystallin, gamma B), PDCD1 (programmed cell death 1), HOXA1 (homeobox A1), ATXN2L (ataxin 2-like), PMS2 (PMS2 postmeiotic segregation increased 2 (*S. cerevisiae*)), GLA (galactosidase, alpha), CBL (Cas-Br-M (murine) ecotropic retroviral transforming sequence), FTH1 (ferritin, heavy polypeptide 1), IL12RB2 (interleukin 12 receptor, beta 2), OTX2 (orthodenticle homeobox 2), HOXA5 (homeobox A5), POLG2 (polymerase (DNA directed), gamma 2, accessory subunit), DLX2 (distal-less homeobox 2), SIRPA (signal-regulatory protein alpha). OTX1 (orthodenticle homeobox 1), AHRR (aryl-hydrocarbon receptor repressor), MANF (mesencephalic astrocyte-derived neurotrophic factor), TMEM158 (transmembrane protein 158 (gene/pseudogene)), and ENSG00000078687.

[0229] Preferred proteins associated with trinucleotide repeat expansion disorders include HTT (Huntingtin), AR (androgen receptor), FXN (frataxin), Atxn3 (ataxin), Atxn1 (ataxin), Atxn2 (ataxin), Atxn7 (ataxin), Atxn10 (ataxin), DMPK (dystrophia myotonica-protein kinase), Atn1 (atrophin 1), CBP (creb binding protein), VLDLR (very low density lipoprotein receptor), and any combination thereof.

[0230] According to another aspect, a method of gene therapy for the treatment of a subject having a mutation in the CFTR gene is provided and comprises administering a therapeutically effective amount of a CRISPR-Cas gene therapy particle, optionally via a biocompatible pharmaceutical carrier, to the cells of a subject. Preferably, the target DNA comprises the mutation deltaF508. In general, it is of preferred that the mutation is repaired to the wildtype. In this case, the mutation is a deletion of the three nucleotides that comprise the codon for phenylalanine (F) at position 508. Accordingly, repair in this instance requires reintroduction of the missing codon into the mutant.

[0231] To implement this Gene Repair Strategy, it is preferred that an adenovirus/AAV vector system is introduced into the host cell, cells or patient. Preferably, the system comprises a Cas9 (or Cas9 nickase) and the guide RNA along with a adenovirus/AAV vector system comprising the homology repair template containing the F508 residue. This may be introduced into the subject via one of the methods of delivery discussed earlier. The CRISPR-Cas system may be guided by the CFTRdelta 508 chimeric guide RNA. It targets a specific site of the CFTR genomic locus to be nicked or cleaved. After cleavage, the repair template is inserted into the cleavage site via homologous recombination correcting the deletion that results in cystic fibrosis or causes cystic fibrosis related symptoms. This strategy to direct delivery and provide systemic introduction of CRISPR systems with appropriate guide RNAs can be employed to target genetic mutations to edit or otherwise manipulate genes that cause metabolic, liver, kidney and protein diseases and disorders such as those in Table B.

[0232] The CRISPR/Cas9 systems of the present invention can be used to correct genetic mutations that were previously attempted with limited success using TALEN and ZFN. For example, WO2013163628 A2, Genetic Correction of Mutated Genes, published application of Duke University describes efforts to correct, for example, a frameshift mutation which causes a premature stop codon and a truncated gene product that can be corrected via nuclease mediated non-homologous end joining such as those responsible for Duchenne Muscular Dystrophy, ("DMD") a recessive, fatal, X-linked disorder that results in muscle degeneration due to mutations in the dystrophin gene. The majority of dystrophin mutations that cause DMD are deletions of exons that disrupt the reading frame and cause premature translation termination in the dystrophin gene. Dystrophin is a cytoplasmic protein that provides structural stability to the dystroglycan complex of the cell membrane that is responsible for regulating muscle cell integrity and function. The dystrophin gene or "DMD gene" as used interchangeably herein is 2.2 megabases at locus Xp21. The primary transcription measures about 2,400 kb with the mature mRNA being about 14 kb. 79 exons code for the protein which is over 3500 amino acids. Exon 51 is frequently adjacent to frame-disrupting deletions in DMD patients and has been targeted in clinical trials for oligonucleotide-based exon skipping. A clinical trial for the exon 51 skipping compound eteplirsen recently reported a significant functional benefit across 48 weeks, with an average of 47% dystrophin positive fibers compared to baseline. Mutations in exon 51 are ideally suited for permanent correction by NHEJ-based genome editing.

[0233] The methods of US Patent Publication No. 20130145487 assigned to Cellectis, which relates to meganuclease variants to cleave a target sequence from the human dystrophin gene (DMD), may also be modified to for the CRISPR Cas system of the present invention.

[0234] The invention uses nucleic acids to bind target DNA sequences. This is advantageous as nucleic acids are much easier and cheaper to produce than proteins, and the specificity can be varied according to the length of the stretch where homology is sought. Complex 3-D positioning of multiple fingers, for example is not required.

[0235] The terms "polynucleotide", "nucleotide", "nucleotide sequence", "nucleic acid" and "oligonucleotide" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. The term also encompasses nucleic-acid-like structures with synthetic backbones, see, e.g., Eckstein, 1991; Baserga et al., 1992; Milligan, 1993; WO 97/03211; WO 96/39154; Mata, 1997; Strauss-Soukup, 1997; and Samstag, 1996. A polynucleotide may comprise one or more modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

[0236] As used herein the term "wild type" is a term of the art understood by skilled persons and means the typical form of an organism, strain, gene or characteristic as it occurs in nature as distinguished from mutant or variant forms.

[0237] As used herein the term "variant" should be taken to mean the exhibition of qualities that have a pattern that deviates from what occurs in nature.

[0238] The terms "non-naturally occurring" or "engineered" are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature.

[0239] "Complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick base pairing or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence (e.g., 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary). "Perfectly complementary" means that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of contiguous residues in a second nucleic acid sequence. "Substantially complementary" as used herein refers to a degree of complementarity that is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions.

[0240] As used herein, "stringent conditions" for hybridization refer to conditions under which a nucleic acid having complementarity to a target sequence predominantly hybridizes with the target sequence, and substantially does not hybridize to non-target sequences. Stringent conditions are generally sequence-dependent, and vary depending on a number of factors. In general, the longer the sequence, the higher the temperature at which the sequence specifically hybridizes to its target sequence. Non-limiting examples of stringent conditions are described in detail in Tijssen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part I, Second Chapter "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N.Y. Where reference is made to a polynucleotide sequence, then complementary or partially complementary sequences are also envisaged. These are preferably capable of hybridising to the reference sequence under highly stringent conditions. Generally, in order to maximize the hybridization rate, relatively low-stringency hybridization conditions are selected: about 20 to 25° C. lower than the thermal melting point (T_m). The T_m is the temperature at which 50% of specific target sequence hybridizes to a perfectly complementary probe in solution at a defined ionic strength and pH. Generally, in order to require at least about 85% nucleotide complementarity of hybridized sequences, highly stringent washing conditions are selected to be about 5 to 15° C. lower than the T_m. In order to require at least about 70% nucleotide complementarity of hybridized sequences, moderately-stringent washing conditions are selected to be about 15 to 30° C. lower than the T_m. Highly permissive (very low stringency) washing conditions may be as low as 50° C. below the T_m, allowing a high level of mis-matching between hybridized sequences. Those skilled in the art will recognize that other physical and chemical parameters in the hybridization and wash stages can also be altered to affect the outcome of a detectable hybridization signal from a specific level of homology between target and probe sequences. Preferred highly stringent conditions comprise incubation in 50% formamide, 5×SSC, and 1% SDS at

42° C., or incubation in 5×SSC and 1% SDS at 65° C., with wash in 0.2×SSC and 0.1% SDS at 65° C.

[0241] "Hybridization" refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of PCR, or the cleavage of a polynucleotide by an enzyme. A sequence capable of hybridizing with a given sequence is referred to as the "complement" of the given sequence.

[0242] As used herein, the term "genomic locus" or "locus" (plural loci) is the specific location of a gene or DNA sequence on a chromosome. A "gene" refers to stretches of DNA or RNA that encode a polypeptide or an RNA chain that has functional role to play in an organism and hence is the molecular unit of heredity in living organisms. For the purpose of this invention it may be considered that genes include regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions.

[0243] As used herein, "expression of a genomic locus" or "gene expression" is the process by which information from a gene is used in the synthesis of a functional gene product. The products of gene expression are often proteins, but in non-protein coding genes such as rRNA genes or tRNA genes, the product is functional RNA. The process of gene expression is used by all known life—eukaryotes (including multicellular organisms), prokaryotes (bacteria and archaea) and viruses to generate functional products to survive. As used herein "expression" of a gene or nucleic acid encompasses not only cellular gene expression, but also the transcription and translation of nucleic acid(s) in cloning systems and in any other context. As used herein, "expression" also refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

[0244] The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term "amino acid" includes natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics.

[0245] As used herein, the term "domain" or "protein domain" refers to a part of a protein sequence that may exist and function independently of the rest of the protein chain.

[0246] As described in aspects of the invention, sequence identity is related to sequence homology. Homology comparisons may be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs may calculate percent (%) homology between two or more sequences and may also calculate the sequence identity shared by two or more amino acid or nucleic acid sequences. In some preferred embodiments, the capping region of the dTALEs described herein have sequences that are at least 95% identical or share identity to the capping region amino acid sequences provided herein.

[0247] Sequence homologies may be generated by any of a number of computer programs known in the art, for example BLAST or FASTA, etc. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, U.S.A; Devereux et al., 1984, Nucleic Acids Research 12:387). Examples of other software than may perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel et al., 1999 ibid—Chapter 18), FASTA (Atschul et al., 1990, J. Mol. Biol., 403-410) and the GENEWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel et al., 1999 ibid, pages 7-58 to 7-60). However it is preferred to use the GCG Bestfit program.

[0248] Percentage (%) sequence homology may be calculated over contiguous sequences, i.e., one sequence is aligned with the other sequence and each amino acid or nucleotide in one sequence is directly compared with the corresponding amino acid or nucleotide in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues.

[0249] Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion may cause the following amino acid residues to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without unduly penalizing the overall homology or identity score. This is achieved by inserting "gaps" in the sequence alignment to try to maximize local homology or identity.

[0250] However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment with as few gaps as possible—reflecting higher relatedness between the two compared sequences—may achieve a higher score than one with many gaps. "Affinity gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties may, of course, produce optimized alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example, when using the GCG Wisconsin Bestfit package the default gap penalty for amino acid sequences is −12 for a gap and −4 for each extension.

[0251] Calculation of maximum % homology therefore first requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (Devereux et al., 1984 *Nuc. Acids Research* 12 p 387). Examples of other software that may perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel et al., 1999 *Short Protocols in Molecular Biology,* 4 h Ed.—Chapter 18), FASTA (Altschul et al., 1990 *J. Mol. Biol.* 403-410) and the GENEWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel et al., 1999*, Short Protocols in Molecular Biology, pages* 7-58 to 7-60). However, for some applications, it is preferred to use the GCG Bestfit program. A new tool, called BLAST 2 Sequences is also available for comparing protein and nucleotide sequences (see *FEMS Microbiol Lett.* 1999 174(2): 247-50*; FEAMS Microbiol Lett.* 1999 177(1): 187-8 and the website of the National Center for Biotechnology information at the website of the National Institutes for Health).

[0252] Although the final % homology may be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pair-wise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix—the default matrix for the BLAST suite of programs. GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table, if supplied (see user manual for further details). For some applications, it is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62.

[0253] Alternatively, percentage homologies may be calculated using the multiple alignment feature in DNASIS™ (Hitachi Software), based on an algorithm, analogous to CLUSTAL (Higgins D G & Sharp P M (1988), *Gene* 73(1), 237-244). Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

[0254] The sequences may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent substance. Deliberate amino acid substitutions may be made on the basis of similarity in amino acid properties (such as polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues) and it is therefore useful to group amino acids together in functional groups. Amino acids may be grouped together based on the properties of their side chains alone. However, it is more useful to include mutation data as well. The sets of amino acids thus derived are likely to be conserved for structural reasons. These sets may be described in the form of a Venn diagram (Livingstone C. D. and Barton G. J. (1993) "Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation" *Comput. Appl. Biosci.* 9: 745-756) (Taylor W. R. (1986) "The classification of amino acid conservation" *J. Theor. Biol.* 119; 205-218). Conservative substitutions may be made, for example according to the table below which describes a generally accepted Venn diagram grouping of amino acids.

Set		SuB-set	
Hydrophobic	F M I L V A G C	Aromatic	F W Y H
		Aliphatic	I L V
Polar	W Y H K R E D C S T N Q	Charged	H K R E D
		Positively charged	H K R
		Negatively charged	E D
Small	V C A G S P T N D	Tiny	A G S

[0255] Embodiments of the invention include sequences (both polynucleotide or polypeptide) which may comprise homologous substitution (substitution and replacement are both used herein to mean the interchange of an existing amino acid residue or nucleotide, with an alternative residue or nucleotide) that may occur i.e., like-for-like substitution in the case of amino acids such as basic for basic, acidic for acidic, polar for polar, etc. Non-homologous substitution may also occur i.e., from one class of residue to another or alternatively involving the inclusion of unnatural amino acids such as ornithine (hereinafter referred to as Z), diaminobutyric acid ornithine (hereinafter referred to as B), norleucine ornithine (hereinafter referred to as O), pyriylalanine, thienylalanine, naphthylalanine and phenylglycine.

[0256] Variant amino acid sequences may include suitable spacer groups that may be inserted between any two amino acid residues of the sequence including alkyl groups such as methyl, ethyl or propyl groups in addition to amino acid spacers such as glycine or β-alanine residues. A further form of variation, which involves the presence of one or more amino acid residues in peptoid form, may be well understood by those skilled in the art. For the avoidance of doubt, "the peptoid form" is used to refer to variant amino acid residues wherein the α-carbon substituent group is on the residue's nitrogen atom rather than the α-carbon. Processes for preparing peptides in the peptoid form are known in the art, for example Simon R J et al., *PNAS* (1992) 89(20), 9367-9371 and Horwell D C, *Trends Biotechnol.* (1995) 13(4), 132-134.

[0257] The practice of the present invention employs, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTO-

COLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M. J. MacPherson, B. D. Hames and G. R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R.I. Freshney, ed. (1987)).

[0258] In one aspect, the invention provides for vectors that are used in the engineering and optimization of CRISPR-Cas systems.

[0259] A used herein, a "vector" is a tool that allows or facilitates the transfer of an entity from one environment to another. It is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements. In general, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses (AAVs)). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0260] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). With regards to recombination and cloning methods, mention is made of U.S. patent application Ser. No. 10/815,730, published Sep. 2, 2004 as US 2004-0171156 A1, the contents of which are herein incorporated by reference in their entirety.

[0261] Aspects of the invention relate to bicistronic vectors for chimeric RNA and Cas9. Bicistronic expression vectors for chimeric RNA and Cas9 are preferred. In general and particularly in this embodiment Cas9 is preferably driven by the CBh promoter. The chimeric RNA may preferably be driven by a U6 promoter. Ideally the two are combined. The chimeric guide RNA typically consists of a 20 bp guide sequence (Ns) and this may be joined to the tracr sequence (running from the first "U" of the lower strand to the end of the transcript). The tracr sequence may be truncated at various positions as indicated. The guide and tracr sequences are separated by the tracr-mate sequence, which may be GUUUUAGAGCUA. This may be followed by the loop sequence GAAA as shown. Both of these are preferred examples. Applicants have demonstrated Cas9-mediated indels at the human EMX1 and PVALB loci by SURVEYOR assays. ChiRNAs are indicated by their "+n" designation, and crRNA refers to a hybrid RNA where guide and tracr sequences are expressed as separate transcripts. Throughout this application, chimeric RNA may also be called single guide, or synthetic guide RNA (sgRNA). The loop is preferably GAAA, but it is not limited to this sequence or indeed to being only 4 bp in length. Indeed, preferred loop forming sequences for use in hairpin structures are four nucleotides in length, and most preferably have the sequence GAAA. However, longer or shorter loop sequences may be used, as may alternative sequences. The sequences preferably include a nucleotide triplet (for example, AAA), and an additional nucleotide (for example C or G). Examples of loop forming sequences include CAAA and AAAG.

[0262] The term "regulatory element" is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g. transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g. liver, pancreas), or particular cell types (e.g. lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector comprises one or more pol III promoter (e.g. 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g. 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g. 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) [see, e.g., Boshart et al, Cell, 41:521-530 (1985)], the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter. Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β-globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc. A vector can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., clustered regularly interspersed short palindromic repeats (CRISPR) transcripts, proteins, enzymes, mutant forms thereof, fusion proteins thereof, etc.). With regards to regulatory sequences, mention is made of U.S. patent application Ser. No. 10/491,026, the contents of which are incorporated by reference herein in their entirety. With regards to promoters, mention is made of PCT publication WO 2011/028929 and U.S. application Ser. No. 12/511,940, the contents of which are incorporated by reference herein in their entirety.

[0263] Vectors can be designed for expression of CRISPR transcripts (e.g. nucleic acid transcripts, proteins, or enzymes) in prokaryotic or eukaryotic cells. For example, CRISPR transcripts can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

[0264] Vectors may be introduced and propagated in a prokaryote or prokaryotic cell. In some embodiments, a prokaryote is used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a eukaryotic cell (e.g. amplifying a plasmid as part of a viral vector packaging system). In some embodiments, a prokaryote is used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism. Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, such as to the amino terminus of the recombinant protein. Such fusion vectors may serve one or more purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Example fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

[0265] Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann et al., (1988) Gene 69:301-315) and pET 11d (Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

[0266] In some embodiments, a vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, et al., 1987. EMBO J. 6: 229-234), pMFa (Kuijan and Herskowitz, 1982. Cell 30: 933-943), pJRY88 (Schultz et al., 1987. Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

[0267] In some embodiments, a vector drives protein expression in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, et al., 1983. Mol. Cell. Biol. 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. Virology 170: 31-39).

[0268] In some embodiments, a vector is capable of driving expression of one or more sequences in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. Nature 329: 840) and pMT2PC (Kaufman, et al., 1987. EMBO J. 6: 187-195). When used in mammalian cells, the expression vector's control functions are typically provided by one or more regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, et al., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[0269] In some embodiments, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, et al., 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Baneiji, et al., 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, et al., 1985. *Science* 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264, 166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546). With regards to these prokaryotic and eukaryotic vectors, mention is made of U.S. Pat. No. 6,750,059, the contents of which are incorporated by reference herein in their entirety. Other embodiments of the invention may relate to the use of viral vectors, with regards to which mention is made of U.S. patent application Ser. No. 13/092,085, the contents of which are incorporated by reference herein in their entirety. Tissue-specific regulatory elements are known in the art and in this regard, mention is made of U.S. Pat. No. 7,776,321, the contents of which are incorporated by reference herein in their entirety.

[0270] In some embodiments, a regulatory element is operably linked to one or more elements of a CRISPR system so as to drive expression of the one or more elements of the CRISPR system. In general, CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats), also known as SPIDRs (SPacer Interspersed Direct Repeats), constitute a family of DNA loci that are usually specific to a particular bacterial species. The CRISPR locus comprises a distinct class of interspersed short sequence repeats (SSRs) that were recognized in *E. coli* (Ishino et al., J. Bacteriol., 169:5429-5433 [1987]; and Nakata et al., J. Bacteriol., 171:3553-3556 [1989]), and associated genes. Similar interspersed SSRs have been identified in *Haloferax mediterranei, Streptococcus pyogenes, Anabaena*, and *Mycobacterium tuberculosis* (See, Groenen et al., Mol. Microbiol., 10:1057-1065 [1993]; Hoe et al., Emerg. Infect. Dis., 5:254-263 [1999]; Masepohl et al., Biochim. Biophys. Acta 1307:26-30 [1996]; and Mojica et al., Mol. Microbiol., 17:85-93 [1995]). The CRISPR loci typically differ from other SSRs by the structure of the repeats, which have been termed short regularly spaced repeats (SRSRs) (Janssen et al., OMICS J. Integ. Biol., 6:23-33 [2002]; and Mojica et al., Mol. Microbiol., 36:244-246 [2000]). In general, the repeats are short elements that occur in clusters that are regularly spaced by unique intervening sequences with a substantially constant length (Mojica et al., [2000], supra). Although the repeat sequences are highly conserved between strains, the number of interspersed repeats and the sequences of the spacer regions typically differ from strain to strain (van Embden et al., J. Bacteriol., 182:2393-2401 [2000]). CRISPR loci have been identified in more than 40 prokaryotes (See e.g., Jansen et al., Mol. Microbiol., 43:1565-1575 [2002]; and Mojica et al., [2005]) including, but not limited to *Aeropyrum, Pyrobaculum, Sulfolobus, Archaeoglobus, Halocarcula, Methanobacterium, Methanococcus, Methanosarcina, Methanopyrus, Pyrococcus, Picrophilus, Thermoplasma, Corynebacterium, Mycobacterium, Streptomyces, Aquifex, Porphyromonas, Chlorobium, Thermus, Bacillus, Listeria, Staphylococcus, Clostridium, Thermoanaerobacter, Mycoplasma, Fusobacterium, Azarcus, Chromobacterium, Neisseria, Nitrosomonas, Desulfovibrio, Geobacter, Myxococcus, Canmpylobacter, Wolinella, Acinetobacter, Erwinia, Escherichia, Legionella, Methylococcus, Pasteurella, Photobacterium, Salmonella, Xanthomonas, Yersinia, Treponema*, and *Thermotoga*.

[0271] In general, "CRISPR system" refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or other sequences and transcripts from a CRISPR locus. In embodiments of the invention the terms guide sequence and guide RNA are used interchangeably. In some embodiments, one or more elements of a CRISPR system is derived from a type I, type II, or type III CRISPR system. In some embodiments, one or more elements of a CRISPR system is derived from a particular organism comprising an endogenous CRISPR system, such as *Streptococcus pyogenes*. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell.

[0272] In some embodiments, direct repeats may be identified in silico by searching for repetitive motifs that fulfill any or all of the following criteria:

[0273] 1. found in a 2 Kb window of genomic sequence flanking the type II CRISPR locus;

[0274] 2. span from 20 to 50 bp; and

[0275] 3. interspaced by 20 to 50 bp.

[0276] In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[0277] In some embodiments, candidate tracrRNA may be subsequently predicted by sequences that fulfill any or all of the following criteria:

[0278] 1. sequence homology to direct repeats (motif search in Geneious with up to 18-bp mismatches);

[0279] 2. presence of a predicted Rho-independent transcriptional terminator in direction of transcription; and

[0280] 3. stable hairpin secondary structure between tracrRNA and direct repeat.

[0281] In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[0282] In some embodiments, chimeric synthetic guide RNAs (sgRNAs) designs may incorporate at least 12 bp of duplex structure between the direct repeat and tracrRNA.

[0283] In preferred embodiments of the invention, the CRISPR system is a type II CRISPR system and the Cas enzyme is Cas9, which catalyzes DNA cleavage. Enzymatic action by Cas9 derived from *Streptococcus pyogenes* or any closely related Cas9 generates double stranded breaks at target site sequences which hybridize to 20 nucleotides of the guide sequence and that have a protospacer-adjacent motif (PAM) sequence (examples include NGG/NRG or a PAM that can be determined as described herein) following the 20 nucleotides of the target sequence. CRISPR activity through Cas9 for site-specific DNA recognition and cleavage is defined by the guide sequence, the tracr sequence that hybridizes in part to the guide sequence and the PAM sequence. More aspects of the CRISPR system are described in Karginov and Hannon, The CRISPR system: small RNA-guided defense in bacteria and archaea, Mole Cell 2010, Jan. 15; 37(1): 7.

[0284] The type II CRISPR locus from *Streptococcus pyogenes* SF370 contains a cluster of four genes Cas9, Cas1, Cas2, and Csn1, as well as two non-coding RNA elements, tracrRNA and a characteristic array of repetitive sequences (direct repeats) interspaced by short stretches of non-repetitive sequences (spacers, about 30 bp each). In this system, targeted DNA double-strand break (DSB) is generated in four sequential steps (FIG. **2**A). First, two non-coding RNAs, the pre-crRNA array and tracrRNA, are transcribed from the

CRISPR locus. Second, tracrRNA hybridizes to the direct repeats of pre-crRNA, which is then processed into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas9 to the DNA target consisting of the protospacer and the corresponding PAM via heteroduplex formation between the spacer region of the crRNA and the protospacer DNA. Finally, Cas9 mediates cleavage of target DNA upstream of PAM to create a DSB within the protospacer (FIG. **2**A). FIG. **2**B demonstrates the nuclear localization of the codon optimized Cas9. To promote precise transcriptional initiation, the RNA polymerase III-based U6 promoter was selected to drive the expression of tracrRNA (FIG. **2**C). Similarly, a U6 promoter-based construct was developed to express a pre-crRNA array consisting of a single spacer flanked by two direct repeats (DRs, also encompassed by the term "tracr-mate sequences"; FIG. **2**C). The initial spacer was designed to target a 33-base-pair (bp) target site (30-bp protospacer plus a 3-bp CRISPR motif (PAM) sequence satisfying the NGG recognition motif of Cas9) in the human EMX1 locus (FIG. **2**C), a key gene in the development of the cerebral cortex.

[0285] Typically, in the context of an endogenous CRISPR system, formation of a CRISPR complex (comprising a guide sequence hybridized to a target sequence and complexed with one or more Cas proteins) results in cleavage of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. Without wishing to be bound by theory, the tracr sequence, which may comprise or consist of all or a portion of a wild-type tracr sequence (e.g. about or more than about 20, 26, 32, 45, 48, 54, 63, 67, 85, or more nucleotides of a wild-type tracr sequence), may also form part of a CRISPR complex, such as by hybridization along at least a portion of the tracr sequence to all or a portion of a tracr mate sequence that is operably linked to the guide sequence. In some embodiments, one or more vectors driving expression of one or more elements of a CRISPR system are introduced into a host cell such that expression of the elements of the CRISPR system direct formation of a CRISPR complex at one or more target sites. For example, a Cas enzyme, a guide sequence linked to a tracr-mate sequence, and a tracr sequence could each be operably linked to separate regulatory elements on separate vectors. Alternatively, two or more of the elements expressed from the same or different regulatory elements, may be combined in a single vector, with one or more additional vectors providing any components of the CRISPR system not included in the first vector. CRISPR system elements that are combined in a single vector may be arranged in any suitable orientation, such as one element located 5' with respect to ("upstream" of) or 3' with respect to ("downstream" of) a second element. The coding sequence of one element may be located on the same or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In some embodiments, a single promoter drives expression of a transcript encoding a CRISPR enzyme and one or more of the guide sequence, tracr mate sequence (optionally operably linked to the guide sequence), and a tracr sequence embedded within one or more intron sequences (e.g. each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the CRISPR enzyme, guide sequence, tracr mate sequence, and tracr sequence are operably linked to and expressed from the same promoter.

[0286] In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a "cloning site"). In some embodiments, one or more insertion sites (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors. In some embodiments, a vector comprises an insertion site upstream of a tracr mate sequence, and optionally downstream of a regulatory element operably linked to the tracr mate sequence, such that following insertion of a guide sequence into the insertion site and upon expression the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell. In some embodiments, a vector comprises two or more insertion sites, each insertion site being located between two tracr mate sequences so as to allow insertion of a guide sequence at each site. In such an arrangement, the two or more guide sequences may comprise two or more copies of a single guide sequence, two or more different guide sequences, or combinations of these. When multiple different guide sequences are used, a single expression construct may be used to target CRISPR activity to multiple different, corresponding target sequences within a cell. For example, a single vector may comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide sequences. In some embodiments, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-sequence-containing vectors may be provided, and optionally delivered to a cell.

[0287] In some embodiments, a vector comprises a regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, such as a Cas protein. Non-limiting examples of Cas proteins include Cas1, Cas 1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologues thereof, or modified versions thereof. In some embodiments, the unmodified CRISPR enzyme has DNA cleavage activity, such as Cas9. In some embodiments, the CRISPR enzyme directs cleavage of one or both strands at the location of a target sequence, such as within the target sequence and/or within the complement of the target sequence. In some embodiments, the CRISPR enzyme directs cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence. In some embodiments, a vector encodes a CRISPR enzyme that is mutated to with respect to a corresponding wild-type enzyme such that the mutated CRISPR enzyme lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence. For example, an aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of Cas9 from *S. pyogenes* converts Cas9 from a nuclease that cleaves both strands to a nickase (cleaves a single strand). Other examples of mutations that render Cas9 a nickase include, without limitation, H840A, N854A, and N863A. As a further example, two or more catalytic domains of Cas9 (RuvC I, RuvC II, and RuvC III or the HNH domain) may be mutated to produce a mutated Cas9 substantially lacking all DNA cleavage activity. In some embodiments, a D10A mutation is combined with one or more of H840A, N854A, or N863A mutations to produce a Cas9 enzyme substantially lacking all DNA cleavage activity. In some embodiments, a CRISPR enzyme is considered to substantially lack all DNA cleavage activity when the DNA cleavage activity of the mutated enzyme is less than about 25%, 10%, 5%, 1%, 0.1%, 0.01%, or lower with respect to its non-mutated form. Where the enzyme is not SpCas9, mutations may be made at any or all residues corresponding to positions 10, 762, 840, 854, 863 and/or 986 of SpCas9 (which may be ascertained for instance by standard sequence comparison tools). In particular, any or all of the following mutations are preferred in SpCas9: D10A, E762A, H840A, N854A, N863A and/or D986A; as well as conservative substitution for any of the replacement amino acids is also envisaged. The same (or conservative substitutions of these mutations) at corresponding positions in other Cas9s are also preferred. Particularly preferred are D10 and H840 in SpCas9. However, in other Cas9s, residues corresponding to SpCas9 D10 and H840 are also preferred.

[0288] An aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of SpCas9 was engineered to convert the nuclease into a nickase (SpCas9n) (see e.g. Sapranauskas et al., 2011, Nucleic Acis Research, 39: 9275; Gasiunas et al., 2012. Proc. Natl. Acad. Sci. USA, 109:E2579), such that nicked genomic DNA undergoes the high-fidelity homology-directed repair (HDR). Surveyor assay confirmed that SpCas9n does not generate indels at the EMX1 protospacer target. Co-expression of EMX1-targeting chimeric crRNA (having the tracrRNA component as well) with SpCas9 produced indels in the target site, whereas co-expression with SpCas9n did not (n=3). Moreover, sequencing of 327 amplicons did not detect any indels induced by SpCas9n. The same locus was selected to test CRISPR-mediated HR by co-transfecting HEK 293FT cells with the chimeric RNA targeting EMX1, hSpCas9 or hSpCas9n, as well as a HR template to introduce a pair of restriction sites (HindIII and NheI) near the protospacer.

[0289] Preferred orthologs are described herein. A Cas enzyme may be identified as Cas9 as this can refer to the general class of enzymes that share homology to the biggest nuclease with multiple nuclease domains from the type II CRISPR system. Most preferably, the Cas9 enzyme is from, or is derived from, spCas9 or saCas9. By derived, Applicants mean that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as described herein.

[0290] It will be appreciated that the terms Cas and CRISPR enzyme are generally used herein interchangeably, unless otherwise apparent. As mentioned above, many of the residue numberings used herein refer to the Cas9 enzyme from the type II CRISPR locus in *Streptococcus pyogenes*. However, it will be appreciated that this invention includes many more Cas9s from other species of microbes, such as SpCas9, SaCa9, St1Cas9 and so forth.

[0291] An example of a codon optimized sequence, in this instance optimized for humans (i.e. being optimized for expression in humans) is provided herein, see the SaCas9 human codon optimized sequence. Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species is known.

[0292] In some embodiments, an enzyme coding sequence encoding a CRISPR enzyme is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a mammal, including but not limited to human, mouse, rat, rabbit, dog, or non-human mammal or primate. In some embodiments, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded.

[0293] In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g. about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at www.kazusa.orjp/codon/ (visited Jul. 9, 2002), and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, Pa.), are also available. In some embodiments, one or more codons (e.g. 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a CRISPR enzyme correspond to the most frequently used codon for a particular amino acid.

[0294] In some embodiments, a vector encodes a CRISPR enzyme comprising one or more nuclear localization sequences (NLSs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs. In some embodiments, the CRISPR enzyme comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g. one or more NLS at the amino-terminus and one or more NLS at the carboxy terminus). When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies. In a preferred embodiment of the invention, the CRISPR enzyme comprises at most 6 NLSs. In some embodiments, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus. Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV; the NLS from nucleoplasmin (e.g. the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK); the c-myc NLS having the amino acid sequence PAAKRVKLD or RQRRNELKRSP; the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY; the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV of the IBB domain from importin-alpha; the sequences VSRKRPRP and PPKKARED of the myoma T protein; the sequence POPKKKPL of human p53; the sequence SALIKKKKKMAP of mouse c-abl IV; the sequences DRLRR and PKQKKRK of the influenza virus NS1; the sequence RKLKKKIKKL of the Hepatitis virus delta antigen; the sequence REKKKFLKRR of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK of the steroid hormone receptors (human) glucocorticoid.

[0295] In general, the one or more N LSs are of sufficient strength to drive accumulation of the CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the CRISPR enzyme, the particular NLS(s) used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the CRISPR enzyme, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g. a stain specific for the nucleus such as DAPI). Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay for the effect of CRISPR complex formation (e.g. assay for DNA cleavage or mutation at the target sequence, or assay for altered gene expression activity affected by CRISPR complex formation and/or CRISPR enzyme activity), as compared to a control no exposed to the CRISPR enzyme or complex, or exposed to a CRISPR enzyme lacking the one or more NLSs.

[0296] In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraf.com), ELAND (Illumina, San Diego, Calif.), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. The ability of a guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[0297] A guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a genome of a cell. Exemplary target sequences include those that are unique in the target genome. For example, for the *S. pyogenes* Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXGG where NNNNNNNNNNNNXGG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. pyogenes* Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNXGG where NNNNNNNNNNNXGG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. For the *S. thermophilus* CRISPR1 Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXXAGAAW where NNNNNNNNNNNNXXAGAAW (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. thermophilus* CRISPR1Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNXXAGAAW where NNNNNNNNNNNXXAGAAW (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. For the *S. pyogenes* Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXGGXG where NNNNNNNNNNNNXGGXG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. pyogenes* Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNXGGXG where NNNNNNNNNNNXGGXG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. In each of these sequences "M" may be A, G, T, or C, and need not be considered in identifying a sequence as unique.

[0298] In some embodiments, a guide sequence is selected to reduce the degree of secondary structure within the guide sequence. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the guide sequence participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g. A. R. Gruber et al., 2008, *Cell* 106(1): 23-24; and P A Carr and G M Church, 2009, *Nature Biotechnology* 27(12): 1151-62).

[0299] In general, a tracr mate sequence includes any sequence that has sufficient complementarity with a tracr sequence to promote one or more of: (1) excision of a guide sequence flanked by tracr mate sequences in a cell containing the corresponding tracr sequence; and (2) formation of a CRISPR complex at a target sequence, wherein the CRISPR complex comprises the tracr mate sequence hybridized to the tracr sequence. In general, degree of complementarity is with reference to the optimal alignment of the tracr mate sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the tracr sequence or tracr mate sequence. In some embodiments, the degree of complementarity between the tracr sequence and tracr mate sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and tracr mate sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. In an embodiment of the invention, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In preferred embodiments, the transcript has two, three, four or five hairpins. In a further embodiment of the invention, the transcript has at most five hairpins. In a hairpin structure the portion of the sequence 5' of the final "N" and upstream of the loop corresponds to the tracr mate sequence, and the portion of the sequence 3' of the loop corresponds to the tracr sequence Further non-limiting examples of single polynucleotides comprising a guide sequence, a tracr mate sequence, and a tracr sequence are as follows (listed 5' to 3'), where "N" represents a base of a guide sequence, the first block of lower case letters represent the tracr mate sequence, and the second block of lower case letters represent the tracr sequence, and the final poly-T sequence represents the transcription terminator:

```
(1)
NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaagatttaGAAA
taaatcttgcagaagctacaaagataaggcttcatgccgaaatcaa
caccctgtcattttatggcagggtgttttcgttatttaaTTTTTT;

(2)
NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaa
gctacaaagataaggcttcatgccgaaatcaacaccctgtcatttt
atggcagggtgttttcgttatttaaTTTTT;

(3)
NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaa
gctacaaagataaggcttcatgccgaaatcaacaccctgtcatttt
atggcagggtgtTTTTTT;

(4)
NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaagtta
aaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcg
gtgcTTTTT;
```

-continued

```
(5)
NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagtta
aaataaggctagtccgttatcaacttgaaaaagtgTTTTTTT;
and

(6)
NNNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaagtta
aaataaggctagtccgttatcaTTTTTTTT.
```

[0300] In some embodiments, sequences (1) to (3) are used in combination with Cas9 from *S. thermophilus* CRISPR1. In some embodiments, sequences (4) to (6) are used in combination with Cas9 from *S. pyogenes*. In some embodiments, the tracr sequence is a separate transcript from a transcript comprising the tracr mate sequence.

[0301] In some embodiments, a recombination template is also provided. A recombination template may be a component of another vector as described herein, contained in a separate vector, or provided as a separate polynucleotide. In some embodiments, a recombination template is designed to serve as a template in homologous recombination, such as within or near a target sequence nicked or cleaved by a CRISPR enzyme as a part of a CRISPR complex. A template polynucleotide may be of any suitable length, such as about or more than about 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, or more nucleotides in length. In some embodiments, the template polynucleotide is complementary to a portion of a polynucleotide comprising the target sequence. When optimally aligned, a template polynucleotide might overlap with one or more nucleotides of a target sequences (e.g. about or more than about 1, 5, 10, 15, 20, or more nucleotides). In some embodiments, when a template sequence and a polynucleotide comprising a target sequence are optimally aligned, the nearest nucleotide of the template polynucleotide is within about 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, or more nucleotides from the target sequence.

[0302] In some embodiments, the CRISPR enzyme is part of a fusion protein comprising one or more heterologous protein domains (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more domains in addition to the CRISPR enzyme). A CRISPR enzyme fusion protein may comprise any additional protein sequence, and optionally a linker sequence between any two domains. Examples of protein domains that may be fused to a CRISPR enzyme include, without limitation, epitope tags, reporter gene sequences, and protein domains having one or more of the following activities: methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity and nucleic acid binding activity. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to, glutathione-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). A CRISPR enzyme may be fused to a gene sequence encoding a protein or a fragment of a protein that bind DNA molecules or bind other cellular molecules, including but not limited to maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4A DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions. Additional domains that may form part of a fusion protein comprising a CRISPR enzyme are described in US20110059502, incorporated herein by reference. In some embodiments, a tagged CRISPR enzyme is used to identify the location of a target sequence.

[0303] In some embodiments, a CRISPR enzyme may form a component of an inducible system. The inducible nature of the system would allow for spatiotemporal control of gene editing or gene expression using a form of energy. The form of energy may include but is not limited to electromagnetic radiation, sound energy, chemical energy and thermal energy. Examples of inducible system include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc), or light inducible systems (Phytochrome, LOV domains, or cryptochrome). In one embodiment, the CRISPR enzyme may be a part of a Light Inducible Transcriptional Effector (LITE) to direct changes in transcriptional activity in a sequence-specific manner. The components of a light may include a CRISPR enzyme, a light-responsive cytochrome heterodimer (e.g. from *Arabidopsis thaliana*), and a transcriptional activation/repression domain. Further examples of inducible DNA binding proteins and methods for their use are provided in U.S. 61/736,465 and U.S. 61/721,283, which is hereby incorporated by reference in its entirety.

[0304] In some aspects, the invention provides methods comprising delivering one or more polynucleotides, such as or one or more vectors as described herein, one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell. In some aspects, the invention further provides cells produced by such methods, and animals comprising or produced from such cells. In some embodiments, a CRISPR enzyme in combination with (and optionally complexed with) a guide sequence is delivered to a cell. Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding components of a CRISPR system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, Science 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada et al., in Current Topics in Microbiology and Immunology Doerfler and Böhm (eds) (1995); and Yu et al., Gene Therapy 1:13-26 (1994).

[0305] Methods of non-viral delivery of nucleic acids include lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424; WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration).

[0306] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[0307] The use of RNA or DNA viral based systems for the delivery of nucleic acids take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro, and the modified cells may optionally be administered to patients (ex vivo). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0308] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommnerfelt et al., Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); PCT/US94/05700).

[0309] In another embodiment, Cocal vesiculovirus envelope pseudotyped retroviral vector particles are contemplated (see, e.g., US Patent Publication No. 20120164118 assigned to the Fred Hutchinson Cancer Research Center). Cocal virus is in the Vesiculovirus genus, and is a causative agent of vesicular stomatitis in mammals. Cocal virus was originally isolated from mites in Trinidad (Jonkers et al., Am. J. Vet. Res. 25:236-242 (1964)), and infections have been identified in Trinidad, Brazil, and Argentina from insects, cattle, and horses. Many of the vesiculoviruses that infect mammals have been isolated from naturally infected arthropods, suggesting that they are vector-borne. Antibodies to vesiculoviruses are common among people living in rural areas where the viruses are endemic and laboratory-acquired; infections in humans usually result in influenza-like symptoms. The Cocal virus envelope glycoprotein shares 71.5% identity at the amino acid level with VSV-G Indiana, and phylogenetic comparison of the envelope gene of vesiculoviruses shows that Cocal virus is serologically distinct from, but most closely related to, VSV-G Indiana strains among the vesiculoviruses. Jonkers et al., Am. J. Vet. Res. 25:236-242 (1964) and Travassos da Rosa et al., Am. J. Tropical Med. & Hygiene 33:999-1006 (1984). The Cocal vesiculovirus envelope pseudotyped retroviral vector particles may include for example, lentiviral, alpharetroviral, betaretroviral, gammaretroviral, deltaretroviral, and epsilonretroviral vector particles that may comprise retroviral Gag, Pol, and/or one or more accessory protein(s) and a Cocal vesiculovirus envelope protein. Within certain aspects of these embodiments, the Gag, Pol, and accessory proteins are lentiviral and/or gammaretroviral.

[0310] In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors may also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., Virology 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989).

[0311] Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producer a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line may also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

[0312] Accordingly, AAV is considered an ideal candidate for use as a transducing vector. Such AAV transducing vectors can comprise sufficient cis-acting functions to replicate in the presence of adenovirus or herpesvirus or poxvirus (e.g., vaccinia virus) helper functions provided in trans. Recombinant AAV (rAAV) can be used to carry exogenous genes into cells of a variety of lineages. In these vectors, the AAV cap and/or rep genes are deleted from the viral genome and replaced with a DNA segment of choice. Current AAV vectors may accommodate up to 4300 bases of inserted DNA.

[0313] There are a number of ways to produce rAAV, and the invention provides rAAV and methods for preparing rAAV. For example, plasmid(s) containing or consisting essentially of the desired viral construct are transfected into AAV-infected cells. In addition, a second or additional helper plasmid is cotransfected into these cells to provide the AAV rep and/or cap genes which are obligatory for replication and packaging of the recombinant viral construct. Under these conditions, the rep and/or cap proteins of AAV act in trans to stimulate replication and packaging of the rAAV construct. Two to Three days after transfection, rAAV is harvested. Traditionally rAAV is harvested from the cells along with adenovirus. The contaminating adenovirus is then inactivated by heat treatment. In the instant invention, rAAV is advantageously harvested not from the cells themselves, but from cell supernatant. Accordingly, in an initial aspect the invention provides for preparing rAAV, and in addition to the foregoing, rAAV can be prepared by a method that comprises or consists essentially of: infecting susceptible cells with a rAAV containing exogenous DNA including DNA for expression, and helper virus (e.g., adenovirus, herpesvirus, poxvirus such as vaccinia virus) wherein the rAAV lacks functioning cap and/or rep (and the helper virus (e.g., adenovirus, herpesvirus, poxvirus such as vaccinia virus) provides the cap and/or rev function that the rAAV lacks); or infecting susceptible cells with a rAAV containing exogenous DNA including DNA for expression, wherein the recombinant lacks functioning cap and/or rep, and transfecting said cells with a plasmid supplying cap and/or rep function that the rAAV lacks; or infecting susceptible cells with a rAAV containing exogenous DNA including DNA for expression, wherein the recombinant lacks functioning cap and/or rep, wherein said cells supply cap and/or rep function that the recombinant lacks; or transfecting the susceptible cells with an AAV lacking functioning cap and/or rep and plasmids for inserting exogenous DNA into the recombinant so that the exogenous DNA is expressed by the recombinant and for supplying rep and/or cap functions whereby transfection results in an rAAV containing the exogenous DNA including DNA for expression that lacks functioning cap and/or rep.

[0314] The rAAV can be from an AAV as herein described, and advantageously can be an rAAV1, rAAV2, AAV5 or rAAV having hybrid or capsid which may comprise AAV1, AAV2, AAV5 or any combination thereof. One can select the AAV of the rAAV with regard to the cells to be targeted by the rAAV; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid or capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue.

[0315] In addition to 293 cells, other cells that can be used in the practice of the invention and the relative infectivity of certain AAV serotypes in vitro as to these cells (see Grimm. D. et al, J. Virol. 82: 5887-5911 (2008)) are as follows:

Cell Line	AAV-1	AAV-2	AAV-3	AAV-4	AAV-5	AAV-6	AAV-8	AAV-9
Huh-7	13	100	2.5	0.0	0.1	10	0.7	0.0
HEK293	25	100	2.5	0.1	0.1	5	0.7	0.1
HeLa	3	100	2.0	0.1	6.7	1	0.2	0.1
HepG2	3	100	16.7	0.3	1.7	5	0.3	ND
Hep1A	20	100	0.2	1.0	0.1	1	0.2	0.0
911	17	100	11	0.2	0.1	17	0.1	ND
CHO	100	100	14	1.4	333	50	10	1.0
COS	33	100	33	3.3	5.0	14	2.0	0.5
MeWo	10	100	20	0.3	6.7	10	1.0	0.2
NIH3T3	10	100	2.9	2.9	0.3	10	0.3	ND
A549	14	100	20	ND	0.5	10	0.5	0.1
HT1180	20	100	10	0.1	0.3	33	0.5	0.1
Monocytes	1111	100	ND	ND	125	1429	ND	ND
Immature DC	2500	100	ND	ND	222	2857	ND	ND
Mature DC	2222	100	ND	ND	333	3333	ND	ND

[0316] The invention provides rAAV that contains or consists essentially of an exogenous nucleic acid molecule encoding a CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) system, e.g., a plurality of cassettes comprising or consisting a first cassette comprising or consisting essentially of a promoter, a nucleic acid molecule encoding a CRISPR-associated (Cas) protein (putative nuclease or helicase proteins), e.g., Cas9 and a terminator, and a two, or more, advantageously up to the packaging size limit of the vector, e.g., in total (including the first cassette) five, cassettes comprising or consisting essentially of a promoter, nucleic acid molecule encoding guide RNA (gRNA) and a terminator (e.g., each cassette schematically represented as Promoter-gRNA1-terminator, Promoter-gRNA2-terminator . . . Promoter-gRNA(N)-terminator (where N is a number that can be inserted that is at an upper limit of the packaging size limit of the vector), or two or more individual rAAVs, each containing one or more than one cassette of a CRISPR system, e.g., a first rAAV containing the first cassette comprising or consisting essentially of a promoter, a nucleic acid molecule encoding Cas, e.g., Cas9 and a terminator, and a second rAAV containing a plurality, four, cassettes comprising or consisting essentially of a promoter, nucleic acid molecule encoding guide RNA (gRNA) and a terminator (e.g., each cassette schematically represented as Promoter-gRNA1-terminator, Promoter-gRNA2-terminator . . . Promoter-gRNA(N)-terminator (where N is a number that can be inserted that is at an upper limit of the packaging size limit of the vector). As rAAV is a DNA virus, the nucleic acid molecules in the herein discussion concerning AAV or rAAV are advantageously DNA. The promoter is in some embodiments advantageously human Synapsin I promoter (hSyn).

[0317] Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

[0318] In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panel, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr −/−, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN/OPCT cell lines, Peer, PNT-1A/PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In some embodiments, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the components of a CRISPR system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a CRISPR complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[0319] In some embodiments, one or more vectors described herein are used to produce a non-human transgenic animal or transgenic plant. In some embodiments, the transgenic animal is a mammal, such as a mouse, rat, or rabbit. Methods for producing transgenic animals and plants are known in the art, and generally begin with a method of cell transfection, such as described herein.

[0320] In another embodiment, a fluid delivery device with an array of needles (see, e.g., US Patent Publication No. 20110230839 assigned to the Fred Hutchinson Cancer Research Center) may be contemplated for delivery of CRISPR Cas to solid tissue. A device of US Patent Publication No. 20110230839 for delivery of a fluid to a solid tissue may comprise a plurality of needles arranged in an array; a plurality of reservoirs, each in fluid communication with a respective one of the plurality of needles; and a plurality of actuators operatively coupled to respective ones of the plurality of reservoirs and configured to control a fluid pressure within the reservoir. In certain embodiments each of the plurality of actuators may comprise one of a plurality of plungers, a first end of each of the plurality of plungers being received in a respective one of the plurality of reservoirs, and in certain further embodiments the plungers of the plurality of plungers are operatively coupled together at respective second ends so as to be simultaneously depressable. Certain still further embodiments may comprise a plunger driver configured to depress all of the plurality of plungers at a selectively variable rate. In other embodiments each of the plurality of actuators may comprise one of a plurality of fluid transmission lines having first and second ends, a first end of each of the plurality of fluid transmission lines being coupled to a respective one of the plurality of reservoirs. In other embodiments the device may comprise a fluid pressure source, and each of the plurality of actuators comprises a fluid coupling between the fluid pressure source and a respective one of the plurality of reservoirs. In further embodiments the fluid pressure source may comprise at least one of a compressor, a vacuum accumulator, a peristaltic pump, a master cylinder, a microfluidic pump, and a valve. In another embodiment, each of the plurality of needles may comprise a plurality of ports distributed along its length.

[0321] In one aspect, the invention provides for methods of modifying a target polynucleotide in a eukaryotic cell, which may be in vivo, ex vivo or in vitro. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal, or a plant, and modifying the cell or cells. Culturing may occur at any stage ex vivo. The cell or cells may even be re-introduced into the non-human animal or plant. For re-introduced cells it is particularly preferred that the cells are stem cells.

[0322] In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence.

[0323] In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. Similar considerations and conditions apply as above for methods of modifying a target polynucleotide. In fact, these sampling, culturing and re-introduction options apply across the aspects of the present invention.

[0324] Indeed, in any aspect of the invention, the CRISPR complex may comprise a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence, wherein said guide sequence may be linked to a tracr mate sequence which in turn may hybridize to a tracr sequence. Similar considerations and conditions apply as above for methods of modifying a target polynucleotide.

[0325] In one aspect, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. Elements may be provided individually or in combinations, and may be provided in any suitable container, such as a vial, a bottle, or a tube. In some embodiments, the kit includes instructions in one or more languages, for example in more than one language.

[0326] In some embodiments, a kit comprises one or more reagents for use in a process utilizing one or more of the elements described herein. Reagents may be provided in any suitable container. For example, a kit may provide one or more reaction or storage buffers. Reagents may be provided in a form that is usable in a particular assay, or in a form that requires addition of one or more other components before use (e.g. in concentrate or lyophilized form). A buffer can be any buffer, including but not limited to a sodium carbonate buffer, a sodium bicarbonate buffer, a borate buffer, a Tris buffer, a MOPS buffer, a HEPES buffer, and combinations thereof. In some embodiments, the buffer is alkaline. In some embodiments, the buffer has a pH from about 7 to about 10. In some embodiments, the kit comprises one or more oligonucleotides corresponding to a guide sequence for insertion into a vector so as to operably link the guide sequence and a regulatory element. In some embodiments, the kit comprises a homologous recombination template polynucleotide. In some embodiments, the kit comprises one or more of the vectors and/or one or more of the polynucleotides described herein. The kit may advantageously allows to provide all elements of the systems of the invention.

[0327] In one aspect, the invention provides methods for using one or more elements of a CRISPR system. The CRISPR complex of the invention provides an effective means for modifying a target polynucleotide. The CRISPR complex of the invention has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target polynucleotide in a multiplicity of cell types. As such the CRISPR complex of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis. An exemplary CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within the target polynucleotide. The guide sequence is linked to a tracr mate sequence, which in turn hybridizes to a tracr sequence.

[0328] In one embodiment, this invention provides a method of cleaving a target polynucleotide. The method comprises modifying a target polynucleotide using a CRISPR complex that binds to the target polynucleotide and effect cleavage of said target polynucleotide. Typically, the CRISPR complex of the invention, when introduced into a cell, creates a break (e.g., a single or a double strand break) in the genome sequence. For example, the method can be used to cleave a disease gene in a cell.

[0329] The break created by the CRISPR complex can be repaired by a repair processes such as the error prone non-homologous end joining (NHEJ) pathway or the high fidelity homology-directed repair (HDR) (FIG. **29**). During these repair process, an exogenous polynucleotide template can be introduced into the genome sequence. In some methods, the HDR process is used to modify the genome sequence. For example, an exogenous polynucleotide template comprising a sequence to be integrated flanked by an upstream sequence and a downstream sequence is introduced into a cell. The upstream and downstream sequences share sequence similarity with either side of the site of integration in the chromosome.

[0330] Where desired, a donor polynucleotide can be DNA, e.g., a DNA plasmid, a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), a viral vector, a linear piece of DNA, a PCR fragment, a naked nucleic acid, or a nucleic acid complexed with a delivery vehicle such as a liposome or poloxamer.

[0331] The exogenous polynucleotide template comprises a sequence to be integrated (e.g., a mutated gene). The sequence for integration may be a sequence endogenous or exogenous to the cell. Examples of a sequence to be integrated include polynucleotides encoding a protein or a non-coding RNA (e.g., a microRNA). Thus, the sequence for integration may be operably linked to an appropriate control sequence or sequences. Alternatively, the sequence to be integrated may provide a regulatory function.

[0332] The upstream and downstream sequences in the exogenous polynucleotide template are selected to promote recombination between the chromosomal sequence of interest and the donor polynucleotide. The upstream sequence is a nucleic acid sequence that shares sequence similarity with the genome sequence upstream of the targeted site for integration. Similarly, the downstream sequence is a nucleic acid sequence that shares sequence similarity with the chromosomal sequence downstream of the targeted site of integration. The upstream and downstream sequences in the exogenous polynucleotide template can have 75%, 80%, 85%, 90%, 95%, or 100% sequence identity with the targeted genome sequence. Preferably, the upstream and downstream sequences in the exogenous polynucleotide template have about 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with the targeted genome sequence. In some methods, the upstream and downstream sequences in the exogenous polynucleotide template have about 99% or 100% sequence identity with the targeted genome sequence.

[0333] An upstream or downstream sequence may comprise from about 20 bp to about 2500 bp, for example, about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, or 2500 bp. In some methods, the exemplary upstream or downstream sequence have about 200 bp to about 2000 bp, about 600 bp to about 1000 bp, or more particularly about 700 bp to about 1000 bp.

[0334] In some methods, the exogenous polynucleotide template may further comprise a marker. Such a marker may make it easy to screen for targeted integrations. Examples of suitable markers include restriction sites, fluorescent proteins, or selectable markers. The exogenous polynucleotide template of the invention can be constructed using recombinant techniques (see, for example, Sambrook et al., 2001 and Ausubel et al., 1996).

[0335] In an exemplary method for modifying a target polynucleotide by integrating an exogenous polynucleotide template, a double stranded break is introduced into the genome sequence by the CRISPR complex, the break is repaired via homologous recombination an exogenous polynucleotide template such that the template is integrated into the genome. The presence of a double-stranded break facilitates integration of the template.

[0336] In other embodiments, this invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. The method comprises increasing or decreasing expression of a target polynucleotide by using a CRISPR complex that binds to the polynucleotide.

[0337] In some methods, a target polynucleotide can be inactivated to effect the modification of the expression in a cell. For example, upon the binding of a CRISPR complex to a target sequence in a cell, the target polynucleotide is inactivated such that the sequence is not transcribed, the coded protein is not produced, or the sequence does not function as the wild-type sequence does. For example, a protein or microRNA coding sequence may be inactivated such that the protein is not produced.

[0338] In some methods, a control sequence can be inactivated such that it no longer functions as a control sequence. As used herein, "control sequence" refers to any nucleic acid sequence that effects the transcription, translation, or accessibility of a nucleic acid sequence. Examples of a control sequence include, a promoter, a transcription terminator, and an enhancer are control sequences.

[0339] The inactivated target sequence may include a deletion mutation (i.e., deletion of one or more nucleotides), an insertion mutation (i.e., insertion of one or more nucleotides), or a nonsense mutation (i.e., substitution of a single nucleotide for another nucleotide such that a stop codon is introduced). In some methods, the inactivation of a target sequence results in "knock-out" of the target sequence.

[0340] A method of the invention may be used to create a plant, an animal or cell that may be used as a disease model. As used herein, "disease" refers to a disease, disorder, or indication in a subject. For example, a method of the invention may be used to create an animal or cell that comprises a modification in one or more nucleic acid sequences associated with a disease, or a plant, animal or cell in which the expression of one or more nucleic acid sequences associated with a disease are altered. Such a nucleic acid sequence may encode a disease associated protein sequence or may be a disease associated control sequence. Accordingly, it is understood that in embodiments of the invention, a plant, subject, patient, organism or cell can be a non-human subject, patient, organism or cell. Thus, the invention provides a plant, animal or cell, produced by the present methods, or a progeny thereof. The progeny may be a clone of the produced plant or animal, or may result from sexual reproduction by crossing with other individuals of the same species to introgress further desirable traits into their offspring. The cell may be in vivo or ex vivo in the cases of multicellular organisms, particularly animals or plants. In the instance where the cell is in cultured, a cell line may be established if appropriate culturing conditions are met and preferably if the cell is suitably adapted for this purpose (for instance a stem cell). Bacterial cell lines produced by the invention are also envisaged. Hence, cell lines are also envisaged.

[0341] In some methods, the disease model can be used to study the effects of mutations on the animal or cell and development and/or progression of the disease using measures commonly used in the study of the disease. Alternatively, such a disease model is useful for studying the effect of a pharmaceutically active compound on the disease.

[0342] In some methods, the disease model can be used to assess the efficacy of a potential gene therapy strategy. That is, a disease-associated gene or polynucleotide can be modified such that the disease development and/or progression is inhibited or reduced. In particular, the method comprises modifying a disease-associated gene or polynucleotide such that an altered protein is produced and, as a result, the animal or cell has an altered response. Accordingly, in some methods, a genetically modified animal may be compared with an animal predisposed to development of the disease such that the effect of the gene therapy event may be assessed.

[0343] In another embodiment, this invention provides a method of developing a biologically active agent that modulates a cell signaling event associated with a disease gene. The method comprises contacting a test compound with a cell comprising one or more vectors that drive expression of one or more of a CRISPR enzyme, a guide sequence linked to a tracr mate sequence, and a tracr sequence; and detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with, e.g., a mutation in a disease gene contained in the cell.

[0344] A cell model or animal model can be constructed in combination with the method of the invention for screening a cellular function change. Such a model may be used to study the effects of a genome sequence modified by the CRISPR complex of the invention on a cellular function of interest. For example, a cellular function model may be used to study the effect of a modified genome sequence on intracellular signaling or extracellular signaling. Alternatively, a cellular function model may be used to study the effects of a modified genome sequence on sensory perception. In some such models, one or more genome sequences associated with a signaling biochemical pathway in the model are modified.

[0345] Several disease models have been specifically investigated. These include de novo autism risk genes CHD8, KATNAL2, and SCN2A; and the syndromic autism (Angelman Syndrome) gene UBE3A. These genes and resulting autism models are of course preferred, but serve to show the broad applicability of the invention across genes and corresponding models.

[0346] An altered expression of one or more genome sequences associated with a signaling biochemical pathway can be determined by assaying for a difference in the mRNA levels of the corresponding genes between the test model cell and a control cell, when they are contacted with a candidate agent. Alternatively, the differential expression of the sequences associated with a signaling biochemical pathway is determined by detecting a difference in the level of the encoded polypeptide or gene product.

[0347] To assay for an agent-induced alteration in the level of mRNA transcripts or corresponding polynucleotides, nucleic acid contained in a sample is first extracted according to standard methods in the art. For instance, mRNA can be isolated using various lytic enzymes or chemical solutions according to the procedures set forth in Sambrook et al. (1989), or extracted by nucleic-acid-binding resins following the accompanying instructions provided by the manufacturers. The mRNA contained in the extracted nucleic acid sample is then detected by amplification procedures or conventional hybridization assays (e.g. Northern blot analysis) according to methods widely known in the art or based on the methods exemplified herein.

[0348] For purpose of this invention, amplification means any method employing a primer and a polymerase capable of replicating a target sequence with reasonable fidelity. Amplification may be carried out by natural or recombinant DNA polymerases such as TaqGold™, T7 DNA polymerase, Klenow fragment of *E. coli* DNA polymerase, and reverse transcriptase. A preferred amplification method is PCR. In particular, the isolated RNA can be subjected to a reverse transcription assay that is coupled with a quantitative polymerase chain reaction (RT-PCR) in order to quantify the expression level of a sequence associated with a signaling biochemical pathway.

[0349] Detection of the gene expression level can be conducted in real time in an amplification assay. In one aspect, the amplified products can be directly visualized with fluorescent DNA-binding agents including but not limited to DNA intercalators and DNA groove binders. Because the amount of the intercalators incorporated into the double-stranded DNA molecules is typically proportional to the amount of the amplified DNA products, one can conveniently determine the amount of the amplified products by quantifying the fluorescence of the intercalated dye using conventional optical systems in the art. DNA-binding dye suitable for this application include SYBR green, SYBR blue, DAPI, propidium iodine, Hoeste, SYBR gold, ethidium bromide, acridines, proflavine, acridine orange, acriflavine, fluorcoumanin, ellipticine, daunomycin, chloroquine, distamycin D, chromomycin, homidium, mithramycin, ruthenium polypyridyls, anthramycin, and the like.

[0350] In another aspect, other fluorescent labels such as sequence specific probes can be employed in the amplification reaction to facilitate the detection and quantification of the amplified products. Probe-based quantitative amplification relies on the sequence-specific detection of a desired amplified product. It utilizes fluorescent, target-specific probes (e.g., TaqMan® probes) resulting in increased specificity and sensitivity. Methods for performing probe-based quantitative amplification are well established in the art and are taught in U.S. Pat. No. 5,210,015.

[0351] In yet another aspect, conventional hybridization assays using hybridization probes that share sequence homology with sequences associated with a signaling biochemical pathway can be performed. Typically, probes are allowed to form stable complexes with the sequences associated with a signaling biochemical pathway contained within the biological sample derived from the test subject in a hybridization reaction. It will be appreciated by one of skill in the art that where antisense is used as the probe nucleic acid, the target polynucleotides provided in the sample are chosen to be complementary to sequences of the antisense nucleic acids. Conversely, where the nucleotide probe is a sense nucleic acid, the target polynucleotide is selected to be complementary to sequences of the sense nucleic acid.

[0352] Hybridization can be performed under conditions of various stringency. Suitable hybridization conditions for the practice of the present invention are such that the recognition interaction between the probe and sequences associated with a signaling biochemical pathway is both sufficiently specific and sufficiently stable. Conditions that increase the stringency of a hybridization reaction are widely known and published in the art. See, for example, (Sambrook, et al., (1989);

Nonradioactive In Situ Hybridization Application Manual, Boehringer Mannheim, second edition). The hybridization assay can be formed using probes immobilized on any solid support, including but are not limited to nitrocellulose, glass, silicon, and a variety of gene arrays. A preferred hybridization assay is conducted on high-density gene chips as described in U.S. Pat. No. 5,445,934.

[0353] For a convenient detection of the probe-target complexes formed during the hybridization assay, the nucleotide probes are conjugated to a detectable label. Detectable labels suitable for use in the present invention include any composition detectable by photochemical, biochemical, spectroscopic, immunochemical, electrical, optical or chemical means. A wide variety of appropriate detectable labels are known in the art, which include fluorescent or chemiluminescent labels, radioactive isotope labels, enzymatic or other ligands. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as digoxigenin, β-galactosidase, urease, alkaline phosphatase or peroxidase, avidin/biotin complex.

[0354] The detection methods used to detect or quantify the hybridization intensity will typically depend upon the label selected above. For example, radiolabels may be detected using photographic film or a phosphoimager. Fluorescent markers may be detected and quantified using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and measuring the reaction product produced by the action of the enzyme on the substrate; and finally colorimetric labels are detected by simply visualizing the colored label.

[0355] An agent-induced change in expression of sequences associated with a signaling biochemical pathway can also be determined by examining the corresponding gene products. Determining the protein level typically involves a) contacting the protein contained in a biological sample with an agent that specifically bind to a protein associated with a signaling biochemical pathway; and (b) identifying any agent:protein complex so formed. In one aspect of this embodiment, the agent that specifically binds a protein associated with a signaling biochemical pathway is an antibody, preferably a monoclonal antibody.

[0356] The reaction is performed by contacting the agent with a sample of the proteins associated with a signaling biochemical pathway derived from the test samples under conditions that will allow a complex to form between the agent and the proteins associated with a signaling biochemical pathway. The formation of the complex can be detected directly or indirectly according to standard procedures in the art. In the direct detection method, the agents are supplied with a detectable label and unreacted agents may be removed from the complex; the amount of remaining label thereby indicating the amount of complex formed. For such method, it is preferable to select labels that remain attached to the agents even during stringent washing conditions. It is preferable that the label does not interfere with the binding reaction. In the alternative, an indirect detection procedure may use an agent that contains a label introduced either chemically or enzymatically. A desirable label generally does not interfere with binding or the stability of the resulting agent:polypeptide complex. However, the label is typically designed to be accessible to an antibody for an effective binding and hence generating a detectable signal.

[0357] A wide variety of labels suitable for detecting protein levels are known in the art. Non-limiting examples include radioisotopes, enzymes, colloidal metals, fluorescent compounds, bioluminescent compounds, and chemiluminescent compounds.

[0358] The amount of agent:polypeptide complexes formed during the binding reaction can be quantified by standard quantitative assays. As illustrated above, the formation of agent:polypeptide complex can be measured directly by the amount of label remained at the site of binding. In an alternative, the protein associated with a signaling biochemical pathway is tested for its ability to compete with a labeled analog for binding sites on the specific agent. In this competitive assay, the amount of label captured is inversely proportional to the amount of protein sequences associated with a signaling biochemical pathway present in a test sample.

[0359] A number of techniques for protein analysis based on the general principles outlined above are available in the art. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), “sandwich” immunoassays, immunoradiometric assays, in situ immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunofluorescent assays, and SDS-PAGE.

[0360] Antibodies that specifically recognize or bind to proteins associated with a signaling biochemical pathway are preferable for conducting the aforementioned protein analyses. Where desired, antibodies that recognize a specific type of post-translational modifications (e.g., signaling biochemical pathway inducible modifications) can be used. Post-translational modifications include but are not limited to glycosylation, lipidation, acetylation, and phosphorylation. These antibodies may be purchased from commercial vendors. For example, anti-phosphotyrosine antibodies that specifically recognize tyrosine-phosphorylated proteins are available from a number of vendors including Invitrogen and Perkin Elmer. Anti-phosphotyrosine antibodies are particularly useful in detecting proteins that are differentially phosphorylated on their tyrosine residues in response to an ER stress. Such proteins include but are not limited to eukaryotic translation initiation factor 2 alpha (eIF-2α). Alternatively, these antibodies can be generated using conventional polyclonal or monoclonal antibody technologies by immunizing a host animal or an antibody-producing cell with a target protein that exhibits the desired post-translational modification.

[0361] In practicing the subject method, it may be desirable to discern the expression pattern of an protein associated with a signaling biochemical pathway in different bodily tissue, in different cell types, and/or in different subcellular structures. These studies can be performed with the use of tissue-specific, cell-specific or subcellular structure specific antibodies capable of binding to protein markers that are preferentially expressed in certain tissues, cell types, or subcellular structures.

[0362] An altered expression of a gene associated with a signaling biochemical pathway can also be determined by examining a change in activity of the gene product relative to a control cell. The assay for an agent-induced change in the activity of a protein associated with a signaling biochemical pathway will dependent on the biological activity and/or the signal transduction pathway that is under investigation. For example, where the protein is a kinase, a change in its ability to phosphorylate the downstream substrate(s) can be determined by a variety of assays known in the art. Representative assays include but are not limited to immunoblotting and immunoprecipitation with antibodies such as anti-phosphotyrosine antibodies that recognize phosphorylated proteins. In addition, kinase activity can be detected by high throughput chemiluminescent assays such as AlphaScreen™ (available from Perkin Elmer) and eTag™ assay (Chan-Hui, et al. (2003) Clinical Immunology 111: 162-174).

[0363] Where the protein associated with a signaling biochemical pathway is part of a signaling cascade leading to a fluctuation of intracellular pH condition, pH sensitive molecules such as fluorescent pH dyes can be used as the reporter molecules. In another example where the protein associated with a signaling biochemical pathway is an ion channel, fluctuations in membrane potential and/or intracellular ion concentration can be monitored. A number of commercial kits and high-throughput devices are particularly suited for a rapid and robust screening for modulators of ion channels. Representative instruments include FLIPR™ (Molecular Devices, Inc.) and VIPR (Aurora Biosciences). These instruments are capable of detecting reactions in over 1000 sample wells of a microplate simultaneously, and providing real-time measurement and functional data within a second or even a minisecond.

[0364] In practicing any of the methods disclosed herein, a suitable vector can be introduced to a cell or an embryo via one or more methods known in the art, including without limitation, microinjection, electroporation, sonoporation, biolistics, calcium phosphate-mediated transfection, cationic transfection, liposome transfection, dendrimer transfection, heat shock transfection, nucleofection transfection, magnetofection, lipofection, impalefection, optical transfection, proprietary agent-enhanced uptake of nucleic acids, and delivery via liposomes, immunoliposomes, virosomes, or artificial virions. In some methods, the vector is introduced into an embryo by microinjection. The vector or vectors may be microinjected into the nucleus or the cytoplasm of the embryo. In some methods, the vector or vectors may be introduced into a cell by nucleofection.

[0365] The target polynucleotide of a CRISPR complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA).

[0366] Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A "disease-associated" gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at a normal or abnormal level.

[0367] The target polynucleotide of a CRISPR complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA). Without wishing to be bound by theory, it is believed that the target sequence should be associated with a PAM (protospacer adjacent motif); that is, a short sequence recognized by the CRISPR complex. The precise sequence and length requirements for the PAM differ depending on the CRISPR enzyme used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence) Examples of PAM sequences are given in the examples section below, and the skilled person will be able to identify further PAM sequences for use with a given CRISPR enzyme.

[0368] The target polynucleotide of a CRISPR complex may include a number of disease-associated genes and polynucleotides as well as signaling biochemical pathway-associated genes and polynucleotides as listed in U.S. provisional patent applications 61/736,527 and 61/748,427 having Broad reference BI-2011/008/WSGR Docket No. 44063-701.101 and BI-2011/008/WSGR Docket No. 44063-701.102 respectively, both entitled SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION filed on Dec. 12, 2012 and Jan. 2, 2013, respectively, the contents of all of which are herein incorporated by reference in their entirety.

[0369] Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A "disease-associated" gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at a normal or abnormal level.

[0370] Examples of disease-associated genes and polynucleotides are listed in Tables A and B. Disease specific information is available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web. Examples of signaling biochemical pathway-associated genes and polynucleotides are listed in Table C.

[0371] Mutations in these genes and pathways can result in production of improper proteins or proteins in improper amounts which affect function. Further examples of genes, diseases and proteins are hereby incorporated by reference from U.S. Provisional applications 61/736,527 filed Dec. 12, 2012 and 61/748,427 filed on Jan. 2, 2013. Such genes, proteins and pathways may be the target polynucleotide of a CRISPR complex.

TABLE A

DISEASE/DISORDERS	GENE(S)
Neoplasia	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF; HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alpha; PPAR gamma; WT1 (Wilms Tumor); FGF Receptor Family members (5 members: 1, 2, 3, 4, 5); CDKN2a; APC; RB (retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR (Androgen Receptor); TSG101; IGF; IGF Receptor; Igf1(4 variants); Igf2 (3 variants); Igf 1 Receptor; Igf 2 Receptor; Bax; Bcl2; caspases family (9 members: 1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
Age-related Macular Degeneration	Abcr; Ccl2; Cc2; cp (ceruloplasmin); Timp3; cathepsinD; Vldlr; Ccr2
Schizophrenia	Neuregulin1 (Nrg1); Erb4 (receptor for Neuregulin); Complexin1 (Cplx1); Tph1 Tryptophan hydroxylase; Tph2 Tryptophan hydroxylase 2; Neurexin 1; GSK3; GSK3a; GSK3b
Disorders	5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA; DTNBP1; Dao (Dao1)
Trinucleotide Repeat Disorders	HTT (Huntington's Dx); SBMA/SMAX1/AR (Kennedy's Dx); FXN/X25 (Friedrich's Ataxia); ATX3 (Machado-Joseph's Dx); ATXN1 and ATXN2 (spinocerebellar ataxias); DMPK (myotonic dystrophy); Atrophin-1 and Atn1 (DRPLA Dx); CBP (Creb-BP-global instability); VLDLR (Alzheimer's); Atxn7; Atxn10
Fragile X Syndrome	FMR2; FXR1; FXR2; mGLUR5
Secretase Related Disorders	APH-1 (alpha and beta); Presenilin (Psen1); nicastrin (Ncstn); PEN-2
Others	Nos1; Parp1; Nat1; Nat2
Prion-related disorders	Prp
ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a; VEGF-b; -VEGF-c)
Drug addiction	Prkce (alcohol); Drd2; Drd4; ABAT (alcohol); GRIA2; Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1 (alcohol)
Autism	Mecp2; BZRAP1; MDGA2; Sema5A; Neurexin 1; Fragile X (FMR2 (AFF2); FXR1; FXR2; Mglur5)
Alzheimer's Disease	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterin; PS1; SORL1; CR1; Vldlr; Uba1; Uba3; CHIP28 (Aqp1, Aquaporin 1); Uchl1; Urhl3; APP
Inflammation	IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL-17b; IL-17c; IL-17d; IL-17f); Il-23; Cx3cr1; ptpn22; TNFa; NOD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b): CTLA4; Cx3cl1
Parkinson's Disease	x-Synuclein; DJ-1; LRRK2; Parkin; PINK1

TABLE B

Blood and coagulation diseases and disorders	Anemia (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); Bare lymphocyte syndrome (TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5), Bleeding disorders (TBXA2R, P2RX1, P2X1); Factor H and factor H-like 1 (HF1, CFH, HUS); Factor V and factor VIII (MCFD2); Factor VII deficiency (F7); Factor X deficiency (F10); Factor XI deficiency (F11); Factor XII deficiency (F12, HAF); Factor XIIIA deficiency (F13A1, F13A); Factor XIIIB deficiency (F13B); Fanconi anemia (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCJ, PHF9, FANCL, FANCM, KIAA1596); Hemophagocytic lymphohistiocytosis disorders (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemophilia A (F8, F8C, HEMA); Hemophilia B (F9, HEMB), Hemorrhagic disorders (PI, ATT, F5); Leukocyde deficiencies and disorders (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); Sickle cell anemia (HBB); Thalassemia (HBA2, HBB, HBD, LCRB, HBA1).
Cell dysregulation and oncology diseases and disorders	B-cell non-Hodgkin lymphoma (BCL7A, BCL7); Leukemia, (TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3,

TABLE B-continued

	FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL11, ARLTS1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN).
Inflammation and immune related diseases and disorders	AIDS (KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1); Autoimmune lymphoproliferative syndrome (TNFRSF6, APT1, FAS, CD95, ALPS1A); Combined immunodeficiency, (IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228), HIV susceptibility or infection (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCKR5 (CCR5)); Immunodeficiencies (CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI); Inflammation (IL-10, IL-1 (IL-1a, IL-1b), IL-13, IL-17 (IL-17a (CTLA8), IL-17b, IL-17c, IL-17d, IL-17F), II-23, Cx3cr1, ptpn22, TNFa, NOD2/CARD15 for IBD, IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3cl1); Severe combined immunodeficiencies (SCIDs)(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).
Metabolic, liver, kidney and protein diseases and disorders	Amyloid neuropathy (TTR, PALB); Amyloidosis (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); Cirrhosis (KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); Cystic fibrosis (CFTR, ABCC7, CF, MRP7); Glycogen storage diseases (SLC2A2, GLUT2, G6PC, G6PT, G6PT1 , GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Hepatic adenoma, 142330 (TCF1, HNF1A, MODY3), Hepatic failure, early onset, and neurologic disorder (SCOD1, SCO1), Hepatic lipase deficiency (LIPC), Hepatoblastoma, cancer and carcinomas (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; Medullary cystic kidney disease (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); Phenylketonuria (PAH, PKU1, QDPR, DHPR, PTS); Polycystic kidney and hepatic disease (FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63).
Muscular/Skeletal diseases and disorders	Becker muscular dystrophy (DMD, BMD, MYF6), Duchenne Muscular Dystrophy (DMD, BMD); Emery-Dreifuss muscular dystrophy (LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); Facioscapulohumeral muscular dystrophy (FSHM1A, FSHD1A); Muscular dystrophy (FKRP, MDC1C, LUMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3, LGMD1C, SEPN1, SELN, RSMD1, PLEC1, PLTN, EBS1); Osteopetrosis (LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); Muscular atrophy (VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1).
Neurological and neuronal diseases and disorders	ALS (SOD1, ALS2, STEX, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c); Alzheimer disease (APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLAU, URK, ACE, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); Autism (Mecp2, BZRAP1, MDGA2, Sema5A, Neurexin 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); Fragile X Syndrome (FMR2, FXR1, FXR2, mGLUR5); Huntington's disease and disease like disorders (HD, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); Parkinson disease (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK5, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); Rett syndrome (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Synuclein, DJ-1); Schizophrenia (Neuregulin1 (Nrg1), Erb4 (receptor for Neuregulin), Complexin1 (Cplx1), Tph1 Tryptophan hydroxylase, Tph2, Tryptophan hydroxylase 2, Neurexin 1, GSK3, GSK3a, GSK3b, 5-HTT (Slc6a4), COMT, DRD (Drd1a), SLC6A3, DAOA, DTNBP1, Dao (Dao1)); Secretase Related Disorders (APH-1 (alpha and beta), Presenilin (Psen1), nicastrin, (Ncstn), PEN-2, Nos1, Parp1, Nat1, Nat2); Trinucleotide Repeat Disorders (HTT (Huntington's Dx), SBMA/SMAX1/AR (Kennedy's Dx), FXN/X25 (Friedrich's Ataxia), ATX3 (Machado-Joseph's Dx), ATXN1 and ATXN2 (spinocerebellar ataxias), DIVITK (myotonic dystrophy), Atrophin-1 and Atn1 (DRPLA Dx), CBP (Crete-BP—global instability), VLDLR (Alzheimer's), Atxn7, Atxn10).

TABLE B-continued

Occular diseases and disorders	Age-related macular degeneration (Abcr, Ccl2, Cc2, cp ceruloplasmin), Timp3, cathepsinD, Vldlr, Ccr2); Cataract (CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); Corneal clouding and dystrophy (APOA1, TGFBI, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD); Cornea plana congenital (KERA, CNA2); Glaucoma (MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A); Leber congenital amaurosis (CRB1, RP12, CRX, CORD2, CRD, RPGRIP1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3); Macular dystrophy (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2).

TABLE C

CELLULAR FUNCTION	GENES
PI3K/AKT Signaling	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2; PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1; AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2; PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2; ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3; PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7; YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A; CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1; PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2; TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK; HSP90AA1; RPS6KB1
ERK/MAPK Signaling	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1; PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A; PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN; EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ; PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1; PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1; CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
Glucocorticoid Receptor Signaling	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1; MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I; PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2; MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1; MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13; RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1; PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3; MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP; CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1; ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1; STAT1; IL6; HSP90AA1
Axonal Guidance Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12; IGF1; RAC1; RAP1A; EIF4E; PRKCZ; NRP1; NTRK2; ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2; CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11; PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1; GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B; AKT3; PRKCA
Ephrin Receptor Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1; PRKAA2; EIF2AK2; RAC1 ; RAP1A; GRK6; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2; DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14; CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1;

TABLE C-continued

CELLULAR FUNCTION	GENES
	KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; A1F4; AKT3; SGK
Actin Cytoskeleton Signaling	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1; PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6; ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8; F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7; PPP1CC; PXN; VIL2; RAF1; GSN; DYRK1A; ITGB1 MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3; ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL; BRAF; VAV3; SGK
Huntington's Disease Signaling	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2; MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2; PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST; GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1; GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2; HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A; HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1; PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX; ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3
Apoptosis Signaling	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1; BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB; CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8; BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF; RAF1; IKBKG; RFLB; CASP9; DYRK1A; MAP2K2; CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2; BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK; CASP3; BIRC3; PARP1
B Cell Receptor Signaling	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11; AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A; MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1; MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9; EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1; NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN; GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Leukocyte Extravasation Signaling	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA; RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11; MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12; PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB; MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK; MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2; CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK; CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
Integrin Signaling	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A; TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2; CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8; CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA; SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1; TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2; CRKL; BRAF; GSK3B; AKT3
Acute Phase Response Signaling	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11; AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14; PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1; TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN; AKT3; IL1R1; IL6
PTEN Signaling	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11; MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA; CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1; MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR; RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2; AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1; NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2; GSK3B; AKT3; FOXO1; CASP3; RPS6KB1

TABLE C-continued

CELLULAR FUNCTION	GENES
p53 Signaling	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A; BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2; PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1; PMAIP1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9; CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A; HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1; SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN; SNAI2; GSK3B; BAX; AKT3
Aryl Hydrocarbon Receptor Signaling	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1; NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1; SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1; MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1; SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF; CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1; CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1; HSP90AA1
Xenobiotic Metabolism Signaling	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1; NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A; PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1; ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD; GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL; NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1; CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1; NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1; HSP90AA1
SAPK/JNK Signaling	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1; GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA; FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1; GNB2L1 ; IRS1; MAPK3; MAPK10; DAXX; KRAS; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2; PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1 ; CRKL; BRAF; SGK
PPAr/RXR Signaling	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN; RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2; ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8; IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A; NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1; TGFBR1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1; ADIPOQ
NF-KB Signaling	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAF6; TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2; MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2; KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGRB; TNF; INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1; PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10; GSK3B; AKT3; TNFAIP3; IL1R1
Neuregulin Signaling	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1; MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI; CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS; PRKCD, STAT5A, SRC; ITGB7; RAF1; ITGB1; MAP2K2; ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3; EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL; AKT3; PRKCA; HSP90AA1; RPS6KB1
Wnt & Beta catenin Signaling	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO; AKT2, PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK; LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1 PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1; CCND1; GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B AKT3; SOX2
Insulin Receptor Signaling	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1; PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3; MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1; SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; MAP2K1; GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK; RPS6KB1
IL-6 Signaling	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3; MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3; MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6

TABLE C-continued

CELLULAR FUNCTION	GENES
Hepatic Cholestasis	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA; RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8; PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1; TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8; CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4; JUN; IL1R1; PRKCA; IL6
IGF-1 Signaling	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2; PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8; IGF1R; IRS1; MAPK3; IGFBP7; FRAS; PIK3C2A; YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3; FOXO1; SRF; CTGF; RPS6KB1
NRF2-mediated Oxidative Stress Response	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1; NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL; NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP; MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1; GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
Hepatic Fibrosis/Hepatic Stellate Cell Activation	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF; SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9; IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8; PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX; IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
PPAR Signaling	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB; NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3; NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2; PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA; MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Fc Epsilon RI Signaling	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD; MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3; VAV3; PRKCA
G-Protein Coupled Receptor Signaling	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB; PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB; PIK3C3: MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1; IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK; PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3; PRKCA
Inositol Phosphate Metabolism	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6; MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3; MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1; MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
PDGF Signaling	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC; PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2; PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC; JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
VEGF Signaling	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF; AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3; BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN; RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN; VEGFA; AKT3; FOXO1; PRKCA
Natural Killer Cell Signaling	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11; KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6; PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1; PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
Cell Cycle: Gl/S Checkpoint Regulation	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC; ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11; HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1; E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1; GSK3B; RBL1; HDAC6
T Cell Receptor Signaling	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS; NFKB2; PIK3GB; PIK3C3; MAPK8; MAPK3; KRAS; RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN; MAP2K2; PIK3R1; CHUK; MAP2K1; NFKBI; ITK; BCL10; JUN; VAV3
Death Receptor Signaling	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD; FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8; DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB;

TABLE C-continued

CELLULAR FUNCTION	GENES
	CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3; BIRC3
FGF Signaling	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11; AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8; MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1; AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4; AKT3; PRKCA; HGF
GM-CSF Signaling	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A; STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3; ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3; STAT1
Amyotrophic Lateral Sclerosis Signaling	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2; PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1; PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1; APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
JAK/Stat Signaling	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B; PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A; PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3; STAT1
Nicotinate and Nicotinamide Metabolism	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1; PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1; PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2; MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
Chemokine Signaling	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ; CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13; RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1; MAP2K2; MAP2K1; JUN; CCL2; PRKCA
IL-2 Signaling	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS; STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2; JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
Synaptic Long Term Depression	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS; PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3; KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA; YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Estrogen Receptor Signaling	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2; SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1; HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP; MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Protein Ubiquitination Pathway	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4; CBL; UBE21; BTRC; HSPA5; USP7; USP10; FBXW7; USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8; USP1; VHL; HSP90AA1; BIRC3
IL-10 Signaling	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF; IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1; JUN; IL1R1; IL6
VDR/RXR Activation	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1; NCOR2; SP1; PRKCI; CDKN1B; PRKD1, PRKCD; RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1; LRP5; CEBPB; FOXO1; PRKCA
TGF-beta Signaling	EP300; SMAD2; SMURE1; MAPK1; SMAD3; SMAD1; FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2; SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2; MAP2K1; TGFBRI; SMAD4; JUN; SMAD5
Toll-like Receptor Signaling	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK; NFKB1; TLR2; JUN
p38 MAPK Signaling	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS; CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2; MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1; SRF; STAT1
Neurotrophin/TRK Signaling	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A; RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; CDC42; JUN; ATF4
FXR/RXR Activation	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8; APOB; MAPK10; PPARG; MTTP; MAPK9; PPARGC1A; TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1

TABLE C-continued

CELLULAR FUNCTION	GENES
Synaptic Long Term Potentiation	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1; PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS; PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1; ATF4; PRKCA
Calcium Signaling	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1; CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11; HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4; HDAC6
EGF Signaling	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1; STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Hypoxia Signaling in the Cardiovascular System	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT; HIF1A; SLC2A4; NOS3; TP53; LDHA; AKTI ; ATM; VEGFA; JUN; ATF4; VHL; HSP90AA1
LPS/IL-1 Mediated Inhibition of RXR Function	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1; MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2; TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
LXRIRXR Activation	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA; NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1; SREBF1; IL1R1; CCL2; IL6; MMP9
Amyloid Processing	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2; CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1; PSEN1; CSNK1A1; GSK3B; AKT3; APP
IL-4 Signaling	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1; PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1; FRAP1; AKT3; RPS6KB1
Cell Cycle: G2/M DNA Damage Checkpoint Regulation	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC; CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A; PRKDC; ATM; SFN; CDKN2A
Nitric Oxide Signaling in the Cardiovascular System	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3; CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1; VEGFA; AKT3; HSP90AA1
Purine Metabolism	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4; PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C; NT5E; POLD1; NME1
cAMP-Mediated Signaling	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3; SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Mitochondrial Dysfunction	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9; PARK7; PSEN1; PARK2; APP; CASP3
Notch Signaling	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2; PSEN1; NOTCH3; NOTCH1; DLL4
Endoplasmic Reticulum Stress Pathway	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4; EIF2AK3; CASP3
Pyrimidine Metabolism	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B; NT5E; POLD1; NME1
Parkinson's Signaling	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7; PARK2; CASP3
Cardiac & Beta Adrenergic Signaling	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC; PPP2R5C
Glycolysis/Gluconeogenesis Interferon Signaling	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1 IFR1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Sonic Hedgehog Signaling	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B
Glycerophospholipid Metabolism	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Phospholipid Degradation	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Tryptophan Metabolism	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Lysine Degradation	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Nucleotide Excision Repair Pathway	ERCC5; ERCC4; XPA; XPC; ERCC1
Starch and Sucrose Metabolism	UCHL1; HK2; GCK; GPI; HK1
Aminosugars Metabolism	NQO1; HK2; GCK; HK1
Arachidonic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Circadian Rhythm Signaling	CSNK1E; CREB1; ATF4; NR1D1
Coagulation System	BDKRB1; F2R; SERPINE1; F3
Dopamine Receptor Signaling	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Glutathione Metabolism	IDH2; GSTP1; ANPEP; IDH1
Glycerolipid Metabolism	ALDH1A1; GPAM; SPHK1; SPHK2

TABLE C-continued

CELLULAR FUNCTION	GENES
Linoleic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Methionine Metabolism	DNMT1; DNMT3B; AHCY; DNMT3A
Pyruvate Metabolism	GLO1; ALDR1A1; PKM2; LDHA
Arginine and Proline Metabolism	ALDH1A1; NOS3; NOS2A
Eicosanoid Signaling	PRDX6; GRN; YWHAZ
Fructose and Mannose Metabolism	HK2; GCK; HK1
Galactose Metabolism	HK2; GCK; HK1
Stilbene, Cournarine and Lignin Biosynthesis	PRDX6; PRDX1; TYR
Antigen Presentation Pathway	CALR; B2M
Biosynthesis of Steroids	NQO1; DFICR7
Butanoate Metabolism	ALDH1A1; NLGN1
Citrate Cycle	IDH2; IDH1
Fatty Acid Metabolism	ALDH1A1; CYP1B1
Glycerophospholipid Metabolism	PRDX6; CHKA
Histidine Metabolism	PRMT5; ALDH1A1
inositol Metabolism	ERO1L; APEX1
Metabolism of Xenobiotics by Cytochrome p450	GSTP1; CYP1B1
Methane Metabolism	PRDX6; PRDX1
Phenylalanine Metabolism	PRDX6; PRDX1
Propanoate Metabolism	ALDHI1A1; LDHA
Selenoamino Acid Metabolism	PRMT5; AHCY
Sphingolipid Metabolism	SPHK1; SPFIK2
Aminophosphonate Metabolism	PRMT5
Androgen and Estrogen Metabolism	PRMT5
Ascorbate and Aldarate Metabolism	ALDH1A1
Bile Acid Biosynthesis	ALDH1A1
Cysteine Metabolism	LDHA
Fatty Acid Biosynthesis	FASN
Glutamate Receptor Signaling	GNB2L1
NRF2-mediated Oxidative Stress Response	PRDX1
Pentose Phosphate Pathway	GPI
Pentose and Glucuronate Interconversions	UCHL1
Retinol Metabolism	ALDH1A1
Riboflavin Metabolism	TYR
Tyrosine Metabolism	PRMT5, TYR
Ubiquinone Biosynthesis	PRMT5
Valine, Leucine and Isoleucine Degradation	ALDH1A1
Glycine, Serine and Threonine Metabolism	CHKA
Lysine Degradation	ALDH1A1
Pain/Taste	TRPM5; TRPA1
Pain	TRPM7; TRPC5; TRPC6; TRPC1; Cnrl; cnr2; Grk2; Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca; Prkacb; Prkar1a; Prkar2a
Mitochondrial Function	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Developmental Neurology	BMP-4; Chordin (Chrd); Noggin (Nog); WNT (Wnt2; Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b; Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenin; Dkk-1; Frizzled related proteins; Otx-2; Gbx2; FGF-8; Reelin; Dab1; unc-86 (Pou4fl or Brn3a); Numb; Reln

[0372] Embodiments of the invention also relate to methods and compositions related to knocking out genes, amplifying genes and repairing particular mutations associated with DNA repeat instability and neurological disorders (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Second Edition, Academic Press, Oct. 13, 2011—Medical). Specific aspects of tandem repeat sequences have been found to be responsible for more than twenty human diseases (New insights into repeat instability: role of RNA*DNA hybrids. Mclvor E I, Polak U. Napierala M. RNA Biol. 2010 September-October; 7(5):551-8). The CRISPR-Cas system may be harnessed to correct these defects of genomic instability.

[0373] A further aspect of the invention relates to utilizing the CRISPR-Cas system for correcting defects in the EMP2A and EMP2B genes that have been identified to be associated with Lafora disease. Lafora disease is an autosomal recessive condition which is characterized by progressive myoclonus epilepsy which may start as epileptic seizures in adolescence. A few cases of the disease may be caused by mutations in genes yet to be identified. The disease causes seizures, muscle spasms, difficulty walking, dementia, and eventually death. There is currently no therapy that has proven effective against disease progression. Other genetic abnormalities associated with epilepsy may also be targeted by the CRISPR-Cas system and the underlying genetics is further described in Genetics of Epilepsy and Genetic Epilepsies, edited by Giuliano Avanzini, Jeffrey L. Noebels, Mariani Foundation Paediatric Neurology: 20; 2009).

[0374] The methods of US Patent Publication No. 20110158957 assigned to Sangamo BioSciences, Inc. involved in inactivating T cell receptor (TCR) genes may also be modified to the CRISPR Cas system of the present invention. In another example, the methods of US Patent Publication No. 20100311124 assigned to Sangamo BioSciences. Inc. and US Patent Publication No. 20110225664 assigned to Cellectis, which are both involved in inactivating glutamine synthetase gene expression genes may also be modified to the CRISPR Cas system of the present invention.

[0375] Several further aspects of the invention relate to correcting defects associated with a wide range of genetic diseases which are further described on the website of the National Institutes of Health under the topic subsection Genetic Disorders (website at health.nih.gov/topic/GeneticDisorders). The genetic brain diseases may include but are not limited to Adrenoleukodystrophy, Agenesis of the Corpus Callosum, Aicardi Syndrome, Alpers' Disease, Alzheimer's Disease, Barth Syndrome, Batten Disease, CADASIL, Cerebellar Degeneration, Fabry's Disease, Gerstmann-Straussler-Scheinker Disease, Huntington's Disease and other Triplet Repeat Disorders, Leigh's Disease, Lesch-Nyhan Syndrome, Menkes Disease, Mitochondrial Myopathies and NINDS Colpocephaly. These diseases are further described on the website of the National Institutes of Health under the subsection Genetic Brain Disorders.

[0376] In some embodiments, the condition may be neoplasia. In some embodiments, where the condition is neoplasia, the genes to be targeted are any of those listed in Table A (in this case PTEN and so forth). In some embodiments, the condition may be Age-related Macular Degeneration. In some embodiments, the condition may be a Schizophrenic Disorder. In some embodiments, the condition may be a Trinucleotide Repeat Disorder. In some embodiments, the condition may be Fragile X Syndrome. In some embodiments, the condition may be a Secretase Related Disorder. In some embodiments, the condition may be a Prion-related disorder. In some embodiments, the condition may be ALS. In some embodiments, the condition may be a drug addiction. In some embodiments, the condition may be Autism. In some embodiments, the condition may be Alzheimer's Disease. In some embodiments, the condition may be inflammation. In some embodiments, the condition may be Parkinson's Disease.

[0377] For example, US Patent Publication No. 20110023145, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with autism spectrum disorders (ASD). Autism spectrum disorders (ASDs) are a group of disorders characterized by qualitative impairment in social interaction and communication, and restricted repetitive and stereotyped patterns of behavior, interests, and activities. The three disorders, autism, Asperger syndrome (AS) and pervasive developmental disorder-not otherwise specified (PDD-NOS) are a continuum of the same disorder with varying degrees of severity, associated intellectual functioning and medical conditions. ASDs are predominantly genetically determined disorders with a heritability of around 90%.

[0378] US Patent Publication No. 20110023145 comprises editing of any chromosomal sequences that encode proteins associated with ASD which may be applied to the CRISPR Cas system of the present invention. The proteins associated with ASD are typically selected based on an experimental association of the protein associated with ASD to an incidence or indication of an ASD. For example, the production rate or circulating concentration of a protein associated with ASD may be elevated or depressed in a population having an ASD relative to a population lacking the ASD. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the proteins associated with ASD may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[0379] Non limiting examples of disease states or disorders that may be associated with proteins associated with ASD include autism, Asperger syndrome (AS), pervasive developmental disorder-not otherwise specified (PDD-NOS), Rett's syndrome, tuberous sclerosis, phenylketonuria, Smith-Lemli-Opitz syndrome and fragile X syndrome. By way of non-limiting example, proteins associated with ASD include but are not limited to the following proteins: ATP10C aminophospholipid-MET MET receptor transporting ATPase tyrosine kinase (ATP10C) BZRAP1 MGLUR5 (GRM5) Metabotropic glutamate receptor 5 (MGLUR5) CDH10 Cadherin-10 MGLUR6 (GRM6) Metabotropic glutamate receptor 6 (MGLUR6) CDH9 Cadherin-9 NLGN1 Neuroligin-1 CNTN4 Contactin-4 NLGN2 Neuroligin-2 CNTNAP2 Contactin-associated SEMA5A Neuroligin-3 protein-like 2 (CNTNAP2) DHCR7 7-dehydrocholesterol NLGN4X Neuroligin-4X-reductase (DHCR7) linked DOC2A Double C2-like domain—NLGN4Y Neuroligin-4 Y—containing protein alpha linked DPP6 Dipeptidyl NLGN5 Neuroligin-5 aminopeptidase-like protein 6 EN2 engrailed 2 (EN2) NRCAM Neuronal cell adhesion molecule (NRCAM) MDGA2 fragile X mental retardation NRXN1 Neurexin-1 1

(MDGA2) FMR2 (AFF2) AF4/FMR2 family member 20R4M2 Olfactory receptor (AFF2) 4M2 FOXP2 Forkhead box protein P2 OR4N4 Olfactory receptor (FOXP2) 4N4 FXR1Fragile X mental OXTR oxytocin receptor retardation, autosomal (OXTR) homolog 1 (FXR1) FXR2 Fragile X mental PAH phenylalanine retardation, autosomal hydroxylase (PAH) homolog 2 (FXR2) GABRA1 Gamma-aminobutyric acid PTEN Phosphatase and receptor subunit alpha-1 tensin homologue (GABRA1) (PTEN) GABRA5 GABAA (.gamma.-aminobutyric PTPRZ1 Receptor-type acid) receptor alpha 5 tyrosine-protein subunit (GABRAS) phosphatase zeta (PTPRZ1) GABRB1 Gamma-aminobutyric acid RELN Reelin receptor subunit beta-1 (GABRB1) GABRB3 GABAA (.gamma.-aminobutyric RPL10 60S ribosomal acid) receptor .beta.3 subunit protein L10 (GABRB3) GABRG1 Gamma-aminobutyric acid SEMA5A Semaphorin-5A receptor subunit gamma-1 (SEMA5A) (GABRG1) HIRIP3 HIRA-interacting protein 3 SEZ6L2 seizure related 6 homolog (mouse)-like 2 HOXA1 Homeobox protein Hox-A1 SHANK3 SH3 and multiple (HOXA1) ankyrin repeat domains 3 (SHANK3) IL6 Interleukin-6 SHBZRAP1 SH3 and multiple ankyrin repeat domains 3 (SHBZRAP1) LAMB1 Laminin subunit beta-1 SLC6A4 Serotonin (LAMB1) transporter (SERT) MAPK3 Mitogen-activated protein TAS2R1Taste receptor kinase 3 type 2 member 1 TAS2R1 MAZ Myc-associated zinc finger TSC1 Tuberous sclerosis protein protein 1 MDGA2 MAM domain containing TSC2 Tuberous sclerosis glycosylphosphatidylinositol protein 2 anchor 2 (MDGA2) MECP2 Methyl CpG binding UBE3A Ubiquitin protein protein 2 (MECP2) ligase E3A (UBE3A) MECP2 methyl CpG binding WNT2 Wingless-type protein 2 (MECP2) MMTV integration site family, member 2 (WNT2)

[0380] The identity of the protein associated with ASD whose chromosomal sequence is edited can and will vary. In preferred embodiments, the proteins associated with ASD whose chromosomal sequence is edited may be the benzodiazapine receptor (peripheral) associated protein 1 (BZRAP1) encoded by the BZRAP1 gene, the AF4/FMR2 family member 2 protein (AFF2) encoded by the AFF2 gene (also termed MFR2), the fragile X mental retardation autosomal homolog 1 protein (FXR1) encoded by the FXR1 gene, the fragile X mental retardation autosomal homolog 2 protein (FXR2) encoded by the FXR2 gene, the MAM domain containing glycosylphosphatidylinositol anchor 2 protein (MDGA2) encoded by the MDGA2 gene, the methyl CpG binding protein 2 (MECP2) encoded by the MECP2 gene, the metabotropic glutamate receptor 5 (MGLURS) encoded by the MGLUR5-1 gene (also termed GRM5), the neurexin 1 protein encoded by the NRXN1 gene, or the semaphorin-5A protein (SEMA5A) encoded by the SEMA5A gene. In an exemplary embodiment, the genetically modified animal is a rat, and the edited chromosomal sequence encoding the protein associated with ASD is as listed below: BZRAP1 benzodiazapine receptor XM_002727789, (peripheral) associated XM_213427, protein 1 (BZRAP1) XM_002724533, XM_001081125 AFF2 (FMR2) AF4/FMR2 family member 2 XM_219832, (AFF2) XM_001054673 FXR1Fragile X mental NM_001012179 retardation, autosomal homolog 1 (FXR1) FXR2Fragile X mental NM_001100647 retardation, autosomal homolog 2 (FXR2) MDGA2 MAM domain containing NM_199269 glycosylphosphatidylinositol anchor 2 (MDGA2) MECP2 Methyl CpG binding NM_022673 protein 2 (MECP2) MGLUR5Metabotropic glutamate NM_017012 (GRM5) receptor 5 (MGLUR5) NRXN1 Neurexin-1 NM_021767 SEMA5A Semaphorin-5A (SEMA5A) NM_001107659

[0381] Exemplary animals or cells may comprise one, two, three, four, five, six, seven, eight, or nine or more inactivated chromosomal sequences encoding a protein associated with ASD, and zero, one, two, three, four, five, six, seven, eight, nine or more chromosomally integrated sequences encoding proteins associated with ASD. The edited or integrated chromosomal sequence may be modified to encode an altered protein associated with ASD. Non-limiting examples of mutations in proteins associated with ASD include the L18Q mutation in neurexin 1 where the leucine at position 18 is replaced with a glutamine, the R451c mutation in neuroligin 3 where the arginine at position 451 is replaced with a cysteine, the R87W mutation in neuroligin 4 where the arginine at position 87 is replaced with a tryptophan, and the I425V mutation in serotonin transporter where the isoleucine at position 425 is replaced with a valine. A number of other mutations and chromosomal rearrangements in ASD-related chromosomal sequences have been associated with ASD and are known in the art. See, for example, Freitag et al. (2010) Eur. Child. Adolesc. Psychiatry 19:169-178, and Bucan et al. (2009) PLoS Genetics 5: e1000536, the disclosure of which is incorporated by reference herein in its entirety.

[0382] Examples of proteins associated with Parkinson's disease include but are not limited to α-synuclein, DJ-1, LRRK2, PINK1, Parkin, UCHL1, Synphilin-1, and NURR1.

[0383] Examples of addiction-related proteins may include ABAT for example.

[0384] Examples of inflammation-related proteins may include the monocyte chemoattractant protein-1 (MCP1) encoded by the Ccr2 gene, the C-C chemokine receptor type 5 (CCR5) encoded by the Ccr5 gene, the IgG receptor IIB (FCGR2b, also termed CD32) encoded by the Fcgr2b gene, or the Fc epsilon R1g (FCER1g) protein encoded by the Fcer1g gene, for example.

[0385] Examples of cardiovascular diseases associated proteins may include IL1B (interleukin 1, beta), XDH (xanthine dehydrogenase), TP53 (tumor protein p53), PTGIS (prostaglandin 12 (prostacyclin) synthase). MB (myoglobin), IL4 (interleukin 4), ANGPT1 (angiopoietin 1), ABCG8 (ATP-binding cassette, sub-family G (WHITE), member 8), or CTSK (cathepsin K), for example.

[0386] For example, US Patent Publication No. 20110023153, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with Alzheimer's Disease. Once modified cells and animals may be further tested using known methods to study the effects of the targeted mutations on the development and/or progression of AD using measures commonly used in the study of AD—such as, without limitation, learning and memory, anxiety, depression, addiction, and sensory motor functions as well as assays that measure behavioral, functional, pathological, metaboloic and biochemical function.

[0387] The present disclosure comprises editing of any chromosomal sequences that encode proteins associated with AD. The AD-related proteins are typically selected based on an experimental association of the AD-related protein to an AD disorder. For example, the production rate or circulating concentration of an AD-related protein may be elevated or depressed in a population having an AD disorder relative to a population lacking the AD disorder. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the AD-related proteins may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[0388] Examples of Alzheimer's disease associated proteins may include the very low density lipoprotein receptor protein (VLDLR) encoded by the VLDLR gene, the ubiquitin-like modifier activating enzyme 1 (UBA1) encoded by the UBA1 gene, or the NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) encoded by the UBA3 gene, for example.

[0389] By way of non-limiting example, proteins associated with AD include but are not limited to the proteins listed as follows: Chromosomal Sequence Encoded Protein ALAS2 Delta-aminolevulinate synthase 2 (ALAS2) ABCA1 ATP-binding cassette transporter (ABCA1) ACE Angiotensin 1-converting enzyme (ACE) APOE Apolipoprotein E precursor (APOE) APP amyloid precursor protein (APP) AQP1 aquaporin 1 protein (AQP1) BIN1 Myc box-dependent-interacting protein 1 or bridging integrator 1 protein (BIN1) BDNF brain-derived neurotrophic factor (BDNF) BTNL8 Butyrophilin-like protein 8 (BTNL8) C1ORF49 chromosome 1 open reading frame 49 CDH4 Cadherin-4 CHRNB2 Neuronal acetylcholine receptor subunit beta-2 CKLFSF2 CKLF-like MARVEL transmembrane domain-containing protein 2 (CKLFSF2) CLEC4E C-type lectin domain family 4, member e (CLEC4E) CLU clusterin protein (also known as apoplipoprotein J) CR1 Erythrocyte complement receptor 1 (CR1, also known as CD35, C3b/C4b receptor and immune adherence receptor) CR1L Erythrocyte complement receptor 1 (CR1L) CSF3R granulocyte colony-stimulating factor 3 receptor (CSF3R) CST3 Cystatin C or cystatin 3 CYP2C Cytochrome P450 2C DAPK1 Death-associated protein kinase 1 (DAPK1) ESR1 Estrogen receptor 1 FCAR Fc fragment of IgA receptor (FCAR, also known as CD89) FCGR3B Fc fragment of IgG, low affinity IIIb, receptor (FCGR3B or CD16b) FFA2 Free fatty acid receptor 2 (FFA2) FGA Fibrinogen (Factor I) GAB2 GRB2-associated-binding protein 2 (GAB2) GAB2 GRB2-associated-binding protein 2 (GAB2) GALP Galanin-like peptide GAPDHS Glyceraldehyde-3-phosphate dehydrogenase, spermatogenic (GAPDHS) GMPB GMBP HP Haptoglobin (HP) HTR75-hydroxytryptamine (serotonin) receptor 7 (adenylate cyclase-coupled) IDE Insulin degrading enzyme IF127 IF127 IF16 Interferon, alpha-inducible protein 6 (IF16) IFIT2 Interferon-induced protein with tetratricopeptide repeats 2 (IFIT2) IL1RN interleukin-1 receptor antagonist (IL-RA) IL8RA Interleukin 8 receptor, alpha (IL8RA or CD181) IL8RB Interleukin 8 receptor, beta (IL8RB) JAG1 Jagged 1 (JAG1) KCNJ15 Potassium inwardly-rectifying channel, subfamily J, member 15 (KCNJ15) LRP6 Low-density lipoprotein receptor-related protein 6 (LRP6) MAPT microtubule-associated protein tau (MAPT) MARK4 MAP/microtubule affinity-regulating kinase 4 (MARK4) MPHOSPH1 M-phase phosphoprotein 1 MTHFR 5,10-methylenetetrahydrofolate reductase MX2 Interferon-induced GTP-binding protein Mx2 NBN Nibrin, also known as NBN NCSTN Nicastrin NIACR2 Niacin receptor 2 (NIACR2, also known as GPR109B) NMNAT3 nicotinamide nucleotide adenylyltransferase 3 NTM Neurotrimin (or HNT) ORM1 Orosmucoid 1 (ORM1) or Alpha-1-acid glycoprotein 1 P2RY13 P2Y purinoceptor 13 (P2RY13) PBEF1 Nicotinamide phosphoribosyltransferase (NAmPRTase or Nampt) also known as pre-B-cell colony-enhancing factor 1 (PBEF1) or visfatin PCK1 Phosphoenolpyruvate carboxykinase PICALM phosphatidylinositol binding clathrin assembly protein (PICALM) PLAU Urokinase-type plasminogen activator (PLAU) PLXNC1 Plexin C1 (PLXNC1) PRNP Prion protein PSEN1 presenilin 1 protein (PSEN1) PSEN2 presenilin 2 protein (PSEN2) PTPRA protein tyrosine phosphatase receptor type A protein (PTPRA) RALGPS2 RAl GEF with PH domain and SH3 binding motif 2 (RALGPS2) RGSL2 regulator of G-protein signaling like 2 (RGSL2) SELENBP1 Selenium binding protein 1 (SELNBP1) SLC25A37 Mitoferrin-1 SORL1 sortilin-related receptor L(DLR class) A repeats-containing protein (SORL1) TF Transferrin TFAM Mitochondrial transcription factor A TNF Tumor necrosis factor TNFRSF10C Tumor necrosis factor receptor superfamily member 10C (TNFRSF10C) TNFSF10 Tumor necrosis factor receptor superfamily, (TRAIL) member 10a (TNFSF10) UBA1 ubiquitin-like modifier activating enzyme 1 (UBA1) UBA3 NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) UBB ubiquitin B protein (UBB) UBQLN1 Ubiquilin-1 UCHL1 ubiquitin carboxyl-terminal esterase L1 protein (UCHL1) UCHL3 ubiquitin carboxyl-terminal hydrolase isozyme L3 protein (UCHL3) VLDLR very low density lipoprotein receptor protein (VLDLR)

[0390] In exemplary embodiments, the proteins associated with AD whose chromosomal sequence is edited may be the very low density lipoprotein receptor protein (VLDLR) encoded by the VLDLR gene, the ubiquitin-like modifier activating enzyme 1 (UBA1) encoded by the UBA1 gene, the NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) encoded by the UBA3 gene, the aquaporin 1 protein (AQP1) encoded by the AQP1 gene, the ubiquitin carboxyl-terminal esterase L1 protein (UCHL1) encoded by the UCHL1 gene, the ubiquitin carboxyl-terminal hydrolase isozyme L3 protein (UCHL3) encoded by the UCHL3 gene, the ubiquitin B protein (UBB) encoded by the UBB gene, the microtubule-associated protein tau (MAPT) encoded by the MAPT gene, the protein tyrosine phosphatase receptor type A protein (PTPRA) encoded by the PTPRA gene, the phosphatidylinositol binding clathrin assembly protein (PICALM) encoded by the PICALM gene, the clusterin protein (also known as apoplipoprotein J) encoded by the CLU gene, the presenilin 1 protein encoded by the PSEN1 gene, the presenilin 2 protein encoded by the PSEN2 gene, the sortilin-related receptor L(DLR class) A repeats-containing protein (SORL1) protein encoded by the SORL1 gene, the amyloid precursor protein (APP) encoded by the APP gene, the Apolipoprotein E precursor (APOE) encoded by the APOE gene, or the brain-derived neurotrophic factor (BDNF) encoded by the BDNF gene. In an exemplary embodiment, the genetically modified animal is a rat, and the edited chromosomal sequence encoding the protein associated with AD is as as follows: APP amyloid precursor protein (APP) NM_019288 AQP1 aquaporin 1 protein (AQP1) NM_012778 BDNF Brain-derived neurotrophic factor NM_012513 CLU clusterin protein (also known as NM_053021 apoplipoprotein J) MAPT microtubule-associated protein NM_017212 tau (MAPT) PICALM phosphatidylinositol binding NM_053554 clathrin assembly protein (PICALM) PSEN1 presenilin 1 protein (PSEN1) NM_019163 PSEN2 presenilin 2 protein (PSEN2) NM_031087 PTPRA protein tyrosine phosphatase NM_012763 receptor type A protein (PTPRA) SOR1 sortilin-related receptor L(DLR NM_053519, class) A repeats-containing XM_001065506, protein (SORL1) XM_217115 UBA1 ubiquitin-like modifier activating NM_001014080 enzyme 1 (UBA1) UBA3 NEDD8-activating enzyme E1 NM_057205 catalytic subunit protein (UBE1C) UBB ubiquitin B protein (UBB) NM_138895 UCHL1 ubiquitin carboxyl-terminal NM_017237 esterase L1 protein (UCHL1) UCHL3 ubiquitin carboxyl-terminal NM_001110165 hydrolase isozyme L3 protein (UCHL3) VLDLR very low density lipoprotein NM_013155 receptor protein (VLDLR)

[0391] The animal or cell may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more disrupted chromosomal sequences encoding a protein associated with AD and zero, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more chromosomally integrated sequences encoding a protein associated with AD.

[0392] The edited or integrated chromosomal sequence may be modified to encode an altered protein associated with AD. A number of mutations in AD-related chromosomal sequences have been associated with AD. For instance, the V7171 (i.e. valine at position 717 is changed to isoleucine) missense mutation in APP causes familial AD. Multiple mutations in the presenilin-1 protein, such as H163R (i.e. histidine at position 163 is changed to arginine), A246E (i.e. alanine at position 246 is changed to glutamate), L286V (i.e. leucine at position 286 is changed to valine) and C410Y (i.e. cysteine at position 410 is changed to tyrosine) cause familial Alzheimer's type 3. Mutations in the presenilin-2 protein, such as N141 I (i.e. asparagine at position 141 is changed to isoleucine), M239V (i.e. methionine at position 239 is changed to valine), and D439A (i.e. aspartate at position 439 is changed to alanine) cause familial Alzheimer's type 4. Other associations of genetic variants in AD-associated genes and disease are known in the art. See, for example, Waring et al. (2008) Arch. Neurol. 65:329-334, the disclosure of which is incorporated by reference herein in its entirety.

[0393] Examples of proteins associated Autism Spectrum Disorder may include the benzodiazapine receptor (peripheral) associated protein 1 (BZRAP1) encoded by the BZRAP1 gene, the AF4/FMR2 family member 2 protein (AFF2) encoded by the AFF2 gene (also termed MFR2), the fragile X mental retardation autosomal homolog 1 protein (FXR1) encoded by the FXR1 gene, or the fragile X mental retardation autosomal homolog 2 protein (FXR2) encoded by the FXR2 gene, for example.

[0394] Examples of proteins associated Macular Degeneration may include the ATP-binding cassette, sub-family A (ABC1) member 4 protein (ABCA4) encoded by the ABCR gene, the apolipoprotein E protein (APOE) encoded by the APOE gene, or the chemokine (C-C motif) Ligand 2 protein (CCL2) encoded by the CCL2 gene, for example.

[0395] Examples of proteins associated Schizophrenia may include NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, DISC1, GSK3B, and combinations thereof.

[0396] Examples of proteins involved in tumor suppression may include ATM (ataxia telangiectasia mutated), ATR (ataxia telangiectasia and Rad3 related), EGFR (epidermal growth factor receptor), ERBB2 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 2), ERBB3 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 3), ERBB4 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 4), Notch 1, Notch2, Notch 3, or Notch 4, for example.

[0397] Examples of proteins associated with a secretase disorder may include PSENEN (presenilin enhancer 2 homolog (*C. elegans*)), CTSB (cathepsin B), PSEN1 (presenilin 1), APP (amyloid beta (A4) precursor protein), APH1B (anterior pharynx defective I homolog B (*C. elegans*)), PSEN2 (presenilin 2 (Alzheimer disease 4)), or BACE (beta-site APP-cleaving enzyme 1), for example.

[0398] For example, US Patent Publication No. 20110023146, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with secretase-associated disorders. Secretases are essential for processing pre-proteins into their biologically active forms. Defects in various components of the secretase pathways contribute to many disorders, particularly those with hallmark amyloidogenesis or amyloid plaques, such as Alzheimer's disease (AD).

[0399] A secretase disorder and the proteins associated with these disorders are a diverse set of proteins that effect susceptibility for numerous disorders, the presence of the disorder, the severity of the disorder, or any combination thereof. The present disclosure comprises editing of any chromosomal sequences that encode proteins associated with a secretase disorder. The proteins associated with a secretase disorder are typically selected based on an experimental association of the secretase-related proteins with the development of a secretase disorder. For example, the production rate or circulating concentration of a protein associated with a secretase disorder may be elevated or depressed in a population with a secretase disorder relative to a population without a secretase disorder. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the protein associated with a secretase disorder may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[0400] By way of non-limiting example, proteins associated with a secretase disorder include PSENEN (presenilin enhancer 2 homolog (*C. elegans*)), CTSB (cathepsin B), PSEN1 (presenilin 1), APP (amyloid beta (A4) precursor protein), APH1B (anterior pharynx defective 1 homolog B (*C. elegans*)), PSEN2 (presenilin 2 (Alzheimer disease 4)), BACE1 (beta-site APP-cleaving enzyme 1), ITM2B (integral membrane protein 2B), CTSD (cathepsin D), NOTCH1 (Notch homolog 1, translocation-associated (*Drosophila*)), TNF (tumor necrosis factor (TNF superfamily, member 2)), INS (insulin), DYT10 (dystonia 10), ADAM17 (ADAM metallopeptidase domain 17), APOE (apolipoprotein E), ACE (angiotensin I converting enzyme (peptidyl-dipeptidase A) 1), STN (statin), TP53 (tumor protein p53), IL6 (interleukin 6 (interferon, beta 2)), NGFR (nerve growth factor receptor (TNFR superfamily, member 16)), IL1B (interleukin 1, beta), ACHE (acetylcholinesterase (Yt blood group)), CTNNB1 (catenin (cadherin-associated protein), beta 1, 88 kDa), IGF1 (insulin-like growth factor 1 (somatomedin C)), IFNG (interferon, gamma), NRG1 (neuregulin 1), CASP3 (caspase 3, apoptosis-related cysteine peptidase), MAPK1 (mitogen-activated protein kinase 1), CDH1 (cadherin 1, type 1, E-cadherin (epithelial)), APBB1 (amyloid beta (A4) precursor protein-binding, family B, member 1 (Fe65)), HMGCR (3-hydroxy-3-methylglutaryl-Coenzyme A reductase).

CREB1 (cAMP responsive element binding protein 1), PTGS2 (prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)), HES1 (hairy and enhancer of split 1, (*Drosophila*)), CAT (catalase), TGFB1 (transforming growth factor, beta 1), ENO2 (enolase 2 (gamma, neuronal)), ERBB4 (v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)), TRAPPC10 (trafficking protein particle complex 10), MAOB (monoamine oxidase B), NGF (nerve growth factor (beta polypeptide)), MMP12 (matrix metallopeptidase 12 (macrophage elastase)), JAG1 (jagged 1 (Alagille syndrome)), CD40LG (CD40 ligand), PPARG (peroxisome proliferator-activated receptor gamma), FGF2 (fibroblast growth factor 2 (basic)), IL3 (interleukin 3 (colony-stimulating factor, multiple)), LRP1 (low density lipoprotein receptor-related protein 1), NOTCH4 (Notch homolog 4 (*Drosophila*)), MAPK8 (mitogen-activated protein kinase 8), PREP (prolyl endopeptidase), NOTCH3 (Notch homolog 3 (*Drosophila*)), PRNP (prion protein), CTSG (cathepsin G), EGF (epidermal growth factor (beta-urogastrone)), REN (renin), CD44 (CD44 molecule (Indian blood group)), SELP (selectin P (granule membrane protein 140 kDa, antigen CD62)), GHR (growth hormone receptor), ADCYAP1 (adenylate cyclase activating polypeptide 1 (pituitary)), INSR (insulin receptor), GFAP (glial fibrillary acidic protein), MMP3 (matrix metallopeptidase 3 (stromelysin 1, progelatinase)), MAPK10 (mitogen-activated protein kinase 10), SP (Sp 1 transcription factor), MYC (v-myc myelocytomatosis viral oncogene homolog (avian)), CTSE (cathepsin E), PPARA (peroxisome proliferator-activated receptor alpha), JUN (jun oncogene), TIMP1 (TIMP metallopeptidase inhibitor 1), IL5 (interleukin 5 (colony-stimulating factor, eosinophil)), IL1A (interleukin 1, alpha), MMP9 (matrix metallopeptidase 9 (gelatinase B, 92 kDa gelatinase, 92 kDa type IV collagenase)), HTR4 (5-hydroxytryptamine (serotonin) receptor 4), HSPG2 (heparan sulfate proteoglycan 2), KRAS (v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), CYCS (cytochrome c, somatic), SMG1 (SMG1 homolog, phosphatidylinositol 3-kinase-related kinase (*C. elegans*)), IL1R1 (interleukin 1 receptor, type 1), PROK1 (prokineticin 1), MAPK3 (mitogen-activated protein kinase 3), NTRK1 (neurotrophic tyrosine kinase, receptor, type 1), IL13 (interleukin 13), MME (membrane metallo-endopeptidase), TKT (transketolase), CXCR2 (chemokine (C-X-C motif) receptor 2), IGF1R (insulin-like growth factor 1 receptor), RARA (retinoic acid receptor, alpha), CREBBP (CREB binding protein), PTGS1 (prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)), GALT (galactose-1-phosphate uridylyltransferase), CHRM1 (cholinergic receptor, muscarinic 1), ATXN1 (ataxin 1), PAWR (PRKC, apoptosis, WT1, regulator), NOTCH2 (Notch homolog 2 (*Drosophila*)), M6PR (mannose-6-phosphate receptor (cation dependent)), CYP46A1 (cytochrome P450, family 46, subfamily A, polypeptide 1), CSNK1 D (casein kinase 1, delta), MAPK14 (mitogen-activated protein kinase 14), PRG2 (proteoglycan 2, bone marrow (natural killer cell activator, eosinophil granule major basic protein)), PRKCA (protein kinase C, alpha), L1 CAM (L1 cell adhesion molecule), CD40 (CD40 molecule, TNF receptor superfamily member 5), NR1I2 (nuclear receptor subfamily 1, group I, member 2), JAG2 (jagged 2), CTNND1 (catenin (cadherin-associated protein), delta 1), CDH2 (cadherin 2, type 1, N-cadherin (neuronal)), CMA1 (chymase 1, mast cell), SORT1 (sortilin 1), DLK1 (delta-like 1 homolog (*Drosophila*)), THEM4 (thioesterase superfamily member 4), JUP (junction plakoglobin), CD46 (CD46 molecule, complement regulatory protein), CCL11 (chemokine (C-C motif) ligand 11), CAV3 (caveolin 3), RNASE3 (ribonuclease, RNase A family, 3 (eosinophil cationic protein)), HSPA8 (heat shock 70 kDa protein 8), CASP9 (caspase 9, apoptosis-related cysteine peptidase), CYP3A4 (cytochrome P450, family 3, subfamily A, polypeptide 4), CCR3 (chemokine (C-C motif) receptor 3), TFAP2A (transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)), SCP2 (sterol carrier protein 2), CDK4 (cyclin-dependent kinase 4), HIF1A (hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)), TCF7L2 (transcription factor 7-like 2 (T-cell specific, HMG-box)), IL1R2 (interleukin 1 receptor, type II), B3GALTL (beta 1,3-galactosyltransferase-like), MDM2 (Mdm2 p53 binding protein homolog (mouse)), RELA (v-rel reticuloendotheliosis viral oncogene homolog A (avian)), CASP7 (caspase 7, apoptosis-related cysteine peptidase), IDE (insulin-degrading enzyme), FABP4 (fatty acid binding protein 4, adipocyte), CASK (calcium/calmodulin-dependent serine protein kinase (MAGUK family)), ADCYAP1R1 (adenylate cyclase activating polypeptide 1 (pituitary) receptor type I), ATF4 (activating transcription factor 4 (tax-responsive enhancer element B67)), PDGFA (platelet-derived growth factor alpha polypeptide), C21 or f33 (chromosome 21 open reading frame 33), SCG5 (secretogranin V (7B2 protein)), RNF123 (ring finger protein 123), NFKB1 (nuclear factor of kappa light polypeptide gene enhancer in B-cells 1), ERBB2 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)), CAV1 (caveolin 1, caveolae protein, 22 kDa), MMP7 (matrix metallopeptidase 7 (matrilysin, uterine)), TGFA (transforming growth factor, alpha), RXRA (retinoid X receptor, alpha), STX1A (syntaxin 1A (brain)), PSMC4 (proteasome (prosome, macropain) 26S subunit, ATPase, 4), P2RY2 (purinergic receptor P2Y, G-protein coupled, 2), TNFRSF21 (tumor necrosis factor receptor superfamily, member 21). DLG1 (discs, large homolog 1 (*Drosophila*)), NUMBL (numb homolog (*Drosophila*)-like), SPN (sialophorin), PLSCR1 (phospholipid scramblase 1), UBQLN2 (ubiquilin 2), UBQLN1 (ubiquilin 1), PCSK7 (proprotein convertase subtilisin/kexin type 7), SPON1 (spondin 1, extracellular matrix protein), SILV (silver homolog (mouse)), QPCT (glutaminyl-peptide cyclotransferase), HESS (hairy and enhancer of split 5 (*Drosophila*)), GCC1 (GRIP and coiled-coil domain containing 1), and any combination thereof.

[0401] The genetically modified animal or cell may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more disrupted chromosomal sequences encoding a protein associated with a secretase disorder and zero, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more chromosomally integrated sequences encoding a disrupted protein associated with a secretase disorder.

[0402] Examples of proteins associated with Amyotrophic Lateral Sclerosis may include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

[0403] For example, US Patent Publication No. 20110023144, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with amyotrophyic lateral sclerosis (ALS) disease. ALS is characterized by the gradual steady degeneration of certain nerve cells in the brain cortex, brain stem, and spinal cord involved in voluntary movement.

[0404] Motor neuron disorders and the proteins associated with these disorders are a diverse set of proteins that effect susceptibility for developing a motor neuron disorder, the presence of the motor neuron disorder, the severity of the motor neuron disorder or any combination thereof. The present disclosure comprises editing of any chromosomal sequences that encode proteins associated with ALS disease, a specific motor neuron disorder. The proteins associated with ALS are typically selected based on an experimental association of ALS-related proteins to ALS. For example, the production rate or circulating concentration of a protein associated with ALS may be elevated or depressed in a population with ALS relative to a population without ALS. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the proteins associated with ALS may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[0405] By way of non-limiting example, proteins associated with ALS include but are not limited to the following proteins: SOD1 superoxide dismutase 1, ALS3 amyotrophic lateral soluble sclerosis 3 SETX senataxin ALS5 amyotrophic lateral sclerosis 5 FUS fused in sarcoma ALS7 amyotrophic lateral sclerosis 7 ALS2 amyotrophic lateral DPP6 Dipeptidyl-peptidase 6 sclerosis 2 NEFH neurofilament, heavy PTGS1 prostaglandin-polypeptide endoperoxide synthase 1 SLC1A2 solute carrier family 1 TNFRSF10B tumor necrosis factor (glial high affinity receptor superfamily, glutamate transporter), member 10b member 2 PRPH peripherin HSP90AA1 heat shock protein 90 kDa alpha (cytosolic), class A member 1 GRIA2 glutamate receptor, IFNG interferon, gamma ionotropic, AMPA 2 S100B S100 calcium binding FGF2 fibroblast growth factor 2 protein B AOX1 aldehyde oxidase 1 CS citrate synthase TARDBP TAR DNA binding protein TXN thioredoxin RAPH1 Ras association MAP3K5 mitogen-activated protein (RalGDS/AF-6) and kinase 5 pleckstrin homology domains 1 NBEAL1 neurobeachin-like 1 GPX1 glutathione peroxidase 1 ICA1L islet cell autoantigen RAC1 ras-related C3 botulinum 1.69 kDa-like toxin substrate 1 MAPT microtubule-associated ITPR2 inositol 1,4,5-protein tau triphosphate receptor, type 2 ALS2CR4 amyotrophic lateral GLS glutaminase sclerosis 2 (juvenile) chromosome region, candidate 4 ALS2CR8 amyotrophic lateral CNTFR ciliary neurotrophic factor sclerosis 2 (juvenile) receptor chromosome region, candidate 8 ALS2CR11 amyotrophic lateral FOLH1 folate hydrolase 1 sclerosis 2 (juvenile) chromosome region, candidate 11 FAM117B family with sequence P4HB prolyl 4-hydroxylase, similarity 117, member B beta polypeptide CNTF ciliary neurotrophic factor SQSTM1 sequestosome 1 STRADB STE20-related kinase NAIP NLR family, apoptosis adaptor beta inhibitory protein YWHAQ tyrosine 3-SLC33A1 solute carrier family 33 monooxygenase/tryptoph (acetyl-CoA transporter), an 5-monooxygenase member 1 activation protein, theta polypeptide TRAK2 trafficking protein, FIG. **4** FIG. **4** homolog, SAC1 kinesin binding 2 lipid phosphatase domain containing NIF3L1 NIF3 NGG1 interacting INA internexin neuronal factor 3-like 1 intermediate filament protein, alpha PARD3B par-3 partitioning COX8A cytochrome c oxidase defective 3 homolog B subunit VIIIA CDK15 cyclin-dependent kinase HECW1 HECT, C2 and WW 15 domain containing E3 ubiquitin protein ligase 1 NOS1 nitric oxide synthase 1 MET met proto-oncogene SOD2 superoxide dismutase 2, HSPB1 heat shock 27 kDa mitochondrial protein 1 NEFL neurofilament, light CTSB cathepsin B polypeptide ANG angiogenin, HSPA8 heat shock 70 kDa ribonuclease, RNase A protein 8 family, 5 VAPB VAMP (vesicle-ESR1 estrogen receptor 1 associated membrane protein)-associated protein B and C SNCA synuclein, alpha HGF hepatocyte growth factor CAT catalase ACTB actin, beta NEFM neurofilament, medium TH tyrosine hydroxylase polypeptide BCL2 B-cell CLL/lymphoma 2 FAS Fas (TNF receptor superfamily, member 6) CASP3 caspase 3, apoptosis-CLU clusterin related cysteine peptidase SMN1 survival of motor neuron G6PD glucose-6-phosphate 1, telomeric dehydrogenase BAX BCL2-associated X HSF1 heat shock transcription protein factor 1 RNF19A ring finger protein 19A JUN jun oncogene ALS2CR12 amyotrophic lateral HSPA5 heat shock 70 kDa sclerosis 2 (juvenile) protein 5 chromosome region, candidate 12 MAPK14 mitogen-activated protein IL10 interleukin 10 kinase 14 APEX1 APEX nuclease TXNRD1 thioredoxin reductase 1 (multifunctional DNA repair enzyme) 1 NOS2 nitric oxide synthase 2, TIMP1 TIMP metallopeptidase inducible inhibitor 1 CASP9 caspase 9, apoptosis-XIAP X-linked inhibitor of related cysteine apoptosis peptidase GLG1 golgi glycoprotein 1 EPO erythropoietin VEGFA vascular endothelial ELN elastin growth factor A GDNF glial cell derived NFE2L2 nuclear factor (erythroid-neurotrophic factor derived 2)-like 2 SLC6A3 solute carrier family 6 HSPA4 heat shock 70 kDa (neurotransmitter protein 4 transporter, dopamine), member 3 APOE apolipoprotein E PSMB8 proteasome (prosome, macropain) subunit, beta type, 8 DCTN1 dynactin 1 TIMP3 TIMP metallopeptidase inhibitor 3 KIFAP3 kinesin-associated SLC1A1 solute carrier family 1 protein 3 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1 SMN2 survival of motor neuron CCNC cyclin C 2, centromeric MPP4 membrane protein, STUB1 STIP1 homology and U-palmitoylated 4 box containing protein 1 ALS2 amyloid beta (A4) PRDX6 peroxiredoxin 6 precursor protein SYP synaptophysin CABIN1 calcineurin binding protein 1 CASP1 caspase 1, apoptosis-GART phosphoribosylglycinami related cysteine de formyltransferase, peptidase phosphoribosylglycinami de synthetase, phosphoribosylaminoimidazole synthetase CDK5 cyclin-dependent kinase 5 ATXN3 ataxin 3 RTN4 reticulon 4 C1QB complement component 1, q subcomponent, B chain VEGFC nerve growth factor HTT huntingtin receptor PARK7 Parkinson disease 7 XDH xanthine dehydrogenase GFAP glial fibrillary acidic MAP2 microtubule-associated protein protein 2 CYCS cytochrome c, somatic FCGR3B Fc fragment of IgG, low affinity IIIb, CCS copper chaperone for UBL5 ubiquitin-like 5 superoxide dismutase MMP9 matrix metallopeptidase SLC18A3 solute carrier family 18 9 ((vesicular acetylcholine), member 3 TRPM7 transient receptor HSPB2 heat shock 27 kDa potential cation channel, protein 2 subfamily M, member 7 AKT1 v-akt murine thymoma DERL1 Der1-like domain family, viral oncogene homolog 1 member 1 CCL2 chemokine (C-C motif) NGRN neugrin, neurite ligand 2 outgrowth associated GSR glutathione reductase TPPP3 tubulin polymerization-promoting protein family member 3 APAF1 apoptotic peptidase BTBD10 BTB (POZ) domain activating factor 1 containing 10 GLUD1 glutamate CXCR4 chemokine (C-X-C motif) dehydrogenase 1 receptor 4 SLCIA3 solute carrier family 1 FLT1 fms-related tyrosine (glial high affinity glutamate transporter), member 3 kinase 1 PON1 paraoxonase 1 AR androgen receptor LIF leukemia inhibitory factor ERBB3 v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 LGALS1 lectin, galactoside-CD44 CD44 molecule binding, soluble, 1 TP53 tumor protein p53 TLR3 toll-like receptor 3 GRIA1 glutamate receptor, GAPDH glyceraldehyde-3-ionotropic, AMPA 1 phosphate dehydrogenase GRIK1 glutamate receptor, DES desmin ionotropic, kainate 1 CHAT choline acetyltransferase FLT4 fins-related tyrosine kinase 4 CHMP2B chromatin modifying BAG1 BCL2-associated protein 2B athanogene MT3 metallothionein 3 CHRNA4 cholinergic receptor, nicotinic, alpha 4 GSS glutathione synthetase BAK1 BCL2-antagonist/killer 1 KDR kinase insert domain GSTP1 glutathione S-transferase receptor (a type III pi 1 receptor tyrosine kinase) OGG1 8-oxoguanine DNA IL6 interleukin 6 (interferon, glycosylase beta 2).

[0406] The animal or cell may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more disrupted chromosomal sequences encoding a protein associated with ALS and zero, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more chromosomally integrated sequences encoding the disrupted protein associated with ALS. Preferred proteins associated with ALS include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

[0407] Examples of proteins associated with prion diseases may include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

[0408] Examples of proteins related to neurodegenerative conditions in prion disorders may include A2M (Alpha-2-Macroglobulin), AATF (Apoptosis antagonizing transcription factor), ACPP (Acid phosphatase prostate), ACTA2 (Actin alpha 2 smooth muscle aorta), ADAM22 (ADAM metallopeptidase domain), ADORA3 (Adenosine A3 receptor), or ADRA1D (Alpha-1D adrenergic receptor for Alpha-1D adrenoreceptor), for example.

[0409] Examples of proteins associated with Immunodeficiency may include A2M [alpha-2-macroglobulin]; AANAT [arylalkylamine N-acetyltransferase]; ABCA1 [ATP-binding cassette, sub-family A (ABC1), member 1]; ABCA2 [ATP-binding cassette, sub-family A (ABC1), member 2]; or ABCA3 [ATP-binding cassette, sub-family A (ABC1), member 3]; for example.

[0410] Examples of proteins associated with Trinucleotide Repeat Disorders include AR (androgen receptor), FMR1 (fragile X mental retardation 1), HTT (huntingtin), or DMPK (dystrophia myotonica-protein kinase), FXN (frataxin), ATXN2 (ataxin 2), for example.

[0411] Examples of proteins associated with Neurotransmission Disorders include SST (somatostatin), NOS1 (nitric oxide synthase 1 (neuronal)), ADRA2A (adrenergic, alpha-2A-, receptor), ADRA2C (adrenergic, alpha-2C-, receptor), TACR1 (tachykinin receptor 1), or HTR2c (5-hydroxytryptamine (serotonin) receptor 2C), for example.

[0412] Examples of neurodevelopmental-associated sequences include A2BP1 [ataxin 2-binding protein 1], AADAT [aminoadipate aminotransferase], AANAT [arylalkylamine N-acetyltransferase], ABAT [4-aminobutyrate aminotransferase], ABCA1 [ATP-binding cassette, sub-family A (ABC1), member 1], or ABCA13 [ATP-binding cassette, sub-family A (ABC1), member 13], for example.

[0413] Further examples of preferred conditions treatable with the present system include may be selected from: Aicardi-Goutières Syndrome; Alexander Disease; Allan-Herndon-Dudley Syndrome; POLG-Related Disorders; Alpha-Mannosidosis (Type II and III); Alström Syndrome; Angelman; Syndrome; Ataxia-Telangiectasia; Neuronal Ceroid-Lipofuscinoses; Beta-Thalassemia; Bilateral Optic Atrophy and (Infantile) Optic Atrophy Type 1; Retinoblastoma (bilateral); Canavan Disease; Cerebrooculofacioskeletal Syndrome 1 [COFS1]; Cerebrotendinous Xanthomatosis; Cornelia de Lange Syndrome; MAPT-Related Disorders; Genetic Prion Diseases; Dravet Syndrome; Early-Onset Familial Alzheimer Disease; Friedreich Ataxia [FRDA]; Fryns Syndrome; Fucosidosis; Fukuyama Congenital Muscular Dystrophy; Galactosialidosis; Gaucher Disease; Organic Acidemias; Hemophagocytic Lymphohistiocytosis; Hutchinson-Gilford Progeria Syndrome; Mucolipidosis II; Infantile Free Sialic Acid Storage Disease; PLA2G6-Associated Neurodegeneration; Jervell and Lange-Nielsen Syndrome; Junctional Epidermolysis Bullosa; Huntington Disease; Krabbe Disease (Infantile); Mitochondrial DNA-Associated Leigh Syndrome and NARP; Lesch-Nyhan Syndrome; LIS1-Associated Lissencephaly; Lowe Syndrome; Maple Syrup Urine Disease; MECP2 Duplication Syndrome; ATP7A-Related Copper Transport Disorders: LAMA2-Related Muscular Dystrophy; Arylsulfatase A Deficiency; Mucopolysaccharidosis Types I, II or III; Peroxisome Biogenesis Disorders, Zellweger Syndrome Spectrum; Neurodegeneration with Brain Iron Accumulation Disorders; Acid Sphingomyelinase Deficiency; Niemann-Pick Disease Type C; Glycine Encephalopathy; ARX-Related Disorders; Urea Cycle Disorders; COL1A1/2-Related Osteogenesis Imperfecta; Mitochondrial DNA Deletion Syndromes; PLP1-Related Disorders; Perry Syndrome; Phelan-McDermid Syndrome; Glycogen Storage Disease Type II (Pompe Disease) (Infantile); MAPT-Related Disorders; MECP2-Related Disorders; Rhizomelic Chondrodysplasia Punctata Type 1; Roberts Syndrome; Sandhoff Disease; Schindler Disease—Type 1; Adenosine Deaminase Deficiency; Smith-Lemli-Opitz Syndrome; Spinal Muscular Atrophy; Infantile-Onset Spinocerebellar Ataxia; Hexosaminidase A Deficiency; Thanatophoric Dysplasia Type 1; Collagen Type VI-Related Disorders; Usher Syndrome Type I; Congenital Muscular Dystrophy; Wolf-Hirschhorn Syndrome; Lysosomal Acid Lipase Deficiency; and Xeroderma Pigmentosum.

[0414] As will be apparent, it is envisaged that the present system can be used to target any polynucleotide sequence of interest. Some examples of conditions or diseases that might be usefully treated using the present system are included in the Tables above and examples of genes currently associated with those conditions are also provided there. However, the genes exemplified are not exhaustive.

[0415] For example, "wild type StCas9" refers to wild type Cas9 from *S. thermophilus*, the protein sequence of which is given in the SwissProt database under accession number G3ECR1. Similarly, *S. pyogenes* Cas9 is included in SwissProt under accession number Q99ZW2.

[0416] The ability to use CRISPR-Cas systems to perform efficient and cost effective gene editing and manipulation will allow the rapid selection and comparison of single and multiplexed genetic manipulations to transform such genomes for improved production and enhanced traits. In this regard reference is made to US patents and publications: U.S. Pat. No. 6,603,061—*Agrobacterium*-Mediated Plant Transformation Method; U.S. Pat. No. 7,868,149-Plant Genome Sequences and Uses Thereof and US 2009/0100536—Transgenic Plants with Enhanced Agronomic Traits, all the contents and disclosure of each of which are herein incorporated by reference in their entirety. In the practice of the invention, the contents and disclosure of Morrell et al "Crop genomics: advances and applications" Nat Rev Genet. 2011 Dec. 29; 13(2):85-96 are also herein incorporated by reference in their entirety.

EXAMPLES

[0417] The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

Example 1

CRISPR Complex Activity in the Nucleus of a Eukaryotic Cell

[0418] An example type II CRISPR system is the type II CRISPR locus from *Streptococcus pyogenes* SF370, which contains a cluster of four genes Cas9, Cas1, Cas2, and Csn1, as well as two non-coding RNA elements, tracrRNA and a characteristic array of repetitive sequences (direct repeats) interspaced by short stretches of non-repetitive sequences (spacers, about 30 bp each). In this system, targeted DNA double-strand break (DSB) is generated in four sequential steps (FIG. **2**A). First, two non-coding RNAs, the pre-crRNA array and tracrRNA, are transcribed from the CRISPR locus. Second, tracrRNA hybridizes to the direct repeats of pre-crRNA, which is then processed into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas9 to the DNA target consisting of the protospacer and the corresponding PAM via heteroduplex formation between the spacer region of the crRNA and the protospacer DNA. Finally, Cas9 mediates cleavage of target DNA upstream of PAM to create a DSB within the protospacer (FIG. **2**A). This example describes an example process for adapting this RNA-programmable nuclease system to direct CRISPR complex activity in the nuclei of eukaryotic cells.

[0419] Cell Culture and Transfection

[0420] Human embryonic kidney (HEK) cell line HEK 293FT (Life Technologies) was maintained in Dulbecco's modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (HyClone), 2 mM GlutaMAX (Life Technologies), 100 U/mL penicillin, and 100 μg/mL streptomycin at 37° C. with 5% CO_2 incubation. Mouse neuro2A (N2A) cell line (ATCC) was maintained with DMEM supplemented with 5% fetal bovine serum (HyClone), 2 mM GlutaMAX (Life Technologies), 100 U/mL penicillin, and 100 μg/mL streptomycin at 37° C. with 5% CO_2.

[0421] HEK 293FT or N2A cells were seeded into 24-well plates (Corning) one day prior to transfection at a density of 200,000 cells per well. Cells were transfected using Lipofectamine 2000 (Life Technologies) following the manufacturer's recommended protocol. For each well of a 24-well plate a total of 800 ng of plasmids were used.

[0422] Surveyor Assay and Sequencing Analysis for Genome Modification

[0423] HEK 293FT or N2A cells were transfected with plasmid DNA as described above. After transfection, the cells were incubated at 37° C. for 72 hours before genomic DNA extraction. Genomic DNA was extracted using the QuickExtract DNA extraction kit (Epicentre) following the manufacturer's protocol. Briefly, cells were resuspended in QuickExtract solution and incubated at 65° C. for 15 minutes and 98° C. for 10 minutes. Extracted genomic DNA was immediately processed or stored at −20° C.

[0424] The genomic region surrounding a CRISPR target site for each gene was PCR amplified, and products were purified using QiaQuick Spin Column (Qiagen) following manufacturer's protocol. A total of 400 ng of the purified PCR products were mixed with 2 μl 10× Taq polymerase PCR buffer (Enzymatics) and ultrapure water to a final volume of 20 μl, and subjected to a re-annealing process to enable heteroduplex formation: 95° C. for 10 min, 95° C. to 85° C. ramping at −2° C./s, 85° C. to 25° C. at −0.25° C./s, and 25° C. hold for 1 minute. After re-annealing, products were treated with Surveyor nuclease and Surveyor enhancer S (Transgenomics) following the manufacturer's recommended protocol, and analyzed on 4-20% Novex TBE poly-acrylamide gels (Life Technologies). Gels were stained with SYBR Gold DNA stain (Life Technologies) for 30 minutes and imaged with a Gel Doc gel imaging system (Bio-rad). Quantification was based on relative band intensities, as a measure of the fraction of cleaved DNA. FIG. **7** provides a schematic illustration of this Surveyor assay.

[0425] Restriction fragment length polymorphism assay for detection of homologous recombination.

[0426] HEK 293FT and N2A cells were transfected with plasmid DNA, and incubated at 37° C. for 72 hours before genomic DNA extraction as described above. The target genomic region was PCR amplified using primers outside the homology arms of the homologous recombination (HR) template. PCR products were separated on a 1% agarose gel and extracted with MinElute GelExtraction Kit (Qiagen). Purified products were digested with HindIII (Fermentas) and analyzed on a 6% Novex TBE poly-acrylamide gel (Life Technologies).

[0427] RNA Secondary Structure Prediction and Analysis

[0428] RNA secondary structure prediction was performed using the online webserver RNAfold developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g. A. R. Gruber et al., 2008, Cell 106(1): 23-24; and P A Carr and G M Church, 2009, Nature Biotechnology 27(12): 1151-62).

[0429] RNA Purification

[0430] HEK 293FT cells were maintained and transfected as stated above. Cells were harvested by trypsinization followed by washing in phosphate buffered saline (PBS). Total cell RNA was extracted with TRI reagent (Sigma) following manufacturer's protocol. Extracted total RNA was quantified using Naonodrop (Thermo Scientific) and normalized to same concentration.

[0431] Northern Blot Analysis of crRNA and tracrRNA Expression in Mammalian Cells

[0432] RNAs were mixed with equal volumes of 2× loading buffer (Ambion), heated to 95° C. for 5 min, chilled on ice for 1 min, and then loaded onto 8% denaturing polyacrylamide gels (SequaGel, National Diagnostics) after pre-running the gel for at least 30 minutes. The samples were electrophoresed for 1.5 hours at 40 W limit. Afterwards, the RNA was transferred to Hybond N+ membrane (GE Healthcare) at 300 mA in a semi-dry transfer apparatus (Bio-rad) at room temperature for 1.5 hours. The RNA was crosslinked to the membrane using autocrosslink button on Stratagene UV Crosslinker the Stratalinker (Stratagene). The membrane was pre-hybridized in ULTRAhyb-Oligo Hybridization Buffer (Ambion) for 30 min with rotation at 42° C., and probes were then added and hybridized overnight. Probes were ordered from IDT and labeled with [gamma-^{32}P] ATP (Perkin Elmer) with T4 polynucleotide kinase (New England Biolabs). The membrane was washed once with pre-warmed (42° C.) 2×SSC, 0.5% SDS for 1 min followed by two 30 minute washes at 42° C. The membrane was exposed to a phosphor screen for one hour or overnight at room temperature and then scanned with a phosphorimager (Typhoon).

[0433] Bacterial CRISPR System Construction and Evaluation

[0434] CRISPR locus elements, including tracrRNA, Cas9, and leader were PCR amplified from *Streptococcus pyogenes* SF370 genomic DNA with flanking homology arms for Gibson Assembly. Two BsaI type IIS sites were introduced in between two direct repeats to facilitate easy insertion of spacers (FIG. **8**). PCR products were cloned into EcoRV-digested pACYC184 downstream of the tet promoter using Gibson Assembly Master Mix (NEB). Other endogenous CRISPR system elements were omitted, with the exception of the last 50 bp of Csn2. Oligos (Integrated DNA Technology) encoding spacers with complimentary overhangs were cloned into the BsaI-digested vector pDC000 (NEB) and then ligated with T7 ligase (Enzymatics) to generate pCRISPR plasmids. Challenge plasmids containing spacers with PAM

[0435] expression in mammalian cells (expression constructs illustrated in FIG. **6**A, with functionality as determined by results of the Surveyor assay shown in FIG. **6**B). Transcription start sites are marked as +1, and transcription terminator and the sequence probed by northern blot are also indicated. Expression of processed tracrRNA was also confirmed by Northern blot. FIG. **6**C shows results of a Northern blot analysis of total RNA extracted from 293FT cells transfected with U6 expression constructs carrying long or short tracrRNA, as well as SpCas9 and DR-EMX1(1)-DR. Left and right panels are from 293FT cells transfected without or with SpRNase III, respectively. U6 indicate loading control blotted with a probe targeting human U6 snRNA. Transfection of the short tracrRNA expression construct led to abundant levels of the processed form of tracrRNA (~75 bp). Very low amounts of long tracrRNA are detected on the Northern blot.

[0436] To promote precise transcriptional initiation, the RNA polymerase III-based U6 promoter was selected to drive the expression of tracrRNA (FIG. **2**C). Similarly, a U6 promoter-based construct was developed to express a pre-crRNA array consisting of a single spacer flanked by two direct repeats (DRs, also encompassed by the term "tracr-mate sequences"; FIG. **2**C). The initial spacer was designed to target a 33-base-pair (bp) target site (30-bp protospacer plus a 3-bp CRISPR motif (PAM) sequence satisfying the NGG recognition motif of Cas9) in the human EMX1 locus (FIG. **2**C), a key gene in the development of the cerebral cortex.

[0437] To test whether heterologous expression of the CRISPR system (SpCas9, SpRNase III, tracrRNA, and pre-crRNA) in mammalian cells can achieve targeted cleavage of mammalian chromosomes, HEK 293FT cells were transfected with combinations of CRISPR components. Since DSBs in mammalian nuclei are partially repaired by the non-homologous end joining (NHEJ) pathway, which leads to the formation of indels, the Surveyor assay was used to detect potential cleavage activity at the target EMX1 locus (FIG. **7**) (see e.g. Guschin et al., 2010, Methods Mol Biol 649: 247). Co-transfection of all four CRISPR components was able to induce up to 5.0% cleavage in the protospacer (see FIG. **2**D). Co-transfection of all CRISPR components minus SpRNase III also induced up to 4.7% indel in the protospacer, suggesting that there may be endogenous mammalian RNases that are capable of assisting with crRNA maturation, such as for example the related Dicer and Drosha enzymes. Removing any of the remaining three components abolished the genome cleavage activity of the CRISPR system (FIG. **2**D). Sanger sequencing of amplicons containing the target locus verified the cleavage activity: in 43 sequenced clones, 5 mutated alleles (11.6%) were found. Similar experiments using a variety of guide sequences produced indel percentages as high as 29% (see FIGS. **3**-**6**, **10**, and **11**). These results define a three-component system for efficient CRISPR-mediated genome modification in mammalian cells. To optimize the cleavage efficiency, Applicants also tested whether different isoforms of tracrRNA affected the cleavage efficiency and found that, in this example system, only the short (89-bp) transcript form was able to mediate cleavage of the human EMX1 genomic locus (FIG. **6**B).

[0438] FIG. **12** provides an additional Northern blot analysis of crRNA processing in mammalian cells. FIG. **12**A illustrates a schematic showing the expression vector for a single spacer flanked by two direct repeats (DR-EMX1(1)-DR). The 30 bp spacer targeting the human EMX1 locus protospacer 1 (see FIG. **6**) and the direct repeat sequences are shown in the sequence beneath FIG. **12**A. The line indicates the region whose reverse-complement sequence was used to generate Northern blot probes for EMX1(1) crRNA detection. FIG. **12**B shows a Northern blot analysis of total RNA extracted from 293FT cells transfected with U6 expression constructs carrying DR-EMX1(1)-DR. Left and right panels are from 293FT cells transfected without or with SpRNase III respectively. DR-EMX1(1)-DR was processed into mature crRNAs only in the presence of SpCas9 and short tracrRNA and was not dependent on the presence of SpRNase III. The mature crRNA detected from transfected 293FT total RNA is −33 bp and is shorter than the 39-42 bp mature crRNA from *S. pyogenes*. These results demonstrate that a CRISPR system can be transplanted into eukaryotic cells and reprogrammed to facilitate cleavage of endogenous mammalian target polynucleotides.

[0439] FIG. **2** illustrates the bacterial CRISPR system described in this example. FIG. **2**A illustrates a schematic showing the CRISPR locus 1 from *Streptococcus pyogenes* SF370 and a proposed mechanism of CRISPR-mediated DNA cleavage by this system. Mature crRNA processed from the direct repeat-spacer array directs Cas9 to genomic targets consisting of complimentary protospacers and a protospacer-adjacent motif (PAM). Upon target-spacer base pairing, Cas9 mediates a double-strand break in the target DNA. FIG. **2**B illustrates engineering of *S. pyogenes* Cas9 (SpCas9) and RNase III (SpRNase III) with nuclear localization signals (NLSs) to enable import into the mammalian nucleus. FIG. **2**C illustrates mammalian expression of SpCas9 and SpRNase III driven by the constitutive EF1a promoter and tracrRNA and pre-crRNA array (DR-Spacer-DR) driven by the RNA Pol3 promoter U6 to promote precise transcription initiation and termination. A protospacer from the human EMX1 locus with a satisfactory PAM sequence is used as the spacer in the pre-crRNA array. FIG. **2**D illustrates surveyor nuclease assay for SpCas9-mediated minor insertions and deletions. SpCas9 was expressed with and without SpRNase III, tracrRNA, and a pre-crRNA array carrying the EMX1-target spacer. FIG. **2**E illustrates a schematic representation of base pairing between target locus and EMX1-targeting crRNA, as well as an example chromatogram showing a micro deletion adjacent to the SpCas9 cleavage site. FIG. **2**F illustrates mutated alleles identified from sequencing analysis of 43 clonal amplicons showing a variety of micro insertions and deletions. Dashes indicate deleted bases, and non-aligned or mismatched bases indicate insertions or mutations. Scale bar=10 μm.

[0440] To further simplify the three-component system, a chimeric crRNA-tracrRNA hybrid design was adapted, where a mature crRNA (comprising a guide sequence) may be fused to a partial tracrRNA via a stem-loop to mimic the natural crRNA:tracrRNA duplex. To increase co-delivery efficiency, a bicistronic expression vector was created to drive co-expression of a chimeric RNA and SpCas9 in transfected cells. In parallel, the bicistronic vectors were used to express a pre-crRNA (DR-guide sequence-DR) with SpCas9, to induce processing into crRNA with a separately expressed tracrRNA (compare FIG. **11**B top and bottom). FIG. **8** provides schematic illustrations of bicistronic expression vectors for pre-crRNA array (FIG. **8**A) or chimeric crRNA (represented by the short line downstream of the guide sequence insertion site and upstream of the EF1α promoter in FIG. **8**B) with hSpCas9, showing location of various elements and the point of guide sequence insertion. The expanded sequence around the location of the guide sequence insertion site in FIG. **8**B also shows a partial DR sequence (GTTTAGAGCTA) and a partial tracrRNA sequence (TAGCAAGTTAAAATAAGGCTAGTCCGTTTTT). Guide sequences can be inserted between BbsI sites using annealed oligonucleotides. Sequence design for the oligonucleotides are shown below the schematic illustrations in FIG. **8**, with appropriate ligation adapters indicated. WPRE represents the Woodchuck hepatitis virus post-transcriptional regulatory element. The efficiency of chimeric RNA-mediated cleavage was tested by targeting the same EMX1 locus described above. Using both Surveyor assay and Sanger sequencing of amplicons, Applicants confirmed that the chimeric RNA design facilitates cleavage of human EMX1 locus with approximately a 4.7% modification rate (FIG. **3**).

[0441] Generalizability of CRISPR-mediated cleavage in eukaryotic cells was tested by targeting additional genomic loci in both human and mouse cells by designing chimeric RNA targeting multiple sites in the human EMX1 and PVALB, as well as the mouse Th loci. FIG. **13** illustrates the selection of some additional targeted protospacers in human PVALB (FIG. **13**A) and mouse Th (FIG. **13**B) loci. Schematics of the gene loci and the location of three protospacers within the last exon of each are provided. The underlined sequences include 30 bp of protospacer sequence and 3 bp at the 3' end corresponding to the PAM sequences. Protospacers on the sense and anti-sense strands are indicated above and below the DNA sequences, respectively. A modification rate of 6.3% and 0.75% was achieved for the human PVALB and mouse Th loci respectively, demonstrating the broad applicability of the CRISPR system in modifying different loci across multiple organisms (FIG. **5**). While cleavage was only detected with one out of three spacers for each locus using the chimeric constructs, all target sequences were cleaved with efficiency of indel production reaching 27% when using the co-expressed pre-crRNA arrangement (FIGS. **6** and **13**).

[0442] FIG. **11** provides a further illustration that SpCas9 can be reprogrammed to target multiple genomic loci in mammalian cells. FIG. **11**A provides a schematic of the human EMX1 locus showing the location of five protospacers, indicated by the underlined sequences. FIG. **11**B provides a schematic of the pre-crRNA/trcrRNA complex showing hybridization between the direct repeat region of the pre-crRNA and tracrRNA (top), and a schematic of a chimeric RNA design comprising a 20 bp guide sequence, and tracr mate and tracr sequences consisting of partial direct repeat and tracrRNA sequences hybridized in a hairpin structure (bottom). Results of a Surveyor assay comparing the efficacy of Cas9-mediated cleavage at five protospacers in the human EMX1 locus is illustrated in FIG. **11**C. Each protospacer is targeted using either processed pre-crRNA/tracrRNA complex (crRNA) or chimeric RNA (chiRNA).

[0443] Since the secondary structure of RNA can be crucial for intermolecular interactions, a structure prediction algorithm based on minimum free energy and Boltzmann-weighted structure ensemble was used to compare the putative secondary structure of all guide sequences used in the genome targeting experiment (see e.g. Gruber et al., 2008, Nucleic Acids Research, 36: W70). Analysis revealed that in most cases, the effective guide sequences in the chimeric crRNA context were substantially free of secondary structure motifs, whereas the ineffective guide sequences were more likely to form internal secondary structures that could prevent base pairing with the target protospacer DNA. It is thus possible that variability in the spacer secondary structure might impact the efficiency of CRISPR-mediated interference when using a chimeric crRNA.

[0444] Further vector designs for SpCas9 are shown in FIG. **22**, which illustrates single expression vectors incorporating a U6 promoter linked to an insertion site for a guide oligo, and a Cbh promoter linked to SpCas9 coding sequence. The vector shown in FIG. **22***b* includes a tracrRNA coding sequence linked to an H1 promoter.

[0445] In the bacterial assay, all spacers facilitated efficient CRISPR interference (FIG. **3**C). These results suggest that there may be additional factors affecting the efficiency of CRISPR activity in mammalian cells.

[0446] To investigate the specificity of CRISPR-mediated cleavage, the effect of single-nucleotide mutations in the guide sequence on protospacer cleavage in the mammalian genome was analyzed using a series of EMX1-targeting chimeric crRNAs with single point mutations (FIG. **3**A). FIG. **3**B illustrates results of a Surveyor nuclease assay comparing the cleavage efficiency of Cas9 when paired with different mutant chimeric RNAs. Single-base mismatch up to 12-bp 5' of the PAM substantially abrogated genomic cleavage by

SpCas9, whereas spacers with mutations at farther upstream positions retained activity against the original protospacer target (FIG. **3**B). In addition to the PAM, SpCas9 has single-base specificity within the last 12-bp of the spacer. Furthermore, CRISPR is able to mediate genomic cleavage as efficiently as a pair of TALE nucleases (TALEN) targeting the same EMX1 protospacer. FIG. **3**C provides a schematic showing the design of TALENs targeting EMX1, and FIG. **3**D shows a Surveyor gel comparing the efficiency of TALEN and Cas9 (n=3).

[0447] Having established a set of components for achieving CRISPR-mediated gene editing in mammalian cells through the error-prone NHEJ mechanism, the ability of CRISPR to stimulate homologous recombination (HR), a high fidelity gene repair pathway for making precise edits in the genome, was tested. The wild type SpCas9 is able to mediate site-specific DSBs, which can be repaired through both NHEJ and HR. In addition, an aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of SpCas9 was engineered to convert the nuclease into a nickase (SpCas9n; illustrated in FIG. **4**A) (see e.g. Sapranausaks et al., 2011, Nucleic Acids Resch, 39: 9275; Gasiunas et al., 2012, Proc. Natl. Acad. Sci. USA, 109:E2579), such that nicked genomic DNA undergoes the high-fidelity homology-directed repair (HDR). Surveyor assay confirmed that SpCas9n does not generate indels at the EMX1 protospacer target. As illustrated in FIG. **4**B, co-expression of EMX1-targeting chimeric crRNA with SpCas9 produced indels in the target site, whereas co-expression with SpCas9n did not (n=3). Moreover, sequencing of 327 amplicons did not detect any indels induced by SpCas9n. The same locus was selected to test CRISPR-mediated HR by co-transfecting HEK 293FT cells with the chimeric RNA targeting EMX1, hSpCas9 or hSpCas9n, as well as a HR template to introduce a pair of restriction sites (HindIII and NheI) near the protospacer. FIG. **4**C provides a schematic illustration of the HR strategy, with relative locations of recombination points and primer annealing sequences (arrows). SpCas9 and SpCas9n indeed catalyzed integration of the HR template into the EMX1 locus. PCR amplification of the target region followed by restriction digest with HindIII revealed cleavage products corresponding to expected fragment sizes (arrows in restriction fragment length polymorphism gel analysis shown in FIG. **4**D), with SpCas9 and SpCas9n mediating similar levels of HR efficiencies. Applicants further verified HR using Sanger sequencing of genomic amplicons (FIG. **4**E). These results demonstrate the utility of CRISPR for facilitating targeted gene insertion in the mammalian genome. Given the 14-bp (12-bp from the spacer and 2-bp from the PAM) target specificity of the wild type SpCas9, the availability of a nickase can significantly reduce the likelihood of off-target modifications, since single strand breaks are not substrates for the error-prone NHEJ pathway.

[0448] Expression constructs mimicking the natural architecture of CRISPR loci with arrayed spacers (FIG. **2**A) were constructed to test the possibility of multiplexed sequence targeting. Using a single CRISPR array encoding a pair of EMX1- and PVALB-targeting spacers, efficient cleavage at both loci was detected (FIG. **4**F, showing both a schematic design of the crRNA array and a Surveyor blot showing efficient mediation of cleavage). Targeted deletion of larger genomic regions through concurrent DSBs using spacers against two targets within EMX1 spaced by 119 bp was also tested, and a 1.6% deletion efficacy (3 out of 182 amplicons; FIG. **4**G) was detected. This demonstrates that the CRISPR system can mediate multiplexed editing within a single genome.

Example 2

CRISPR System Modifications and Alternatives

[0449] The ability to use RNA to program sequence-specific DNA cleavage defines a new class of genome engineering tools for a variety of research and industrial applications. Several aspects of the CRISPR system can be further improved to increase the efficiency and versatility of CRISPR targeting. Optimal Cas9 activity may depend on the availability of free Mg^{2+} at levels higher than that present in the mammalian nucleus (see e.g. Jinek et al., 2012, Science, 337:816), and the preference for an NGG motif immediately downstream of the protospacer restricts the ability to target on average every 12-bp in the human genome (FIG. **9**, evaluating both plus and minus strands of human chromosomal sequences). Some of these constraints can be overcome by exploring the diversity of CRISPR loci across the microbial metagenome (see e.g. Makarova et al., 2011, Nat Rev Microbiol, 9:467). Other CRISPR loci may be transplanted into the mammalian cellular milieu by a process similar to that described in Example 1. For example, FIG. **10** illustrates adaptation of the Type II CRISPR system from CRISPR 1 of *Streptococcus thermophilus* LMD-9 for heterologous expression in mammalian cells to achieve CRISPR-mediated genome editing. FIG. **10**A provides a Schematic illustration of CRISPR 1 from *S. thermophilus* LMD-9. FIG. **10**B illustrates the design of an expression system for the *S. thermophilus* CRISPR system. Human codon-optimized hStCas9 is expressed using a constitutive EF1α promoter. Mature versions of tracrRNA and crRNA are expressed using the U6 promoter to promote precise transcription initiation. Sequences from the mature crRNA and tracrRNA are illustrated. A single base indicated by the lower case "a" in the crRNA sequence is used to remove the polyU sequence, which serves as a RNA polIII transcriptional terminator. FIG. **10**C provides a schematic showing guide sequences targeting the human EMX1 locus. FIG. **10**D shows the results of hStCas9-mediated cleavage in the target locus using the Surveyor assay. RNA guide spacers 1 and 2 induced 14% and 6.4%, respectively. Statistical analysis of cleavage activity across biological replica at these two protospacer sites is also provided in FIG. **5**. FIG. **14** provides a schematic of additional protospacer and corresponding PAM sequence targets of the *S. thermophilus* CRISPR system in the human EMX1 locus. Two protospacer sequences are highlighted and their corresponding PAM sequences satisfying NNAGAAW motif are indicated by underlining 3' with respect to the corresponding highlighted sequence. Both protospacers target the anti-sense strand.

Example 3

Sample Target Sequence Selection Algorithm

[0450] A software program is designed to identify candidate CRISPR target sequences on both strands of an input DNA sequence based on desired guide sequence length and a CRISPR motif sequence (PAM) for a specified CRISPR enzyme. For example, target sites for Cas9 from *S. pyogenes*, with PAM sequences NGG, may be identified by searching for 5'-N_x-NGG-3' both on the input sequence and on the reverse-complement of the input. Likewise, target sites for Cas9 of *S. thermophilus* CRISPR1, with PAM sequence NNAGAAW, may be identified by searching for 5'-N_x-NNAGAAW-3' both on the input sequence and on the reverse-complement of the input. Likewise, target sites for Cas9 of *S. thermophilus* CRISPR3, with PAM sequence NGGNG, may be identified by searching for 5'-N_x-NGGNG-3' both on the input sequence and on the reverse-complement of the input. The value "x" in N_x may be fixed by the program or specified by the user, such as 20.

[0451] Since multiple occurrences in the genome of the DNA target site may lead to nonspecific genome editing, after identifying all potential sites, the program filters out sequences based on the number of times they appear in the relevant reference genome. For those CRISPR enzymes for which sequence specificity is determined by a 'seed' sequence, such as the 11-12 bp 5' from the PAM sequence, including the PAM sequence itself, the filtering step may be based on the seed sequence. Thus, to avoid editing at additional genomic loci, results are filtered based on the number of occurrences of the seed:PAM sequence in the relevant genome. The user may be allowed to choose the length of the seed sequence. The user may also be allowed to specify the number of occurrences of the seed:PAM sequence in a genome for purposes of passing the filter. The default is to screen for unique sequences. Filtration level is altered by changing both the length of the seed sequence and the number of occurrences of the sequence in the genome. The program may in addition or alternatively provide the sequence of a guide sequence complementary to the reported target sequence(s) by providing the reverse complement of the identified target sequence(s). An example visualization of some target sites in the human genome is provided in FIG. **18**.

[0452] Further details of methods and algorithms to optimize sequence selection can be found in U.S. application Ser. No. 61/064,798 (Attorney docket 44790.11.2022; Broad Reference BI-2012/084); incorporated herein by reference.

Example 4

Evaluation of Multiple Chimeric crRNA-tracrRNA Hybrids

[0453] This example describes results obtained for chimeric RNAs (chiRNAs; comprising a guide sequence, a tracr mate sequence, and a tracr sequence in a single transcript) having tracr sequences that incorporate different lengths of wild-type tracrRNA sequence. FIG. **16***a* illustrates a schematic of a bicistronic expression vector for chimeric RNA and Cas9. Cas9 is driven by the CBh promoter and the chimeric RNA is driven by a U6 promoter. The chimeric guide RNA consists of a 20 bp guide sequence (Ns) joined to the tracr sequence (running from the first "U" of the lower strand to the end of the transcript), which is truncated at various positions as indicated. The guide and tracr sequences are separated by the tracr-mate sequence GUUUUAGAGCUA followed by the loop sequence GAAA. Results of SURVEYOR assays for Cas9-mediated indels at the human EMX1 and PVALB loci are illustrated in FIGS. **16***b* and **16***c*, respectively. Arrows indicate the expected SURVEYOR fragments. ChiRNAs are indicated by their "+n" designation, and crRNA refers to a hybrid RNA where guide and tracr sequences are expressed as separate transcripts. Quantification of these results, performed in triplicate, are illustrated by histogram in FIGS. **17***a* and **17***b*, corresponding to FIGS. **16***b* and **16***c*, respectively ("N.D." indicates no indels detected). Protospacer IDs and their corresponding genomic target, protospacer sequence, PAM sequence, and strand location are provided in Table D. Guide sequences were designed to be complementary to the entire protospacer sequence in the case of separate transcripts in the hybrid system, or only to the underlined portion in the case of chimeric RNAs.

TABLE D

protospacer ID	genomic target	protospacer sequence (5' to 3')	PAM	strand
1	EMX1	GGACATCGATGTCACCTCCAATGACTAGCG	TGG	+
2	EMX1	CATTGGAGGTGACATCGATGTCCTCCCCAT	TGG	–
3	EMX1	GGAAGGGCCTGAGTCCGAGCAGAAGAAGAA	GGG	+
4	PVALB	GGTGGCGAGAGGGGCCGAGATTGGGTGTTC	AGG	+
5	PVALB	ATGCAGGAGGGTGGCGAGAGGGGCCGAGAT	TGG	+

[0454] Further details to optimize guide sequences can be found in U.S. application Ser. No. 61/836,127 (Attorney docket 44790.08.2022; Broad Reference BI-2013/004G); incorporated herein by reference.

[0455] Initially, three sites within the EMX1 locus in human HEK 293FT cells were targeted. Genome modification efficiency of each chiRNA was assessed using the SURVEYOR nuclease assay, which detects mutations resulting from DNA double-strand breaks (DSBs) and their subsequent repair by the non-homologous end joining (NHEJ) DNA damage repair pathway. Constructs designated chiRNA(+n) indicate that up to the +n nucleotide of wild-type tracrRNA is included in the chimeric RNA construct, with values of 48, 54, 67, and 85 used for n. Chimeric RNAs containing longer fragments of wild-type tracrRNA (chiRNA(+67) and chiRNA(+85)) mediated DNA cleavage at all three EMX1 target sites, with chiRNA(+85) in particular demonstrating significantly higher levels of DNA cleavage than the corresponding crRNA/tracrRNA hybrids that expressed guide and tracr sequences in separate transcripts (FIGS. **16***b* and **17***a*). Two sites in the PVALB locus that yielded no detectable cleavage using the hybrid system (guide sequence and tracr sequence expressed as separate transcripts) were also targeted using chiRNAs. chiRNA(+67) and chiRNA(+85) were able to mediate significant cleavage at the two PVALB protospacers (FIGS. **16***c* and **17***b*).

[0456] For all five targets in the EMX1 and PVALB loci, a consistent increase in genome modification efficiency with increasing tracr sequence length was observed. Without wishing to be bound by any theory, the secondary structure formed by the 3' end of the tracrRNA may play a role in enhancing the rate of CRISPR complex formation.

Example 5

Cas9 Diversity

[0457] The CRISPR-Cas system is an adaptive immune mechanism against invading exogenous DNA employed by diverse species across bacteria and archaea. The type II CRISPR-Cas9 system consists of a set of genes encoding proteins responsible for the "acquisition" of foreign DNA into the CRISPR locus, as well as a set of genes encoding the "execution" of the DNA cleavage mechanism; these include the DNA nuclease (Cas9), a non-coding transactivating cr-RNA (tracrRNA), and an array of foreign DNA-derived spacers flanked by direct repeats (crRNAs). Upon maturation by Cas9, the tracRNA and crRNA duplex guide the Cas9 nuclease to a target DNA sequence specified by the spacer guide sequences, and mediates double-stranded breaks in the DNA near a short sequence motif in the target DNA that is required for cleavage and specific to each CRISPR-Cas system. The type II CRISPR-Cas systems are found throughout the bacterial kingdom and highly diverse in in Cas9 protein sequence and size, tracrRNA and crRNA direct repeat sequence, genome organization of these elements, and the motif requirement for target cleavage. One species may have multiple distinct CRISPR-Cas systems.

[0458] Applicants evaluated 207 putative Cas9s from bacterial species identified based on sequence homology to known Cas9s and structures orthologous to known subdomains, including the HNH endonuclease domain and the RuvC endonuclease domains [information from the Eugene Koonin and Kira Makarova]. Phylogenetic analysis based on the protein sequence conservation of this set revealed five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids) (see FIGS. **19** and **20**A-F).

[0459] Further details of Cas9s and mutations of the Cas9 enzyme to convert into a nickase or DNA binding protein and use of same with altered functionality can be found in U.S. application Ser. Nos. 61/836,101 and 61/835,936 (Attorney docket 44790.09.2022 and 4790.07.2022 and Broad Reference BI-2013/004E and BI-2013/004F respectively) incorporated herein by reference.

Example 6

Cas9 Orthologs

[0460] Applicants analyzed Cas9 orthologs to identify the relevant PAM sequences and the corresponding chimeric guide RNA. Having an expanded set of PAMs provides broader targeting across the genome and also significantly increases the number of unique target sites and provides potential for identifying novel Cas9s with increased levels of specificity in the genome.

[0461] The specificity of Cas9 orthologs can be evaluated by testing the ability of each Cas9 to tolerate mismatches between the guide RNA and its DNA target. For example, the specificity of SpCas9 has been characterized by testing the effect of mutations in the guide RNA on cleavage efficiency. Libraries of guide RNAs were made with single or multiple mismatches between the guide sequence and the target DNA. Based on these findings, target sites for SpCas9 can be selected based on the following guidelines:

[0462] To maximize SpCas9 specificity for editing a particular gene, one should choose a target site within the locus of interest such that potential 'off-target' genomic sequences abide by the following four constraints: First and foremost, they should not be followed by a PAM with either 5'-NGG or NAG sequences. Second, their global sequence similarity to the target sequence should be minimized. Third, a maximal number of mismatches should lie within the PAM-proximal region of the off-target site. Finally, a maximal number of mismatches should be consecutive or spaced less than four bases apart.

[0463] Similar methods can be used to evaluate the specificity of other Cas9 orthologs and to establish criteria for the selection of specific target sites within the genomes of target species. As mentioned previously phylogenetic analysis based on the protein sequence conservation of this set revealed five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids) (see FIGS. **19** and **20**A-F). Further details on Cas orthologs can be found in U.S. application Ser. Nos. 61/836,101 and 61/835,936 (Attorney docket 44790.09.2022 and 4790.07.2022 and Broad Reference BI-2013/004E and BI-2013/004F respectively) incorporated herein by reference.

Example 7

Methodological Improvement to Simplify Cloning and Delivery

[0464] Rather than encoding the U6-promoter and guide RNA on a plasmid, Applicants amplified the U6 promoter with a DNA oligo to add on the guide RNA. The resulting PCR product may be transfected into cells to drive expression of the guide RNA.

[0465] Example primer pair that allows the generation a PCR product consisting of U6-promoter::guideRNA targeting human Emx1 locus:

[0466] Forward Primer: AAACTCTAGAgagggcctatttcccatgattc

[0467] Reverse Primer (carrying the guide RNA, which is underlined):

acctctagAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTT
GATAACGGACTAGCATTTTAACTTGCTATGCCTTTGTTTTGTTT
CCAAAACAGCATAGCTCTAAAACCCCTAGTCATTGGAGGTGACG
GTGTTTCGTCCTTTCCACaag

Example 8

Methodological Improvement to Improve Activity

[0468] Rather than use pol3 promoters, in particular RNA polymerase III (e.g. U6 or H1 promoters), to express guide RNAs in eukaryotic cells, Applicants express the T7 polymerase in eukaryotic cells to drive expression of guide RNAs using the T7 promoter.

[0469] One example of this system may involve introduction of three pieces of DNA:

[0470] 1. expression vector for Cas9

[0471] 2. expression vector for T7 polymerase

[0472] 3. expression vector containing guideRNA fused to the T7 promoter

Example 9

Methodological Improvement to Reduce Toxicity of Cas9: Delivery of Cas9 in the Form of mRNA

[0473] Delivery of Cas9 in the form of mRNA enables transient expression of Cas9 in cells, to reduce toxicity. For example, humanized SpCas9 may be amplified using the following primer pair:

[0474] Forward Primer (to add on T7 promoter for in vitro transcription):

```
TAATACGACTCACTATAGGAAGTGCGCCACCATGGC

CCCAAAGAAGAAGCGG
```

[0475] Reverse Primer (to add on polyA tail):

```
GGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTttcttaCT

TTTTCTTTTTTTGCCTGGCCG
```

[0476] Applicants transfect the Cas9 mRNA into cells with either guide RNA in the form of RNA or DNA cassettes to drive guide RNA expression in eukaryotic cells.

Example 10

Methodological Improvement to Reduce Toxicity of Cas9: Use of an Inducible Promoter

[0477] Applicants transiently turn on Cas9 expression only when it is needed for carrying out genome modification. Examples of inducible system include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc), or light inducible systems (Phytochrome, LOV domains, or cryptochrome).

Example 11

Improvement of the Cas9 System for In Vivo Application

[0478] Applicants conducted a Metagenomic search for a Cas9 with small molecular weight. Most Cas9 homologs are fairly large. For example the SpCas9 is around 1368aa long, which is too large to be easily packaged into viral vectors for delivery. A graph representing the length distribution of Cas9 homologs is generated from sequences deposited in GenBank (FIG. **23**). Some of the sequences may have been mis-annotated and therefore the exact frequency for each length may not necessarily be accurate. Nevertheless it provides a glimpse at distribution of Cas9 proteins and suggest that there are shorter Cas9 homologs.

[0479] Through computational analysis, Applicants found that in the bacterial strain *Campylobacter*, there are two Cas9 proteins with less than 1000 amino acids. The sequence for one Cas9 from *Campylobacter jejuni* is presented below. At this length, CjCas9 can be easily packaged into AAV, lentiviruses, Adenoviruses, and other viral vectors for robust delivery into primary cells and in vivo in animal models. In a preferred embodiment of the invention, the Cas9 protein from *S. aureus* is used.

```
>Campylobacter jejuni Cas9 (CjCas9)
MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGES

LALPRRLARSARKRLARRKARLNHLKHLIANEFKLNYEDYQSF

DESLAKAYKGSLISPYELRFRALNELLSKQDFARVILHIAKRR

GYDDIKNSDDKEKGMLKAIKQNEEKLANYQSVGEYLYKEYFQK

FKENSKEFTNVRNKKESYERCIAQSFLKDELKIAFKKQREFGF

SFSKKFEEEVLSVAFYKRALKDFSHINGNCSFFTDEKRAPKNS

PLAFMFVALTRIINLINNLKNTEGILYTKDDLNALLNEVLKNG

TUFYKQTKKLLGLSDDYEFKGEKGTYHEFKKYKEFIKALGEHN

LSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLE

FKUHLNISFKALKLNTPLMLEGKKYDEACNELNLKVAIINEDK

KDFLPAFNETYYKDEVTNPVVLRAIIKEYRKVLNALLKKYGKV

MINIEIAREVGKNFISQRAKIEKEQNENYKAKKDAELECEKLG

LKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQDEKMLEIDH

IYPYSRSFDDSYMNKVINFTKQNQEKLNQTPFEAFGNDSAKWQ

KIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIAR

LVLNYTKDYLDFITLSDDENTKLNDTQKGSKVHVEAKSCIMLT

SALRHTWGFSAICDRNNHLHHAIDAVIIAYANNSIVKAFSDFK

KEQESNSAELYAKKISTLDYKNKRKFTEPFSGERQKVIDKIDE

IFVSKPERKKPSCIALFIEETFRKEEEFYQSYGGKEGVLKALE

LGKIRKVNGKIVKNGDMFRVDIFKEIKKTNKFYAVPIYTMDFA

LKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQT

KDMQEPEFVYYNAFTSSTVSLIVSKIADNKFETLSKNQKILFK

NANEKEVIAKSIGIQNLKVFEKYWSALGEVTKAEFRQREDFKK
```

[0480] The putative tracrRNA element for this CjCas9 is:

```
TATAATCTCATAAGAAATTTAAAAAGGGACTAAAATAAAGAGTTTG

CGGGACTCTGCGGGGTTACAATCCCCTAAAACCGCTTTTAAAATT
```

[0481] The Direct Repeat sequence is:

```
ATTTTACCATAAAGAAATTTAAAAAGGGACTAAAAC
```

[0482] An example of a chimeric guideRNA for CjCas9 is:

```
NNNNNNNNNNNNNNNNNNNNGUUUUAGUCCCGAAAGGGACUAAAAU

AAAGAGUUUGCGGGACUCUGCGGGGUUACAAUCCCCUAAAACCGCU

UUU
```

Example 12

Cas9 Optimization

[0483] For enhanced function or to develop new functions, Applicants generate chimeric Cas9 proteins by combining fragments from different Cas9 homologs. For example, two example chimeric Cas9 proteins:

[0484] For example, Applicants fused the N-term of St1Cas9 (fragment from this protein is in bold) with C-term of SpCas9 (fragment from this protein is underlined).

>St1(N)Sp(C)Cas9

MSDLVLGLDIGIGSVGVGIILNKVTGEHHKNSRIFPAAQAENNLVRRTN

RQGRRLARRKKHRRVRLNRLFEESGLITDVIKISINLNPYQLRVKGLID

ELSNEELHALKNININKTIRGISYLDDASDDGNSSVGDYAQIWKENSKQ

LETKTPGIMLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALR

ILQTQQEFNPQITDEHNRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSG

ETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETK

KLSKEQKNQIINYVKNEKAMGPAKLFKYIAKILLSCDVADIKGYRIDKS

GKAMITFEAYRKMKTILETLDIEQMDRETIDKLAYNTLTLNTEREGIQE

ALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMEIWELY

ETSKEQMTELTRILGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRO

AIKIVINAAIKEYGDFDNIVIEMARENQTTQKGOKNSRERNIKRIEEGI

NKELGSQILKEHPVETQLIONEKINLYYLQNGRDMYVDQELDINRISDY

DVDIIIVPQSFIXDDSIDNKVILTRSDKNRGKSDNVPSEEVVKKMKNYW

RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKEIV

AQIIDSMNTKIDENDKLIREVKVITIKSKINSDRKDIFQFYKYREINNY

HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG

KATAKYFFYSNIMNFFICTEITLANGEIRKRPLIETNGETGEIVWDKGR

DFATVRKVLSMPQVMVKKTEVQTGGFSKESILPKRINSDKIIARKKDWD

PKKYGGFDSPTVAYSVINVAKVEKGKSICKLKSVKELLGITIMERSSFE

KNPIDFLEAKGYKEVKKDLIIKLPKYSIFELENGRKRMILASAGELQKG

NEALPSKYVNFLYLASIIYEKLKGSPEDNEQKQLEVEQHKHYLDEIIEQ

ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP

AAFKYFDTTIDRKRYTSTKEVLDATLIFIQSITGLYETRIDLSQLGGD

>Sp(N)St1(C)Cas9

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKITIQINQTYNQUEENPI

NASGVDAKAIISARISKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPN

FKSNFDLAEDAKLOLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL

LSDILRVNTEETKAPLSASMIKRYDEHFIQDLTLLKALVRQQLPEKYKEI

FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

-continued

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

NLPNEKVLPKHSLLYEYFTVYNELTKYKYVTEGMRKRAFLSGFKKAIVDL

IFKTNRKVINKKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKD

KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR

RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQUHDDSEKE

DIQKAQVSGQGDSUIEFTIANLAGSPAIKKGILQTVKVVDEINKVMGRHK

PENIVIEMARE**TNEDDEKKMQMQKANKDEKDAAMLKAANQYNGKAELPHS**

VFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLS

ITFDDSLANKVINYATANQEKGQRTPWALDSMDDAWSFRELKAFVRESKT

LSNKKKENLLTEEDISKHARKKFIERNLVDTRYASIZVVLNALQERFRAR

KIDTKVSVVRGQFTSQLRRHWGIEKTRIYIYHHHAVDALIINASSQLNLW

KKQKNTLVSYSEDQLLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKE

FEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYYLGKIKDIY

TQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEFILENYPNKQINEKGK

EVPCNPFLKYKEEHIGYIRKYSKKGNGPEIKSLKYYDSKLGNIIIDFIPI

KDSNNKVAILQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTCTYKI

SQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTM

PKQKHYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLCKSNISIYIK

VRTDVLGNQIIIIKNEGDKPKLDF

[0485] The benefit of making chimeric Cas9 include:

[0486] reduce toxicity

[0487] improve expression in eukaryotic cells

[0488] enhance specificity

[0489] reduce molecular weight of protein, make protein smaller by combining the smallest domains from different Cas9 homologs.

[0490] Altering the PAM sequence requirement

Example 13

Utilization of Cas9 as a Generic DNA Binding Protein

[0491] Applicants used Cas9 as a generic DNA binding protein by mutating the two catalytic domains (D10 and H840) responsible for cleaving both strands of the DNA target. In order to upregulate gene transcription at a target locus Applicants fused the transcriptional activation domain (VP64) to Cas9. Applicants hypothesized that it would be important to see strong nuclear localization of the Cas9-VP64 fusion protein because transcription factor activation strength is a function of time spent at the target. Therefore, Applicants cloned a set of Cas9-VP64-GFP constructs, transfected them into 293 cells and assessed their localization under a fluorescent microscope 12 hours post-transfection.

[0492] The same constructs were cloned as a 2A-GFP rather than a direct fusion in order to functionally test the constructs without a bulky GFP present to interfere. Applicants elected to target the Sox2 locus with the Cas9 transactivator because it could be useful for cellular reprogram and the locus has already been validated as a target for TALE-TF mediated transcriptional activation. For the Sox2 locus Applicants chose eight targets near the transcriptional start site (TSS). Each target was 20 bp long with a neighboring NGG protospacer adjacent motif (PAM). Each Cas9-VP64 construct was co-transfected with each PCR generated chimeric crispr RNA (chiRNA) in 293 cells. 72 hours post transfection the transcriptional activation was assessed using RT-qPCR.

[0493] To further optimize the transcriptional activator, Applicants titrated the ratio of chiRNA (Sox2.1 and Sox2.5) to Cas9 (NLS-VP64-NLS-hSpCas9-NLS-VP64-NLS), transfected into 293 cells, and quantified using RT-qPCR. These results indicate that Cas9 can be used as a generic DNA binding domain to upregulate gene transcription at a target locus.

[0494] Applicants designed a second generation of constructs. (Table below).

pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-hSpCsn1(D10A, H840A)-NLS
pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-hSpCsn1(D10A, H840A)
pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-NLS-hSpCsn1(D10A, H840A)
pLenti-EF1a-GFP-2A-6xHis-NLS-hSpCsn1(D10A, H840A)-NLS
pLenti-EF1a-GFP-2A-6xHis-NLS-hSpCsn1(D10A, H840A)
pLenti-EF1a-GFP-2A-6xHis-NLS-NLS-hSpCsn1(D10A, H840A)

[0495] Applicants use these constructs to assess transcriptional activation (VP64 fused constructs) and repression (Cas9 only) by RT-qPCR. Applicants assess the cellular localization of each construct using anti-His antibody, nuclease activity using a Surveyor nuclease assay, and DNA binding affinity using a gel shift assay. In a preferred embodiment of the invention, the gel shift assay is an EMSA gel shift assay.

Example 14

Cas9 Transgenic and Knock in Mice

[0496] To generate a mouse that expresses the Cas9 nuclease Applicants submit two general strategies, transgenic and knock in. These strategies may be applied to generate any other model organism of interest, for e.g. Rat. For each of the general strategies Applicants made a constitutively active Cas9 and a Cas9 that is conditionally expressed (Cre recombinase dependent). The constitutively active Cas9 nuclease is expressed in the following context: pCAG-NLS-Cas9-NLS-P2A-EGFP-WPRE-bGHpolyA. pCAG is the promoter, NLS is a nuclear localization signal, P2A is the peptide cleavage sequence, EGFP is enhanced green fluorescent protein, WPRE is the woodchuck hepatitis virus posttranscriptional regulatory element, and bGHpolyA is the bovine growth hormone poly-A signal sequence (FIGS. **25**A-B). The conditional version has one additional stop cassette element, loxP-SV40 polyA x3-loxP, after the promoter and before NLS-Cas9-NLS (i.e. pCAG-loxP-SV40polyAx3-loxP-NLS-Cas9-NLS-P2A-EGFP-WPRE-bGHpolyA). The important expression elements can be visualized as in FIG. **26**. The constitutive construct should be expressed in all cell types throughout development, whereas, the conditional construct will only allow Cas9 expression when the same cell is expressing the Cre recombinase. This latter version will allow for tissue specific expression of Cas9 when Cre is under the expression of a tissue specific promoter. Moreover, Cas9 expression could be induced in adult mice by putting Cre under the expression of an inducible promoter such as the TET on or off system.

[0497] Validation of Cas9 constructs: Each plasmid was functionally validated in three ways: 1) transient transfection in 293 cells followed by confirmation of GFP expression; 2) transient transfection in 293 cells followed by immunofluorescence using an antibody recognizing the P2A sequence; and 3) transient transfection followed by Surveyor nuclease assay. The 293 cells may be 293FT or 293 T cells depending on the cells that are of interest. In a preferred embodiment the cells are 293FT cells. The results of the Surveyor were run out on the top and bottom row of the gel for the conditional and constitutive constructs, respectively. Each was tested in the presence and absence of chimeric RNA targeted to the hEMX1 locus (chimeric RNA hEMX1.1). The results indicate that the construct can successfully target the hEMX1 locus only in the presence of chimeric RNA (and Cre in the conditional case). The gel was quantified and the results are presented as average cutting efficiency and standard deviation for three samples.

[0498] Transgenic Cas9 mouse: To generate transgenic mice with constructs, Applicants inject pure, linear DNA into the pronucleus of a zygote from a pseudo pregnant CB56 female. Founders are identified, genotyped, and backcrossed to CB57 mice. The constructs were successfully cloned and verified by Sanger sequencing.

[0499] Knock in Cas9 mouse: To generate Cas9 knock in mice Applicants target the same constitutive and conditional constructs to the Rosa26 locus. Applicants did this by cloning each into a Rosa26 targeting vector with the following elements: Rosa26 short homology arm—constitutive/conditional Cas9 expression cassette—pPGK-Neo-Rosa26 long homology arm—pPGK-DTA. pPGK is the promoter for the positive selection marker Neo, which confers resistance to neomycin, a 1 kb short arm, a 4.3 kb long arm, and a negative selection diphtheria toxin (DTA) driven by PGK.

[0500] The two constructs were electroporated into R1 mESCs and allowed to grow for 2 days before neomycin selection was applied. Individual colonies that had survived by days 5-7 were picked and grown in individual wells. 5-7 days later the colonies were harvested, half were frozen and the other half were used for genotyping. Genotyping was done by genomic PCR, where one primer annealed within the donor plasmid (AttpF) and the other outside of the short homology arm (Rosa26-R) Of the 22 colonies harvested for the conditional case, 7 were positive (Left). Of the 27 colonies harvested for the constitutive case, zero were positive (Right). It is likely that Cas9 causes some level of toxicity in the mESC and for this reason there were no positive clones. To test this Applicants introduced a Cre expression plasmid into correctly targeted conditional Cas9 cells and found very low toxicity after many days in culture. The reduced copy number of Cas9 in correctly targeted conditional Cas9 cells (1-2 copies per cell) is enough to allow stable expression and relatively no cytotoxicity. Moreover, this data indicates that the Cas9 copy number determines toxicity. After electroporation each cell should get several copies of Cas9 and this is likely why no positive colonies were found in the case of the constitutive Cas9 construct. This provides strong evidence that utilizing a conditional, Cre-dependent strategy should show reduced toxicity. Applicants inject correctly targeted cells into a blastocyst and implant into a female mouse. Chimerics are identified and backcrossed. Founders are identified and genotyped.

[0501] Utility of the conditional Cas9 mouse: Applicants have shown in 293 cells that the Cas9 conditional expression construct can be activated by co-expression with Cre. Applicants also show that the correctly targeted R1 mESCs can have active Cas9 when Cre is expressed. Because Cas9 is followed by the P2A peptide cleavage sequence and then EGFP Applicants identify successful expression by observing EGFP. This same concept is what makes the conditional Cas9 mouse so useful. Applicants may cross their conditional Cas9 mouse with a mouse that ubiquitously expresses Cre (ACTB-Cre line) and may arrive at a mouse that expresses Cas9 in every cell. It should only take the delivery of chimeric RNA to induce genome editing in embryonic or adult mice. Interestingly, if the conditional Cas9 mouse is crossed with a mouse expressing Cre under a tissue specific promoter, there should only be Cas9 in the tissues that also express Cre. This approach may be used to edit the genome in only precise tissues by delivering chimeric RNA to the same tissue.

Example 15

Cas9 Diversity and Chimeric RNAs

[0502] The CRISPR-Cas system is an adaptive immune mechanism against invading exogenous DNA employed by diverse species across bacteria and archaea. The type II CRISPR-Cas system consists of a set of genes encoding proteins responsible for the "acquisition" of foreign DNA into the CRISPR locus, as well as a set of genes encoding the "execution" of the DNA cleavage mechanism; these include the DNA nuclease (Cas9), a non-coding transactivating crRNA (tracrRNA), and an array of foreign DNA-derived spacers flanked by direct repeats (crRNAs). Upon maturation by Cas9, the tracrRNA and crRNA duplex guide the Cas9 nuclease to a target DNA sequence specified by the spacer guide sequences, and mediates double-stranded breaks in the DNA near a short sequence motif in the target DNA that is required for cleavage and specific to each CRISPR-Cas system. The type II CRISPR-Cas systems are found throughout the bacterial kingdom and highly diverse in in Cas9 protein sequence and size, tracrRNA and crRNA direct repeat sequence, genome organization of these elements, and the motif requirement for target cleavage. One species may have multiple distinct CRISPR-Cas systems.

[0503] Applicants evaluated 207 putative Cas9s from bacterial species identified based on sequence homology to known Cas9s and structures orthologous to known subdomains, including the HNH endonuclease domain and the RuvC endonuclease domains [information from the Eugene Koonin and Kira Makarova]. Phylogenetic analysis based on the protein sequence conservation of this set revealed five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids) (FIGS. **19**A-D and **20**A-F).

[0504] Applicants have also optimized Cas9 guide RNA using in vitro methods.

Example 16

Cas9 Mutations

[0505] In this example, Applicants show that the following mutations can convert SpCas9 into a nicking enzyme: D10A, E762A, H840A, N854A, N863A, D986A.

[0506] Applicants provide sequences showing where the mutation points are located within the SpCas9 gene (FIG. **24**A-M). Applicants also show that the nickases are still able to mediate homologous recombination. Furthermore, Applicants show that SpCas9 with these mutations (individually) do not induce double strand break.

[0507] Cas9 orthologs all share the general organization of 3-4 RuvC domains and a HNH domain. The 5' most RuvC domain cleaves the non-complementary strand, and the HNH domain cleaves the complementary strand. All notations are in reference to the guide sequence.

[0508] The catalytic residue in the 5' RuvC domain is identified through homology comparison of the Cas9 of interest with other Cas9 orthologs (from *S. pyogenes* type II CRISPR locus, *S. thermophilus* CRISPR locus 1, *S. thermophilus* CRISPR locus 3, and *Franciscilla novicida* type II CRISPR locus), and the conserved Asp residue is mutated to alanine to convert Cas9 into a complementary-strand nicking enzyme. Similarly, the conserved His and Asn residues in the HNH domains are mutated to Alanine to convert Cas9 into a non-complementary-strand nicking enzyme.

Example 17

Cas9 Transcriptional Activation and Cas9 Repressor

[0509] Cas9 Transcriptional Activation

[0510] A second generation of constructs were designed and tested (Table 1). These constructs are used to assess transcriptional activation (VP64 fused constructs) and repression (Cas9 only) by RT-qPCR. Applicants assess the cellular localization of each construct using anti-H is antibody, nuclease activity using a Surveyor nuclease assay, and DNA binding affinity using a gel shift assay.

[0511] Cas Repressor

[0512] It has been shown previously that dCas9 can be used as a generic DNA binding domain to repress gene expression. Applicants report an improved dCas9 design as well as dCas9 fusions to the repressor domains KRAB and SID4x. From the plasmid library created for modulating transcription using Cas9 in Table 1, the following repressor plasmids were functionally characterized by qPCR: pXRP27, pXRP28, pXRP29, pXRP48, pXRP49, pXRP50, pXRP51, pXRP52, pXRP53, pXRP56, pXRP58, pXRP59, pXRP61, and pXRP62.

[0513] Each dCas9 repressor plasmid was co-transfected with two guide RNAs targeted to the coding strand of the beta-catenin gene. RNA was isolated 72 hours after transfection and gene expression was quantified by RT-qPCR. The endogenous control gene was GAPDH. Two validated shRNAs were used as positive controls. Negative controls were certain plasmids transfected without gRNA, these are denoted as "pXRP## control". The plasmids pXRP28, pXRP29, pXRP48, and pXRP49 could repress the beta-catenin gene when using the specified targeting strategy. These plasmids correspond to dCas9 without a functional domain (pXRP28 and pXRP28) and dCas9 fused to SID4x (pXRP48 and pXRP49).

[0514] Further work investigates: repeating the above experiment, targeting different genes, utilizing other gRNAs to determine the optimal targeting position, and multiplexed repression.

TABLE 1

pXRP024-pLenti2-EF1a-VP64-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP025-pLenti2-EF1a-VP64-NLS-GGGGS$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP026-pLenti2-EF1a-VP64-NLS-EAAAK$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP027-pLenti2-EF1a-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP028-pLenti2-EF1a-NLS-GGGGS$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP029-pLenti2-EF1a-NLS-EAAAK$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP030-pLenti2-pSV40-VP64-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP031-pLenti2-pPGK-VP64-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP032-pLenti2-LTR-VP64-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP033-pLenti2-pSV40-VP64-NLS-GGGGS$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP034-pLenti2-pPGK-VP64-NLS-GGGGS$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP035-pLenti2-LTR-VP64-NLS-GGGGS$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP036-pLenti2-pSV40-VP64-NLS-EAAAK$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP037-pLenti2-pPGK-VP64-NLS-EAAAK$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP038-pLenti2-LTR-VP64-NLS-EAAAK$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP048-pLenti2-EF1a-SID4x-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP049-pLenti2-EF1a-SID4X-NLS-GGGGS$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP050-pLenti2-EF1a-SID4X-NLS-EAAAK$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP051-pLenti2-EF1a-KRAR-NLS-FLAG-Liniker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP052-pLenti2-EF1a-KRAB-NLS-GGGGS$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP053-pLenti2-EF1a-KRAB-NLS-EAAAK$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP054-pLenti2-EF1a-dCas9-Liniker-FLAG-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP055-pLenti2-EF1a-dCas9-Linker-FLAG-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP056-pLenti2-EF1a-dCas9-Linker-FLAG-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP057-pLenti2-EF1a-dCas9-GGGGGS$_3$-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP058-pLenti2-EF1a-dCas9-GGGGGS$_3$-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP059-pLenti2-EF1a-dCas9-GGGGGS$_3$-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP060-pLenti2-EF1a-dCas9-EAAAK$_3$-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP061-pLenti2-EF1a-dCas9-EAAAK$_3$-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP062-pLenti2-EF1a-dCas9-EAAAK$_3$-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP024-pLenti2-EF1a-VP64-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP025-pLenti2-EF1a-VP64-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP026-pLenti2-EF1a-VP64-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP027-pLenti2-EF1a-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP028-pLenti2-EF1a-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP029-pLenti2-EF1a-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP030-pLenti2-pSV40-VP64-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP031-pLenti2-pPGK-VP64-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP032-pLenti2-ETR-VP64-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP033-pLenti2-pSV40-VP64-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP034-pLenti2-pPGK-VP64-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP035-pLenti2-LTR-VP64-NLS-GUGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP036-pLenti2-pSV40-VP64-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP037-pLenti2-pPGK-VP64-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP038-pLenti2-LTR-VP64-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP048-pLenti2-EF1a-SID4x-NLS-FEAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP049-pLenti2-EF1a-SID4X-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP050-pLenti2-EF1a-SID4X-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP051-pLenti2-EF1a-KRAB-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP052-pLenti2-EF1a-KRAB-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GEP-WPRE
pXRP053-pLenti2-FF1a-KRAB-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP054-pLenti2-EF1a-Cas9-Linker-FLAG-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP055-pLenti2-EF1a-Cas9-Linker-FLAG-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP056-pLenti2-EF1a-Cas9-Linker-FLAG-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP057-pLenti2-EF1a-Cas9-GGGGGS$_3$-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP058-pLenti2-EF1a-Cas9-GGGGGS$_3$-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP059-pLenti2-EF1a-Cas9-GGGGGS$_3$-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP060-pLenti2-FF1a-Cas9-EAAAK$_3$-NES-VP64-gLuc-2A-GFP-WPRE
pXRP061-pLenti2-EF1a-Cas9-EAAAK$_3$-NES-SID4X-gLuc-2A-GFP-WPRE
pXRP062-pLenti2-EF1a-Cas9-EAAAK$_3$-NES-KRAB-gLuc-2A-GFP-WPRE

Example 18

Targeted Deletion of Genes Involved in Cholesterol Biosynthesis, Fatty Acid Biosynthesis, and Other Metabolic Disorders, Genes Encoding Mis-Folded Proteins Involved in Amyloid and Other Diseases, Oncogenes Leading to Cellular Transformation, Latent Viral Genes, and Genes Leading to Dominant-Negative Disorders, Amongst Other Disorders

[0515] Applicants demonstrate gene delivery of a CRISPR-Cas system in the liver, brain, ocular, epithelial, hematopoetic, or another tissue of a subject or a patient in need thereof, suffering from metabolic disorders, amyloidosis and protein-aggregation related diseases, cellular transformation arising from genetic mutations and translocations, dominant negative effects of gene mutations, latent viral infections, and other related symptoms, using either viral or nanoparticle delivery system.

[0516] Study Design:

[0517] Subjects or patients in need thereof suffering from metabolic disorders, amyloidosis and protein aggregation related disease which include but are not limited to human, non-primate human, canine, feline, bovine, equine, other domestic animals and related mammals. The CRISPR-Cas system is guided by a chimeric guide RNA and targets a specific site of the human genomic loci to be cleaved. After cleavage and non-homologous end-joining mediated repair, frame-shift mutation results in knock out of genes.

[0518] Applicants select guide-RNAs targeting genes involved in above-mentioned disorders to be specific to endogenous loci with minimal off-target activity. Two or more guide RNAs may be encoded into a single CRISPR array to induce simultaneous double-stranded breaks in DNA leading to micro-deletions of affected genes or chromosomal regions.

[0519] Identification and Design of Gene Targets

[0520] For each candidate disease gene, Applicants select DNA sequences of interest include protein-coding exons, sequences including and flanking known dominant negative mutation sites, sequences including and flanking pathological repetitive sequences. For gene-knockout approaches, early coding exons closest to the start codon offer best options for achieving complete knockout and minimize possibility of truncated protein products retaining partial function.

[0521] Applicants analyze sequences of interest for all possible targetable 20-bp sequences immediately 5' to a NGG motif (for SpCas9 system) or a NNAGAAW (for St1Cas9 system). Applicants choose sequences for unique, single RNA-guided Cas9 recognition in the genome to minimize off-target effects based on computational algorithm to determine specificity.

[0522] Cloning of Guide Sequences into a Delivery System

[0523] Guide sequences are synthesized as double-stranded 20-24 bp oligonucleotides. After 5'-phosphorylation treatment of oligos and annealing to form duplexes, oligos are ligated into suitable vector depending on the delivery method:

[0524] Virus-Based Delivery Methods

[0525] AAV-based vectors (PX260, 330, 334, 335) have been described elsewhere

[0526] Lentiviral-based vectors use a similar cloning strategy of directly ligating guide sequences into a single vector carrying a U6 promoter-driven chimeric RNA scaffold and a EF1a promoter-driven Cas9 or Cas9 nickase.

[0527] Virus production is described elsewhere.

[0528] Nanoparticle-Based RNA Delivery Methods

[0529] 1. Guide sequences are synthesized as an oligonucleotide duplex encoding T7 promoter-guide sequence-chimeric RNA. A T7 promoter is added 5' of Cas9 by PCR method.

[0530] 2. T7-driven Cas9 and guide-chimeric RNAs are transcribed in vitro, and Cas9 mRNA is further capped and A-tailed using commercial kits. RNA products are purified per kit instructions.

[0531] Hydrodynamic Tail Vein Delivery Methods (for Mouse)

[0532] Guide sequences are cloned into AAV plasmids as described above and elsewhere in this application.

[0533] In Vitro Validation on Cell Lines

[0534] Transfection

[0535] 1. DNA Plasmid Transfection

[0536] Plasmids carrying guide sequences are transfected into human embryonic kidney (HEK293T) or human embryonic stem (hES) cells, other relevant cell types using lipid-, chemical-, or electroporation-based methods. For a 24-well transfection of HEK293T cells (~260,000 cells), 500 ng of total DNA is transfected into each single well using Lipofectamine 2000. For a 12-well transfection of hES cells, 1 ug of total DNA is transfected into a single well using Fugene HD.

[0537] 2. RNA Transfection

[0538] Purified RNA described above is used for transfection into HEK293T cells. 1-2 ug of RNA may be transfected into ~260,000 using Lipofectamine 2000 per manufacturer's instruction. RNA delivery of Cas9 and chimeric RNA is shown in FIG. **28**.

[0539] Assay of Indel Formation In Vitro

[0540] Cells are harvested 72-hours post-transfection and assayed for indel formation as an indication of double-stranded breaks.

[0541] Briefly, genomic region around target sequence is PCR amplified (~400-600 bp amplicon size) using high-fidelity polymerase. Products are purified, normalized to equal concentration, and slowly annealed from 95° C. to 4° C. to allow formation of DNA heteroduplexes. Post annealing, the Cel-I enzyme is used to cleave heteroduplexes, and resulting products are separated on a polyacrylamide gel and indel efficiency calculated.

[0542] In Vivo Proof of Principle in Animal

[0543] Delivery Mechanisms

[0544] AAV or Lentivirus production is described elsewhere.

[0545] Nanoparticle Formulation: RNA Mixed into Nanoparticle Formulation

[0546] Hydrodynamic tail vein injections with DNA plasmids in mice are conducted using a commercial kit

[0547] Cas9 and guide sequences are delivered as virus, nanoparticle-coated RNA mixture, or DNA plasmids, and injected into subject animals. A parallel set of control animals is injected with sterile saline, Cas9 and GFP, or guide sequence and GFP alone.

[0548] Three weeks after injection, animals are tested for amelioration of symptoms and sacrificed. Relevant organ systems analyzed for indel formation. Phenotypic assays include blood levels of HDL, LDL, lipids,

[0549] Assay for Indel Formation

[0550] DNA is extracted from tissue using commercial kits; indel assay will be performed as described for in vitro demonstration.

[0551] Therapeutic applications of the CRISPR-Cas system are amenable for achieving tissue-specific and temporally controlled targeted deletion of candidate disease genes. Examples include genes involved in cholesterol and fatty acid metabolism, amyloid diseases, dominant negative diseases, latent viral infections, among other disorders.

[0552] Examples of a single guide-RNA to introduce targeted indels at a gene locus

Disease	GENE	SPACER	PAM	Mechanism	References
Hypercholesterolemia	HMG-CR	GCCAAATTGGACGACCCTCG	CGG	Knockout	Fluvastatin: a review of its pharmacology and use in the management of hypercholesterolaemia.(Plosker

-continued

Disease	GENE	SPACER	PAM	Mechanism	References
					GL et al. Drugs 1996, 51(3): 433-459)
Hypercholesterolemia	SQLE	CGAGGAGAC CCCCGTTTC GG	TGG	Knockout	Potential role of nonstatin cholesterol lowering agents (Trapani et al. IUBMB Life, Volume 63, Issue 11, pages 964-971, November 2011)
Hyperlipidemia	DGAT1	CCCGCCGCC GCCGTGGCT CG	AGG	Knockout	DGAT1 inhibitors as anti-obesity and anti-diabetic agents. (Birch AM et al. Current Opinion in Drug Discovery & Development
Leukemia	BCR-ABL	TGAGCTCTA CGAGATCCA CA	AGG	Knockout	Killing of leukemic cells with a BCR/ABL fusion gene by RNA interference (RNAi.).(Fuchs et al. Oncogene 2002, 21(37): 5716-5724)

[0553] Examples of a pair of guide-RNA to introduce chromosomal microdeletion at a gene locus

Disease	GENE	SPACER	PAM	Mechanism	References
Hyperlipidemia	PLIN2 guide1	CTCAAAATT CATACCGGT TG	TGG	Micro-deletion	Perilipin-2 Null Mice are Protected Against Diet-Induced Obesity, Adipose Inflammation and Fatty Liver Disease (McManaman JL et al. The Journal of Lipid Research, jlr.M035063. First Published on Feb. 12, 2013)
Hyperlipidemia	PLIN2 guide2	CGTTAAACA ACAACCGGA CT	TGG	Micro-deletion	
Hyperlipidemia	SREBP guide1	TTCACCCCCG CGGCGCTGA AT	ggg	Micro-deletion	Inhibition of SREBP by a Small Molecule, Betulin, Improves Hyperlipidemia and Insulin Resistance and Reduces Atherosclerotic Plaques (Tang J et al. Cell Metabolism, Volume 13, issue 1, 44-56, 5 Jan. 2011)
Hyperlipidemia	SREBP guide2	ACCACTACC AGTCCGTCC AC	agg	Micro-deletion	

Example 19

Targeted Integration of Repair for Genes Carrying Disease-Causing Mutations; Reconstitution of Enzyme Deficiencies and Other Related Diseases

[0554] Study Design

[0555] I. Identification and Design of Gene Targets

[0556] Described in Example 22

[0557] II. Cloning of Guide Sequences and Repair Templates into a Delivery System

[0558] Described above in Example 22

[0559] Applicants clone DNA repair templates to include homology arms with diseased allele as well a wild-type repair template

[0560] III. In Vitro Validation on Cell Lines

[0561] a. Transfection is described above in Example 22; Cas9, guide RNAs, and repair template are co-transfected into relevant cell types.

[0562] b. Assay for repair in vitro

[0563] i. Applicants harvest cells 72-hours post-transfection and assay for repair

[0564] ii. Briefly, Applicants amplify genomic region around repair template PCR using high-fidelity polymerase. Applicants sequence products for decreased incidence of mutant allele.

[0565] IV. In Vivo Proof of Principle in Animal

[0566] a. Delivery mechanisms are described above Examples 22 and 34.

[0567] b. Assay for repair in vivo

[0568] i. Applicants perform the repair assay as described in the in vitro demonstration.

[0569] V. Therapeutic Applications

[0570] The CRISPR-Cas system is amenable for achieving tissue-specific and temporally controlled targeted deletion of candidate disease genes. Examples include genes involved in cholesterol and fatty acid metabolism, amyloid diseases, dominant negative diseases, latent viral infections, among other disorders.

[0571] Example of one single missense mutation with repair template:

Disease	GENE	SPACER	PAM	Mechanism	Reference
Familial amyloid polyneuropathy	TTR	AGCCTTTCTGA ACACATGCA	CGG	V30M repair	Transthyretin mutations in health and disease (Joao et al. Human Mutation, Volume 5, Issue 3, pages 191-196, 1995)
				V30M	CCTGCCATCAATGTGGCCATGCATGTGTTCAGAAAGGCT
				allele WT allele	CCTGCCATCAATGTGGCCGTGCATGTGTTCAGAAAGGCT

Example 20

Therapeutic Application of the CRISPR-Cas System in Glaucoma, Amyloidosis, and Huntington's Disease

[0572] Glaucoma: Applicants design guide RNAs to target the first exon of the mycilin (MYOC) gene. Applicants use adenovirus vectors (Ad5) to package both Cas9 as well as a guide RNA targeting the MYOC gene. Applicants inject adenoviral vectors into the trabecular meshwork where cells have been implicated in the pathophysiology of glaucoma. Applicants initially test this out in mouse models carrying the mutated MYOC gene to see whether they improve visual acuity and decrease pressure in the eyes. Therapeutic application in humans employ a similar strategy.

[0573] Amyloidosis: Applicants design guide RNAs to target the first exon of the transthyretin (TTR) gene in the liver. Applicants use AAV8 to package Cas9 as well as guide RNA targeting the first exon of the TTR gene. AAV8 has been shown to have efficient targeting of the liver and will be administered intravenously. Cas9 can be driven either using liver specific promoters such as the albumin promoter, or using a constitutive promoter. A pol3 promoter drives the guide RNA.

[0574] Alternatively, Applicants utilize hydrodynamic delivery of plasmid DNA to knockout the TTR gene. Applicants deliver a plasmid encoding Cas9 and the guideRNA targeting Exon1 of TTR.

[0575] As a further alternative approach, Applicants administer a combination of RNA (mRNA for Cas9, and guide RNA). RNA can be packaged using liposomes such as Invivofectamine from Life Technologies and delivered intravenously. To reduce RNA-induced immunogenicity, increase the level of Cas9 expression and guide RNA stability, Applicants modify the Cas9 mRNA using 5' capping. Applicants also incorporate modified RNA nucleotides into Cas9 mRNA and guide RNA to increase their stability and reduce immunogenicity (e.g. activation of TLR). To increase efficiency, Applicants administer multiple doses of the virus, DNA, or RNA.

[0576] Huntington's Disease: Applicants design guide RNA based on allele specific mutations in the HTT gene of patients. For example, in a patient who is heterozygous for HTT with expanded CAG repeat, Applicants identify nucleotide sequences unique to the mutant HTT allele and use it to design guideRNA. Applicants ensure that the mutant base is located within the last 9 bp of the guide RNA (which Applicants have ascertained has the ability to discriminate between single DNA base mismatches between the target size and the guide RNA).

[0577] Applicants package the mutant HTT allele specific guide RNA and Cas9 into AAV9 and deliver into the striatum of Huntington's patients. Virus is injected into the striatum stereotactically via a craniotomy. AAV9 is known to transduce neurons efficiently. Applicants drive Cas9 using a neuron specific promoter such as human Synapsin 1.

Example 21

Therapeutic Application of the CRISPR-Cas System in HIV

[0578] Chronic viral infection is a source of significant morbidity and mortality. While there exists for many of these viruses conventional antiviral therapies that effectively target various aspects of viral replication, current therapeutic modalities are usually non-curative in nature due to "viral latency." By its nature, viral latency is characterized by a dormant phase in the viral life cycle without active viral production. During this period, the virus is largely able to evade both immune surveillance and conventional therapeutics allowing for it to establish long-standing viral reservoirs within the host from which subsequent re-activation can permit continued propagation and transmission of virus. Key to viral latency is the ability to stably maintain the viral genome, accomplished either through episomal or proviral latency, which stores the viral genome in the cytoplasm or integrates it into the host genome, respectively. In the absence of effective vaccinations which would prevent primary infection, chronic viral infections characterized by latent reservoirs and episodes of lytic activity can have significant consequences: human papilloma virus (HPV) can result in cervical cancer, hepatitis C virus (HCV) predisposes to hepatocellular carcinoma, and human immunodeficiency virus eventually destroys the host immune system resulting in susceptibility to opportunistic infections. As such, these infections require life-long use of currently available antiviral therapeutics. Further complicating matters is the high mutability of many of these viral genomes which lead to the evolution of resistant strains for which there exists no effective therapy.

[0579] The CRISPR-Cas system is a bacterial adaptive immune system able to induce double-stranded DNA breaks (DSB) in a multiplex-able, sequence-specific manner and has been recently re-constituted within mammalian cell systems. It has been shown that targeting DNA with one or numerous guide-RNAs can result in both indels and deletions of the intervening sequences, respectively. As such, this new technology represents a means by which targeted and multiplexed DNA mutagenesis can be accomplished within a single cell with high efficiency and specificity. Consequently, delivery of the CRISPR-Cas system directed against viral DNA sequences could allow for targeted disruption and deletion of latent viral genomes even in the absence of ongoing viral production.

[0580] As an example, chronic infection by HIV-1 represents a global health issue with 33 million individuals infected and an annual incidence of 2.6 million infections. The use of the multimodal highly active antiretroviral therapy (HAART), which simultaneously targets multiple aspects of viral replication, has allowed HIV infection to be largely managed as a chronic, not terminal, illness. Without treatment, progression of HIV to AIDS occurs usually within 9-10 years resulting in depletion of the host immune system and occurrence of opportunistic infections usually leading to death soon thereafter. Secondary to viral latency, discontinuation of HAART invariably leads to viral rebound. Moreover, even temporary disruptions in therapy can select for resistant strains of HIV uncontrollable by available means. Additionally, the costs of HAART therapy are significant: within the US $10,000-15,0000 per person per year. As such, treatment approaches directly targeting the HIV genome rather than the process of viral replication represents a means by which eradication of latent reservoirs could allow for a curative therapeutic option.

[0581] Development and delivery of an HIV-1 targeted CRISPR-Cas system represents a unique approach differentiable from existing means of targeted DNA mutagenesis, i.e. ZFN and TALENs, with numerous therapeutic implications. Targeted disruption and deletion of the HIV-1 genome by CRISPR-mediated DSB and indels in conjunction with HAART could allow for simultaneous prevention of active viral production as well as depletion of latent viral reservoirs within the host.

[0582] Once integrated within the host immune system, the CRISPR-Cas system allows for generation of a HIV-1 resistant sub-population that, even in the absence of complete viral eradication, could allow for maintenance and re-constitution of host immune activity. This could potentially prevent primary infection by disruption of the viral genome preventing viral production and integration, representing a means to "vaccination". Multiplexed nature of the CRISPR-Cas system allows targeting of multiple aspects of the genome simultaneously within individual cells.

[0583] As in HAART, viral escape by mutagenesis is minimized by requiring acquisition of multiple adaptive mutations concurrently. Multiple strains of HIV-1 can be targeted simultaneously which minimizes the chance of super-infection and prevents subsequent creation of new recombinants strains. Nucleotide, rather than protein, mediated sequence-specificity of the CRISPR-Cas system allows for rapid generation of therapeutics without need for significantly altering delivery mechanism.

[0584] In order to accomplish this, Applicants generate CRISPR-Cas guide RNAs that target the vast majority of the HIV-1 genome while taking into account HIV-1 strain variants for maximal coverage and effectiveness. Sequence analyses of genomic conservation between HIV-1 subtypes and variants should allow for targeting of flanking conserved regions of the genome with the aims of deleting intervening viral sequences or induction of frame-shift mutations which would disrupt viral gene functions.

[0585] Applicants accomplish delivery of the CRISPR-Cas system by conventional adenoviral or lentiviral-mediated infection of the host immune system. Depending on approach, host immune cells could be a) isolated, transduced with CRISPR-Cas, selected, and re-introduced in to the host or b) transduced in vivo by systemic delivery of the CRISPR-Cas system. The first approach allows for generation of a resistant immune population whereas the second is more likely to target latent viral reservoirs within the host.

Examples of potential HIV-1 targeted spacers adapted from Mcintyre et al, which generated shRNAs against HIV-1 optimized for maximal coverage of HIV-1 variants.
CACTGCTTAAGCCTCGCTCGAGG
TCACCAGCAATATTCGCTCGAGG
CACCAGCAATATTCCGCTCGAGG
TAGCAACAGACATACGCTCGAGG
GGGCAGTAGTAATACGCTCGAGG
CCAATTCCCATACATTATTGTAC

Example 22

Targeted Correction of deltaF508 or Other Mutations in Cystic Fibrosis

[0586] An aspect of the invention provides for a pharmaceutical composition that may comprise an CRISPR-Cas gene therapy particle and a biocompatible pharmaceutical carrier. According to another aspect, a method of gene therapy for the treatment of a subject having a mutation in the CFTR gene comprises administering a therapeutically effective amount of a CRISPR-Cas gene therapy particle to the cells of a subject.

[0587] This Example demonstrates gene transfer or gene delivery of a CRISPR-Cas system in airways of subject or a patient in need thereof, suffering from cystic fibrosis or from cystic fibrosis related symptoms, using adeno-associated virus (AAV) particles.

[0588] Study Design: Subjects or patients in need there of: Human, non-primate human, canine, feline, bovine, equine and other domestic animals, related. This study tests efficacy of gene transfer of a CRISPR-Cas system by a AAV vector.

Applicants determine transgene levels sufficient for gene expression and utilize a CRISPR-Cas system comprising a Cas9 enzyme to target deltaF508 or other CFTR-inducing mutations.

[0589] The treated subjects receive pharmaceutically effective amount of aerosolized AAV vector system per lung endobronchially delivered while spontaneously breathing. The control subjects receive equivalent amount of a pseudotyped AAV vector system with an internal control gene. The vector system may be delivered along with a pharmaceutically acceptable or biocompatible pharmaceutical carrier. Three weeks or an appropriate time interval following vector administration, treated subjects are tested for amelioration of cystic fibrosis related symptoms.

[0590] Applicants use an adenovirus or an AAV particle.

[0591] Applicants clone the following gene constructs, each operably linked to one or more regulatory sequences (Cbh or EF1a promoter for Cas9, U6 or H1 promoter for chimeric guide RNA), into one or more adenovirus or AAV vectors or any other compatible vector: A CFTRdelta508 targeting chimeric guide RNA (FIG. **31**B), a repair template for deltaF508 mutation (FIG. **31**C) and a codon optimized Cas9 enzyme with optionally one or more nuclear localization signal or sequence(s) (NLS(s)), e.g., two (2) NLSs.

[0592] Identification of Cas9 Target Site

[0593] Applicants analyzed the human CFTR genomic locus and identified the Cas9 target site (FIG. **31**A). (PAM may contain a NGG or a NNAGAAW motif).

[0594] Gene Repair Strategy

[0595] Applicants introduce an adenovirus/AAV vector system comprising a Cas9 (or Cas9 nickase) and the guide RNA along with a adenovirus/AAV vector system comprising the homology repair template containing the F508 residue into the subject via one of the methods of delivery discussed earlier. The CRISPR-Cas system is guided by the CFTRdelta 508 chimeric guide RNA and targets a specific site of the CFTR genomic locus to be nicked or cleaved. After cleavage, the repair template is inserted into the cleavage site via homologous recombination correcting the deletion that results in cystic fibrosis or causes cystic fibrosis related symptoms. This strategy to direct delivery and provide systemic introduction of CRISPR systems with appropriate guide RNAs can be employed to target genetic mutations to edit or otherwise manipulate genes that cause metabolic, liver, kidney and protein diseases and disorders such as those in Table B.

Example 23

Generation of Gene Knockout Cell Library

[0596] This example demonstrates how to generate a library of cells where each cell has a single gene knocked out:

[0597] Applicants make a library of ES cells where each cell has a single gene knocked out, and the entire library of ES cells will have every single gene knocked out. This library is useful for the screening of gene function in cellular processes as well as diseases.

[0598] To make this cell library, Applicants integrate Cas9 driven by an inducible promoter (e.g. doxycycline inducible promoter) into the ES cell. In addition, Applicants integrate a single guide RNA targeting a specific gene in the ES cell. To make the ES cell library, Applicants simply mix ES cells with a library of genes encoding guide RNAs targeting each gene in the human genome. Applicants first introduce a single BxB1 attB site into the AAVS1 locus of the human ES cell. Then Applicants use the BxB1 integrase to facilitate the integration of individual guide RNA genes into the BxB1 attB site in AAVS1 locus. To facilitate integration, each guide RNA gene is contained on a plasmid that carries of a single attP site. This way BxB1 will recombine the attB site in the genome with the attP site on the guide RNA containing plasmid.

[0599] To generate the cell library, Applicants take the library of cells that have single guide RNAs integrated and induce Cas9 expression. After induction, Cas9 mediates double strand break at sites specified by the guide RNA. To verify the diversity of this cell library, Applicants carry out whole exome sequencing to ensure that Applicants are able to observe mutations in every single targeted gene. This cell library can be used for a variety of applications, including whole library-based screens, or can be sorted into individual cell clones to facilitate rapid generation of clonal cell lines with individual human genes knocked out.

Example 24

Engineering of Microalgae Using Cas9

[0600] Methods of Delivering Cas9

[0601] Method 1: Applicants deliver Cas9 and guide RNA using a vector that expresses Cas9 under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2-tubulin.

[0602] Method 2: Applicants deliver Cas9 and T7 polymerase using vectors that expresses Cas9 and T7 polymerase under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2-tubulin. Guide RNA will be delivered using a vector containing T7 promoter driving the guide RNA.

[0603] Method 3: Applicants deliver Cas9 mRNA and in vitro transcribed guide RNA to algae cells. RNA can be in vitro transcribed. Cas9 mRNA will consist of the coding region for Cas9 as well as 3'UTR from Cop1 to ensure stabilization of the Cas9 mRNA.

[0604] For Homologous recombination, Applicants provide an additional homology directed repair template.

[0605] Sequence for a cassette driving the expression of Cas9 under the control of beta-2 tubulin promoter, followed by the 3' UTR of Cop1.

```
TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGA

GACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGA

AGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGC

TGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCA

AAGCCATATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCG

AGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAG

TCACAACCCGCAAACATGTACCCATACGATGTTCCAGATTACGCTTCGCC

GAAGAAAAAGCGCAAGGTCGAAGCGTCCGACAAGAAGTACAGCATCGGCC

TGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTAC

AAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAG

CATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAG

CCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGG
```

-continued

AAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAA

GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAG

AGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAG

GTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACT

GGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGG

CCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAAC

CCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTA

CAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCA

AGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTG

ATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGAT

TGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGG

CCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTG

GACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGC

CGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGA

ACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATAC

GACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCA

GCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCT

ACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTC

ATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAA

GCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCA

GCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGG

CAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAA

GATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAA

ACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCC

TGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCAT

CGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGC

CCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACC

AAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGG

CGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG

TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTC

GACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGG

CACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACA

ATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTG

TTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCT

GTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCT

GGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCC

GGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAA

CTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCC

AGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCC

-continued

AATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA

GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACA

TCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAG

AACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGG

CAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACG

AGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGAC

CAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGT

GCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCA

GAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTC

GTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGAT

TACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGA

GCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGG

CAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAA

GTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGA

AGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTG

CGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGT

CGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCG

TGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGC

GAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACAT

CATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGA

AGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGAT

AAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGT

GAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGT

CTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGAC

TGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTC

TGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGA

GTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAG

AAGAATCCCATGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCG

GAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGG

CCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAG

AAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGA

ACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCT

CCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCC

TACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCAT

CCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACT

TTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTG

GACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGAT

CGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGG

AGGCCAGCTAAGGATCCGGCAAGACTGGCCCCGCTTGGCAACGCAACAGT

-continued

GAGCCCCTCCCTAGTGTGTTTGGGGATGTGACTATGTATTCGTGTGTTGG

CCAACGGGTCAACCCGAACAGATTGATACCCGCCTTGGCATTTCCTGTCA

GAATGTAACGTCAGTTGATGGTACT

[0606] Sequence for a cassette driving the expression of T7 polymerase under the control of beta-2 tubulin promoter, followed by the 3' UTR of Cop1:

TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGA

GACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGA

AGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGC

TGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCA

AAGCCATATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCG

AGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAG

TCACAACCCGCAAACatgcctaagaagaagaggaaggttaacacgattaa

catcgctaagaacgacttctctgacatcgaactggctgctatcccgttca

acactctggctgaccattacggtgagcgtttagctcgcgaacagttggcc

cttgagcatgagtcttacgagatgggtgaagcacgcttccgcaagatgta

gagcgtcaacttaaagctggtgaggttgcggataacgctgccgccaagcc

tctcatcactaccctactccctaagatgattgcacgcatcaacgactggt

ttgaggaagtgaaagctaagcgcggcaagcgcccgacagccttccagttc

ctgcaagaaatcaagccggaagccgtagcgtacatcaccattaagaccac

tctggcttgcctaaccagtgctgacaatacaaccgttcaggctgtagcaa

gcgcaatcggtcgggccattgaggacgaggctcgcttcggtcgtatccgt

gaccttgaagctaagcacttcaagaaaaacgttgaggaacaactcaacaa

gcgcgtagggcacgtctacaagaaagcatttatgcaagttgtcgaggctg

acatgctctctaagggtctactcggtggcgaggcgtggtcttcgtggcat

aaggaagactctattcatgtaggagtacgcgcatcgagatgctcattgag

tcaaccggaatggttagcttacaccgccaaaatgctggcgtagtaggtca

agactctgagactatcgaactcgcacctgaatacgctgaggctatcgcaa

cccgtgcaggtgcgctggctggcatctctccgatgttccaaccttgcgta

gttcctcctaagccgtggactggcattactggtggtggctattgggctaa

cggtcgtcgtcctctggcgctggtgcgtactcacagtaagaaagcactga

tgcgctacgaagacgtttacatgcctgaggtgtacaaagcgattaacatt

gcgcaaaacaccgcatggaaaatcaacaagaaagtcctagcggtcgccaa

cgtaatcaccaagtggaagcattgtccggtcgaggacatccctgcgattg

agcgtgaagaactcccgatgaaaccggaagacatcgacatgaatcctgag

gctctcaccgcgtggaaacgtgctgccgctgctgtgtaccgcaaggacaa

ggctcgcaagtctcgccgtatcagccttgagttcatgcttgagcaagcca

ataagtttgctaaccataaggccatctggttcccttacaacatggactgg

cgcggtcgtgtttacgctgtgtcaatgttcaacccgcaaggtaacgatat

-continued

gaccaaaggactgcttacgctggcgaaaggtaaaccaatcggtaaggaag

gttactactggctgaaaatccacggtgcaaactgtgcgggtgtcgacaag

gttccgttccctgagcgcatcaagttcattgaggaaaaccacgagaacat

catggcttgcgctaagtctccactggagaacacttggtgggctgagcaag

attctccgttctgcttccttgcgttctgctttgagtacgctggggtacag

caccacggcctgagctataactgctcccttccgctggcgtttgacgggtc

ttgctctggcatccagcacttctccgcgatgctccgagatgaggtaggtg

gtcgcgcggttaacttgcttcctagtgaaaccgttcaggacatctacggg

attgttgctaagaaagtcaacgagattctacaagcagacgcaatcaatgg

gaccgataacgaagtagttaccgtgaccgatgagaacactggtgaaatct

ctgagaaagtcaagctgggcactaaggcactggctggtcaatggctggct

tacggtgttactcgcagtgtgactaagcgttcagtcatgacgctggctta

cgggtccaaagagttcggcttccgtcaacaagtgctggaagataccattc

agccagctattgattccggcaagggtctgatgttcactcagccgaatcag

gctgctggatacatggctaagctgatttgggaatctgtgagcgtgacggt

ggtagctgcggttgaagcaatgaactggcttaagtctgctgctaagctgc

tggctgctgaggtcaaagataagaagactggagagattcttcgcaagcgt

tgcgctgtgcattgggtaactcctgatggtttccctgtgtggcaggaata

caagaagcctattcagacgcgcttgaacctgatgttcctcggtcagttcc

gcttacagcctaccattaacaccaacaaagatagcgagattgatgcacac

aaacaggagtctggtatcgctcctaactttgtacacagccaagacggtag

ccaccttcgtaagactgtagtgtgggcacacgagaagtacggaatcgaat

cttttgcactgattcacgactccttcggtacgattccggctgacgctgcg

aacctgttcaaagcagtgcgcgaaactatggttgacacatatgagtcttg

tgatgtactggctgatttctacgaccagttcgctgaccagttgcacgagt

ctcaattggacaaaatgccagcacttccggctaaaggtaacttgaacctc

cgtgacatcttagagtcggacttcgcgttcgcgtaaGGATCCGGCAAGAC

TGGCCCCGCTTGGCAACGCAACAGTGAGCCCCTCCCTAGTGTGTTTGGGG

ATGTGACTATGTATTCGTGTGTTGGCCAACGGGTCAACCCGAACAGATTG

ATACCCGCCTTGGCATTTCCTGTCAGAATGTAACGTCAGTTGATGGTACT

[0607] Sequence of guide RNA driven by the T7 promoter (T7 promoter, Ns represent targeting sequence):

gaaat**TAATACGACTCACTATA**NNNNNNNNNNNNNNNNNNNNgttttaga

gctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaaa

gtggcaccgagtcggtgcttttttt

[0608] Gene Delivery:

[0609] *Chlamydomonas reinhardtii* strain CC-124 and CC-125 from the *Chlamydomonas* Resource Center will be used for electroporation. Electroporation protocol follows standard recommended protocol from the GeneArt *Chlamydomonas* Engineering kit.

[0610] Also, Applicants generate a line of *Chlamydomonas reinhardtii* that expresses Cas9 constitutively. This can be done by using pChlamy1 (linearized using PvuI) and selecting for hygromycin resistant colonies. Sequence for pChlamy1 containing Cas9 is below. In this way to achieve gene knockout one simply needs to deliver RNA for the guideRNA. For homologous recombination Applicants deliver guideRNA as well as a linearized homologous recombination template.

[0611] pChlamy1-Cas9:

TGCGGTATTTCACACCGCATCAGGTGGCACTTTTCGGGGAAATGTGC

GCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCG

CTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAA

ATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACA

GTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTT

CGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACG

GGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCAC

GCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC

GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAA

TTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCA

ACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGT

ATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATC

CCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTG

TCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTG

CATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGG

TGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTT

GCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACT

TTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAG

GATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCA

ACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAA

ACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATG

TTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGG

GTTATTGTCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGA

GCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTT

TCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGG

TGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT

GGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTA

GTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTC

TGCTAATCCTGTTACCAGTGGCTGTTGCCAGTGGCGATAAGTCGTGTCTT

ACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGG

CTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACA

CCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCC

-continued

GAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGG

AGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTC

CTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCG

TCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACG

GTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTAT

CCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACC

GCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC

GGTCGCTGAGGCTTGACATGATTGGTGCGTATGTTTGTATGAAGCTACAG

GACTGATTTGGCGGGCTATGAGGGCGGGGGAAGCTCTGGAAGGGCCGCGA

TGGGGCGCGCGGCGTCCAGAAGGCGCCATACGGCCCGCTGGCGGCACCCA

TCCGGTATAAAAGCCCGCGACCCCGAACGGTGACCTCCACTTTCAGCGAC

AAACGAGCACTTATACATACGCGACTATTCTGCCGCTATACATAACCACT

CAGCTAGCTTAAGATCCCATCAAGCTTGCATGCCGGGCGCGCCAGAAGGA

GCGCAGCCAAACCAGGATGATGTTTGATGGGGTATTTGAGCACTTGCAAC

CCTTATCCGGAAGCCCCCTGGCCCACAAAGGCTAGGCGCCAATGCAAGCA

GTTCGCATGCAGCCCCTGGAGCGGTGCCCTCCTGATAAACCGGCCAGGGG

GCCTATGTTCTTTACTTTTTTACAAGAGAAGTCACTCAACATCTTAAAAT

GGCCAGGTGAGTCGACGAGCAAGCCCGGCGGATCAGGCAGCGTGCTTGCA

GATTTGACTTGCAACGCCCGCATTGTGTCGACGAAGGCTTTTGGCTCCTC

TGTCGCTGTCTCAAGCAGCATCTAACCCTGCGTCGCCGTTTCCATTTGCA

GGAGATTCGAGGTACCATGTACCCATACGATGTTCCAGATTACGCTTCGC

CGAAGAAAAAGCGCAAGGTCGAAGCGTCCGACAAGAAGTACAGCATCGGC

CTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTA

CAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACA

GCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACA

GCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACG

GAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCA

AGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAA

GAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGA

GGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAAC

TGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTG

GCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAA

CCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCT

ACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCC

AAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCG

ATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGAT

TGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGG

CCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTG

GACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGC

-continued

CGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGA

ACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATAC

GACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCA

GCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCT

ACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTC

ATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAA

GCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCA

GCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGG

CAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAA

GATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAA

ACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCC

TGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCAT

CGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGC

CCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACC

AAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGG

CGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG

TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTC

GACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGG

CACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACA

ATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTG

TTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCT

GTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCT

GGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCC

GGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAA

CTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCC

AGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCC

AATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAA

GGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACA

TCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAG

AACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGG

CAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACG

AGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGAC

CAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGT

GCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCA

GAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTC

GTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGAT

TACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGA

GCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGG

CAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAA

-continued

GTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGA

AGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTG

CGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGT

CGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCG

TGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGC

GAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACAT

CATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGA

AGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGAT

AAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGT

GAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGT

CTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGAC

TGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTC

TGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGA

GTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAG

AAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAA

GGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACG

GCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAA

CTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTA

TGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTG

TGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAG

TTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTC

CGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATA

TCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAG

TACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGT

GCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACAC

GGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAG

GTGGAGGCCAGCTAACATATGATTCGAATGTCTTTCTTGCGCTATGACAC

TTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATG

CAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGG

CGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCG

ATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGA

TCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTA

AGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACATGACAC

AAGAATCCCTGTTACTTCTCGACCGTATTGATTCGGATGATTCCTACGCG

AGCCTGCGGAACGACCAGGAATTGTGGGAGGTGAGTCGACGAGCAAGCCC

GGCGGATCAGGCAGCGTGCTTGCAGATTTGACTTGCAACGCCCGCATTGT

GTCGACGAAGGCTTTTGGCTCCTCTGTCGCTGTCTCAAGCAGCATCTAAC

CCTGCGTCGCCGTTTCCATTTGCAGCCGCTGGCCCGCCGAGCCCTGGAGG

AGCTCGGGCTGCCGGTGCCGCCGGTGCTGCGGGTGCCCGGCGAGAGCACC

-continued

AACCCCGTACTGGTCGGCGAGCCCGGCCCGGTGATCAAGCTGTTCGGCGA

GCACTGGTGCGGTCCGGAGAGCCTCGCGTCGGAGTCGGAGGCGTACGCGG

TCCTGGCGGACGCCCCGGTGCCGGTGCCCCGCCTCCTCGGCCGCGGCGAG

CTGCGGCCCGGCACCGGAGCCTGGCCGTGGCCCTACCTGGTGATGAGCCG

GATGACCGGCACCACCTGGCGGTCCGCGATGGACGGCACGACCGACCGGA

ACGCGCTGCTCGCCCTGGCCCGCGAACTCGGCCGGGTGCTCGGCCGGCTG

CACAGGGTGCCGCTGACCGGGAACACCGTGCTCACCCCCCATTCCGAGGT

CTTCCCGGAACTGCTGCGGGAACGCCGCGCGGCGACCGTCGAGGACCACC

GCGGGTGGGGCTACCTCTCGCCCCGGCTGCTGGACCGCCTGGAGGACTGG

CTGCCGGACGTGGACACGCTGCTGGCCGGCCGCGAACCCCGGTTCGTCCA

CGGCGACCTGCACGGGACCAACATCTTCGTGGACCTGGCCGCGACCGAGG

TCACCGGGATCGTCGACTTCACCGACGTCTATGCGGGAGACTCCCGCTAC

AGCCTGGTGCAACTGCATCTCAACGCCTTCCGGGGCGACCGCGAGATCCT

GGCCGCGCTGCTCGACGGGGCGCAGTGGAAGCGGACCGAGGACTTCGCCC

GCGAACTGCTCGCCTTCACCTTCCTGCACGACTTCGAGGTGTTCGAGGAG

ACCCCGCTGGATCTCTCCGGCTTCACCGATCCGGAGGAACTGGCGCAGTT

CCTCTGGGGGCCGCCGGACACCGCCCCCGGCGCCTGATAAGGATCCGGCA

AGACTGGCCCCGCTTGGCAACGCAACAGTGAGCCCCTCCCTAGTGTGTTT

GGGGATGTGACTATGTATTCGTGTGTTGGCCAACGGGTCAACCCGAACAG

ATTGATACCCGCCTTGGCATTTCCTGTCAGAATGTAACGTCAGTTGATGG

TACT

[0612] For all modified *Chlamydomonas* reinhardtii cells, Applicants use PCR, SURVEYOR nuclease assay, and DNA sequencing to verify successful modification.

Example 25

Use of Cas9 to Target a Variety of Disease Types

[0613] Diseases that involve mutations in protein coding sequence:

[0614] Dominant disorders may be targeted by inactivating the dominant negative allele. Applicants use Cas9 to target a unique sequence in the dominant negative allele and introduce a mutation via NHEJ. The NHEJ-induced indel may be able to introduce a frame-shift mutation in the dominant negative allele and eliminate the dominant negative protein. This may work if the gene is haplo-sufficient (e.g. MYOC mutation induced glaucoma and Huntington's disease).

[0615] Recessive disorders may be targeted by repairing the disease mutation in both alleles. For dividing cells, Applicants use Cas9 to introduce double strand breaks near the mutation site and increase the rate of homologous recombination using an exogenous recombination template. For dividing cells, this may be achieved using multiplexed nickase activity to catalyze the replacement of the mutant sequence in both alleles via NHEJ-mediated ligation of an exogenous DNA fragment carrying complementary overhangs.

[0616] Applicants also use Cas9 to introduce protective mutations (e.g. inactivation of CCR5 to prevent HIV infection, inactivation of PCSK9 for cholesterol reduction, or introduction of the A673T into APP to reduce the likelihood of Alzheimer's disease).

[0617] Diseases that Involve Non-Coding Sequences

[0618] Applicants use Cas9 to disrupt non-coding sequences in the promoter region, to alter transcription factor binding sites and alter enhancer or repressor elements. For example, Cas9 may be used to excise out the Klf1 enhancer EHS1 in hematopoietic stem cells to reduce BCL11a levels and reactivate fetal globin gene expression in differentiated erythrocytes

[0619] Applicants also use Cas9 to disrupt functional motifs in the 5' or 3' untranslated regions. For example, for the treatment of myotonic dystrophy, Cas9 may be used to remove CTG repeat expansions in the DMPK gene.

Example 26

Multiplexed Nickase

[0620] Aspects of optimization and the teachings of Cas9 detailed in this application may also be used to generate Cas9 nickases. Applicants use Cas9 nickases in combination with pairs of guide RNAs to generate DNA double strand breaks with defined overhangs. When two pairs of guide RNAs are used, it is possible to excise an intervening DNA fragment. If an exogenous piece of DNA is cleaved by the two pairs of guide RNAs to generate compatible overhangs with the genomic DNA, then the exogenous DNA fragment may be ligated into the genomic DNA to replace the excised fragment. For example, this may be used to remove trinucleotide repeat expansion in the huntintin (HTT) gene to treat Huntington's Disease.

[0621] If an exogenous DNA that bears fewer number of CAG repeats is provided, then it may be able to generate a fragment of DNA that bears the same overhangs and can be ligated into the HTT genomic locus and replace the excised fragment.

```
HTT locus    . . . CCGTGCCGGGCGGGAGACCGCCATGG                       GGCCCGGCTGTGGCTGAGGAGC . . .
with         . . . GGCACGGCCCGCCCTCTGGC                           TGGGCCGGGCCGACACCGACTCCTCG . . .
fragment
excised
by Cas9
nickase
and two
pairs of
guide RNAs
```

+

-continued

exogenous DNA fragment with fewer number of CAG repeats also cleaved by Cas9 nicakse and the two pairs of guide RNAs	CGACCCTGGAAA . . . reduced number of CAG repeats . . . CCCCGCCGCCACCC GGTACCGCTGGGACCTTT . . . . . . GGGGCGGCGG

[0622] The ligation of the exogenous DNA fragment into the genome does not require homologous recombination machineries and therefore this method may be used in post-mitotic cells such as neurons.

Example 27

Delivery of CRISPR System

[0623] Cas9 and its chimeric guide RNA, or combination of tracrRNA and crRNA, can be delivered either as DNA or RNA. Delivery of Cas9 and guide RNA both as RNA (normal or containing base or backbone modifications) molecules can be used to reduce the amount of time that Cas9 protein persist in the cell. This may reduce the level of off-target cleavage activity in the target cell. Since delivery of Cas9 as mRNA takes time to be translated into protein, it might be advantageous to deliver the guide RNA several hours following the delivery of Cas9 mRNA, to maximize the level of guide RNA available for interaction with Cas9 protein.

[0624] In situations where guide RNA amount is limiting, it may be desirable to introduce Cas9 as mRNA and guide RNA in the form of a DNA expression cassette with a promoter driving the expression of the guide RNA. This way the amount of guide RNA available will be amplified via transcription.

[0625] A variety of delivery systems can be introduced to introduce Cas9 (DNA or RNA) and guide RNA (DNA or RNA) into the host cell. These include the use of liposomes, viral vectors, electroporation, nanoparticles, nanowires (Shalek et al., Nano Letters, 2012), exosomes. Molecular trojan horses liposomes (Pardridge et al., Cold Spring Harb Protoc; 2010; doi: 10.1101/pdb.prot5407) may be used to deliver Cas9 and guide RNA across the blood brain barrier.

Example 28

Therapeutic Strategies for Trinucleotide Repeat Disorders

[0626] As previously mentioned in the application, the target polynucleotide of a CRISPR complex may include a number of disease-associated genes and polynucleotides and some of these disease associated gene may belong to a set of genetic disorders referred to as Trinucleotide repeat disorders (referred to as also trinucleotide repeat expansion disorders, triplet repeat expansion disorders or codon reiteration disorders).

[0627] These diseases are caused by mutations in which the trinucleotide repeats of certain genes exceed the normal, stable threshold which may usually differ in a gene. The discovery of more repeat expansion disorders has allowed for the classification of these disorders into a number of categories based on underlying similar characteristics. Huntington's disease (HD) and the spinocerebellar ataxias that are caused by a CAG repeat expansion in protein-coding portions of specific genes are included in Category I. Diseases or disorders with expansions that tend to make them phenotypically diverse and include expansions are usually small in magnitude and also found in exons of genes are included in Category II. Category III includes disorders or diseases which are characterized by much larger repeat expansions than either Category I or II and are generally located outside protein coding regions. Examples of Category III diseases or disorders include but are not limited to Fragile X syndrome, myotonic dystrophy, two of the spinocerebellar ataxias, juvenile myoclonic epilepsy, and Friedreich's ataxia.

[0628] Similar therapeutic strategies, like the one mentioned for Friedreich's ataxia below may be adopted to address other trinucleotide repeat or expansion disorders as well. For example, another triple repeat disease that can be treated using almost identical strategy is dystrophia myotonica 1 (DM1), where there is an expanded CTG motif in the 3' UTR. In Friedreich's ataxia, the disease results from expansion of GAA trinucleotides in the first intron of frataxin (FXN). One therapeutic strategy using CRISPR is to excise the GAA repeat from the first intron. The expanded GAA repeat is thought to affect the DNA structure and leads to recruit the formation of heterochrornatin which turn off the frataxin gene (FIG. **32**A).

[0629] Competitive Advantage over other therapeutic strategies are listed below:

[0630] siRNA knockdown is not applicable in this case, as disease is due to reduced expression of frataxin. Viral gene therapy is currently being explored. HSV-1 based vectors were used to deliver the frataxin gene in animal models and have shown therapeutic effect. However, long term efficacy of virus-based frataxin delivery suffer from several problems: First, it is difficult to regulate the expression of frataxin to match natural levels in health individuals, and second, long term over expression of frataxin leads to cell death.

[0631] Nucleases may be used to excise the GAA repeat to restore healthy genotype, but Zinc Finger Nuclease and TALEN strategies require delivery of two pairs of high efficacy nucleases, which is difficult for both delivery as well as nuclease engineering (efficient excision of genomic DNA by ZFN or TALEN is difficult to achieve).

[0632] In contrast to above strategies, the CRISPR-Cas system has clear advantages. The Cas9 enzyme is more efficient and more multiplexible, by which it is meant that one or more targets can be set at the same time. So far, efficient excision of genomic DNA >30% by Cas9 in human cells and may be as high as 30%, and may be improved in the future. Furthermore, with regard to certain trinucleotide repeat disorders like Huntington's disease (HD), trinucleotide repeats in the coding region may be addressed if there are differences between the two alleles. Specifically, if a HD patient is heterozygous for mutant HTT and there are nucleotide differences such as SNPs between the wt and mutant HTT alleles, then Cas9 may be used to specifically target the mutant HTT allele. ZFN or TALENs will not have the ability to distinguish two alleles based on single base differences.

[0633] In adopting a strategy using the CRISPR-Cas 9 enzyme to address Friedreich's ataxia, Applicants design a number of guide RNAs targeting sites flanking the GAA expansion and the most efficient and specific ones are chosen (FIG. **32**B).

[0634] Applicants deliver a combination of guide RNAs targeting the intron I of FXN along with Cas9 to mediate excision of the GAA expansion region. AAV9 may be used to mediate efficient delivery of Cas9 and in the spinal cord.

[0635] If the Alu element adjacent to the GAA expansion is considered important, there may be constraints to the number of sites that can be targeted but Applicants may adopt strategies to avoid disrupting it.

[0636] Alternative Strategies:

[0637] Rather than modifying the genome using Cas9, Applicants may also directly activate the FXN gene using Cas9 (nuclease activity deficient)-based DNA binding domain to target a transcription activation domain to the FXN gene. Applicants may have to address the robustness of the Cas9-mediated artificial transcription activation to ensure that it is robust enough as compared to other methods (Tremblay et al., Transcription Activator-Like Effector Proteins Induce the Expression of the Frataxin Gene; Human Gene Therapy. August 2012, 23(8): 883-890.)

Example 29

Strategies for Minimizing Off Target Cleavage Using Cas9 Nickase

[0638] As previously mentioned in the application, Cas9 may be mutated to mediate single strand cleavage via one or more of the following mutations: D10A, E762A, and H840A.

[0639] To mediate gene knockout via NHEJ, Applicants use a nickase version of Cas9 along with two guide RNAs. Off-target nicking by each individual guide RNA may be primarily repaired without mutation, double strand breaks (which can lead to mutations via NHEJ) only occur when the target sites are adjacent to each other. Since double strand breaks introduced by double nicking are not blunt, co-expression of end-processing enzymes such as TREX1 will increase the level of NHEJ activity.

[0640] The following list of targets in tabular form are for genes involved in the following diseases:

[0641] Lafora's Disease—target GSY1 or PPP1R3C (PTG) to reduce glycogen in neurons.

[0642] Hypercholesterolemia—target PCSK9

[0643] Target sequences are listed in pairs (L and R) with different number of nucleotides in the spacer (0 to 3 bp). Each spacer may also be used by itself with the wild type Cas9 to introduce double strand break at the target locus.

GYS1 (human)	GGCC-L	ACCCTTGTTAGCCACCTCCC
	GGCC-R	GAACGCAGTGCTCTTCGAAG
	GGNCC-L	CTCACGCCCTGCTCCGTGTA
	GGNCC-R	GGCGACAACTACTTCCTGGT
	GGNNCC-L	CTCACGCCCTGCTCCGTGTA
	GGNNCC-R	GGGCGACAACTACTTCCTGG
	GGNNNCC-L	CCTCTTCAGGGCCGGGGTGG
	GGNNNCC-R	GAGGACCCAGGTGGAACTGC
PCSK9 (human)	GGCC-L	TCAGCTCCAGGCGGTCCTGG
	GGCC-R	AGCAGCAGCAGCAGTGGCAG
	GGNCC-L	TGGGCACCGTCAGCTCCAGG
	GGNCC-R	CAGCAGTGGCAGCGGCCACC
	GGNNCC-L	ACCTCTCCCCTGGCCCTCAT
	GGNNCC-R	CCAGGACCGCCTGGAGCTGA
	GGNNNCC-L	CCGTCAGCTCCAGGCGGTCC
	GGNNNCC-R	AGCAGCAGCAGCAGTGGCAG
PPP1R3C (PTG) (human)	GGCC-L	ATGTGCCAAGCAAAGCCTCA
	GGCC-R	TTCGGTCATGCCCGTGGATG
	GGNCC-L	GTCGTTGAAATTCATCGTAC
	GGNCC-R	ACCACCTGTGAAGAGTTTCC
	GGNNCC-L	CGTCGTTGAAATTCATCGTA
	GGNNCC-R	ACCACCTGTGAAGAGTTTCC
Gys1 (mouse)	GGCC-L	GAACGCAGTGCTTTTCGAGG
	GGCC-R	ACCCTTGTTGGCCACCTCCC
	GGNCC-L	GGTGACAACTACTATCTGGT
	GGNCC-R	CTCACACCCTGCTCCGTGTA
	GGNNCC-L	GGGTGACAACTACTATCTGG
	GGNNCC-R	CTCACACCCTGCTCCGTGTA
	GGNNNCC-L	CGAGAACGCAGTGCTTTTCG
	GGNNNCC-R	ACCCTTGTTGGCCACCTCCC
PPP1R3C (PTG) (mouse)	GGCC-L	ATGAGCCAAGCAAATCCTCA
	GGCC-R	TTCCGTCATGCCCGTGGACA
	GGNCC-L	CTTCGTTGAAAACCATTGTA
	GGNCC-R	CCACCTCTGAAGAGTTTCCT
	GGNNCC-L	CTTCGTTGAAAACCATTGTA
	GGNNCC-R	ACCACCTCTGAAGAGTTTCC
	GGNNNCC-L	CTTCCACTCACTCTGCGATT
	GGNNNCC-R	ACCATGTCTCAGTGTCAAGC
PCSK9 (mouse)	GGCC-L	GGCGGCAACAGCGGCAACAG
	GGCC-R	ACTGCTCTGCGTGGCTGCGG
	GGNNCC-L	CCGCAGCCACGCAGAGCAGT
	GGNNCC-R	GCACCTCTCCTCGCCCCGAT

[0644] Alternative Strategies for Improving Stability of Guide RNA and Increasing Specificity

[0645] 1. Nucleotides in the 5' of the guide RNA may be linked via thiolester linkages rather than phosphoester linkage like in natural RNA. Thiolester linkage may prevent the guide RNA from being digested by endogenous RNA degradation machinery.

[0646] 2. Nucleotides in the guide sequence (5' 20 bp) of the guide RNA can use bridged nucleic acids (BNA) as the bases to improve the binding specificity.

Example 30

CRISPR-Cas for Rapid, Multiplex Genome Editing

[0647] Aspects of the invention relate to protocols and methods by which efficiency and specificity of gene modification may be tested within 3-4 days after target design, and modified clonal cell lines may be derived within 2-3 weeks.

[0648] Programmable nucleases are powerful technologies for mediating genome alteration with high precision. The RNA-guided Cas9 nuclease from the microbial CRISPR adaptive immune system can be used to facilitate efficient genome editing in eukaryotic cells by simply specifying a 20-nt targeting sequence in its guide RNA. Applicants describe a set of protocols for applying Cas9 to facilitate efficient genome editing in mammalian cells and generate cell lines for downstream functional studies. Beginning with target design, efficient and specific gene modification can be achieved within 3-4 days, and modified clonal cell lines can be derived within 2-3 weeks.

[0649] The ability to engineer biological systems and organisms holds enormous potential for applications across basic science, medicine, and biotechnology. Programmable sequence-specific endonucleases that facilitate precise editing of endogenous genomic loci are now enabling systematic interrogation of genetic elements and causal genetic variations in a broad range of species, including those that have not been genetically tractable previously. A number of genome editing technologies have emerged in recent years, including zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the RNA-guided CRISPR-Cas nuclease system. The first two technologies use a common strategy of tethering endonuclease catalytic domains to modular DNA-binding proteins for inducing targeted DNA double stranded breaks (DSB) at specific genomic loci. By contrast, Cas9 is a nuclease guided by small RNAs through Watson-Crick base-pairing with target DNA, presenting a system that is easy to design, efficient, and well-suited for high-throughput and multiplexed gene editing for a variety of cell types and organisms. Here Applicants describe a set of protocols for applying the recently developed Cas9 nuclease to facilitate efficient genome editing in mammalian cells and generate cell lines for downstream functional studies.

[0650] Like ZFNs and TALENs, Cas9 promotes genome editing by stimulating DSB at the target genomic loci. Upon cleavage by Cas9, the target locus undergoes one of two major pathways for DNA damage repair, the error-prone non-homologous end joining (NHEJ) or the high-fidelity homology directed repair (HDR) pathway. Both pathways may be utilized to achieve the desired editing outcome.

[0651] NHEJ: In the absence of a repair template, the NHEJ process re-ligates DSBs, which may leave a scar in the form of indel mutations. This process can be harnessed to achieve gene knockouts, as indels occurring within a coding exon may lead to frameshift mutations and a premature stop codon. Multiple DSBs may also be exploited to mediate larger deletions in the genome.

[0652] HDR: Homology directed repair is an alternate major DNA repair pathway to NHEJ. Although HDR typically occurs at lower frequencies than NHEJ, it may be harnessed to generate precise, defined modifications at a target locus in the presence of an exogenously introduced repair template. The repair template may be either in the form of double stranded DNA, designed similarly to conventional DNA targeting constructs with homology arms flanking the insertion sequence, or single-stranded DNA oligonucleotides (ssODNs). The latter provides an effective and simple method for making small edits in the genome, such as the introduction of single nucleotide mutations for probing causal genetic variations. Unlike NHEJ. HDR is generally active only in dividing cells and its efficiency varies depending on the cell type and state.

[0653] Overview of CRISPR: The CRISPR-Cas system, by contrast, is at minimum a two-component system consisting of the Cas9 nuclease and a short guide RNA. Re-targeting of Cas9 to different loci or simultaneous editing of multiple genes simply requires cloning a different 20-bp oligonucleotide. Although specificity of the Cas9 nuclease has yet to be thoroughly elucidated, the simple Watson-Crick base-pairing of the CRISPR-Cas system is likely more predictable than that of ZFN or TA LEN domains.

[0654] The type II CRISPR-Cas (clustered regularly interspaced short palindromic repeats) is a bacterial adaptive immune system that uses Cas9, to cleave foreign genetic elements. Cas9 is guided by a pair of non-coding RNAs, a variable crRNA and a required auxiliary tracrRNA. The crRNA contains a 20-nt guide sequence determines specificity by locating the target DNA via Watson-Crick base-pairing. In the native bacterial system, multiple crRNAs are co-transcribed to direct Cas9 against various targets. In the CRISPR-Cas system derived from *Streptococcus pyogenes*, the target DNA must immediately precede a 5'-NGG/NRG protospacer adjacent motif (PAM), which can vary for other CRISPR systems.

[0655] CRISPR-Cas is reconstituted in mammalian cells through the heterologous expression of human codon-optimized Cas9 and the requisite RNA components. Furthermore, the crRNA and tracrRNA can be fused to create a chimeric, synthetic guide RNA (sgRNA). Cas9 can thus be re-directed toward any target of interest by altering the 20-nt guide sequence within the sgRNA.

[0656] Given its ease of implementation and multiplex capability, Cas9 has been used to generate engineered eukaryotic cells carrying specific mutations via both NHEJ and HDR. In addition, direct injection of sgRNA and mRNA encoding Cas9 into embryos has enabled the rapid generation of transgenic mice with multiple modified alleles; these results hold promise for editing organisms that are otherwise genetically intractable.

[0657] A mutant Cas9 carrying a disruption in one of its catalytic domains has been engineered to nick rather than cleave DNA, allowing for single-stranded breaks and preferential repair through HDR, potentially ameliorating unwanted indel mutations from off-target DSBs. Additionally, a Cas9 mutant with both DNA-cleaving catalytic residues mutated has been adapted to enable transcriptional regulation in *E. coli*, demonstrating the potential of functionalizing Cas9 for diverse applications. Certain aspects of the invention relate to the construction and application of Cas9 for multiplexed editing of human cells.

[0658] Applicants have provided a human codon-optimized, nuclear localization sequence-flanked Cas9 to facilitate eukaryotic gene editing. Applicants describe considerations for designing the 20-nt guide sequence, protocols for rapid construction and functional validation of sgRNAs, and finally use of the Cas9 nuclease to mediate both NHEJ- and HDR-based genome modifications in human embryonic kidney (HEK-293FT) and human stem cell (HUES9) lines. This protocol can likewise be applied to other cell types and organisms.

[0659] Target selection for sgRNA: There are two main considerations in the selection of the 20-nt guide sequence for gene targeting: 1) the target sequence should precede the 5'-NGG PAM for *S. pyogenes* Cas9, and 2) guide sequences should be chosen to minimize off-target activity. Applicants provided an online Cas9 targeting design tool that takes an input sequence of interest and identifies suitable target sites. To experimentally assess off-target modifications for each sgRNA, Applicants also provide computationally predicted off-target sites for each intended target, ranked according to Applicants' quantitative specificity analysis on the effects of base-pairing mismatch identity, position, and distribution.

[0660] The detailed information on computationally predicted off-target sites is as follows:

[0661] Considerations for Off-target Cleavage Activities: Similar to other nucleases, Cas9 can cleave off-target DNA targets in the genome at reduced frequencies. The extent to which a given guide sequence exhibit off-target activity depends on a combination of factors including enzyme concentration, thermodynamics of the specific guide sequence employed, and the abundance of similar sequences in the target genome. For routine application of Cas9, it is important to consider ways to minimize the degree of off-target cleavage and also to be able to detect the presence of off-target cleavage.

[0662] Minimizing off-target activity: For application in cell lines, Applicants recommend following two steps to reduce the degree of off-target genome modification. First, using our online CRISPR target selection tool, it is possible to computationally assess the likelihood of a given guide sequence to have off-target sites. These analyses are performed through an exhaustive search in the genome for off-target sequences that are similar sequences as the guide sequence. Comprehensive experimental investigation of the effect of mismatching bases between the sgRNA and its target DNA revealed that mismatch tolerance is 1) position dependent—the 8-14 bp on the 3' end of the guide sequence are less tolerant of mismatches than the 5' bases, 2) quantity dependent—in general more than 3 mismatches are not tolerated, 3) guide sequence dependent—some guide sequences are less tolerant of mismatches than others, and 4) concentration dependent—off-target cleavage is highly sensitive to the amount of transfected DNA. The Applicants' target site analysis web tool (available at the website genome-engineering.org/tools) integrates these criteria to provide predictions for likely off-target sites in the target genome. Second, Applicants recommend titrating the amount of Cas9 and sgRNA expression plasmid to minimize off-target activity.

[0663] Detection of off-target activities: Using Applicants' CRISPR targeting web tool, it is possible to generate a list of most likely off-target sites as well as primers performing SURVEYOR or sequencing analysis of those sites. For isogenic clones generated using Cas9, Applicants strongly recommend sequencing these candidate off-target sites to check for any undesired mutations. It is worth noting that there may be off target modifications in sites that are not included in the predicted candidate list and full genome sequence should be performed to completely verify the absence of off-target sites. Furthermore, in multiplex assays where several DSBs are induced within the same genome, there may be low rates of translocation events and can be evaluated using a variety of techniques such as deep sequencing.

[0664] The online tool provides the sequences for all oligos and primers necessary for 1) preparing the sgRNA constructs, 2) assaying target modification efficiency, and 3) assessing cleavage at potential off-target sites. It is worth noting that because the U6 RNA polymerase III promoter used to express the sgRNA prefers a guanine (G) nucleotide as the first base of its transcript, an extra G is appended at the 5' of the sgRNA where the 20-nt guide sequence does not begin with G.

[0665] Approaches for sgRNA construction and delivery: Depending on the desired application, sgRNAs may be delivered as either 1) PCR amplicons containing an expression cassette or 2) sgRNA-expressing plasmids. PCR-based sgRNA delivery appends the custom sgRNA sequence onto the reverse PCR primer used to amplify a U6 promoter template. The resulting amplicon may be co-transfected with a plasmid containing Cas9 (PX165). This method is optimal for rapid screening of multiple candidate sgRNAs, as cell transfections for functional testing can be performed mere hours after obtaining the sgRNA-encoding primers. Because this simple method obviates the need for plasmid-based cloning and sequence verification, it is well suited for testing or co-transfecting a large number of sgRNAs for generating large knockout libraries or other scale-sensitive applications. Note that the sgRNA-encoding primers are over 100-bp, compared to the ~20-bp oligos required for plasmid-based sgRNA delivery.

[0666] Construction of an expression plasmid for sgRNA is also simple and rapid, involving a single cloning step with a pair of partially complementary oligonucleotides. After annealing the oligo pairs, the resulting guide sequences may be inserted into a plasmid bearing both Cas9 and an invariant scaffold bearing the remainder of the sgRNA sequence (PX330). The transfection plasmids may also be modified to enable virus production for in vivo delivery.

[0667] In addition to PCR and plasmid-based delivery methods, both Cas9 and sgRNA can be introduced into cells as RNA.

[0668] Design of repair template: Traditionally, targeted DNA modifications have required use of plasmid-based donor repair templates that contain homology arms flanking the site of alteration. The homology arms on each side can vary in length, but are typically longer than 500 bp. This method can be used to generate large modifications, including insertion of reporter genes such as fluorescent proteins or antibiotic resistance markers. The design and construction of targeting plasmids has been described elsewhere.

[0669] More recently, single-stranded DNA oligonucleotides (ssODNs) have been used in place of targeting plasmids for short modifications within a defined locus without cloning. To achieve high HDR efficiencies, ssODNs contain flanking sequences of at least 40 bp on each side that are homologous to the target region, and can be oriented in either the sense or antisense direction relative to the target locus.

[0670] Functional Testing

[0671] SURVEYOR nuclease assay: Applicants detected indel mutations either by the SURVEYOR nuclease assay (or PCR amplicon sequencing. Applicants online CRISPR target design tool provides recommended primers for both approaches. However, SURVEYOR or sequencing primers may also be designed manually to amplify the region of interest from genomic DNA and to avoid non-specific amplicons using NCBI Primer-BLAST. SURVEYOR primers should be designed to amplify 300-400 bp (for a 600-800 bp total amplicon) on either side of the Cas9 target for allowing clear visualization of cleavage bands by gel electrophoresis. To prevent excessive primer dimer formation, SURVEYOR primers should be designed to be typically under 25-nt long with melting temperatures of ~60° C. Applicants recommend testing each pair of candidate primers for specific PCR amplicons as well as for the absence of non-specific cleavage during the SURVEYOR nuclease digestion process.

[0672] Plasmid- or ssODN-mediated HDR: HDR can be detected via PCR-amplification and sequencing of the modified region. PCR primers for this purpose should anneal outside the region spanned by the homology arms to avoid false detection of residual repair template (HDR Fwd and Rev, FIG. **30**). For ssODN-mediated HDR, SURVEYOR PCR primers can be used.

[0673] Detection of indels or HDR by sequencing: Applicants detected targeted genome modifications by either Sanger or next-generation deep sequencing (NGS). For the former, genomic DNA from modified region can be amplified using either SURVEYOR or HDR primers. Amplicons should be subcloned into a plasmid such as pUC19 for transformation; individual colonies can be sequenced to reveal clonal genotype.

[0674] Applicants designed next-generation sequencing (NGS) primers for shorter amplicons, typically in the 100-200 bp size range. For detecting NHEJ mutations, it is important to design primers with at least 10-20 bp between the priming regions and the Cas9 target site to allow detection of longer indels. Applicants provide guidelines for a two-step PCR method to attach barcoded adapters for multiplex deep sequencing. Applicants recommend the Illumina platform, due to its generally low levels of false positive indels. Off-target analysis (described previously) can then be performed through read alignment programs such as ClustalW, Geneious, or simple sequence analysis scripts.

[0675] Materials and Reagents

[0676] sgRNA Preparation:

[0677] UltraPure DNaseRNase-free distilled water (Life Technologies, cat. no. 10977-023)

[0678] Herculase II fusion polymerase (Agilent Technologies, cat. no. 600679)

[0679] CRITICAL. Standard Taq polymerase, which lacks 3'-5' exonuclease proofreading activity, has lower fidelity and can lead to amplification errors. Herculase II is a high-fidelity polymerase (equivalent fidelity to Pfu) that produces high yields of PCR product with minimal optimization. Other high-fidelity polymerases may be substituted.

[0680] Herculase II reaction buffer (5×; Agilent Technologies, included with polymerase)

[0681] dNTP solution mix (25 mM each; Enzymatics, cat. no. N205L)

[0682] MgCl2 (25 mM; ThermoScientific, cat. no. R0971)

[0683] QIAquick gel extraction kit (Qiagen, cat. no. 28704)

[0684] QIAprep spin miniprep kit (Qiagen, cat. no. 27106)

[0685] UltraPure TBE buffer (10×; Life Technologies, cat. no. 15581-028)

[0686] SeaKem LE agarose (Lonza, cat. no. 50004)

[0687] SYBR Safe DNA stain (10,000×; Life Technologies, cat. no. S33102)

[0688] 1-kb Plus DNA ladder (Life Technologies, cat. no. 10787-018)

[0689] TrackIt CyanOrange loading buffer (Life Technologies, cat. no. 10482-028)

[0690] FastDigest BbsI (BpiI) (Fermentas/ThermoScientific, cat. no. FD1014)

[0691] Fermentas Tango Buffer (Fermentas/ThermoScientific, cat. no. BY5)

[0692] DL-dithiothreitol (DTT; Fermentas/ThermoScientific, cat. no. R0862)

[0693] T7 DNA ligase (Enzymatics, cat. no. L602L)

[0694] Critical: Do not substitute the more commonly used T4 ligase. T7 ligase has 1,000-fold higher activity on the sticky ends than on the blunt ends and higher overall activity than commercially available concentrated T4 ligases.

[0695] T7 2X Rapid Ligation Buffer (included with T7 DNA ligase, Enzymatics, cat. no. L602L)

[0696] T4 Polynucleotide Kinase (New England Biolabs, cat. no M0201S)

[0697] T4 DNA Ligase Reaction Buffer (10×; New England Biolabs, cat. no B0202S)

[0698] Adenosine 5'-triphosphate (10 mM; New England Biolabs, cat. no. P0756S)

[0699] PlasmidSafe ATP-dependent DNase (Epicentre, cat. no. E3101K)

[0700] One Shot Stb13 chemically competent *Escherichia coli* (*E. coli*) (Life Technologies, cat. no. C7373-03)

[0701] SOC medium (New England Biolabs, cat. no. B9020S)

[0702] LB medium (Sigma, cat. no. L3022)

[0703] LB agar medium (Sigma, cat. no. L2897)

[0704] Ampicillin, sterile filtered (100 mg ml-1; Sigma, cat. no. A5354)

[0705] Mammalian Cell Culture:

[0706] HEK293FT cells (Life Technologies, cat. no. R700-07)

[0707] Dulbecco's minimum Eagle's medium (DMEM, 1×, high glucose; Life Technologies, cat. no. 10313-039)

[0708] Dulbecco's minimum Eagle's medium (DMEM, 1×, high glucose, no phenol red; Life Technologies, cat. no. 31053-028)

[0709] Dulbecco's phosphate-buffered saline (DPBS, 1×; Life Technologies, cat. no. 14190-250)

[0710] Fetal bovine serum, qualified and heat inactivated (Life Technologies, cat. no. 10438-034)

[0711] Opti-MEM I reduced-serum medium (FBS; Life Technologies, cat. no. 11058-021)

[0712] Penicillin-streptomycin (100×; Life Technologies, cat. no. 15140-163)

[0713] TrypLE™ Express (1×, no Phenol Red; Life Technologies, cat. no. 12604-013)

[0714] Lipofectamine 2000 transfection reagent (Life Technologies, cat. no. 11668027)

[0715] Amaxa SF Cell Line 4D-Nucleofector® X Kit S (32 RCT; Lonza, cat. no V4XC-2032)

[0716] HUES 9 cell line (HARVARD STEM CELL SCIENCE)

[0717] Geltrex LDEV-Free Reduced Growth Factor Basement Membrane Matrix (Life Technologies, cat. no. A1413201)

[0718] mTeSR1 medium (Stemcell Technologies, cat. no. 05850)

[0719] Accutase cell detachment solution (Stemcell Technologies, cat. no. 07920)

[0720] ROCK Inhibitor (Y-27632; Millipore, cat. no. SCM075)

[0721] Amaxa P3 Primary Cell 4D-Nucleofector® X Kit S (32 RCT; Lonza cat. no. V4XP-3032)

[0722] Genotyping Analysis:

[0723] QuickExtract DNA extraction solution (Epicentre, cat. no. QE09050)

[0724] PCR primers for SURVEYOR, RFLP analysis, or sequencing (see Primer table)

[0725] Herculase II fusion polymerase (Agilent Technologies, cat. no. 600679)

[0726] CRITICAL. As Surveyor assay is sensitive to single-base mismatches, it is particularly important to use a high-fidelity polymerase. Other high-fidelity polymerases may be substituted.

[0727] Herculase II reaction buffer (5×; Agilent Technologies, included with polymerase)

[0728] dNTP solution mix (25 mM each; Enzymatics, cat. no. N205L)

[0729] QIAquick gel extraction kit (Qiagen, cat. no. 28704)

[0730] Taq Buffer (10×; Genscript, cat. no. B0005)

[0731] SURVEYOR mutation detection kit for standard gel electrophoresis (Transgenomic, cat. no. 706025)

[0732] UltraPure TBE buffer (10×; Life Technologies, cat. no. 15581-028)

[0733] SeaKem LE agarose (Lonza, cat. no. 50004)

[0734] 4-20% TBE Gels 1.0 mm, 15 Well (Life Technologies, cat. no. EC62255BOX)

[0735] Novex® Hi-Density TBE Sample Buffer (5×; Life Technologies, cat. no. LC6678)

[0736] SYBR Gold Nucleic Acid Gel Stain (10,000×; Life Technologies, cat. no. S-11494)

[0737] 1-kb Plus DNA ladder (Life Technologies, cat. no. 10787-018)

[0738] TrackIt CyanOrange loading buffer (Life Technologies, cat. no. 10482-028)

[0739] FastDigest HindIII (Fermentas/ThermoScientific, cat. no. FD0504)

[0740] Equipment

[0741] Filtered sterile pipette tips (Corning)

[0742] Standard 1.5 ml microcentrifuge tubes (Eppendorf, cat. no. 0030 125.150)

[0743] Axygen 96-well PCR plates (VWR, cat. no. PCR-96M2-HSC)

[0744] Axygen 8-Strip PCR tubes (Fischer Scientific, cat. no. 14-222-250)

[0745] Falcon tubes, polypropylene, 15 ml (BD Falcon, cat. no. 352097)

[0746] Falcon tubes, polypropylene, 50 ml (BD Falcon, cat. no. 352070)

[0747] Round-bottom Tube with cell strainer cap. 5 ml (BD Falcon, cat. no. 352235)

[0748] Petri dishes (60 mm×15 mm; BD Biosciences, cat. no. 351007)

[0749] Tissue culture plate (24 well; BD Falcon, cat. no. 353047)

[0750] Tissue culture plate (96 well, flat bottom: BD Falcon, cat. no. 353075)

[0751] Tissue culture dish (100 mm; 131BD) Falcon, 353003)

[0752] 96-well thermocycler with programmable temperature stepping functionality (Applied Biosystems Veriti, cat. no. 4375786).

[0753] Desktop microcentrifuges 5424, 5804 (Eppendorf)

[0754] Gel electrophoresis system (PowerPac basic power supply, Bio-Rad, cat. no. 164-5050, and Sub-Cell GT System gel tray, Bio-Rad, cat. no. 170-4401)

[0755] Novex XCell SureLock Mini-Cell (Life Technologies, cat. no. EI0001)

[0756] Digital gel imaging system (GelDoc EZ, Bio-Rad, cat. no. 170-8270, and blue sample tray, Bio-Rad, cat. no. 170-8273)

[0757] Blue light transilluminator and orange filter goggles (SafeImager 2.0; Invitrogen, cat. no. G6600)

[0758] Gel quantification software (Bio-Rad, ImageLab, included with GelDoc EZ, or open-source ImageJ from the National Institutes of Health, available at the website rsbweb.nih.gov/ij/) UV spectrophotometer (NanoDrop 2000c, Thermo Scientific)

[0759] Reagent Setup

[0760] Tris-borate EDTA (TBE) electrophoresis solution Dilute TBE buffer in distilled water to 1× working solution for casting agarose gels and for use as a buffer for gel electrophoresis. Buffer may be stored at room temperature (18-22° C.) for at least 1 year.

> **[0761]** ATP, 10 mM Divide 10 mM ATP into 50-μl aliquots and store at −20° C. for up to 1 year; avoid repeated freeze-thaw cycles.
>
> **[0762]** DTT, 10 mM Prepare 10 mM DTT solution in distilled water and store in 20-μl aliquots at −70° C. for up to 2 years; for each reaction, use a new aliquot, as DTT is easily oxidized.
>
> **[0763]** D10 culture medium For culture of HEK293FT cells, prepare D10 culture medium by supplementing DMEM with 1× GlutaMAX and 10% (vol/vol) fetal bovine serum. As indicated in the protocol, this medium can also be supplemented with 1× penicillin-streptomycin. D10 medium can be made in advance and stored at 4° C. for up to 1 month.
>
> **[0764]** mTeSR1 culture medium For culture of human embryonic stem cells, prepare mTeSR1 medium by supplementing the 5× supplement (included with mTeSR1 basal medium), and 100 ug/ml Normocin.

[0765] Procedure

[0766] Design of Targeting Components and Use of the Online Tool • Timing 1 d

[0767] 1| Input target genomic DNA sequence. Applicants provide an online Cas9 targeting design tool that takes an input sequence of interest, identifies and ranks suitable target sites, and computationally predicts off-target sites for each intended target. Alternatively, one can manually select guide sequence by identifying the 20-bp sequence directly upstream of any 5'-NGG.

[0768] 2| Order necessary oligos and primers as specified by the online tool. If the site is chosen manually, the oligos and primers should be designed.

Preparation of sgRNA Expression Construct

[0769] 3| To generate the sgRNA expression construct, either the PCR- or plasmid-based protocol can be used.

[0770] (A) via PCR amplification • Timing 2 h

[0771] (i) Applicants prepare diluted U6 PCR template. Applicants recommend using PX330 as a PCR template, but any U6-containing plasmid may likewise be used as the PCR template. Applicants diluted template with ddH_2O to a concentration of 10 ng/ul. Note that if a plasmid or cassette already containing an U6-driven sgRNA is used as a template, a gel extraction needs to be performed to ensure that the product contains only the intended sgRNA and no trace sgRNA carryover from template.

[0772] (ii) Applicants prepared diluted PCR oligos. U6-Fwd and U6-sgRNA-Rev primers are diluted to a final concentration of 10 uM in ddH_2O (add 10 ul of 100 uM primer to 90 ul ddH_2O).

[0773] (iii) U6-sgRNA PCR reaction. Applicants set up the following reaction for each U6-sgRNA-Rev primer and mastermix as needed:

Component:	Amount (ul)	Final concentration
Herculase II PCR buffer, 5X	10	1X
dNTP, 100 mM (25 mM each)	0.5	1 mM
U6 template (PX330)	1	0.2 ng/ul
U6-Fwd primer	1	0.2 uM
U6-sgRNA-Rev primer (variable)	1	0.2 uM

-continued

Component:	Amount (ul)	Final concentration
Herculase II Fusion polymerase	0.5	
Distilled water	36	
Total	50	

[0774] (iv) Applicants performed PCR reaction on the reactions from step (iii) using the following cycling conditions:

Cycle number	Denature	Anneal	Extend
1	95° C., 2 m		
2-31	95° C., 20 s	60° C., 20 s	72° C., 20 s
32			72° C., 3 m

[0775] (v) After the reaction is completed, Applicants ran the product on a gel to verify successful, single-band amplification. Cast a 2% (wt/vol) agarose gel in 1×TBE buffer with 1× SYBR Safe dye. Run 5 ul of the PCR product in the gel at 15 V cm^{-1} for 20-30 min. Successful amplicons should yield one single 370-bp product and the template should be invisible. It should not be necessary to gel extract the PCR amplicon.

[0776] (vi) Applicants purified the PCR product using the QIAquick PCR purification kit according to the manufacturer's directions. Elute the DNA in 35 ul of Buffer EB or water. Purified PCR products may be stored at 4° C. or −20° C.

[0777] (B) Cloning sgRNA into Cas9-Containing Bicistronic Expression Vector • Timing 3 d

[0778] (i) Prepare the sgRNA oligo inserts. Applicants resuspended the top and bottom strands of oligos for each sgRNA design to a final concentration of 100 uM. Phosphorylate and anneal the oligo as follows:

Oligo 1 (100 uM)	1 ul
Oligo 2 (100 uM)	1 ul
T4 Ligation Buffer, 10X	1 ul
T4 PNK	1 ul
ddH_2O	6 ul
Total	10 ul

[0779] (ii) Anneal in a thermocycler using the following parameters:

[0780] 37° C. for 30 m

[0781] 95° C. for 5 m

[0782] Ramp down to 25° C. at 5° C. per m

[0783] (iii) Applicants diluted phosphorylated and annealed oligos 1:200 by add 1 ul of oligo to 199 ul room temperature ddH_2O.

[0784] (iv) Clone sgRNA oligo into PX330. Applicants set up Golden Gate reaction for each sgRNA. Applicants recommend also setting up a no-insert, PX330 only negative control.

PX330 (100 ng)	x ul
Diluted oligo duplex from step (iii)	2 ul
Tango Buffer, 10X	2 ul
DTT, 10 mM	1 ul
ATP, 10 mM	1 ul
FastDigest BbsI	1 ul
T7 Ligase	0.5 ul
ddH_2O	x ul
Total	20 ul

[0785] (v) Incubate the Golden Gate reaction for a total of 1 h:

Cycle number	Condition
1-6	37° C. for 5 m, 21° C. for 5 m

[0786] (vi) Applicants treated Golden Gate reaction with PlasmidSafe exonuclease to digest any residual linearized DNA. This step is optional but highly recommended.

Golden Gate reaction from step 4	11 ul
10X PlasmidSafe Buffer	1.5 ul
ATP, 10 mM	1.5 ul
PlasmidSafe exonuclease	1 ul
Total	15 ul

[0787] (vii) Applicants incubated the PlasmidSafe reaction at 37° C. for 30 min, followed by inactivation at 70° C. for 30 min. Pause point: after completion, the reaction may be frozen and continued later. The circular DNA should be stable for at least 1 week.

[0788] (viii) Transformation. Applicants transformed the PlasmidSafe-treated plasmid into a competent *E. coli* strain, according to the protocol supplied with the cells. Applicants recommend Stb13 for quick transformation. Briefly, Applicants added 5 ul of the product from step (vii) into 20 ul of ice-cold chemically competent Stb13 cells. This is then incubated on ice for 10 m, heat shocked at 42° C. for 30 s, returned immediately to ice for 2 m, 100 ul of SOC medium is added, and this is plated onto an LB plate containing 100 ug/ml ampicillin with incubation overnight at 37° C.

[0789] (ix) Day 2: Applicants inspected plates for colony growth. Typically, there are no colonies on the negative control plates (ligation of BbsI-digested PX330 only, no annealed sgRNA oligo), and tens to hundreds of colonies on the PX330-sgRNA cloning plates.

[0790] (x) From each plate, Applicants picked 2-3 colonies to check correct insertion of sgRNA. Applicants used a sterile pipette tip to inoculate a single colony into a 3 ml culture of LB medium with 100 ug/ml ampicillin. Incubate and shake at 37° C. overnight.

[0791] (xi) Day 3: Applicants isolated plasmid DNA from overnight cultures using a QiAprep Spin miniprep kit according to the manufacturer's instructions.

[0792] (xii) Sequence validate CRISPR plasmid. Applicants verified the sequence of each colony by sequencing from the U6 promoter using the U6-Fwd primer. Optional: sequence the Cas9 gene using primers listed in the following Primer table.

Primer	Sequence (5' to 3')	Purpose
U6-For	GAGGGCCTATTTCCCATGATTCC	Amplify U6-sgRNA.
U6-Rev	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACNNNNNNNNNNNNNNNNNNNCCGGTGTTTCGTCCTTTCCACAAG	Amplify U6-sgRNA; N is reverse complement of target
sgRNA-top	CACCGNNNNNNNNNNNNNNNNNNN	Clone sgRNA into PX330
sgRNA-bottom	AAACNNNNNNNNNNNNNNNNNNNC	Clone sgRNA into PX330
U6-EMX1-Rev	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCCTAGTCATTGGAGGTGACCGGTGTTTCGTCCTTTCCACAAG	Amplify U6-EMX1 sgRNA
EMX1-top	CACCGTCACCTCCAATGACTAGGG	Clone EMX1 sgRNA into PX330
EMX1-bottom	AAACCCCTAGTCATTGGAGGTGAC	Clone EMX1 sgRNA into PX330
ssODN-sense	CAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGACATCGATGTCACCTCCAATGACAAGCTTGCTAGCGGTGGGCAACCACAAACCCACGAGGGCAGAGTGCTGCTTGCTGCTGGCCAGGCCCCTGCGTGGGCCCAAGCTGGACTCTGGCCACTCCCT	EMX1 HDR (sense; insertion underlined)
ssODN-antisense	AGGGAGTGGCCAGAGTCCAGCTTGGGCCCACGCAGGGGCCTGGCCAGCAGCAAGCAGCACTCTGCCCTCGTGGGTTTGTGGTTGCCCACCGCTAGCAAGCTTGTCATTGGAGGTGACATCGATGTCCTCCCCATTGGCCTGCTTCGTGGCAATGCGCCACCGGTTGATGTGATGGGAGCCCTTCTTCTTCTG	EMX1 HDR (antisense; insertion underlined)
EMX1-SURV-F	CCATCCCCTTCTGTGAATGT	EMX1 SURVEYOR assay PCR, sequencing
EMX1-SURV-R	GGAGATTGGAGACACGGAGA	EMX1 SURVEYOR assay PCR, sequencing
EMX1-HDR-F	GGCTCCCTGGGTTCAAAGTA	EMX1 RFLP analysis PCR, sequencing
EMX1-HDR-R	AGAGGGGTCTGGATGTCGTAA	EMX1 RFLP analysis PCR, sequencing
PUC19-F	CGCCAGGGTTTTCCCAGTCACGAC	pUC19 multiple cloning site F primer, for Sanger sequencing

[0793] Applicants referenced the sequencing results against the PX330 cloning vector sequence to check that the 20 bp guide sequence was inserted between the U6 promoter and the remainder of the sgRNA scaffold. Details and sequence of the PX330 map in GenBank vector map format (*.gb file) can be found at the website crispr.genome-engineering.org.

[0794] (Optional) Design of ssODN Template • Timing 3 d Planning Ahead

[0795] 3| Design and order ssODN. Either the sense or antisense ssODN can be purchased directly from supplier. Applicants recommend designing homology arms of at least 40 bp on either side and 90 bp for optimal HDR efficiency. In Applicants' experience, antisense oligos have slightly higher modification efficiencies.

[0796] 4| Applicants resuspended and diluted ssODN ultramers to a final concentration of 10 uM. Do not combine or anneal the sense and antisense ssODNs. Store at −20° C.

[0797] 5| Note for HDR applications, Applicants recommend cloning sgRNA into the PX330 plasmid.

[0798] Functional Validation of sgRNAs: Cell Culture and Transfections • Timing 3-4 d

[0799] The CRISPR-Cas system has been used in a number of mammalian cell lines. Conditions may vary for each cell line. The protocols below details transfection conditions for HEK239FT cells. Note for ssODN-mediated HDR transfections, the Amaxa SF Cell Line Nucleofector Kit is used for optimal delivery of ssODNs. This is described in the next section.

[0800] 7| HEK293FT maintenance. Cells are maintained according to the manufacturer's recommendations. Briefly, Applicants cultured cells in D10 medium (GlutaMax DMEM supplemented with 10% Fetal Bovine Serum), at 37° C. and 5% CO2.

[0801] 8| To passage, Applicants removed medium and rinsed once by gently adding DPBS to side of vessel, so as not to dislodge cells. Applicants added 2 ml of TrypLE to a T75 flask and incubated for 5 m at 37° C. 10 ml of warm D10 medium is added to inactivate and transferred to a 50 ml Falcon tube. Applicants dissociated cells by triturating gently, and re-seeded new flasks as necessary. Applicants typically passage cells every 2-3 d at a split ratio of 1:4 or 1:8, never allowing cells to reach more than 70% confluency. Cell lines are restarted upon reaching passage number 15.

[0802] 9| Prepare cells for transfection. Applicants plated well-dissociated cells onto 24-well plates in D10 medium without antibiotics 16-24 h before transfection at a seeding density of 1.3×10^5 cells per well and a seeding volume of 500 ul. Scale up or down according to the manufacturer's manual as needed. It is suggested to not plate more cells than recommended density as doing so may reduce transfection efficiency.

[0803] 10| On the day of transfection, cells are optimal at 70-90% confluency. Cells may be transfected with Lipofectamine 2000 or Amaxa SF Cell Line Nucleofector Kit according to the manufacturers' protocols.

[0804] (A) For sgRNAs cloned into PX330, Applicants transfected 500 ng of sequence-verified CRISPR plasmid; if transfecting more than one plasmid, mix at equimolar ratio and no more than 500 ng total.

[0805] (B) For sgRNA amplified by PCR, Applicants mixed the following:

PX165 (Cas9 only)	200 ng
sgRNA amplicon (each)	40 ng
pUC19	fill up total DNA to 500 ng

[0806] Applicants recommend transfecting in technical triplicates for reliable quantification and including transfection controls (e.g. GFP plasmid) to monitor transfection efficiency. In addition, PX330 cloning plasmid and/or sgRNA amplicon may be transfected alone as a negative control for downstream functional assays.

[0807] 11| Applicants added Lipofectamine complex to cells gently as HEK293FT cells may detach easily from plate easily and result in lower transfection efficiency.

[0808] 12| Applicants checked cells 24 h after transfection for efficiency by estimating the fraction of fluorescent cells in the control (e.g., GFP) transfection using a fluorescence microscope. Typically cells are more than 70% transfected.

[0809] 13| Applicants supplemented the culture medium with an additional 500 ul of warm D10 medium. Add D10 very slowly to the side of the well and do not use cold medium, as cells can detach easily.

[0810] 14| Cells are incubated for a total of 48-72 h post-transfection before harvested for indel analysis. Indel efficiency does not increase noticeably after 48 h.

[0811] (Optional) Co-Transfection of CRISPR Plasmids and ssODNs or Targeting Plasmids for HR • Timing 3-4 d

[0812] 15| Linearize targeting plasmid. Targeting vector is linearized if possible by cutting once at a restriction site in the vector backbone near one of the homology arms or at the distal end of either homology arm.

[0813] 16| Applicants ran a small amount of the linearized plasmid alongside uncut plasmid on a 0.8-1% agarose gel to check successful linearization. Linearized plasmid should run above the supercoiled plasmid.

[0814] 17| Applicants purified linearized plasmid with the QIAQuick PCR Purification kit.

[0815] 18| Prepare cells for transfection. Applicants cultured HEK293FT in T75 or T225 flasks. Sufficient cell count before day of transfection is planned for. For the Amaxa strip-cuvette format, 2×10^6 cells are used per transfection.

[0816] 19| Prepare plates for transfection. Applicants added 1 ml of warm D10 medium into each well of a 12 well plate. Plates are placed into the incubator to keep medium warm.

[0817] 20| Nucleofection. Applicants transfected HEK293FT cells according to the Amaxa SF Cell Line Nucleofector 4D Kit manufacturer's instructions, adapted in the steps below.

[0818] a. For ssODN and CRISPR cotransfection, pre-mix the following DNA in PCR tubes:

pCRISPR plasmid (Cas9 + sgRNA)	500 ng
ssODN template (10 uM)	1 ul

[0819] b. For HDR targeting plasmid and CRISPR cotransfection, pre-mix the following DNA in PCR tubes:

CRISPR plasmid (Cas9 + sgRNA)	500 ng
Linearized targeting plasmid	500 ng

[0820] For transfection controls, see previous section. In addition, Applicants recommend transfecting ssODN or targeting plasmid alone as a negative control.

[0821] 21| Dissociate to single cells. Applicants removed medium and rinsed once gently with DPBS, taking care not to dislodge cells. 2 ml of TrypLE is added to a T75 flask and incubated for 5 m at 37° C. 10 ml of warm D10 medium is added to inactivate and triturated gently in a 50 ml Falcon tube. It is recommended that cells are triturated gently and dissociated to single cells. Large clumps will reduce transfection efficiency. Applicants took a 10 ul aliquot from the suspension and diluted into 90 ul of D10 medium for counting. Applicants counted cells and calculated the number of cells and volume of suspension needed for transfection. Applicants typically transfected 2×10^5 cells per condition using the Amaxa Nucleocuvette strips, and recommend calculating for 20% more cells than required to adjust for volume loss in subsequent pipetting steps. The volume needed is transferred into a new Falcon tube.

[0822] 23| Applicants spun down the new tube at 200×g for 5 m.

[0823] Applicants prepared the transfection solution by mixing the SF solution and S supplement as recommended by Amaxa. For Amaxa strip-cuvettes, a total of 20 ul of supplemented SF solution is needed per transfection. Likewise, Applicants recommend calculating for 20% more volume than required.

[0824] 25| Applicants removed medium completely from pelleted cells from step 23 and gently resuspended in appropriate volume (20 ul per 2×10^5 cells) of S1-supplemented SF solution. Do not leave cells in SF solution for extended period of time.

[0825] 26| 20 ul of resuspended cells is pipetted into each DNA pre-mix from step 20. Pipette gently to mix and transfer to Nucleocuvette strip chamber. This is repeated for each transfection condition.

[0826] Electroporate cells using the Nucleofector 4D program recommended by Amaxa, CM-130.

[0827] 28| Applicants gently and slowly pipetted 100 ul of warm D10 medium into each Nucleocuvette strip chamber, and transferred all volume into the pre-warmed plate from step 19. CRITICAL. Cells are very fragile at this stage and harsh pipetting can cause cell death. Incubate for 24 h. At this point, transfection efficiency can be estimated from fraction of fluorescent cells in positive transfection control. Nucleofection typically results in greater than 70-80% transfection efficiency. Applicants slowly added 1 ml warm D10 medium to each well without dislodging the cells. Incubate cells for a total of 72 h.

[0828] Human Embryonic Stem Cell (HUES 9) Culture and Transfection • Timing 3-4 d

[0829] Maintaining hESC (HUES9) line. Applicants routinely maintain HUES9 cell line in feeder-free conditions with mTesR1 medium. Applicants prepared mTeSR1 medium by adding the 5× supplement included with basal medium and 100 ug/ml Normocin. Applicants prepared a 10 ml aliquot of mTeSR1 medium supplemented further with 10 uM Rock Inhibitor. Coat tissue culture plate. Dilute cold GelTrex 1:100 in cold DMEM and coat the entire surface of a 100 mm tissue culture plate.

[0830] Place plate in incubator for at least 30 m at 37° C. Thaw out a vial of cells at 37° C. in a 15 ml Falcon tube, add 5 ml of mTeSR1 medium, and pellet at 200×g for 5 m. Aspirate off GelTrex coating and seed ~1×106 cells with 10 ml mTeSR1 medium containing Rock Inhibitor. Change to normal mTeSR1 medium 24 h after transfection and re-feed daily. Passaging cells. Re-feed cells with fresh mTeSR1 medium daily and passage before reaching 70% confluency. Aspirate off mTeSR1 medium and wash cells once with DPBS. Dissociate cells by adding 2 ml Accutase and incubating at 37° C. for 3-5 m. Add 10 ml mTeSR1 medium to detached cells, transfer to 15 ml Falcon tube and resuspend gently. Re-plate onto GelTrex-coated plates in mTeSR1 medium with 10 uM Rock Inhibitor. Change to normal mTeSR1 medium 24 h after plating.

[0831] Transfection. Applicants recommend culturing cells for at least 1 week post-thaw before transfecting using the Amaxa P3 Primary Cell 4-D Nucleofector Kit (Lonza). Re-feed log-phase growing cells with fresh medium 2 h before transfection. Dissociate to single cells or small clusters of no more than 10 cells with accutase and gentle resuspension. Count the number of cells needed for nucleofection and spin down at 200×g for 5 m. Remove medium completely and resuspend in recommended volume of S1-supplemented P3 nucleofection solution. Gently plate electroporated cells into coated plates in presence of 1× Rock Inhibitor.

[0832] Check transfection success and re-feed daily with regular mTeSR1 medium beginning 24 h after nucleofection. Typically, Applicants observe greater than 70% transfection efficiency with Amaxa Nucleofection. Harvest DNA. 48-72 h post transfection, dissociate cells using accutase and inactivate by adding 5× volume of mTeSR1. Spin cells down at 200×g for 5 m. Pelleted cells can be directed processed for DNA extraction with QuickExtract solution. It is recommended to not mechanically dissociate cells without accutase. It is recommended to not spin cells down without inactivating accutase or above the recommended speed; doing so may cause cells to lyse.

Isolation of Clonal Cell Lines by FACS. Timing • 2-3 h Hands-on; 2-3 Weeks Expansion

[0833] Clonal isolation may be performed 24 h post-transfection by FACS or by serial dilution.

[0834] 54| Prepare FACS buffer. Cells that do not need sorting using colored fluorescence may be sorted in regular D10 medium supplemented with 1× penicillin/streptinomycin. If colored fluorescence sorting is also required, a phenol-free DMEM or DPBS is substituted for normal DMEM. Supplement with 1× penicillin/streptinomycin and filter through a 0.22 um Steriflip filter.

[0835] 51| Prepare 96 well plates. Applicants added 100 ul of D10 media supplemented with 1× penicillin/streptinomycin per well and prepared the number of plates as needed for the desired number of clones.

[0836] 56| Prepare cells for FACS. Applicants dissociated cells by aspirating the medium completely and adding 100 ul TrypLE per well of a 24-well plate. Incubate for 5 m and add 400 ul warm D10 media.

[0837] 57| Resuspended cells are transferred into a 15 ml Falcon tube and gently triturated 20 times. Recommended to check under the microscope to ensure dissociation to single cells.

[0838] 58| Spin down cells at 200×g for 5 minutes.

[0839] 59| Applicants aspirated the media, and resuspended the cells in 200 ul of FACS media.

[0840] 60| Cells are filtered through a 35 um mesh filter into labeled FACS tubes. Applicants recommend using the BD Falcon 12×75 mm Tube with Cell Strainer cap. Place cells on ice until sorting.

[0841] 61| Applicants sorted single cells into 96-well plates prepared from step 55. Applicants recommend that in one single designated well on each plate, sort 100 cells as a positive control.

[0842] NOTE. The remainder of the cells may be kept and used for genotyping at the population level to gauge overall modification efficiency.

[0843] 62| Applicants returned cells into the incubator and allowed them to expand for 2-3 weeks. 100 ul of warm D10 medium is added 5 d post sorting. Change 100 ul of medium every 3-5 d as necessary.

[0844] 63| Colonies are inspected for "clonal" appearance 1 week post sorting: rounded colonies radiating from a central point. Mark off wells that are empty or may have been seeded with doublets or multiplets.

[0845] 64| When cells are more than 60% confluent, Applicants prepared a set of replica plates for passaging. 100 ul of D10 medium is added to each well in the replica plates. Applicants dissociated cells directly by pipetting up and down vigorously 20 times. 20% of the resuspended volume was plated into the prepared replica plates to keep the clonal lines. Change the medium every 2-3 d thereafter and passage accordingly.

[0846] 65| Use the remainder 80% of cells for DNA isolation and genotyping.

[0847] Optional: Isolation of Clonal Cell Lines by Dilution. Timing • 2-3 h Hands-on; 2-3 Weeks Expansion

[0848] 66| Applicants dissociated cells from 24-well plates as described above. Make sure to dissociate to single cells. A cell strainer can be used to prevent clumping of cells.

[0849] 67| The number of cells are counted in each condition. Serially dilute each condition in D10 medium to a final concentration of 0.5 cells per 100 ul. For each 96 well plate, Applicants recommend diluting to a final count of 60 cells in 12 ml of D10. Accurate count of cell number is recommended for appropriate clonal dilution. Cells may be recounted at an intermediate serial dilution stage to ensure accuracy.

[0850] 68| Multichannel pipette was used to pipette 100 ul of diluted cells to each well of a 96 well plate.

[0851] NOTE. The remainder of the cells may be kept and used for genotyping at the population level to gauge overall modification efficiency.

[0852] 69| Applicants inspected colonies for "clonal" appearance ~1 week post plating: rounded colonies radiating from a central point. Mark off wells that may have seeded with doublets or multiplets.

[0853] 70| Applicants returned cells to the incubator and allowed them to expand for 2-3 weeks. Re-feed cells as needed as detailed in previous section.

[0854] SURVEYOR Assay for CRISPR Cleavage Efficiency. Timing • 5-6 h

[0855] Before assaying cleavage efficiency of transfected cells, Applicants recommend testing each new SURVEYOR primer on negative (untransfected) control samples through the step of SURVEYOR nuclease digestion using the protocol described below. Occasionally, even single-band clean SURVEYOR PCR products can yield non-specific SURVEYOR nuclease cleavage bands and potentially interfere with accurate indel analysis.

[0856] 71| Harvest cells for DNA. Dissociate cells and spin down at 200×g for 5 m. NOTE. Replica plate at this stage as needed to keep transfected cell lines.

[0857] 72| Aspirate the supernatant completely.

[0858] 73| Applicants used QuickExtract DNA extraction solution according to the manufacturer's instructions. Applicants typically used 50 ul of the solution for each well of a 24 well plate and 10 ul for a 96 well plate.

[0859] 74| Applicants normalized extracted DNA to a final concentration of 100-200 ng/ul with ddH$_2$O. Pause point: Extracted DNA may be stored at −20° C. for several months.

[0860] 75| Set up the SURVEYOR PCR. Master mix the following using SURVEYOR primers provided by Applicants online/computer algorithm tool:

Component:	Amount (ul)	Final concentration
Herculase II PCR buffer, 5X	10	1X
dNTP, 100 mM (25 mM each)	1	1 mM
SURVEYOR Fwd primer (10 uM)	1	0.2 uM
SURVEYOR Rev primer (10 uM)	1	0.2 uM
Herculase II Fusion polymerase	1	
MgCl$_2$ (25 mM)	2	1 mM
Distilled water	33	
Total	49 (for each reaction)	

[0861] 76| Applicants added 100-200 ng of normalized genomic DNA template from step 74 for each reaction.

[0862] 77| PCR reaction was performed using the following cycling conditions, for no more than 30 amplification cycles:

Cycle number	Denature	Anneal	Extend
1	95° C., 2 min		
2-31	95° C., 20 s	60° C., 20 s	72° C., 30 s
32			72° C., 3 min

[0863] 78| Applicants ran 2-5 ul of PCR product on a 1% gel to check for single-band product. Although these PCR conditions are designed to work with most pairs of SURVEYOR primers, some primers may need additional optimization by adjusting the template concentration, MgCl$_2$ concentration, and/or the annealing temperature.

[0864] 79| Applicants purified the PCR reactions using the QIAQuick PCR purification kit and normalized eluant to 20 ng/ul. Pause point: Purified PCR product may be stored at −20° C.

[0865] 80| DNA heteroduplex formation. The annealing reaction was set up as follows:

Taq PCR buffer, 10X	2 ul
Normalized DNA (20 ng/ul)	18 ul
Total volume	20 ul

[0866] 81| Anneal the reaction using the following conditions:

Cycle number	Condition
1	95° C., 10 mn
2	95° C.-85° C., −2° C./s
3	85° C., 1 min
4	85° C.-75° C., −0.3° C./s
5	75° C., 1 min
6	75° C.-65° C., −0.3° C./s
7	65° C., 1 min
8	65° C.-55° C., −0.3° C./s
9	55° C., 1 min
10	55° C.-45° C., −0.3° C./s
11	45° C., 1 min
12	45° C.-35° C., −0.3° C./s
13	35° C., 1 min
14	35° C.-25° C., −0.3° C./s
15	25° C., 1 min

[0867] 82| SURVEYOR nuclease S digestion. Applicants prepared master-mix and added the following components on ice to annealed heteroduplexes from step 81 for a total final volume of 25 ul:

Component	Amount (ul)	Final Concentration
MgCl$_2$ solution, 0.15 M	2.5	15 mM
ddH$_2$O	0.5	
SURVEYOR nuclease S	1	1X
SURVEYOR enhancer S	1	1X
Total	5	

[0868] 83| Vortex well and spin down. Incubate the reaction at 42° C. for 1 h.

[0869] 84| Optional: 2 ul of the Stop Solution from the SURVEYOR kit may be added. Pause point. The digested product may be stored at −20° C. for analysis at a later time.

[0870] 85| Visualize the SURVEYOR reaction. SURVEYOR nuclease digestion products may be visualized on a 2% agarose gel. For better resolution, products may be run on a 4-20% gradient Polyacrylamide TBE gel. Applicants loaded 10 ul of product with the recommended loading buffer and ran the gel according to manufacturer's instructions. Typically, Applicants run until the bromophenol blue dye has migrated to the bottom of the gel. Include DNA ladder and negative controls on the same gel.

[0871] 86| Applicants stained the gel with 1×SYBR Gold dye diluted in TBE. The gel was gently rocked for 15 m.

[0872] 87| Applicants imaged the gel using a quantitative imaging system without overexposing the bands. The negative controls should have only one band corresponding to the size of the PCR product, but may have occasionally non-specific cleavage bands of other sizes. These will not interfere with analysis if they are different in size from target cleavage bands. The sum of target cleavage band sizes, provided by Applicants online/computer algorithm tool, should be equal to the size of the PCR product.

[0873] 88| Estimate the cleavage intensity. Applicants quantified the integrated intensity of each band using ImageJ or other gel quantification software.

[0874] 89| For each lane, Applicants calculated the fraction of the PCR product cleaved (f_{cut}) using the following formula: $f_{cut}=(b+c)/(a+b+c)$, where a is the integrated intensity of the undigested PCR product and b and c are the integrated intensities of each cleavage product. 901 Cleavage efficiency may be estimated using the following formula, based on the binomial probability distribution of duplex formation:

$$91|\text{indel }(\%)=100\times(1-\sqrt{(1-f_{cut})})$$

[0875] Sanger Sequencing for Assessing CRISPR Cleavage Efficiency. Timing. 3 d

[0876] Initial steps are identical to Steps 71-79 of the SURVEYOR assay. Note: SURVEYOR primers may be used for Sanger sequencing if appropriate restriction sites are appended to the Forward and Reverse primers. For cloning into the recommended pUC19 backbone, EcoRI may be used for the Fwd primer and HindIII for the Rev primer.

[0877] 92| Amplicon digestion. Set up the digestion reaction as follows:

Component	Amount (ul)
Fast Digest buffer, 10X	3
FastDigest EcoRI	1
FastDigest HindIII	1
Normalized DNA (20 ng/ul)	10
ddH_2O	15
Total volume	30

[0878] 93| pUC19 backbone digestion. Set up the digestion reaction as follows:

Component	Amount (ul)
Fast Digest buffer, 10X	3
FastDigest EcoRI	1
FastDigest HindIII	1
FastAP Alkaline Phosphatase	1
pUC19 vector (200 ng/ul)	5
ddH_2O	20
Total volume	30

[0879] 94| Applicants purified the digestion reactions using the QIAQuick PCR purification kit. Pause point: Purified PCR product may be stored at −20° C.

[0880] 95| Applicants ligated the digested pUC19 backbone and Sanger amplicons at a 1:3 vector:insert ratio as follows:

Component	Amount (ul)
Digested pUC19	x (50 ng)
Digested insert	x (1:3 vector:insert molar ratio)
T7 ligase	1
2X Rapid Ligation Buffer	10
ddH_2O	x
Total volume	20

[0881] 96| Transformation. Applicants transformed the PlasmidSafe-treated plasmid into a competent *E. coli* strain, according to the protocol supplied with the cells. Applicants recommend Stb13 for quick transformation. Briefly, 5 ul of the product from step 95 is added into 20 ul of ice-cold chemically competent Stb13 cells, incubated on ice for 10 m, heat shocked at 42° C. for 30 s, returned immediately to ice for 2 m, 100 ul of SOC medium is added, and plated onto an LB plate containing 100 ug/ml ampicillin. This is incubated overnight at 37° C.

[0882] 97| Day 2: Applicants inspected plates for colony growth. Typically, there are no colonies on the negative control plates (ligation of EcoRI-HindIII digested pUC19 only, no Sanger amplicon insert), and tens to hundreds of colonies on the pUC19-Sanger amplicon cloning plates.

[0883] 98| Day 3: Applicants isolated plasmid DNA from overnight cultures using a QIAprep Spin miniprep kit according to the manufacturer's instructions.

[0884] 99| Sanger sequencing. Applicants verified the sequence of each colony by sequencing from the pUC19 backbone using the pUC19-For primer. Applicants referenced the sequencing results against the expected genomic DNA sequence to check for the presence of Cas9-induced NHEJ mutations. % editing efficiency=(# modified clones)/(# total clones). It is important to pick a reasonable number of clones (>24) to generate accurate modification efficiencies.

[0885] Genotyping for Microdeletion. Timing • 2-3 d Hands on; 2-3 Weeks Expansion

[0886] 100| Cells were transfected as described above with a pair of sgRNAs targeting the region to be deleted.

[0887] 101| 24 h post-transfection, clonal lines are isolated by FACS or serial dilution as described above.

[0888] 102| Cells are expanded for 2-3 weeks.

[0889] 103| Applicants harvested DNA from clonal lines as described above using 10 ul QuickExtract solution and normalized genomic DNA with ddH_2O to a final concentration of 50-100 ng/ul.

[0890] 104| PCR Amplify the modified region. The PCR reaction is set up as follows:

Component:	Amount (ul)	Final concentration
Herculase II PCR buffer, 5X	10	1X
dNTP, 100 mM (25 mM each)	1	1 mM
Out Fwd primer (10 uM)	1	0.2 uM
Out Rev primer (10 uM)	1	0.2 uM
Herculase II Fusion polymerase	1	
MgCl2 (25 mM)	2	1 mM
ddH_2O	32	
Total	48 (for each reaction)	

[0891] Note: if deletion size is more than 1 kb, set up a parallel set of PCR reactions with In-Fwd and In-Rev primers to screen for the presence of the wt allele.

[0892] 105| To screen for inversions, a PCR reaction is set up as follows:

Component:	Amount (ul)	Final concentration
Herculase II PCR buffer, 5X	10	1X
dNTP, 100 mM (25 mM each)	1	1 mM
Out Fwd or Out-Rev primer (10 uM)	1	0.2 uM
In Fwd or In-Rev primer (10 uM)	1	0.2 uM
Herculase II Fusion polymerase	1	
$MgCl_2$ (25 mM)	2	1 mM
ddH_2O	32	
Total	48 (for each reaction)	

[0893] Note: primers are paired either as Out-Fwd+In Fwd, or Out-Rev+In-Rev.

[0894] 106| Applicants added 100-200 ng of normalized genomic DNA template from step 103 for each reaction.

[0895] 107| PCR reaction was performed using the following cycling conditions:

Cycle number	Denature	Anneal	Extend
1	95° C., 2 min		
2-31	95° C., 20 s	60° C., 20 s	72° C., 30 s
32			72° C., 3 m

[0896] 108| Applicants run 2-5 ul of PCR product on a 1-2% gel to check for product. Although these PCR conditions are designed to work with most primers, some primers may need additional optimization by adjusting the template concentration, $MgCl_2$ concentration, and/or the annealing temperature.

[0897] Genotyping for Targeted Modifications Via HDR. Timing ∩ 2-3 d, 2-3 h Hands on

[0898] 109| Applicants harvested DNA as described above using QuickExtract solution and normalized genomic DNA with TE to a final concentration of 100-200 ng/ul.

[0899] 110| PCR Amplify the modified region. The PCR reaction is set up as follows:

Component:	Amount (ul)	Final concentration
Herculase II PCR buffer, 5X	10	1X
dNTP, 100 mM (25 mM each)	1	1 mM
HDR Fwd primer (10 uM)	1	0.2 uM
HDR Rev primer (10 uM)	1	0.2 uM
Herculase II Fusion polymerase	1	
$MgCl_2$ (25 mM)	2	1 mM
ddH_2O	33	
Total	49 (for each reaction)	

[0900] 111| Applicants added 100-200 ng of genomic DNA template from step 109 for each reaction and run the following program.

Cycle number	Denature	Anneal	Extend
1	95° C., 2 min		
2-31	95° C., 20 s	60° C., 20 s	72° C., 30-60 s per kb
32			72° C., 3 min

[0901] 112| Applicants ran 5 ul of PCR product on a 0.8-1% gel to check for single-band product. Primers may need additional optimization by adjusting the template concentration, $MgCl_2$ concentration, and/or the annealing temperature.

[0902] 113| Applicants purified the PCR reactions using the QIAQuick PCR purification kit.

[0903] 114| In the HDR example, a HindIII restriction site is inserted into the EMX1 gene. These are detected by a restriction digest of the PCR amplicon:

Component	Amount (ul)
Purified PCR amplicon (200-300 ng)	x
F.D. buffer, Green	1
HindIII	0.5
ddH2O	x
Total	10

[0904] i. The DNA is digested for 10 m at 37° C.:

[0905] ii. Applicants ran 10 ul of the digested product with loading dye on a 4-20% gradient polyacrylamide TBE gel until the xylene cyanol band had migrated to the bottom of the gel.

[0906] iii. Applicants stained the gel with 1×SYBR Gold dye while rocking for 15 m.

[0907] iv. The cleavage products are imaged and quantified as described above in the SURVEYOR assay section. HDR efficiency is estimated by the formula: (b+c)/(a+b+c), where a is the integrated intensity for the undigested HDR PCR product, and b and c are the integrated intensities for the HindIII-cut fragments.

[0908] 115| Alternatively, purified PCR amplicons from step 113 may be cloned and genotyped using Sanger sequencing or NGS.

[0909] Deep Sequencing and Off-Target Analysis • Timing 1-2 d

[0910] The online CRISPR target design tool generates candidate genomic off-target sites for each identified target site. Off-target analysis at these sites can be performed by SURVEYOR nuclease assay, Sanger sequencing, or next-generation deep sequencing. Given the likelihood of low or undetectable modification rates at many of these sites, Applicants recommend deep sequencing with the Illumina Miseq platform for high sensitivity and accuracy. Protocols will vary with sequencing platform; here, Applicants briefly describe a fusion PCR method for attaching sequencing adapters.

[0911] 116| Design deep sequencing primers. Next-generation sequencing (NGS) primers are designed for shorter amplicons, typically in the 100-200 bp size range. Primers may be manually designed using NCBI Primer-Blast or generated with online CRISPR target design tools (website at genome-engineering.org/tools).

[0912] 117| Harvest genomic DNA from Cas9-targeted cells. Normalize QuickExtract genomic DNA to 100-200 ng/ul with ddH2O.

[0913] 118| Initial library preparation PCR. Using the NGS primers from step 116, prepare the initial library preparation PCR

Component:	Amount (ul)	Final concentration
Herculase II PCR buffer, 5X	10	1X
dNTP, 100 mM (25 mM each)	1	1 mM
NGS Fwd primer (10 uM)	1	0.2 uM
NGS Rev primer (10 uM)	1	0.2 uM
Herculase II Fusion polymerase	1	
MgCl2 (25 mM)	2	1 mM
ddH2O	33	
Total	49 (for each reaction)	

[0914] 119| Add 100-200 ng of normalized genomic DNA template for each reaction.

[0915] 120|Perform PCR reaction using the following cycling conditions, for no more than 20 amplification cycles:

Cycle number	Denature	Anneal	Extend
1	95° C., 2 min		
2-21	95° C., 20 s	60° C., 20 s	72° C., 15 s
27			72° C., 3 min

[0916] 121| Run 2-5 ul of PCR product on a 1% gel to check for single-band product. As with all genomic DNA PCRs, NGS primers may require additional optimization by adjusting the template concentration, $MgCl_2$ concentration, and/or the annealing temperature.

[0917] 122| Purify the PCR reactions using the QIAQuick PCR purification kit and normalize eluant to 20 ng/ul. Pause point: Purified PCR product may be stored at −20° C.

[0918] 123| Nextera AT DNA Sample Preparation Kit. Following the manufacturer's protocol, generate Miseq sequencing-ready libraries with unique barcodes for each sample.

[0919] 124| Analyze sequencing data. Off-target analysis may be performed through read alignment programs such as ClustalW, Geneious, or simple sequence analysis scripts.

[0920] Timing

[0921] Steps 1-2 Design and synthesis of sgRNA oligos and ssODNs: 1-5 d, variable depending on supplier

[0922] Steps 3-5 Construction of CRISPR plasmid or PCR expression cassette: 2 h to 3 d

[0923] Steps 6-53 Transfection into cell lines: 3 d (1 h hands-on time)

[0924] Steps 54-70 Optional derivation of clonal lines: 1-3 weeks, variable depending on cell type

[0925] Steps 71-91 Functional validation of NHEJ via SURVEYOR: 5-6 h

[0926] Steps 92-124 Genotyping via Sanger or next-gen deep sequencing: 2-3 d (3-4 h hands on time)

Addressing Situations Concerning Herein Examples

[0927]

Situation	Solution
No amplification of sgRNA	Titrate U6-template concentration
SURVEYOR or HDR PCR dirty or no amplification	Titrate MgCl2; normalize and titrate template concentration; annealing temp gradient; redesign primers
Unequal amplification of alleles in microdeletion PCRs	Set up separate PCRs to detect wildtype and deletion alleles; Redesign primers with similar sized amplicons
Colonies on negative control plate	Increase BbsI; increase Golden Gate reaction cycle number, cut PX330 separately with Antarctic Phosphate treatment
No sgRNA sequences or wrong sequences	Screen additional colonies
Low lipofectamine transfection efficiency	Check cell health and density; titrate DNA; add GFP transfection control
Low nucleofection transfection efficiency	Check cell health and density; titrate DNA; suspend to single cell
Clumps or no cells after FACS	Filter cells before FACS; dissociate to single cells; resuspend in appropriate density
Clumps or no cells in serial dilution	Recount cells; dissociate to single cells and filter through strainer; check serial dilution
High SURVEYOR. background on negative sample	Redesign primers to prime from different locations
Dirty SURVEYOR result on gel	Purify PCR product; reduce input DNA; reduce 42° C. incubation to 30 m
No SURVEYOR cleavage	Purify and normalize PCR product; re-anneal with TaqB buffer; Redesign sgRNAs; sequence verify Cas9 on px330 backbone
Samples do not sink in TBE acrylamide gel	Supplement with MgCl2 to a final concentration of 15 mM or add loading buffer containing glycerol

[0928] Discussion

[0929] CRISPR-Cas may be easily multiplexed to facilitate simultaneous modification of several genes and mediate chromosomal microdeletions at high efficiencies. Applicants used two sgRNAs to demonstrate simultaneous targeting of the human GRIN2B and DYRK1A loci at efficiencies of up to 68% in HEK293FT cells. Likewise, a pair of sgRNAs may be used to mediate microdeletions, such as excision of an exon, which can be genotyped by PCR on a clonal level. Note that the precise location of exon junctions can vary. Applicants also demonstrated the use of ssODNs and targeting vector to mediate HDR with both wildtype and nickase mutant of Cas9 in HEK 293FT and HUES9 cells (FIG. **30**). Note that Applicants have not been able to detect HDR in HUES9 cells using the Cas9 nickase, which may be due to low efficiency or a potential difference in repair activities in HUES9 cells. Although these values are typical, there is some variability in the cleavage efficiency of a given sgRNA, and on rare occasions certain sgRNAs may not work for reasons yet unknown. Applicants recommend designing two sgRNAs for each locus, and testing their efficiencies in the intended cell type.

Example 31

NLSs

[0930] Cas9 Transcriptional Modulator: Applicants set out to turn the Cas9/gRNA CRISPR system into a generalized DNA binding system in which functions beyond DNA cleavage can be executed. For instance, by fusing functional domain(s) onto a catalytically inactive Cas9 Applicants have imparted novel functions, such as transcriptional activation/repression, methylation/demethylation, or chromatin modifications. To accomplish this goal Applicants made a catalytically inactive Cas9 mutant by changing two residues essential for nuclease activity, D10 and H840, to alanine. By mutating these two residues the nuclease activity of Cas9 is abolished while maintaining the ability to bind target DNA. The functional domains Applicants decided to focus on to test Applicants' hypothesis are the transcriptional activator VP64 and the transcriptional repressors SID and KRAB.

[0931] Cas9 Nuclear localization: Applicants hypothesized that the most effective Cas9 transcriptional modulator would be strongly localized to the nucleus where it would have its greatest influence on transcription. Moreover, any residual Cas9 in the cytoplasm could have unwanted effects. Applicants determined that wild-type Cas9 does not localize into the nucleus without including multiple nuclear localization signals (NLSs) (although a CRISPR system need not have one or more NLSs but advantageously has at least one or more NLS(s)). Because multiple NLS sequences were required it was reasoned that it is difficult to get Cas9 into the nucleus and any additional domain that is fused to Cas9 could disrupt the nuclear localization. Therefore, Applicants made four Cas9-VP64-GFP fusion constructs with different NLS sequences (pXRP02—pLenti2-EF1a-NLS-hSpCsn1 (10A, 840A)-NLS-VP64-EGFP, pXRP04—pLenti2-EF1a-NLS-hSpCsn1(10A,840A)-NLS-VP64-2A-EGFP-NLS, pXRP06—pLenti2-EF1a-NLS-EGFP-VP64-NLS-hSpCsn1 (10A,840A)-NLS, pXRP08—pLenti2-EF1a-NLS-VP64-NLS-hSpCsn1(10A,840A)-NLS-VP64-EGFP-NLS). These constructs were cloned into a lenti backbone under the expression of the human EF1a promoter. The WPRE element was also added for more robust protein expression. Each construct was transfected into HEK 293FT cells using Lipofectame 2000 and imaged 24 hours post-transfection. The best nuclear localization is obtained when the fusion proteins have NLS sequences on both the N- and C-term of the fusion protein. The highest observed nuclear localization occurred in the construct with four NLS elements.

[0932] To more robustly understand the influence of NLS elements on Cas9 Applicants made 16 Cas9-GFP fusions by adding the same alpha importin N LS sequence on either the N- or C-term looking at zero to three tandem repeats. Each construct was transfected into HEK 293FT cells using Lipofectame 2000 and imaged 24 hours post-transfection. Notably, the number of NLS elements does not directly correlate with the extent of nuclear localization. Adding an NLS on the C-term has a greater influence on nuclear localization than adding on the N-term.

[0933] Cas9 Transcriptional Activator: Applicants functionally tested the Cas9-VP64 protein by targeting the Sox2 locus and quantifying transcriptional activation by RT-qPCR. Eight DNA target sites were chosen to span the promoter of Sox2. Each construct was transfected into HEK 293FT cells using Lipofectame 2000 and 72 hours post-transfection total RNA was extracted from the cells. 1 ug of RNA was reverse transcribed into cDNA (qScript Supermix) in a 40 ul reaction. 2 ul of reaction product was added into a single 20 ul TaqMan assay qPCR reaction. Each experiment was performed in biological and technical triplicates. No RT control and no template control reactions showed no amplification. Constructs that do not show strong nuclear localization, pXRP02 and pXRP04, result in no activation. For the construct that did show strong nuclear localization, pXRP08, moderate activation was observed. Statistically significant activation was observed in the case of guide RNAs Sox2.4 and Sox2.5.

Example 32

In Vivo Mouse Data

[0934] Material and Reagents

Herculase II fusion polymerase (Agilent Technologies, cat. no. 600679)
10× NEBuffer 4 (NEB, cat. No. B7004S)
BsaI HF (NEB, cat. No. R3535S)
T7 DNA ligase (Enzymatics, cat. no. L602L)
Fast Digest buffer, 10× (ThermoScientific, cat. No. B64)
FastDigest NotI (ThermoScientific, cat. No. FD0594)
FastAP Alkaline Phosphatase (ThermoScientific, cat. No. EF0651)
Lipofectamine-2000 (Life Technologies, cat. No. 11668-019)
Trypsin (Life Technologies, cat. No. 15400054)
Forceps #4 (Sigma, cat. No. Z168777-1EA)
Forceps #5 (Sigma, cat. No. F6521-1EA)
10× Hank's Balanced Salt Solution (Sigma, cat. No. H4641-500M L)
Penicillin/Streptomycin solution (Life Technologies, cat. No. P4333)
Neurobasal (Life Technologies, cat. No. 21103049)
B27 Supplement (Life Technologies, cat. No. 17504044)
L-glutamine (Life Technologies, cat. No. 25030081)
Glutamate (Sigma, cat. No. RES5063G-A7)
β-mercaptoethanol (Sigma, cat. No. M6250-100mL)
HA rabbit antibody (Cell Signaling, cat. No. 3724S)
LIVE/DEAD® Cell Imaging Kit (Life Technologies, cat. No. R37601)
30G World Precision Instrument syringe (World Precision Instruments, cat. No. NANOFIL)
Stereotaxic apparatus (Kopf Instruments)
UltraMicroPump3 (World Precision Instruments, cat. No. UMP3-4)
Sucrose (Sigma, cat. No. S7903)
Calcium chloride (Sigma, cat. No. C1016)
Magnesium acetate (Sigma, cat. No. M0631)

Tris-HCl (Sigma, cat. no T5941)

[0935] EDTA (Sigma, cat. No. E6758)
NP-40 (Sigma, cat. No. NP40)
Phenylmethanesulfonyl fluoride (Sigma, cat. No. 78830)
Magnesium chloride (Sigma, cat. No. M8266)
Potassium chloride (Sigma, cat. No. P9333)
β-glycerophosphate (Sigma, cat. No. G9422)
Glycerol (Sigma, cat. No. G9012)
Vybrant® DyeCycle™ Ruby Stain (Life technologies, cat. No. S4942)
FACS Aria Flu-act-cell sorter (Koch Institute of MIT, Cambridge US)
DNAeasy Blood & Tissue Kit (Qiagen, cat. No. 69504)

[0936] Procedure

[0937] Constructing gRNA Multiplexes for Using In Vivo in the Brain

[0938] Applicants designed and PCR amplified single gRNAs targeting mouse TET and DNMT family members (as described herein) Targeting efficiency was assessed in N2a cell line (FIG. **33**). To obtain simultaneous modification of several genes in vivo, efficient gRNA was multiplexed in AAV-packaging vector (FIG. **34**). To facilitate further analysis of system efficiency applicants added to the system expression cassette consistent of GFP-KASH domain fusion protein under control of human Synapsin I promoter (FIG. **34**). This modification allows for further analysis of system efficiency in neuronal population (more detail procedure in section Sorting nuclei and in vivo results).

[0939] All 4 parts of the system were PCR amplified using Herculase II Fusion polymerase using following primers:

1st U6 7w:
gagggtctcgtccttgcggccgcgctagcgagggcctatttcccatgatt

c

1st gRNA Rv:
ctcggtctcggtAAAAAAgcaccgactcggtgccactttttcaagttgat

aacggactagccttattttaacttgctaTTTCtagctctaaaacNNNNNN

NNNNNNNNNNNNNNGGTGTTTCGTCCTTTCCAC

-continued

```
2nd U6 Fw:
gagggtctaTTTaccggtgagggcctatttcccatgattcc

2nd gRNA Rv:
ctcggtctcctcAAAAAAgcaccgactcggtgccacttttcaagttgat

aacggactagc cttattttaacttgctaTTTCtagctctaaaacNNNNN

NNNNNNNNNNNNNNNGGTGTTTCGTCCTTTCCAC

3rd U6 Fw:
gagggtctcTTTgagctcgagggcctatttcccatgattc

3rd gRNA Rv:
ctcggtctcgcgtAAAAAAgcaccgactcggtgccacttttcaagttga

taacggactag ccttattttaacttgctaTTTCtagctctaaaacNNNN

NNNNNNNNNNNNNNNNGGTGTTTCGTCCTTTCCA

hSyn_GFP-kash Fw:
gagggtctcTTacgcgtgtgtctagac

hSyn_GFP-kash Rv:
ctcggtctcAaggaCAGGGAAGGGAGCAGTGGTTCACGCCTGTAATCCCA

GCAATTTGGGA GGCCAAGGTGGGTAGATCACCTGAGATTAGGAGTTGC
```

(NNNNNNNNNNNNNNNNNNNN is a reverse compliment targeted genomic sequence)

[0940] Applicants used Golden Gate strategy to assemble all parts (1:1 molecular ratio) of the system in a single step reaction:

1^{st} U6_gRNA	18 ng
2^{nd} U6_gRNA	18 ng
3^{rd} U6_gRNA	18 ng
Syn_GFP-kash	100 ng
10x NEBuffer 4	1.0 μl
10x BSA	1.0 μl
10 mM ATP	1.0 μl
BsaI HF	0.75 μl
T7 ligase	0.25 μl
ddH_2O	10 μl
Cycle number	Condition
1-50	37° C. for 5 m, 21° C. for 5 m

[0941] Golden Gate reaction product was PCR amplified using Herculase II fusion polymerase and following primers:

```
Fw 5'cctgtccttgcggccgcgctagcgagggc

Rv 5'cacgcggccgcaaggacagggaagggagcag
```

[0942] PCR product was cloned into AAV backbone, between ITR sequences using NotI restriction sites:

PCR Product Digestion:

[0943]

Fast Digest buffer, 10X	3 μl
FastDigest NotI	1 μl
DNA	1 μg
ddH_2O	up to 30 μl

AAV Backbone Digestion:

[0944]

Fast Digest buffer, 10X	3 μl
FastDigest NotI	1 μl
FastAP Alkaline Phosphatase	1 μl
AAV backbone	1 μg
ddH_2O	up to 30 μl

[0945] After 20 min incubation in 37° C. samples were purified using QiAQuick PCR purification kit. Standardized samples were ligated at a 1:3 vector:insert ratio as follows:

Digested pUC19	50 ng
Digested insert	1:3 vector:insert molar ratio
T7 ligase	1 μl
2X Rapid Ligation Buffer	5 μl
ddH_20	up to 10 μl

[0946] After transformation of bacteria with ligation reaction product, applicants confirmed obtained clones with Sanger sequencing.

[0947] Positive DNA clones were tested in N2a cells after co-transfection with Cas9 construct (FIGS. **35** and **36**).

[0948] Design of New Cas9 Constructs for AAV Delivery

[0949] AAV delivery system despite its unique features has packing limitation—to successfully deliver expressing cassette in vivo it has to be in size < then 4.7 kb. To decrease the size of SpCas9 expressing cassette and facilitate delivery applicants tested several alteration: different promoters, shorter polyA signal and finally a smaller version of Cas9 from *Staphylococcus aureus* (SaCas9) (FIGS. **37** and **38**). All tested promoters were previously tested and published to be active in neurons, including mouse Mecp2 (Gray et al., 2011), rat Map1b and truncated rat Map1b (Liu and Fischer, 1996). Alternative synthetic polyA sequence was previously shown to be functional as well (Levitt et al., 1989; Gray et al., 2011). All cloned constructs were expressed in N2a cells after transfection with Lipofectamine 2000, and tested with Western blotting method (FIG. **39**).

[0950] Testing AAV Multiplex System in Primary Neurons

[0951] To confirm functionality of developed system in neurons, Applicants use primary neuronal cultures in vitro. Mouse cortical neurons was prepared according to the protocol published previously by Banker and Goslin (Banker and Goslin, 1988).

[0952] Neuronal cells are obtained from embryonic day 16. Embryos are extracted from the euthanized pregnant female and decapitated, and the heads are placed in ice-cold HBSS. The brains are then extracted from the skulls with forceps (#4 and #5) and transferred to another change of ice-cold HBSS. Further steps are performed with the aid of a stereoscopic microscope in a Petri dish filled with ice-cold HBSS and #5 forceps. The hemispheres are separated from each other and the brainstem and cleared of meninges. The hippocampi are then very carefully dissected and placed in a 15 ml conical tube filled with ice-cold HBSS. Cortices that remain after hippocampal dissection can be used for further cell isolation using an analogous protocol after removing the brain steam residuals and olfactory bulbs. Isolated hippocampi are washed three times with 10 ml ice-cold HBSS and dissociated by 15 min incubation with trypsin in HBSS (4 ml HBSS with the addition of 10 μl 2.5% trypsin per hippocampus) at 37° C.

After trypsinization, the hippocampi are very carefully washed three times to remove any traces of trypsin with HBSS preheated to 37° C. and dissociated in warm HBSS. Applicants usually dissociate cells obtained from 10-12 embryos in 1 ml HBSS using 1 ml pipette tips and dilute dissociated cells up to 4 ml. Cells are plated at a density of 250 cells/mm2 and cultured at 37° C. and 5% CO2 for up to 3 week

[0953] HBSS

435 ml H20

50 ml 10× Hank's Balanced Salt Solution

16.5 ml 0.3M HEPES pH 7.3

[0954] 5 ml penicillin-streptomycin solution
Filter (0.2 μm) and store 4° C.

[0955] Neuron Plating Medium (100 ml)

97 ml Neurobasal

2 ml B27 Supplement

[0956] 1 ml penicillin-streptomycin solution
250 μl glutamine
125 μl glutamate

Neurons are transduced with concentrated AAV1/2 virus or AAV1 virus from filtered medium of HEK293FT cells, between 4-7 days in culture and keep for at least one week in culture after transduction to allow for delivered gene expression.

[0957] AAV-Driven Expression of the System

[0958] Applicants confirmed expression of SpCas9 and SaCas9 in neuronal cultures after AAV delivery using Western blot method (FIG. **42**). One week after transduction neurons were collected in NuPage SDS loading buffer with β-mercaptoethanol to denaturate proteins in 95° C. for 5 min. Samples were separated on SDS PAGE gel and transferred on PVDF membrane for WB protein detection. Cas9 proteins were detected with HA antibody.

[0959] Expression of Syn-GFP-kash from gRNA multiplex AAV was confirmed with fluorescent microscopy (FIG. **50**).

[0960] Toxicity

[0961] To assess the toxicity of AAV with CRISPR system Applicants tested overall morphology of neurons one week after virus transduction (FIG. **45**). Additionally, Applicants tested potential toxicity of designed system with the LIVE/DEAD® Cell Imaging Kit, which allows to distinguish live and dead cells in culture. It is based on the presence of intracellular esterase activity (as determined by the enzymatic conversion of the non-fluorescent calcein AM to the intensely green fluorescent calcein). On the other hand, the red, cell-impermeant component of the Kit enters cells with damaged membranes only and bind to DNA generating fluorescence in dead cells. Both fluorophores can be easily visualized in living cells with fluorescent microscopy. AAV-driven expression of Cas9 proteins and multiplex gRNA constructs in the primary cortical neurons was well tolerated and not toxic (FIGS. **43** and **44**), what indicates that designed AAV system is suitable for in vivo tests.

[0962] Virus Production

[0963] Concentrated virus was produced according to the methods described in McClure et al., 2011. Supernatant virus production occurred in HEK293FT cells.

[0964] Brain Surgeries

[0965] For viral vector injections 10-15 week old male C57BL/6N mice were anesthetized with a Ketamine/Xylazine cocktail (Ketamine dose of 100 mg/kg and Xylazine dose of 10 mg/kg) by intraperitoneal injection. Intraperitonial administration of Buprenex was used as a pre-emptive analgesic (1 mg/kg). Animals were immobilized in a Kopf stereotaxic apparatus using intra-aural positioning studs and tooth bar to maintain an immobile skull. Using a hand-held drill, a hole (1-2 mm) at −3.0 mm posterior to Bregma and 3.5 mm lateral for injection in the CA1 region of the hippocampus was made. Using 30G World Precision Instrument syringe at a depth of 2.5 mm, the solution of AAV viral particles in a total volume of 1 ul was injected. The injection was monitored by a 'World Precision Instruments UltraMicroPump3' injection pump at a flow rate of 0.5 ul/min to prevent tissue damage. When the injection was complete, the injection needle was removed slowly, at a rate of 0.5 mm/min. After injection, the skin was sealed with 6-0 Ethilon sutures. Animals were postoperatively hydrated with 1 mL lactated Ringer's (subcutaneous) and housed in a temperature controlled (37° C.) environment until achieving an ambulatory recovery. 3 weeks after surgery animals were euthanized by deep anesthesia followed by tissue removal for nuclei sorting or with 4% paraformaldehyde perfusion for immunochemistry.

[0966] Sorting Nuclei and In Vivo Results

[0967] Applicants designed a method to specifically genetically tag the gRNA targeted neuronal cell nuclei with GFP for Fluorescent Activated Cell Sorting (FACS) of the labeled cell nuclei and downstream processing of DNA, RNA and nuclear proteins. To that purpose the applicants' multiplex targeting vector was designed to express both a fusion protein between GFP and the mouse nuclear membrane protein domain KASH (Starr DA, 2011, Current biology) and the 3 gRNAs to target specific gene loci of interest (FIG. **34**). GFP-KASH was expressed under the control of the human Synapsin promoter to specifically label neurons. The amino acid of the fusion protein GFP-KASH was:

```
MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICT

TGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIF

FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHN

VYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNH

YLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKSGLRSREEEEE

TDSRMPHLDSPGSSQPRRSFLSRVIRAALPLQLLLLLLLLLACLLPASED

DYSCTQANNFARSFYPMLRYTNGPPPT
```

[0968] One week after AAV1/2 mediated delivery into the brain a robust expression of GFP-KASH was observed. For FACS and downstream processing of labeled nuclei, the hippocampi were dissected 3 weeks after surgery and processed for cell nuclei purification using a gradient centrifugation step. For that purpose the tissue was homogenized in 320 mM Sucrose, 5 mM CaCl, 3 mM Mg(Ac)2, 10 mM Tris pH 7.8, 0.1 mM EDTA, 0.1% NP40, 0.1 mM Phenylmethanesulfonyl fluoride (PMSF), 1 mM 3-mercaptoethanol using 2 ml Dounce homogenizer (Sigma) The homogenisate was centrifuged on a 25% to 29% Optiprep® gradient according to the manufacture's protocol for 30 min at 3.500 rpm at 4° C. The nuclear pellet was resuspended in 340 mM Sucrose, 2 mM

MgCl2, 25 mM KCl, 65 mM glycerophosphate, 5% glycerol, 0.1 mM PMSF, 1 mM β-mercaptoethanol and Vybrant®; DyeCycle™ Ruby Stain (Life technologies) was added to label cell nuclei (offers near-infrared emission for DNA). The labeled and purified nuclei were sorted by FACS using an Aria Flu-act-cell sorter and BDFACS Diva software. The sorted GFP+ and GFP− nuclei were finally used to purify genomic DNA using DNAeasy Blood & Tissue Kit (Qiagen) for Surveyor assay analysis of the targeted genomic regions. The same approach can be easily used to purify nuclear RNA or protein from targeted cells for downstream processing. Due to the 2-vector system (FIG. **34**) the applicants using in this approach efficient Cas9 mediated DNA cleavage was expected to occur only in a small subset of cells in the brain (cells which were co-infected with both the multiplex targeting vector and the Cas9 encoding vector). The method described here enables the applicants to specifically purify DNA, RNA and nuclear proteins from the cell population expressing the 3 gRNAs of interest and therefore are supposed to undergo Cas9 mediated DNA cleavage. By using this method the applicants were able to visualize efficient DNA cleavage in vivo occurring only in a small subset of cells.

[0969] Essentially, what Applicants have shown here is targeted in vivo cleavage. Furthermore, Applicants used a multiple approach, with several different sequences targeted at the same time, but independently. Presented system can be applied for studying brain pathologic conditions (gene knock out, e.g. Parkinson disease) and also open a field for further development of genome editing tools in the brain. By replacing nuclease activity with gene transcription regulators or epigenetic regulators it will be possible to answer whole spectrum of scientific question about role of gene regulation and epigenetic changes in the brain in not only in the pathologic conditions but also in physiological process as learning and memory formation. Finally, presented technology can be applied in more complex mammalian system as primates, what allows to overcome current technology limitations.

Example 33

Model Data

[0970] Several disease models have been specifically investigated. These include de novo autism risk genes CHD8, KATNAL2, and SCN2A; and the syndromic autism (Angelman Syndrome) gene UBE3A. These genes and resulting autism models are of course preferred, but show that the invention may be applied to any gene and therefore any model is possible.

[0971] Applicants have made these cells lines using Cas9 nuclease in human embryonic stem cells (hESCs). The lines were created by transient transfection of hESCs with Cbh-Cas9-2A-EGFP and pU6-sgRNA. Two sgRNAs are designed for each gene targeting most often the same exons in which patient nonsense (knock-out) mutations have been recently described from whole exome sequencing studies of autistic patients. The Cas9-2A-EGFP and pU6 plasmids were created specifically for this project.

Example 34

AAV Production System or Protocol

[0972] An AAV production system or protocol that was developed for, and works particularly well with, high through put screening uses is provided herein, but it has broader applicability in the present invention as well. Manipulating endogenous gene expression presents various challenges, as the rate of expression depends on many factors, including regulatory elements, mRNA processing, and transcript stability. To overcome this challenge, Applicants developed an adeno-associated virus (AAV)-based vector for the delivery. AAV has an ssDNA-based genome and is therefore less susceptible to recombination.

[0973] AAV1/2 (serotype AAV1/2, i.e., hybrid or mosaic AAV1/AAV2 capsid AAV) heparin purified concentrated virus protocol

[0974] Media: D10+HEPES
500 ml bottle DMEM high glucose+Glutamax (GIBCO)
50 ml Hyclone FBS (heat-inactivated) (Thermo Fischer)
5.5 ml HEPES solution (1M, GIBCO)
Cells: low passage HEK293FT (passage <10 at time of virus production, thaw new cells of passage 2-4 for virus production, grow up for 3-5 passages)

[0975] Transfection Reagent: Polyethylenimine (PEI) "Max"
Dissolve 50 mg PEI "Max" in 50 ml sterile Ultrapure H20

Adjust pH to 7.1

[0976] Filter with 0.22 um fliptop filter
Seal tube and wrap with parafilm
Freeze aliquots at −20° C. (for storage, can also be used immediately)

[0977] Cell Culture
Culture low passage HEK293FT in D10+HEPES
Passage everyday between 1:2 and 1:2.5
Advantageously do not allow cells to reach more than 85% confluency

[0978] For T75

[0979] Warm 10 ml HBSS (—Mg2+, —Ca2+, GIBCO)+1 ml TrypLE Express (GIBCO) per flask to 37° C. (Waterbath)
Aspirate media fully

[0980] Add 10 ml warm HBSS gently (to wash out media completely)

[0981] Add 1 ml TrypLE per Flask

[0982] Place flask in incubator (37° C.) for 1 min

[0983] Rock flask to detach cells

[0984] Add 9 ml D10+HEPES media (37° C.)

[0985] Pipette up and down 5 times to generate single cell suspension

[0986] Split at 1:2-1:2.5 (12 ml media for T75) ratio (if cells are growing more slowly, discard and thaw a new batch, they are not in optimal growth)

[0987] transfer to T225 as soon as enough cells are present (for ease of handling large amounts of cells)

[0988] AAV Production (5*15 Cm Dish Scale Per Construct):
Plate 10 million cells in 21.5 ml media into a 15 cm dish
Incubate for 18-22 hours at 37° C.
Transfection is ideal at 80% confluence

[0989] Per Plate
Prewarm 22 ml media (D10+HEPES)

[0990] Prepare Tube with DNA Mixture (Use Endofree Maxiprep DNA):
5.2 ug vector of interest plasmid
4.35 ug AAV 1 serotype plasmid
4.35 ug AAV 2 serotype plasmid
10.4 ug pDF6 plasmid (adenovirus helper genes) □ Vortex to mix
Add 434 uL DMEM (no serum!)

Add 130 ul PEI solution
Vortex 5-10 seconds
Add DNA/DMEM/PEI mixture to prewarmed media
Vortex briefly to mix
Replace media in 15 cm dish with DNA/DMEM/PEI mixture
Return to 37° C. incubator
Incubate 48 h before harvesting (make sure medium isn't turning too acidic)

[0991] Virus Harvest:

1. aspirate media carefully from 15 cm dish dishes (advantageously do not dislodge cells)
2. Add 25 ml RT DPBS (Invitrogen) to each plate and gently remove cells with a cell scraper. Collect suspension in 50 ml tubes.
3. Pellet cells at 800×g for 10 minutes.
4. Discard supernatant

[0992] Pause Point: Freeze Cell Pellet at −80 C if Desired

5. resuspend pellet in 150 mM NaCl, 20 mM Tris pH 8.0, use 10 ml per tissue culture plate.
6. Prepare a fresh solution of 10% sodium deoxycholate in dH2O. Add 1.25 ml of this per tissue culture plate for a final concentration of 0.5%. Add benzonase nuclease to a final concentration of 50 units per ml. Mix tube thoroughly.
7. Incubate at 37° C. for 1 hour (Waterbath).
8. Remove cellular debris by centrifuging at 3000×g for 15 mins. Transfer to fresh 50 ml tube and ensure all cell debris has been removed to prevent blocking of heparin columns.

[0993] Heparin Column Purification of AAV1/2:

[0994] 1. Set up HiTrap heparin columns using a peristaltic pump so that solutions flow through the column at 1 ml per minute. It is important to ensure no air bubbles are introduced into the heparin column.

[0995] 2. Equilibrate the column with 10 ml 150 mM NaCl, 20 mM Tris, pH 8.0 using the peristaltic pump.

[0996] 3. Binding of virus: Apply 50 ml virus solution to column and allow to flow through.

[0997] 4. Wash step 1: column with 20 ml 100 mM NaCl, 20 mM Tris, pH 8.0. (using the peristaltic pump)

[0998] 5. Wash step 2: Using a 3 ml or 5 ml syringe continue to wash the column with 1 ml 200 mM NaCl, 20 mM Tris, pH 8.0, followed by 1 ml 300 mM NaCl, 20 mM Tris, pH 8.0.

[0999] Discard the flow-through.

[1000] (prepare the syringes with different buffers during the 50 min flow through of virus solution above)

[1001] 6. Elution Using 5 ml syringes and gentle pressure (flow rate of <1 ml/min) elute the virus from the column by applying:

[1002] 1.5 ml 400 mM NaCl, 20 mM Tris, pH 8.0

[1003] 3.0 ml 450 mM NaCl, 20 mM Tris, pH 8.0

[1004] 1.5 ml 500 mM NaCl, 20 mM Tris, pH 8.0

[1005] Collect these in a 15 ml centrifuge tube.

[1006] Concentration of AAV1/2:

[1007] 1. Concentration step 1: Concentrate the eluted virus using Amicon ultra 15 ml centrifugal filter units with a 100,000 molecular weight cutoff. Load column eluate into the concentrator and centrifuge at 2000×g for 2 minutes (at room temperature. Check concentrated volume—it should be approximately 500 μl. If necessary, centrifuge in 1 min intervals until correct volume is reached.

[1008] 2. buffer exchange: Add 1 ml sterile DPBS to filter unit, centrifuge in 1 min intervals until correct volume (500 ul) is reached.

[1009] 3. Concentration step 2: Add 500 ul concentrate to an Amicon Ultra 0.5 ml 100K filter unit. Centrifuge at 6000 g for 2 min. Check concentrated volume—it should be approximately 100 μl. If necessary, centrifuge in 1 min intervals until correct volume is reached.

[1010] 4. Recovery: Invert filter insert and insert into fresh collection tube. Centrifuge at 1000 g for 2 min.

Aliquot and freeze at −80 C
1 μl is typically required per injection site, small aliquots (e.g. 5 ul) are therefore recommended (avoid freeze-thaw of virus).
determine DNaseI-resistant GC particle titer using qPCR (see separate protocol)

[1011] Materials

Amicon Ultra, 0.5 ml, 100K; MILLIPORE; UFC510024

Amicon Ultra, 15 ml, 100K; MILLIPORE; UFC910024

[1012] Benzonase nuclease; Sigma-Aldrich, E1014
HiTrap Heparin cartridge; Sigma-Aldrich; 54836
Sodium deoxycholate; Sigma-Aldrich; D5670

[1013] AAV1 Supernatant Production Protocol

Media: D10+HEPES

[1014] 500 ml bottle DMEM high glucose+Glutamax (Invitrogen)
50 ml Hyclone FBS (heat-inactivated) (Thermo Fischer)
5.5 ml HEPES solution (1M, GIBCO)
Cells: low passage HEK293FT (passage <10 at time of virus production)
Thaw new cells of passage 2-4 for virus production, grow up for 2-5 passages
Transfection reagent: Polyethylenimine (PEI) "Max"
Dissolve 50 mg PEI "Max" in 50 ml sterile Ultrapure H20

Adjust pH to 7.1

[1015] Filter with 0.22 um fliptop filter
Seal tube and wrap with parafilm
Freeze aliquots at −20° C. (for storage, can also be used immediately)

Cell Culture

[1016] Culture low passage HEK293FT in D10+HEPES
Passage everyday between 1:2 and 1:2.5
Advantageously do let cells reach more than 85% confluency

For T75

[1017] Warm 10 ml HBSS (—Mg2+, —Ca2+, GIBCO)+1 ml TrypLE Express (GIBCO) per flask to 37° C. (Waterbath)

[1018] Aspirate media fully

[1019] Add 10 ml warm HBSS gently (to wash out media completely)

[1020] Add 1 ml TrypLE per Flask

[1021] Place flask in incubator (37° C.) for 1 min

[1022] Rock flask to detach cells

[1023] Add 9 ml D10+HEPES media (37° C.)

[1024] Pipette up and down 5 times to generate single cell suspension

[1025] Split at 1:2-1:2.5 (12 ml media for T75) ratio (if cells are growing more slowly, discard and thaw a new batch, they are not in optimal growth)

[1026] transfer to T225 as soon as enough cells are present (for ease of handling large amounts of cells)

AAV production (single 15 cm dish scale)
Plate 10 million cells in 21.5 ml media into a 15 cm dish

Incubate for 18-22 hours at 37° C.
Transfection is ideal at 80% confluence per plate
Prewarm 22 ml media (D10+HEPES)
Prepare tube with DNA mixture (use endofree maxiprep DNA):
5.2 ug vector of interest plasmid
8.7 ug AAV1 serotype plasmid
10.4 ug DF6 plasmid (adenovirus helper genes)

Vortex to mix

[1027] Add 434 uL DMEM (no serum!) Add 130 ul PEI solution
Vortex 5-10 seconds
Add DNA/DMEM/PEI mixture to prewarmed media
Vortex briefly to mix
Replace media in 15 cm dish with DNA/DMEM/PEI mixture
Return to 37° C. incubator
Incubate 48 h before harvesting (advantageously monitor to ensure medium is not turning too acidic)

[1028] Virus Harvest:
Remove supernatant from 15 cm dish
Filter with 0.45 um filter (low protein binding) Aliquot and freeze at −80° C.
Transduction (primary neuron cultures in 24-well format, 5DIV)
Replace complete neurobasal media in each well of neurons to be transduced with fresh neurobasal (usually 400 ul out of 500 ul per well is replaced)
Thaw AAV supernatant in 37° C. waterbath
Let equilibrate in incubator for 30 min
Add 250 ul AAV supernatant to each well

Incubate 24 h at 37° C.

[1029] Remove media/supernatant and replace with fresh complete neurobasal
Expression starts to be visible after 48 h, saturates around 6-7 Days Post Infection
Constructs for pAAV plasmid with GOI should not exceed 4.8 kb including both ITRS.

[1030] Example of a human codon optimized sequence (i.e. being optimized for expression in humans) sequence: SaCas9 is provided below:

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGC

TTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGAAAAGGAACTACA

TTCTGGGGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTATTGAC

TATGAAACAAGGGACGTGATCGACGCAGGCGTCAGACTGTTCAAGGAGGC

CAACGTGGAAAACAATGAGGGACGGAGAAGCAAGAGGGGAGCCAGGCGCC

TGAAACGACGGAGAAGGCACAGAATCCAGAGGGTGAAGAAACTGCTGTTC

GATTACAACCTGCTGACCGACCATTCTGAGCTGAGTGGAATTAATCCTTA

TGAAGCCAGGGTGAAAGGCCTGAGTCAGAAGCTGTCAGAGGAAGAGTTTT

CCGCAGCTCTGCTGCACCTGGCTAAGCGCCGAGGAGTGCATAACGTCAAT

GAGGTGGAAGAGGACACCGGCAACGAGCTGTCTACAAAGGAACAGATCTC

ACGCAATAGCAAAGCTCTGGAAGAGAAGTATGTCGCAGAGCTGCAGCTGG

AACGGCTGAAGAAAGATGGCGAGGTGAGAGGGTCAATTAATAGGTTCAAG

-continued

ACAAGCGACTACGTCAAAGAAGCCAAGCAGCTGCTGAAAGTGCAGAAGGC

TTACCACCAGCTGGATCAGAGCTTCATCGATACTTATATCGACCTGCTGG

AGACTCGGAGAACCTACTATGAGGGACCAGGAGAAGGGAGCCCCTTCGGA

TGGAAAGACATCAAGGAATGGTACGAGATGCTGATGGGACATTGCACCTA

TTTTCCAGAAGAGCTGAGAAGCGTCAAGTACGCTTATAACGCAGATCTGT

ACAACGCCCTGAATGACCTGAACAACCTGGTCATCACCAGGGATGAAAAC

GAGAAACTGGAATACTATGAGAAGTTCCAGATCATCGAAAACGTGTTTAA

GCAGAAGAAAAAGCCTACACTGAAACAGATTGCTAAGGAGATCCTGGTCA

ACGAAGAGGACATCAAGGGCTACCGGGTGACAAGCACTGGAAAACCAGAG

TTCACCAATCTGAAAGTGTATCACGATATTAAGGACATCACAGCACGGAA

AGAAATCATTGAGAACGCCGAACTGCTGGATCAGATTGCTAAGATCCTGA

CTATCTACCAGAGCTCCGAGGACATCCAGGAAGAGCTGACTAACCTGAAC

AGCGAGCTGACCCAGGAAGAGATCGAACAGATTAGTAATCTGAAGGGGTA

CACCGGAACACACAACCTGTCCCTGAAAGCTATCAATCTGATTCTGGATG

AGCTGTGGCATACAAACGACAATCAGATTGCAATCTTTAACCGGCTGAAG

CTGGTCCCAAAAAAGGTGGACCTGAGTCAGCAGAAAGAGATCCCAACCAC

ACTGGTGGACGATTTCATTCTGTCACCCGTGGTCAAGCGGAGCTTCATCC

AGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAAT

GATATCATTATCGAGCTGGCTAGGGAGAAGAACAGCAAGGACGCACAGAA

GATGATCAATGAGATGCAGAAACGAAACCGGCAGACCAATGAACGCATTG

AAGAGATTATCCGAACTACCGGGAAAGAGAACGCAAAGTACCTGATTGAA

AAAATCAAGCTGCACGATATGCAGGAGGGAAAGTGTCTGTATTCTCTGGA

GGCCATCCCCCTGGAGGACCTGCTGAACAATCCATTCAACTACGAGGTCG

ATCATATTATCCCCAGAAGCGTGTCCTTCGACAATTCCTTTAACAACAAG

GTGCTGGTCAAGCAGGAAGAGAACTCTAAAAAGGGCAATAGGACTCCTTT

CCAGTACCTGTCTAGTTCAGATTCCAAGATCTCTTACGAAACCTTTAAAA

AGCACATTCTGAATCTGGCCAAAGGAAAGGGCCGCATCAGCAAGACCAAA

AAGGAGTACCTGCTGGAAGAGCGGGACATCAACAGATTCTCCGTCCAGAA

GGATTTTATTAACCGGAATCTGGTGGACACAAGATACGCTACTCGCGGCC

TGATGAATCTGCTGCGATCCTATTTCCGGGTGAACAATCTGGATGTGAAA

GTCAAGTCCATCAACGGCGGGTTCACATCTTTTCTGAGGCGCAAATGGAA

GTTTAAAAAGGAGCGCAACAAAGGGTACAAGCACCATGCCGAAGATGCTC

TGATTATCGCAAATGCCGACTTCATCTTTAAGGAGTGGAAAAAGCTGGAC

AAAGCCAAGAAAGTGATGGAGAACCAGATGTTCGAAGAGAAGCAGGCCGA

ATCTATGCCCGAAATCGAGACAGAACAGGAGTACAAGGAGATTTTCATCA

CTCCTCACCAGATCAAGCATATCAAGGATTTCAAGGACTACAAGTACTCT

CACCGGGTGGATAAAAAGCCCAACAGAGAGCTGATCAATGACACCCTGTA

TAGTACAAGAAAAGACGATAAGGGGAATACCCTGATTGTGAACAATCTGA

ACGGACTGTACGACAAAGATAATGACAAGCTGAAAAAGCTGATCAACAAA

-continued

```
AGTCCCGAGAAGCTGCTGATGTACCACCATGATCCTCAGACATATCAGAA

ACTGAAGCTGATTATGGAGCAGTACGGCGACGAGAAGAACCCACTGTATA

AGTACTATGAAGAGACTGGGAACTACCTGACCAAGTATAGCAAAAAGGAT

AATGGCCCCGTGATCAAGAAGATCAAGTACTATGGGAACAAGCTGAATGC

CCATCTGGACATCACAGACGATTACCCTAACAGTCGCAACAAGGTGGTCA

AGCTGTCACTGAAGCCATACAGATTCGATGTCTATCTGGACAACGGCGTG

TATAAATTTGTGACTGTCAAGAATCTGGATGTCATCAAAAAGGAGAACTA

CTATGAAGTGAATAGCAAGTGCTACGAAGAGGCTAAAAAGCTGAAAAAGA

TTAGCAACCAGGCAGAGTTCATCGCCTCCTTTTACAACAACGACCTGATT

AAGATCAATGGCGAACTGTATAGGGTCATCGGGGTGAACAATGATCTGCT

GAACCGCATTGAAGTGAATATGATTGACATCACTTACCGAGAGTATCTGG

AAAACATGAATGATAAGCGCCCCCCTCGAATTATCAAAACAATTGCCTCT

AAGACTCAGAGTATCAAAAAGTACTCAACCGACATTCTGGGAAACCTGTA

TGAGGTGAAGAGCAAAAAGCACCCTCAGATTATCAAAAAGGGCTAAGAAT

TC
```

Example 35

Minimizing Off Target Cleavage Using Cas9 Nickase and Two Guide RNAs

[1031] Cas9 is a RNA-guided DNA nuclease that may be targeted to specific locations in the genome with the help of a 20 bp RNA guide. However the guide sequence may tolerate some mismatches between the guide sequence and the DNA-target sequence. The flexibility is undesirable due to the potential for off-target cleavage, when the guide RNA targets Cas9 to a an off-target sequence that has a few bases different from the guide sequence. For all experimental applications (gene targeting, crop engineering, therapeutic applications, etc) it is important to be able to improve the specificity of Cas9 mediated gene targeting and reduce the likelihood of off-target modification by Cas9.

[1032] Applicants developed a method of using a Cas9 nickase mutant in combination with two guide RNAs to facilitate targeted double strand breaks in the genome without off-target modifications. The Cas9 nickase mutant may be generated from a Cas9 nuclease by disabling its cleavage activity so that instead of both strands of the DNA duplex being cleaved only one strand is cleaved. The Cas9 nickase may be generated by inducing mutations in one or more domains of the Cas9 nuclease, e.g. Ruvc1 or HNH. These mutations may include but are not limited to mutations in a Cas9 catalytic domain, e.g in SpCas9 these mutations may be at positions D10 or H840. These mutations may include but are not limited to D10A, E762A, H840A. N854A, N863A or D986A in SpCas9 but nickases may be generated by inducing mutations at corresponding positions in other CRISPR enzymes or Cas9 orthologs. In a most preferred embodiment of the invention the Cas9 nickase mutant is a SpCas9 nickase with a D10A mutation.

[1033] The way this works is that each guide RNA in combination with Cas9 nickase would induce the targeted single strand break of a duplex DNA target. Since each guide RNA nicks one strand, the net result is a double strand break. The reason this method eliminates off-target mutations is because it is very unlikely to have an off-target site that has high degrees of similarity for both guide sequences (20 bp+2 bp(PAM)=22 bp specificity for each guide, and two guides means any off-target site will have to have close to 44 bp of homologous sequence). Although it is still likely that individual guides may have off-targets, but those off-targets will only be nicked, which is unlikely to be repaired by the mutagenic NHEJ process. Therefore the multiplexing of DNA double strand nicking provides a powerful way of introducing targeted DNA double strand breaks without off-target mutagenic effects.

[1034] Applicants carried out experiments involving the co-transfection of HEK293FT cells with a plasmid encoding Cas9(D10A) nickase as well as DNA expression cassettes for one or more guides. Applicants transfected cells using Lipofectamine 2000, and transfected cells were harvested 48 or 72 hours after transfections. Double nicking-induced NHEJ were detected using the SURVEYOR nuclease assay as described previously herein (FIGS. **51**, **52** and **53**).

[1035] Applicants have further identified parameters that relate to efficient cleavage by the Cas9 nickase mutant when combined with two guide RNAs and these parameters include but are not limited to the length of the 5' overhang. Efficient cleavage is reported for 5' overhang of at least 26 base pairs. In a preferred embodiment of the invention, the 5' overhang is at least 30 base pairs and more preferably at least 34 base pairs. Overhangs of up to 200 base pairs may be acceptable for cleavage, while 5' overhangs less than 100 base pairs are preferred and 5' overhangs less than 50 base pairs are most preferred (FIGS. **54** and **55**).

REFERENCES

[1036] Banker G. Goslin K. Developments in neuronal cell culture. Nature. 1988 Nov. 10; 336(6195): 185-6.

[1037] Bedell, V. M. et al. In vivo genome editing using a high-efficiency TALEN system. Nature 491, 114-U133 (2012).

[1038] Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet. 45, 273-297 (2011).

[1039] Bobis-Wozowicz, S., Osiak, A., Rahman, S. H. & Cathomen, T. Targeted genome editing in pluripotent stem cells using zinc-finger nucleases. Methods 53, 339-346 (2011).

[1040] Boch, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326, 1509-1512 (2009).

[1041] Bogenhagen, D. F. & Brown, D. D. Nucleotide sequences in *Xenopus* 5S DNA required for transcription termination. Cell 24, 261-270 (1981).

[1042] Bultmann, S. et al. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. Nucleic Acids Res 40, 5368-5377 (2012).

[1043] Carlson, D. F. et al. Efficient TALEN-mediated gene knockout in livestock. Proc Natl Acad Sci USA 109, 17382-17387 (2012).

[1044] Chen, F. Q. et al. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat Methods 8, 753-U796 (2011).

[1045] Cho, S. W., Kim, S., Kim, J. M. & Kim, J. S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31, 230-232 (2013).

[1046] Christian, M. et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186, 757-761 (2010).

[1047] Cong, L. et al. Multiplex genome engineering using CRISPR-Cas systems. Science 339, 819-823 (2013).

[1048] Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-607 (2011).

[1049] Deveau, H., Garneau, J. E. & Moineau, S. CRISPR-Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol 64, 475-493 (2010).

[1050] Ding, Q. et al. A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell 12, 238-251 (2013).

[1051] Garneau, J. E. et al. The CRISPR-Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67-71 (2010).

[1052] Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys. V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci USA 109, E2579-2586 (2012).

[1053] Geurts, A. M. et al. Knockout Rats via Embryo Microinjection of Zinc-Finger Nucleases. Science 325, 433-433 (2009).

[1054] Gray S J, Foti S B, Schwartz J W. Bachaboina L, Taylor-Blake B, Coleman J. Ehlers M D, Zylka M J, McCown T J, Samulski R J. Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors. Hum Gene Ther. 2011 September; 22(9): 1143-53. doi: 10.1089/hum.2010.245.

[1055] Guschin, D. Y. et al. A rapid and general assay for monitoring endogenous gene modification. Methods Mol Biol 649, 247-256 (2010).

[1056] Hasty, P., Rivera-Perez, J. & Bradley, A. The length of homology required for gene targeting in embryonic stem cells. Mol Cell Biol 11, 5586-5591 (1991).

[1057] Horvath, P. & Barrangou, R. CRISPR-Cas, the immune system of bacteria and archaea. Science 327, 167-170 (2010).

[1058] Hsu, P. D. & Zhang, F. Dissecting neural function using targeted genome engineering technologies. ACS Chem Neurosci 3, 603-610 (2012).

[1059] Hwang, W. Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31, 227-229 (2013).

[1060] Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31, 233-239 (2013).

[1061] Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821 (2012).

[1062] Jinek, M. et al. RNA-programmed genome editing in human cells. eLife 2, e00471 (2013).

[1063] Kaplitt, M. G., et al., Safety and tolerability of gene therapy with an adeno-associated virus (AAV) borne GAD gene for Parkinson's disease: an open label, phase I trial. Lancet. 2007 Jun. 23; 369(9579):2097-105.

[1064] Levitt N. Briggs D. Gil A. Proudfoot N. J. Definition of an efficient synthetic poly(A) site. Genes Dev. 1989; 3:1019-1025.

[1065] Liu D, Fischer I. Two alternative promoters direct neuron-specific expression of the rat microtubule-associated protein 1B gene. J. Neurosci. 1996 Aug. 15; 16(16): 5026-36.

[1066] Lopes, V. S., etc al., Retinal gene therapy with a large MYO7A cDNA using adeno-associated virus. Gene Ther, 2013 Jan 24. doi: 10.1038/gt 2013.3-[Epub ahead of print]

[1067] Mahfouz, M. M. et al. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc Natl Acad Sci USA 108, 2623-2628 (2011).

[1068] Makarova, K. S. et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9, 467-477 (2011).

[1069] Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013).

[1070] McClure C, Cole K L, Wulff P, Klugmann M, Murray A J. Production and titering of recombinant adeno-associated viral vectors. J Vis Exp. 2011 Nov. 27; (57): e3348. doi: 10.3791/3348.

[1071] Michaelis, L. M., Maud "Die kinetik der invertinwirkung.". Biochem. z (1913).

[1072] Miller, J. C. et al. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25, 778-785 (2007).

[1073] Miller, J. C. et al. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29, 143-148 (2011).

[1074] Moscou. M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. Science 326, 1501 (2009). Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. Science 300, 763 (2003).

[1075] Mussolino, C. et al. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic acids research 39, 9283-9293 (2011).

[1076] Nathwani, A. C., et al., Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. N Engl J. Med. 2011 Dec. 22; 365(25):2357-65. doi: 10.1056/NEJMoa1108046. Epub 2011 Dec. 10.

[1077] Oliveira, T. Y. et al. Translocation capture sequencing: a method for high throughput mapping of chromosomal rearrangements. J Immunol Methods 375, 176-181 (2012).

[1078] Perez, E. E. et al. Establishment of HIV-1 resistance in CD4(+) T cells by genome editing using zinc-finger nucleases. Nat Biotechnol 26, 808-816 (2008).

[1079] Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173-1183 (2013).

[1080] REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991)

[1081] Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol 30, 460-465 (2012).

[1082] Saleh-Gohari, N. & Helleday, T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. Nucleic Acids Res 32, 3683-3688 (2004).

[1083] Sander, J. D. et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods 8, 67-69 (2011).

[1084] Sanjana, N. E. et al. A transcription activator-like effector toolbox for genome engineering. Nat Protoc 7, 171-192 (2012).

[1085] Sapranauskas, R. et al. The *Streptococcus thermophilus* CRISPR-Cas system provides immunity in *Escherichia coli*. Nucleic Acids Res 39, 9275-9282 (2011).

[1086] Shen, B. et al. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. Cell Res 23, 720-723 (2013).

[1087] Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature 317, 230-234 (1985).

[1088] Soldner, F. et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 146, 318-331 (2011).

[1089] Takasu, Y. et al. Targeted mutagenesis in the silkworm *Bombyx mori* using zinc finger nuclease mRNA injection. Insect Biochem Molec 40, 759-765 (2010).

[1090] Tangri S, et al., Rationally engineered therapeutic proteins with reduced immunogenicity, J. Immunol. 2005 Mar. 15; 174(6):3187-96.

[1091] Thomas, K. R., Folger, K. R. & Capecchi, M. R. High frequency targeting of genes to specific sites in the mammalian genome. Cell 44, 419-428 (1986).

[1092] Tuschl, T. Expanding small RNA interference. Nat Biotechnol 20, 446-448 (2002).

[1093] Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. Nat Rev Genet. 11, 636-646 (2010).

[1094] Valton, J. et al. Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. J Biol Chem 287, 38427-38432 (2012).

[1095] Wang, H. et al. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR-Cas-Mediated Genome Engineering. Cell 153, 910-918 (2013).

[1096] Watanabe, T. et al. Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. Nat Commun 3 (2012).

[1097] Wilson. E. B. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 22, 209-212 (1927).

[1098] Wood, A. J. et al. Targeted genome editing across species using ZFNs and TALENs. Science 333, 307 (2011).

[1099] Wu, S., Ying, G. X., Wu, Q. & Capecchi, M. R. A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond. Nat Protoc 3, 1056-1076 (2008).

[1100] Zhang, F. et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 29, 149-153 (2011).

SEQUENCE LISTING

```
<160> NUMBER OF SEQ ID NOS: 434

<210> SEQ ID NO 1
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 1

gcactgaggg cctatttccc atgattc                                        27


<210> SEQ ID NO 2
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

gagtccgagc agaagaagaa                                                20


<210> SEQ ID NO 3
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 3

gagtcctagc aggagaagaa                                                20


<210> SEQ ID NO 4
<211> LENGTH: 20
```

-continued

```
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 4

gagtctaagc agaagaagaa                                              20


<210> SEQ ID NO 5
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(44)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (47)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 5

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn nnnnggnnnn nnnnnnnnnn     60


<210> SEQ ID NO 6
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(14)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (17)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 6

nnnnnnnnnn nnnnccnnnn nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 7
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(43)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
```

-continued

```
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (46)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 7

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn nnnggnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 8
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(15)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (18)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 8

nnnnnnnnnn nnnnnccnnn nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 9
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(42)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (45)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 9

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn nnggnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 10
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(15)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
```

-continued

```
<222> LOCATION: (18)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 10

nnnnnnnnnn nnnnnccnnn nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 11
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(41)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (44)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 11

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn nggnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 12
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (20)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 12

nnnnnnnnnn nnnnnnnccn nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 13
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
```

-continued

```
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(40)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (43)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 13

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn ggnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 14
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(18)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (21)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 14

nnnnnnnnnn nnnnnnnncc nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 15
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(39)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (42)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 15

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnng gnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 16
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
```

-continued

```
<222> LOCATION: (1)..(19)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (22)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 16

nnnnnnnnnn nnnnnnnnnc cnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 17
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 17

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 18
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 18

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 19
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
```

-continued

<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(37)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (40)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 19

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnggn nnnnnnnnnn nnnnnnnnnn 60

<210> SEQ ID NO 20
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(21)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (24)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 20

nnnnnnnnnn nnnnnnnnnn nccnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> SEQ ID NO 21
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(36)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (39)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 21

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnggnn nnnnnnnnnn nnnnnnnnnn 60

<210> SEQ ID NO 22
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source

-continued

```
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (25)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 22

nnnnnnnnnn nnnnnnnnnn nnccnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 23
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(35)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (38)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 23

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnggnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 24
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(23)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (26)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 24

nnnnnnnnnn nnnnnnnnnn nnnccnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 25
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
```

-continued

```
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(34)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (37)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 25

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnggnnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 26
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(24)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (27)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 26

nnnnnnnnnn nnnnnnnnnn nnnnccnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 27
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(33)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (36)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 27

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnggnnnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 28
<211> LENGTH: 60
```

-continued

```
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(25)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (28)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 28

nnnnnnnnnn nnnnnnnnnn nnnnnccnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 29
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(32)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (35)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 29

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnggnnnnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 30
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(26)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (29)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 30

nnnnnnnnnn nnnnnnnnnn nnnnnnccnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60
```

-continued

```
<210> SEQ ID NO 31
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(31)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (34)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 31

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nggnnnnnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 32
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(27)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (30)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 32

nnnnnnnnnn nnnnnnnnnn nnnnnnnccn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 33
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(30)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (33)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 33

nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn ggnnnnnnnn nnnnnnnnnn nnnnnnnnnn     60
```

-continued

```
<210> SEQ ID NO 34
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(28)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (31)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 34

nnnnnnnnnn nnnnnnnnnn nnnnnnnncc nnnnnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 35
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(29)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (32)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 35

nnnnnnnnnn nnnnnnnnnn ccnnnnnnng gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 36
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(29)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (32)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 36
```

-continued

```
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnc cnnnnnnngg nnnnnnnnnn nnnnnnnnnn      60


<210> SEQ ID NO 37
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(28)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (31)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 37

nnnnnnnnnn nnnnnnnnnn ccnnnnnngg nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn      60


<210> SEQ ID NO 38
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(30)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (33)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 38

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ccnnnnnngg nnnnnnnnnn nnnnnnnnnn      60


<210> SEQ ID NO 39
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(27)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (30)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
```

-continued

```
<400> SEQUENCE: 39

nnnnnnnnnn nnnnnnnnnn ccnnnnnggn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn      60


<210> SEQ ID NO 40
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(31)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (34)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 40

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nccnnnnngg nnnnnnnnnn nnnnnnnnnn      60


<210> SEQ ID NO 41
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(26)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (29)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 41

nnnnnnnnnn nnnnnnnnnn ccnnnnggnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn      60


<210> SEQ ID NO 42
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(32)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (35)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
```

-continued

```
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 42

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnccnnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 43
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(25)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (28)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 43

nnnnnnnnnn nnnnnnnnnn ccnnnggnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 44
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(33)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (36)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 44

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnccnnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 45
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(24)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
```

-continued

```
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (27)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 45

nnnnnnnnnn nnnnnnnnnn ccnnggnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 46
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(34)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (37)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 46

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnccnngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 47
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(23)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (26)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 47

nnnnnnnnnn nnnnnnnnnn ccnggnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 48
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(35)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
```

-continued

```
<222> LOCATION: (38)..(38)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (41)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 48

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnccngg nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 49
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(21)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (26)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 49

nnnnnnnnnn nnnnnnnnnn nccggnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 50
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(35)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (40)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 50

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnccggn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 51
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(23)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (26)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 51

nnnnnnnnnn nnnnnnnnnn nnnggnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     60
```

-continued

```
<210> SEQ ID NO 52
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(33)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (38)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 52

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnggccnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 53
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (26)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 53

nnnnnnnnnn nnnnnnnnnn nncggnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 54
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(32)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (35)..(35)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (38)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 54

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnggnccnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 55
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
```

```
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(31)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (34)..(35)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (38)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 55

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nggnnccnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 56
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(30)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (33)..(35)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (38)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 56

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ggnnnccnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 57
<211> LENGTH: 60
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(29)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (32)..(35)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (38)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 57

nnnnnnnnnn nnnnnnnnnn nnnnnnnnng gnnnnccnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 58
<211> LENGTH: 60
```

-continued

```
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(28)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (31)..(35)
<223> OTHER INFORMATION: a, c, u, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (38)..(60)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 58

nnnnnnnnnn nnnnnnnnnn nnnnnnnngg nnnnnccnnn nnnnnnnnnn nnnnnnnnnn     60


<210> SEQ ID NO 59
<211> LENGTH: 12
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 59

guuuuagagc ua                                                         12


<210> SEQ ID NO 60
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Simian virus 40

<400> SEQUENCE: 60

Pro Lys Lys Lys Arg Lys Val
1               5


<210> SEQ ID NO 61
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Unknown:
      Nucleoplasmin bipartite NLS sequence"

<400> SEQUENCE: 61

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1               5                   10                  15


<210> SEQ ID NO 62
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Unknown:
      C-myc NLS sequence"

<400> SEQUENCE: 62

Pro Ala Ala Lys Arg Val Lys Leu Asp
```

-continued

```
1               5


<210> SEQ ID NO 63
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Unknown:
      C-myc NLS sequence"

<400> SEQUENCE: 63

Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro
1               5                   10


<210> SEQ ID NO 64
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 64

Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly
1               5                   10                  15

Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro
            20                  25                  30

Arg Asn Gln Gly Gly Tyr
        35


<210> SEQ ID NO 65
<211> LENGTH: 42
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Unknown:
      IBB domain from importin-alpha sequence"

<400> SEQUENCE: 65

Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu
1               5                   10                  15

Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys
            20                  25                  30

Asp Glu Gln Ile Leu Lys Arg Arg Asn Val
        35                  40


<210> SEQ ID NO 66
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Unknown:
      Myoma T protein sequence"

<400> SEQUENCE: 66

Val Ser Arg Lys Arg Pro Arg Pro
1               5


<210> SEQ ID NO 67
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
```

```
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Unknown:
      Myoma T protein sequence"

<400> SEQUENCE: 67

Pro Pro Lys Lys Ala Arg Glu Asp
1               5


<210> SEQ ID NO 68
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 68

Pro Gln Pro Lys Lys Lys Pro Leu
1               5


<210> SEQ ID NO 69
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 69

Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro
1               5                   10


<210> SEQ ID NO 70
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Influenza virus

<400> SEQUENCE: 70

Asp Arg Leu Arg Arg
1               5


<210> SEQ ID NO 71
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Influenza virus

<400> SEQUENCE: 71

Pro Lys Gln Lys Lys Arg Lys
1               5


<210> SEQ ID NO 72
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Hepatitus delta virus

<400> SEQUENCE: 72

Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu
1               5                   10


<210> SEQ ID NO 73
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 73

Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg
1               5                   10
```

-continued

```
<210> SEQ ID NO 74
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 74

Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys
1               5                   10                  15

Lys Ser Lys Lys
            20


<210> SEQ ID NO 75
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 75

Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys
1               5                   10                  15

Lys


<210> SEQ ID NO 76
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, t or g
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (21)..(22)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 76

nnnnnnnnnn nnnnnnnnnn nnagaaw                                          27


<210> SEQ ID NO 77
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(12)
<223> OTHER INFORMATION: a, c, t or g
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13)..(14)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 77

nnnnnnnnnn nnnnagaaw                                                   19


<210> SEQ ID NO 78
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
```

-continued

<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, t or g
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (21)..(22)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 78

nnnnnnnnnn nnnnnnnnnn nnagaaw 27

<210> SEQ ID NO 79
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(11)
<223> OTHER INFORMATION: a, c, t or g
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (12)..(13)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 79

nnnnnnnnnn nnnagaaw 18

<210> SEQ ID NO 80
<211> LENGTH: 137
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic polynucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 80

nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt tagaaataaa tcttgcagaa 60

gctacaaaga taaggcttca tgccgaaatc aacaccctgt cattttatgg cagggtgttt 120

tcgttattta atttttt 137

<210> SEQ ID NO 81
<211> LENGTH: 123
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic polynucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 81

-continued

```
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag     60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttcgt tatttaattt    120

ttt                                                                  123


<210> SEQ ID NO 82
<211> LENGTH: 110
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic polynucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 82

nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag     60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttttt               110


<210> SEQ ID NO 83
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic polynucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 83

nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc     60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tt                       102


<210> SEQ ID NO 84
<211> LENGTH: 88
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 84

nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc     60

cgttatcaac ttgaaaaagt gttttttt                                        88


<210> SEQ ID NO 85
<211> LENGTH: 76
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
```

-continued

```
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 85

nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc      60

cgttatcatt tttttt                                                      76


<210> SEQ ID NO 86
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 86

gttttagagc ta                                                          12


<210> SEQ ID NO 87
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 87

tagcaagtta aaataaggct agtccgtttt t                                     31


<210> SEQ ID NO 88
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 88

nnnnnnnnnn nnnnnnnnnn nnagaaw                                          27


<210> SEQ ID NO 89
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 89

ggacatcgat gtcacctcca atgactaggg tgg                                   33


<210> SEQ ID NO 90
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 90

cattggaggt gacatcgatg tcctccccat tgg                                   33
```

<210> SEQ ID NO 91
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 91

ggaagggcct gagtccgagc agaagaagaa ggg 33

<210> SEQ ID NO 92
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 92

ggtggcgaga ggggccgaga ttgggtgttc agg 33

<210> SEQ ID NO 93
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 93

atgcaggagg gtggcgagag gggccgagat tgg 33

<210> SEQ ID NO 94
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note=“Description of Artificial Sequence:
Synthetic primer”

<400> SEQUENCE: 94

aaactctaga gagggcctat ttcccatgat tc 32

<210> SEQ ID NO 95
<211> LENGTH: 153
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note=“Description of Artificial Sequence:
Synthetic primer”

<400> SEQUENCE: 95

acctctagaa aaaaagcacc gactcggtgc cactttttca agttgataac ggactagcct 60

tattttaact tgctatgctg ttttgtttcc aaaacagcat agctctaaaa cccctagtca 120

ttggaggtga cggtgtttcg tcctttccac aag 153

<210> SEQ ID NO 96
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note=“Description of Artificial Sequence:
Synthetic primer”

<400> SEQUENCE: 96

taatacgact cactatagga agtgcgccac catggcccca aagaagaagc gg 52

-continued

```
<210> SEQ ID NO 97
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 97

ggtttttttt tttttttttt tttttttttt ttttcttact ttttcttttt tgcctggccg     60


<210> SEQ ID NO 98
<211> LENGTH: 984
<212> TYPE: PRT
<213> ORGANISM: Campylobacter jejuni

<400> SEQUENCE: 98

Met Ala Arg Ile Leu Ala Phe Asp Ile Gly Ile Ser Ser Ile Gly Trp
1               5                   10                  15

Ala Phe Ser Glu Asn Asp Glu Leu Lys Asp Cys Gly Val Arg Ile Phe
            20                  25                  30

Thr Lys Val Glu Asn Pro Lys Thr Gly Glu Ser Leu Ala Leu Pro Arg
        35                  40                  45

Arg Leu Ala Arg Ser Ala Arg Lys Arg Leu Ala Arg Arg Lys Ala Arg
    50                  55                  60

Leu Asn His Leu Lys His Leu Ile Ala Asn Glu Phe Lys Leu Asn Tyr
65                  70                  75                  80

Glu Asp Tyr Gln Ser Phe Asp Glu Ser Leu Ala Lys Ala Tyr Lys Gly
                85                  90                  95

Ser Leu Ile Ser Pro Tyr Glu Leu Arg Phe Arg Ala Leu Asn Glu Leu
            100                 105                 110

Leu Ser Lys Gln Asp Phe Ala Arg Val Ile Leu His Ile Ala Lys Arg
        115                 120                 125

Arg Gly Tyr Asp Asp Ile Lys Asn Ser Asp Asp Lys Glu Lys Gly Ala
    130                 135                 140

Ile Leu Lys Ala Ile Lys Gln Asn Glu Glu Lys Leu Ala Asn Tyr Gln
145                 150                 155                 160

Ser Val Gly Glu Tyr Leu Tyr Lys Glu Tyr Phe Gln Lys Phe Lys Glu
                165                 170                 175

Asn Ser Lys Glu Phe Thr Asn Val Arg Asn Lys Lys Glu Ser Tyr Glu
            180                 185                 190

Arg Cys Ile Ala Gln Ser Phe Leu Lys Asp Glu Leu Lys Leu Ile Phe
        195                 200                 205

Lys Lys Gln Arg Glu Phe Gly Phe Ser Phe Ser Lys Lys Phe Glu Glu
    210                 215                 220

Glu Val Leu Ser Val Ala Phe Tyr Lys Arg Ala Leu Lys Asp Phe Ser
225                 230                 235                 240

His Leu Val Gly Asn Cys Ser Phe Phe Thr Asp Glu Lys Arg Ala Pro
                245                 250                 255

Lys Asn Ser Pro Leu Ala Phe Met Phe Val Ala Leu Thr Arg Ile Ile
            260                 265                 270

Asn Leu Leu Asn Asn Leu Lys Asn Thr Glu Gly Ile Leu Tyr Thr Lys
        275                 280                 285
```

-continued

```
Asp Asp Leu Asn Ala Leu Leu Asn Glu Val Leu Lys Asn Gly Thr Leu
    290                 295                 300

Thr Tyr Lys Gln Thr Lys Lys Leu Leu Gly Leu Ser Asp Asp Tyr Glu
305                 310                 315                 320

Phe Lys Gly Glu Lys Gly Thr Tyr Phe Ile Glu Phe Lys Lys Tyr Lys
                325                 330                 335

Glu Phe Ile Lys Ala Leu Gly Glu His Asn Leu Ser Gln Asp Asp Leu
            340                 345                 350

Asn Glu Ile Ala Lys Asp Ile Thr Leu Ile Lys Asp Glu Ile Lys Leu
        355                 360                 365

Lys Lys Ala Leu Ala Lys Tyr Asp Leu Asn Gln Asn Gln Ile Asp Ser
    370                 375                 380

Leu Ser Lys Leu Glu Phe Lys Asp His Leu Asn Ile Ser Phe Lys Ala
385                 390                 395                 400

Leu Lys Leu Val Thr Pro Leu Met Leu Glu Gly Lys Lys Tyr Asp Glu
                405                 410                 415

Ala Cys Asn Glu Leu Asn Leu Lys Val Ala Ile Asn Glu Asp Lys Lys
            420                 425                 430

Asp Phe Leu Pro Ala Phe Asn Glu Thr Tyr Tyr Lys Asp Glu Val Thr
        435                 440                 445

Asn Pro Val Val Leu Arg Ala Ile Lys Glu Tyr Arg Lys Val Leu Asn
    450                 455                 460

Ala Leu Leu Lys Lys Tyr Gly Lys Val His Lys Ile Asn Ile Glu Leu
465                 470                 475                 480

Ala Arg Glu Val Gly Lys Asn His Ser Gln Arg Ala Lys Ile Glu Lys
                485                 490                 495

Glu Gln Asn Glu Asn Tyr Lys Ala Lys Lys Asp Ala Glu Leu Glu Cys
            500                 505                 510

Glu Lys Leu Gly Leu Lys Ile Asn Ser Lys Asn Ile Leu Lys Leu Arg
        515                 520                 525

Leu Phe Lys Glu Gln Lys Glu Phe Cys Ala Tyr Ser Gly Glu Lys Ile
    530                 535                 540

Lys Ile Ser Asp Leu Gln Asp Glu Lys Met Leu Glu Ile Asp His Ile
545                 550                 555                 560

Tyr Pro Tyr Ser Arg Ser Phe Asp Asp Ser Tyr Met Asn Lys Val Leu
                565                 570                 575

Val Phe Thr Lys Gln Asn Gln Glu Lys Leu Asn Gln Thr Pro Phe Glu
            580                 585                 590

Ala Phe Gly Asn Asp Ser Ala Lys Trp Gln Lys Ile Glu Val Leu Ala
        595                 600                 605

Lys Asn Leu Pro Thr Lys Lys Gln Lys Arg Ile Leu Asp Lys Asn Tyr
    610                 615                 620

Lys Asp Lys Glu Gln Lys Asn Phe Lys Asp Arg Asn Leu Asn Asp Thr
625                 630                 635                 640

Arg Tyr Ile Ala Arg Leu Val Leu Asn Tyr Thr Lys Asp Tyr Leu Asp
                645                 650                 655

Phe Leu Pro Leu Ser Asp Asp Glu Asn Thr Lys Leu Asn Asp Thr Gln
            660                 665                 670

Lys Gly Ser Lys Val His Val Glu Ala Lys Ser Gly Met Leu Thr Ser
        675                 680                 685

Ala Leu Arg His Thr Trp Gly Phe Ser Ala Lys Asp Arg Asn Asn His
```

-continued

```
    690                 695                 700

Leu His His Ala Ile Asp Ala Val Ile Ile Ala Tyr Ala Asn Asn Ser
705                 710                 715                 720

Ile Val Lys Ala Phe Ser Asp Phe Lys Lys Glu Gln Glu Ser Asn Ser
                725                 730                 735

Ala Glu Leu Tyr Ala Lys Lys Ile Ser Glu Leu Asp Tyr Lys Asn Lys
            740                 745                 750

Arg Lys Phe Phe Glu Pro Phe Ser Gly Phe Arg Gln Lys Val Leu Asp
        755                 760                 765

Lys Ile Asp Glu Ile Phe Val Ser Lys Pro Glu Arg Lys Lys Pro Ser
    770                 775                 780

Gly Ala Leu His Glu Glu Thr Phe Arg Lys Glu Glu Glu Phe Tyr Gln
785                 790                 795                 800

Ser Tyr Gly Gly Lys Glu Gly Val Leu Lys Ala Leu Glu Leu Gly Lys
                805                 810                 815

Ile Arg Lys Val Asn Gly Lys Ile Val Lys Asn Gly Asp Met Phe Arg
            820                 825                 830

Val Asp Ile Phe Lys His Lys Lys Thr Asn Lys Phe Tyr Ala Val Pro
        835                 840                 845

Ile Tyr Thr Met Asp Phe Ala Leu Lys Val Leu Pro Asn Lys Ala Val
    850                 855                 860

Ala Arg Ser Lys Lys Gly Glu Ile Lys Asp Trp Ile Leu Met Asp Glu
865                 870                 875                 880

Asn Tyr Glu Phe Cys Phe Ser Leu Tyr Lys Asp Ser Leu Ile Leu Ile
                885                 890                 895

Gln Thr Lys Asp Met Gln Glu Pro Glu Phe Val Tyr Tyr Asn Ala Phe
            900                 905                 910

Thr Ser Ser Thr Val Ser Leu Ile Val Ser Lys His Asp Asn Lys Phe
        915                 920                 925

Glu Thr Leu Ser Lys Asn Gln Lys Ile Leu Phe Lys Asn Ala Asn Glu
    930                 935                 940

Lys Glu Val Ile Ala Lys Ser Ile Gly Ile Gln Asn Leu Lys Val Phe
945                 950                 955                 960

Glu Lys Tyr Ile Val Ser Ala Leu Gly Glu Val Thr Lys Ala Glu Phe
                965                 970                 975

Arg Gln Arg Glu Asp Phe Lys Lys
            980


<210> SEQ ID NO 99
<211> LENGTH: 91
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 99

tataatctca taagaaattt aaaaagggac taaaataaag agtttgcggg actctgcggg     60

gttacaatcc cctaaaaccg cttttaaaat t                                    91


<210> SEQ ID NO 100
<211> LENGTH: 36
<212> TYPE: DNA
```

-continued

```
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 100

attttaccat aaagaaattt aaaaagggac taaaac                               36


<210> SEQ ID NO 101
<211> LENGTH: 95
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 101

nnnnnnnnnn nnnnnnnnnn guuuuagucc cgaaagggac uaaaauaaag aguuugcggg     60

acucugcggg guuacaaucc ccuaaaaccg cuuuu                                95


<210> SEQ ID NO 102
<211> LENGTH: 1115
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic polypeptide"

<400> SEQUENCE: 102

Met Ser Asp Leu Val Leu Gly Leu Asp Ile Gly Ile Gly Ser Val Gly
1               5                   10                  15

Val Gly Ile Leu Asn Lys Val Thr Gly Glu Ile Ile His Lys Asn Ser
            20                  25                  30

Arg Ile Phe Pro Ala Ala Gln Ala Glu Asn Asn Leu Val Arg Arg Thr
        35                  40                  45

Asn Arg Gln Gly Arg Arg Leu Ala Arg Arg Lys Lys His Arg Arg Val
    50                  55                  60

Arg Leu Asn Arg Leu Phe Glu Glu Ser Gly Leu Ile Thr Asp Phe Thr
65                  70                  75                  80

Lys Ile Ser Ile Asn Leu Asn Pro Tyr Gln Leu Arg Val Lys Gly Leu
                85                  90                  95

Thr Asp Glu Leu Ser Asn Glu Glu Leu Phe Ile Ala Leu Lys Asn Met
            100                 105                 110

Val Lys His Arg Gly Ile Ser Tyr Leu Asp Asp Ala Ser Asp Asp Gly
        115                 120                 125

Asn Ser Ser Val Gly Asp Tyr Ala Gln Ile Val Lys Glu Asn Ser Lys
    130                 135                 140

Gln Leu Glu Thr Lys Thr Pro Gly Gln Ile Gln Leu Glu Arg Tyr Gln
145                 150                 155                 160

Thr Tyr Gly Gln Leu Arg Gly Asp Phe Thr Val Glu Lys Asp Gly Lys
                165                 170                 175

Lys His Arg Leu Ile Asn Val Phe Pro Thr Ser Ala Tyr Arg Ser Glu
```

-continued

```
            180                 185                 190

Ala Leu Arg Ile Leu Gln Thr Gln Gln Glu Phe Asn Pro Gln Ile Thr
        195                 200                 205

Asp Glu Phe Ile Asn Arg Tyr Leu Glu Ile Leu Thr Gly Lys Arg Lys
    210                 215                 220

Tyr Tyr His Gly Pro Gly Asn Glu Lys Ser Arg Thr Asp Tyr Gly Arg
225                 230                 235                 240

Tyr Arg Thr Ser Gly Glu Thr Leu Asp Asn Ile Phe Gly Ile Leu Ile
                245                 250                 255

Gly Lys Cys Thr Phe Tyr Pro Asp Glu Phe Arg Ala Ala Lys Ala Ser
            260                 265                 270

Tyr Thr Ala Gln Glu Phe Asn Leu Leu Asn Asp Leu Asn Asn Leu Thr
        275                 280                 285

Val Pro Thr Glu Thr Lys Lys Leu Ser Lys Glu Gln Lys Asn Gln Ile
    290                 295                 300

Ile Asn Tyr Val Lys Asn Glu Lys Ala Met Gly Pro Ala Lys Leu Phe
305                 310                 315                 320

Lys Tyr Ile Ala Lys Leu Leu Ser Cys Asp Val Ala Asp Ile Lys Gly
                325                 330                 335

Tyr Arg Ile Asp Lys Ser Gly Lys Ala Glu Ile His Thr Phe Glu Ala
            340                 345                 350

Tyr Arg Lys Met Lys Thr Leu Glu Thr Leu Asp Ile Glu Gln Met Asp
        355                 360                 365

Arg Glu Thr Leu Asp Lys Leu Ala Tyr Val Leu Thr Leu Asn Thr Glu
    370                 375                 380

Arg Glu Gly Ile Gln Glu Ala Leu Glu His Glu Phe Ala Asp Gly Ser
385                 390                 395                 400

Phe Ser Gln Lys Gln Val Asp Glu Leu Val Gln Phe Arg Lys Ala Asn
                405                 410                 415

Ser Ser Ile Phe Gly Lys Gly Trp His Asn Phe Ser Val Lys Leu Met
            420                 425                 430

Met Glu Leu Ile Pro Glu Leu Tyr Glu Thr Ser Glu Glu Gln Met Thr
        435                 440                 445

Ile Leu Thr Arg Leu Gly Lys Gln Lys Thr Thr Ser Ser Ser Asn Lys
    450                 455                 460

Thr Lys Tyr Ile Asp Glu Lys Leu Leu Thr Glu Glu Ile Tyr Asn Pro
465                 470                 475                 480

Val Val Ala Lys Ser Val Arg Gln Ala Ile Lys Ile Val Asn Ala Ala
                485                 490                 495

Ile Lys Glu Tyr Gly Asp Phe Asp Asn Ile Val Ile Glu Met Ala Arg
            500                 505                 510

Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met
        515                 520                 525

Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys
    530                 535                 540

Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu
545                 550                 555                 560

Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp
                565                 570                 575

Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser
            580                 585                 590
```

-continued

```
Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp
        595                 600                 605

Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys
    610                 615                 620

Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr
625                 630                 635                 640

Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser
                645                 650                 655

Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
            660                 665                 670

Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr
        675                 680                 685

Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr
    690                 695                 700

Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr
705                 710                 715                 720

Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu
                725                 730                 735

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
            740                 745                 750

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met
        755                 760                 765

Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe
    770                 775                 780

Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
785                 790                 795                 800

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
                805                 810                 815

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys
            820                 825                 830

Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln
        835                 840                 845

Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp
    850                 855                 860

Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly
865                 870                 875                 880

Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val
                885                 890                 895

Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly
            900                 905                 910

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe
        915                 920                 925

Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
    930                 935                 940

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met
945                 950                 955                 960

Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro
                965                 970                 975

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
            980                 985                 990
```

-continued

```
Lys Gly Ser Pro Glu Asp Asn Glu  Gln Lys Gln Leu Phe  Val Glu Gln
        995                 1000                 1005

His Lys  His Tyr Leu Asp Glu  Ile Ile Glu Gln Ile  Ser Glu Phe
    1010                 1015                 1020

Ser Lys  Arg Val Ile Leu Ala  Asp Ala Asn Leu Asp  Lys Val Leu
    1025                 1030                 1035

Ser Ala  Tyr Asn Lys His Arg  Asp Lys Pro Ile Arg  Glu Gln Ala
    1040                 1045                 1050

Glu Asn  Ile Ile His Leu Phe  Thr Leu Thr Asn Leu  Gly Ala Pro
    1055                 1060                 1065

Ala Ala  Phe Lys Tyr Phe Asp  Thr Thr Ile Asp Arg  Lys Arg Tyr
    1070                 1075                 1080

Thr Ser  Thr Lys Glu Val Leu  Asp Ala Thr Leu Ile  His Gln Ser
    1085                 1090                 1095

Ile Thr  Gly Leu Tyr Glu Thr  Arg Ile Asp Leu Ser  Gln Leu Gly
    1100                 1105                 1110

Gly Asp
    1115


<210> SEQ ID NO 103
<211> LENGTH: 1374
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic polypeptide"

<400> SEQUENCE: 103

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

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
```

-continued

```
        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
        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
```

-continued

```
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 Thr Asn
        755                 760                 765

Glu Asp Asp Glu Lys Lys Ala Ile Gln Lys Ile Gln Lys Ala Asn Lys
    770                 775                 780

Asp Glu Lys Asp Ala Ala Met Leu Lys Ala Ala Asn Gln Tyr Asn Gly
785                 790                 795                 800

Lys Ala Glu Leu Pro His Ser Val Phe His Gly His Lys Gln Leu Ala
                805                 810                 815

Thr Lys Ile Arg Leu Trp His Gln Gln Gly Glu Arg Cys Leu Tyr Thr
            820                 825                 830

Gly Lys Thr Ile Ser Ile His Asp Leu Ile Asn Asn Ser Asn Gln Phe
        835                 840                 845

Glu Val Asp His Ile Leu Pro Leu Ser Ile Thr Phe Asp Asp Ser Leu
    850                 855                 860

Ala Asn Lys Val Leu Val Tyr Ala Thr Ala Asn Gln Glu Lys Gly Gln
865                 870                 875                 880

Arg Thr Pro Tyr Gln Ala Leu Asp Ser Met Asp Asp Ala Trp Ser Phe
                885                 890                 895

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

Lys Lys Glu Tyr Leu Leu Thr Glu Glu Asp Ile Ser Lys Phe Asp Val
        915                 920                 925

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

Arg Val Val Leu Asn Ala Leu Gln Glu His Phe Arg Ala His Lys Ile
945                 950                 955                 960

Asp Thr Lys Val Ser Val Val Arg Gly Gln Phe Thr Ser Gln Leu Arg
                965                 970                 975

Arg His Trp Gly Ile Glu Lys Thr Arg Asp Thr Tyr His His His Ala
            980                 985                 990

Val Asp Ala Leu Ile Ile Ala Ala  Ser Ser Gln Leu Asn  Leu Trp Lys
        995                 1000                 1005
```

-continued

```
Lys Gln  Lys Asn Thr Leu Val  Ser Tyr Ser Glu Asp  Gln Leu Leu
    1010                 1015                 1020

Asp Ile  Glu Thr Gly Glu Leu  Ile Ser Asp Asp Glu  Tyr Lys Glu
    1025                 1030                 1035

Ser Val  Phe Lys Ala Pro Tyr  Gln His Phe Val Asp  Thr Leu Lys
    1040                 1045                 1050

Ser Lys  Glu Phe Glu Asp Ser  Ile Leu Phe Ser Tyr  Gln Val Asp
    1055                 1060                 1065

Ser Lys  Phe Asn Arg Lys Ile  Ser Asp Ala Thr Ile  Tyr Ala Thr
    1070                 1075                 1080

Arg Gln  Ala Lys Val Gly Lys  Asp Lys Ala Asp Glu  Thr Tyr Val
    1085                 1090                 1095

Leu Gly  Lys Ile Lys Asp Ile  Tyr Thr Gln Asp Gly  Tyr Asp Ala
    1100                 1105                 1110

Phe Met  Lys Ile Tyr Lys Lys  Asp Lys Ser Lys Phe  Leu Met Tyr
    1115                 1120                 1125

Arg His  Asp Pro Gln Thr Phe  Glu Lys Val Ile Glu  Pro Ile Leu
    1130                 1135                 1140

Glu Asn  Tyr Pro Asn Lys Gln  Ile Asn Glu Lys Gly  Lys Glu Val
    1145                 1150                 1155

Pro Cys  Asn Pro Phe Leu Lys  Tyr Lys Glu Glu His  Gly Tyr Ile
    1160                 1165                 1170

Arg Lys  Tyr Ser Lys Lys Gly  Asn Gly Pro Glu Ile  Lys Ser Leu
    1175                 1180                 1185

Lys Tyr  Tyr Asp Ser Lys Leu  Gly Asn His Ile Asp  Ile Thr Pro
    1190                 1195                 1200

Lys Asp  Ser Asn Asn Lys Val  Val Leu Gln Ser Val  Ser Pro Trp
    1205                 1210                 1215

Arg Ala  Asp Val Tyr Phe Asn  Lys Thr Thr Gly Lys  Tyr Glu Ile
    1220                 1225                 1230

Leu Gly  Leu Lys Tyr Ala Asp  Leu Gln Phe Glu Lys  Gly Thr Gly
    1235                 1240                 1245

Thr Tyr  Lys Ile Ser Gln Glu  Lys Tyr Asn Asp Ile  Lys Lys Lys
    1250                 1255                 1260

Glu Gly  Val Asp Ser Asp Ser  Glu Phe Lys Phe Thr  Leu Tyr Lys
    1265                 1270                 1275

Asn Asp  Leu Leu Leu Val Lys  Asp Thr Glu Thr Lys  Glu Gln Gln
    1280                 1285                 1290

Leu Phe  Arg Phe Leu Ser Arg  Thr Met Pro Lys Gln  Lys His Tyr
    1295                 1300                 1305

Val Glu  Leu Lys Pro Tyr Asp  Lys Gln Lys Phe Glu  Gly Gly Glu
    1310                 1315                 1320

Ala Leu  Ile Lys Val Leu Gly  Asn Val Ala Asn Ser  Gly Gln Cys
    1325                 1330                 1335

Lys Lys  Gly Leu Gly Lys Ser  Asn Ile Ser Ile Tyr  Lys Val Arg
    1340                 1345                 1350

Thr Asp  Val Leu Gly Asn Gln  His Ile Ile Lys Asn  Glu Gly Asp
    1355                 1360                 1365

Lys Pro  Lys Leu Asp Phe
    1370
```

-continued

```
<210> SEQ ID NO 104
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic 6xHis tag"

<400> SEQUENCE: 104

His His His His His His
1               5


<210> SEQ ID NO 105
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic peptide"

<400> SEQUENCE: 105

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1               5                   10                  15


<210> SEQ ID NO 106
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic peptide"

<400> SEQUENCE: 106

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys
1               5                   10                  15


<210> SEQ ID NO 107
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic peptide"

<400> SEQUENCE: 107

Gly Gly Gly Gly Gly Ser Gly Gly Gly Gly Gly Ser Gly Gly Gly Gly
1               5                   10                  15

Gly Ser


<210> SEQ ID NO 108
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 108

gccaaattgg acgaccctcg cgg                                             23
```

-continued

```
<210> SEQ ID NO 109
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 109

cgaggagacc cccgtttcgg tgg                                             23


<210> SEQ ID NO 110
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 110

cccgccgccg ccgtggctcg agg                                             23


<210> SEQ ID NO 111
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 111

tgagctctac gagatccaca agg                                             23


<210> SEQ ID NO 112
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 112

ctcaaaattc ataccggttg tgg                                             23


<210> SEQ ID NO 113
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 113

cgttaaacaa caaccggact tgg                                             23


<210> SEQ ID NO 114
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
```

-continued

```
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 114

ttcaccccgc ggcgctgaat ggg                                           23


<210> SEQ ID NO 115
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 115

accactacca gtccgtccac agg                                           23


<210> SEQ ID NO 116
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 116

agcctttctg aacacatgca cgg                                           23


<210> SEQ ID NO 117
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 117

cctgccatca atgtggccat gcatgtgttc agaaaggct                          39


<210> SEQ ID NO 118
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 118

cctgccatca atgtggccgt gcatgtgttc agaaaggct                          39


<210> SEQ ID NO 119
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 119

cactgcttaa gcctcgctcg agg                                           23


<210> SEQ ID NO 120
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
```

-continued

<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 120

tcaccagcaa tattcgctcg agg 23

<210> SEQ ID NO 121
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 121

caccagcaat attccgctcg agg 23

<210> SEQ ID NO 122
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 122

tagcaacaga catacgctcg agg 23

<210> SEQ ID NO 123
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 123

gggcagtagt aatacgctcg agg 23

<210> SEQ ID NO 124
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 124

ccaattccca tacattattg tac 23

<210> SEQ ID NO 125
<211> LENGTH: 4677
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic polynucleotide"

<400> SEQUENCE: 125

tctttcttgc gctatgacac ttccagcaaa aggtagggcg ggctgcgaga cggcttcccg 60

-continued

gcgctgcatg caacaccgat gatgcttcga ccccccgaag ctccttcggg gctgcatggg 120

cgctccgatg ccgctccagg gcgagcgctg tttaaatagc caggcccccg attgcaaaga 180

cattatagcg agctaccaaa gccatattca aacacctaga tcactaccac ttctacacag 240

gccactcgag cttgtgatcg cactccgcta agggggcgcc tcttcctctt cgtttcagtc 300

acaacccgca aacatgtacc catacgatgt tccagattac gcttcgccga agaaaaagcg 360

caaggtcgaa gcgtccgaca agaagtacag catcggcctg gacatcggca ccaactctgt 420

gggctgggcc gtgatcaccg acgagtacaa ggtgcccagc aagaaattca aggtgctggg 480

caacaccgac cggcacagca tcaagaagaa cctgatcgga gccctgctgt tcgacagcgg 540

cgaaacagcc gaggccaccc ggctgaagag aaccgccaga agaagataca ccagacggaa 600

gaaccggatc tgctatctgc aagagatctt cagcaacgag atggccaagg tggacgacag 660

cttcttccac agactggaag agtccttcct ggtggaagag gataagaagc acgagcggca 720

ccccatcttc ggcaacatcg tggacgaggt ggcctaccac gagaagtacc ccaccatcta 780

ccacctgaga aagaaactgg tggacagcac cgacaaggcc gacctgcggc tgatctatct 840

ggccctggcc cacatgatca agttccgggg ccacttcctg atcgagggcg acctgaaccc 900

cgacaacagc gacgtggaca agctgttcat ccagctggtg cagacctaca accagctgtt 960

cgaggaaaac cccatcaacg ccagcggcgt ggacgccaag gccatcctgt ctgccagact 1020

gagcaagagc agacggctgg aaaatctgat cgcccagctg cccggcgaga agaagaatgg 1080

cctgttcggc aacctgattg ccctgagcct gggcctgacc cccaacttca agagcaactt 1140

cgacctggcc gaggatgcca aactgcagct gagcaaggac acctacgacg acgacctgga 1200

caacctgctg gcccagatcg gcgaccagta cgccgacctg tttctggccg ccaagaacct 1260

gtccgacgcc atcctgctga gcgacatcct gagagtgaac accgagatca ccaaggcccc 1320

cctgagcgcc tctatgatca agagatacga cgagcaccac caggacctga ccctgctgaa 1380

agctctcgtg cggcagcagc tgcctgagaa gtacaaagag attttcttcg accagagcaa 1440

gaacggctac gccggctaca ttgacggcgg agccagccag gaagagttct acaagttcat 1500

caagcccatc ctggaaaaga tggacggcac cgaggaactg ctcgtgaagc tgaacagaga 1560

ggacctgctg cggaagcagc ggaccttcga caacggcagc atcccccacc agatccacct 1620

gggagagctg cacgccattc tgcggcggca ggaagatttt tacccattcc tgaaggacaa 1680

ccgggaaaag atcgagaaga tcctgacctt ccgcatcccc tactacgtgg gccctctggc 1740

caggggaaac agcagattcg cctggatgac cagaaagagc gaggaaacca tcaccccctg 1800

gaacttcgag gaagtggtgg acaagggcgc ttccgcccag agcttcatcg agcggatgac 1860

caacttcgat aagaacctgc ccaacgagaa ggtgctgccc aagcacagcc tgctgtacga 1920

gtacttcacc gtgtataacg agctgaccaa agtgaaatac gtgaccgagg gaatgagaaa 1980

gcccgccttc ctgagcggcg agcagaaaaa ggccatcgtg gacctgctgt tcaagaccaa 2040

ccggaaagtg accgtgaagc agctgaaaga ggactacttc aagaaaatcg agtgcttcga 2100

ctccgtggaa atctccggcg tggaagatcg gttcaacgcc tccctgggca cataccacga 2160

tctgctgaaa attatcaagg acaaggactt cctggacaat gaggaaaacg aggacattct 2220

ggaagatatc gtgctgaccc tgacactgtt tgaggacaga gagatgatcg aggaacggct 2280

gaaaacctat gcccacctgt tcgacgacaa agtgatgaag cagctgaagc ggcggagata 2340

-continued

caccggctgg ggcaggctga gccggaagct gatcaacggc atccgggaca agcagtccgg 2400

caagacaatc ctggatttcc tgaagtccga cggcttcgcc aacagaaact tcatgcagct 2460

gatccacgac gacagcctga cctttaaaga ggacatccag aaagcccagg tgtccggcca 2520

gggcgatagc ctgcacgagc acattgccaa tctggccggc agccccgcca ttaagaaggg 2580

catcctgcag acagtgaagg tggtggacga gctcgtgaaa gtgatgggcc ggcacaagcc 2640

cgagaacatc gtgatcgaaa tggccagaga gaaccagacc acccagaagg gacagaagaa 2700

cagccgcgag agaatgaagc ggatcgaaga gggcatcaaa gagctgggca gccagatcct 2760

gaaagaacac cccgtggaaa acacccagct gcagaacgag aagctgtacc tgtactacct 2820

gcagaatggg cgggatatgt acgtggacca ggaactggac atcaaccggc tgtccgacta 2880

cgatgtggac catatcgtgc ctcagagctt tctgaaggac gactccatcg acaacaaggt 2940

gctgaccaga agcgacaaga accggggcaa gagcgacaac gtgccctccg aagaggtcgt 3000

gaagaagatg aagaactact ggcggcagct gctgaacgcc aagctgatta cccagagaaa 3060

gttcgacaat ctgaccaagg ccgagagagg cggcctgagc gaactggata aggccggctt 3120

catcaagaga cagctggtgg aaacccggca gatcacaaag cacgtggcac agatcctgga 3180

ctcccggatg aacactaagt acgacgagaa tgacaagctg atccgggaag tgaaagtgat 3240

caccctgaag tccaagctgg tgtccgattt ccggaaggat ttccagtttt acaaagtgcg 3300

cgagatcaac aactaccacc acgcccacga cgcctacctg aacgccgtcg tgggaaccgc 3360

cctgatcaaa aagtacccta agctggaaag cgagttcgtg tacggcgact acaaggtgta 3420

cgacgtgcgg aagatgatcg ccaagagcga gcaggaaatc ggcaaggcta ccgccaagta 3480

cttcttctac agcaacatca tgaacttttt caagaccgag attaccctgg ccaacggcga 3540

gatccggaag cggcctctga tcgagacaaa cggcgaaacc ggggagatcg tgtgggataa 3600

gggccgggat tttgccaccg tgcggaaagt gctgagcatg ccccaagtga atatcgtgaa 3660

aaagaccgag gtgcagacag gcggcttcag caaagagtct atcctgccca agaggaacag 3720

cgataagctg atcgccagaa agaaggactg ggaccctaag aagtacggcg gcttcgacag 3780

ccccaccgtg gcctattctg tgctggtggt ggccaaagtg gaaaagggca agtccaagaa 3840

actgaagagt gtgaaagagc tgctggggat caccatcatg gaaagaagca gcttcgagaa 3900

gaatcccatc gactttctgg aagccaaggg ctacaaagaa gtgaaaaagg acctgatcat 3960

caagctgcct aagtactccc tgttcgagct ggaaaacggc cggaagagaa tgctggcctc 4020

tgccggcgaa ctgcagaagg gaaacgaact ggccctgccc tccaaatatg tgaacttcct 4080

gtacctggcc agccactatg agaagctgaa gggctccccc gaggataatg agcagaaaca 4140

gctgtttgtg gaacagcaca agcactacct ggacgagatc atcgagcaga tcagcgagtt 4200

ctccaagaga gtgatcctgg ccgacgctaa tctggacaaa gtgctgtccg cctacaacaa 4260

gcaccgggat aagcccatca gagagcaggc cgagaatatc atccacctgt ttaccctgac 4320

caatctggga gcccctgccg ccttcaagta ctttgacacc accatcgacc ggaagaggta 4380

caccagcacc aaagaggtgc tggacgccac cctgatccac cagagcatca ccggcctgta 4440

cgagacacgg atcgacctgt ctcagctggg aggcgacagc cccaagaaga agagaaaggt 4500

ggaggccagc taaggatccg gcaagactgg ccccgcttgg caacgcaaca gtgagcccct 4560

ccctagtgtg tttggggatg tgactatgta ttcgtgtgtt ggccaacggg tcaacccgaa 4620

-continued

```
cagattgata cccgccttgg catttcctgt cagaatgtaa cgtcagttga tggtact      4677


<210> SEQ ID NO 126
<211> LENGTH: 3150
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic polynucleotide"

<400> SEQUENCE: 126

tctttcttgc gctatgacac ttccagcaaa aggtagggcg ggctgcgaga cggcttcccg     60

gcgctgcatg caacaccgat gatgcttcga ccccccgaag ctccttcggg gctgcatggg    120

cgctccgatg ccgctccagg gcgagcgctg tttaaatagc caggcccccg attgcaaaga    180

cattatagcg agctaccaaa gccatattca aacacctaga tcactaccac ttctacacag    240

gccactcgag cttgtgatcg cactccgcta agggggcgcc tcttcctctt cgtttcagtc    300

acaacccgca aacatgccta agaagaagag gaaggttaac acgattaaca tcgctaagaa    360

cgacttctct gacatcgaac tggctgctat cccgttcaac actctggctg accattacgg    420

tgagcgttta gctcgcgaac agttggccct tgagcatgag tcttacgaga tgggtgaagc    480

acgcttccgc aagatgtttg agcgtcaact taaagctggt gaggttgcgg ataacgctgc    540

cgccaagcct ctcatcacta ccctactccc taagatgatt gcacgcatca acgactggtt    600

tgaggaagtg aaagctaagc gcggcaagcg cccgacagcc ttccagttcc tgcaagaaat    660

caagccggaa gccgtagcgt acatcaccat taagaccact ctggcttgcc taaccagtgc    720

tgacaataca accgttcagg ctgtagcaag cgcaatcggt cgggccattg aggacgaggc    780

tcgcttcggt cgtatccgtg accttgaagc taagcacttc aagaaaaacg ttgaggaaca    840

actcaacaag cgcgtagggc acgtctacaa gaaagcattt atgcaagttg tcgaggctga    900

catgctctct aagggtctac tcggtggcga ggcgtggtct tcgtggcata aggaagactc    960

tattcatgta ggagtacgct gcatcgagat gctcattgag tcaaccggaa tggttagctt   1020

acaccgccaa aatgctggcg tagtaggtca agactctgag actatcgaac tcgcacctga   1080

atacgctgag gctatcgcaa cccgtgcagg tgcgctggct ggcatctctc cgatgttcca   1140

accttgcgta gttcctccta agccgtggac tggcattact ggtggtggct attgggctaa   1200

cggtcgtcgt cctctggcgc tggtgcgtac tcacagtaag aaagcactga tgcgctacga   1260

agacgtttac atgcctgagg tgtacaaagc gattaacatt gcgcaaaaca ccgcatggaa   1320

aatcaacaag aaagtcctag cggtcgccaa cgtaatcacc aagtggaagc attgtccggt   1380

cgaggacatc cctgcgattg agcgtgaaga actcccgatg aaaccggaag acatcgacat   1440

gaatcctgag gctctcaccg cgtggaaacg tgctgccgct gctgtgtacc gcaaggacaa   1500

ggctcgcaag tctcgccgta tcagccttga gttcatgctt gagcaagcca ataagtttgc   1560

taaccataag gccatctggt tcccttacaa catggactgg cgcggtcgtg tttacgctgt   1620

gtcaatgttc aacccgcaag gtaacgatat gaccaaagga ctgcttacgc tggcgaaagg   1680

taaaccaatc ggtaaggaag gttactactg gctgaaaatc cacggtgcaa actgtgcggg   1740

tgtcgacaag gttccgttcc ctgagcgcat caagttcatt gaggaaaacc acgagaacat   1800

catggcttgc gctaagtctc cactggagaa cacttggtgg gctgagcaag attctccgtt   1860
```

-continued

ctgcttcctt gcgttctgct ttgagtacgc tggggtacag caccacggcc tgagctataa 1920

ctgctccctt ccgctggcgt ttgacgggtc ttgctctggc atccagcact tctccgcgat 1980

gctccgagat gaggtaggtg gtcgcgcggt taacttgctt cctagtgaaa ccgttcagga 2040

catctacggg attgttgcta agaaagtcaa cgagattcta caagcagacg caatcaatgg 2100

gaccgataac gaagtagtta ccgtgaccga tgagaacact ggtgaaatct ctgagaaagt 2160

caagctgggc actaaggcac tggctggtca atggctggct tacggtgtta ctcgcagtgt 2220

gactaagcgt tcagtcatga cgctggctta cgggtccaaa gagttcggct tccgtcaaca 2280

agtgctggaa gataccattc agccagctat tgattccggc aagggtctga tgttcactca 2340

gccgaatcag gctgctggat acatggctaa gctgatttgg gaatctgtga gcgtgacggt 2400

ggtagctgcg gttgaagcaa tgaactggct taagtctgct gctaagctgc tggctgctga 2460

ggtcaaagat aagaagactg gagagattct tcgcaagcgt tgcgctgtgc attgggtaac 2520

tcctgatggt ttccctgtgt ggcaggaata caagaagcct attcagacgc gcttgaacct 2580

gatgttcctc ggtcagttcc gcttacagcc taccattaac accaacaaag atagcgagat 2640

tgatgcacac aaacaggagt ctggtatcgc tcctaacttt gtacacagcc aagacggtag 2700

ccaccttcgt aagactgtag tgtgggcaca cgagaagtac ggaatcgaat cttttgcact 2760

gattcacgac tccttcggta cgattccggc tgacgctgcg aacctgttca aagcagtgcg 2820

cgaaactatg gttgacacat atgagtcttg tgatgtactg gctgatttct acgaccagtt 2880

cgctgaccag ttgcacgagt ctcaattgga caaaatgcca gcacttccgg ctaaaggtaa 2940

cttgaacctc cgtgacatct tagagtcgga cttcgcgttc gcgtaaggat ccggcaagac 3000

tggccccgct tggcaacgca acagtgagcc cctccctagt gtgtttgggg atgtgactat 3060

gtattcgtgt gttggccaac gggtcaaccc gaacagattg atacccgcct tggcatttcc 3120

tgtcagaatg taacgtcagt tgatggtact 3150

<210> SEQ ID NO 127
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic polynucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (23)..(42)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 127

gaaattaata cgactcacta tannnnnnnn nnnnnnnnnn nngttttaga gctagaaata 60

gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga gtcggtgctt 120

ttttt 125

<210> SEQ ID NO 128
<211> LENGTH: 8452
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic polynucleotide"

-continued

<400> SEQUENCE: 128

tgcggtattt cacaccgcat caggtggcac ttttcgggga aatgtgcgcg gaacccctat 60

ttgtttattt ttctaaatac attcaaatat gtatccgctc atgagattat caaaaaggat 120

cttcacctag atccttttaa attaaaaatg aagttttaaa tcaatctaaa gtatatatga 180

gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacctatct cagcgatctg 240

tctatttcgt tcatccatag ttgcctgact ccccgtcgtg tagataacta cgatacggga 300

gggcttacca tctggcccca gtgctgcaat gataccgcga gacccacgct caccggctcc 360

agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg gtcctgcaac 420

tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa gtagttcgcc 480

agttaatagt ttgcgcaacg ttgttgccat tgctacaggc atcgtggtgt cacgctcgtc 540

gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc 600

catgttgtgc aaaaaagcgg ttagctcctt cggtcctccg atcgttgtca gaagtaagtt 660

ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta ctgtcatgcc 720

atccgtaaga tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg 780

tatgcggcga ccgagttgct cttgcccggc gtcaatacgg gataataccg cgccacatag 840

cagaacttta aaagtgctca tcattggaaa acgttcttcg gggcgaaaac tctcaaggat 900

cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact gatcttcagc 960

atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa 1020

aaagggaata agggcgacac ggaaatgttg aatactcata ctcttccttt ttcaatatta 1080

ttgaagcatt tatcagggtt attgtctcat gaccaaaatc ccttaacgtg agttttcgtt 1140

ccactgagcg tcagaccccg tagaaaagat caaaggatct tcttgagatc cttttttct 1200

gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc 1260

ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc 1320

aaatactgtt cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc 1380

gcctacatac ctcgctctgc taatcctgtt accagtggct gttgccagtg gcgataagtc 1440

gtgtcttacc gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg 1500

aacggggggt tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata 1560

cctacagcgt gagctatgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta 1620

tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc 1680

ctggtatctt tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg 1740

atgctcgtca ggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt 1800

cctggccttt tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt 1860

ggataaccgt attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga 1920

gcgcagcgag tcagtgagcg aggaagcggt cgctgaggct tgacatgatt ggtgcgtatg 1980

tttgtatgaa gctacaggac tgatttggcg ggctatgagg gcgggggaag ctctggaagg 2040

gccgcgatgg ggcgcgcggc gtccagaagg cgccatacgg cccgctggcg gcacccatcc 2100

ggtataaaag cccgcgaccc cgaacggtga cctccacttt cagcgacaaa cgagcactta 2160

tacatacgcg actattctgc cgctatacat aaccactcag ctagcttaag atcccatcaa 2220

-continued

```
gcttgcatgc cgggcgcgcc agaaggagcg cagccaaacc aggatgatgt ttgatggggt   2280
atttgagcac ttgcaaccct tatccggaag ccccctggcc cacaaaggct aggcgccaat   2340
gcaagcagtt cgcatgcagc ccctggagcg gtgccctcct gataaaccgg ccagggggcc   2400
tatgttcttt actttttac aagagaagtc actcaacatc ttaaaatggc caggtgagtc   2460
gacgagcaag cccggcggat caggcagcgt gcttgcagat ttgacttgca acgcccgcat   2520
tgtgtcgacg aaggcttttg gctcctctgt cgctgtctca agcagcatct aaccctgcgt   2580
cgccgtttcc atttgcagga gattcgaggt accatgtacc catacgatgt tccagattac   2640
gcttcgccga agaaaaagcg caaggtcgaa gcgtccgaca agaagtacag catcggcctg   2700
gacatcggca ccaactctgt gggctgggcc gtgatcaccg acgagtacaa ggtgcccagc   2760
aagaaattca aggtgctggg caacaccgac cggcacagca tcaagaagaa cctgatcgga   2820
gccctgctgt tcgacagcgg cgaaacagcc gaggccaccc ggctgaagag aaccgccaga   2880
agaagataca ccagacggaa gaaccggatc tgctatctgc aagagatctt cagcaacgag   2940
atggccaagg tggacgacag cttcttccac agactggaag agtccttcct ggtggaagag   3000
gataagaagc acgagcggca ccccatcttc ggcaacatcg tggacgaggt ggcctaccac   3060
gagaagtacc ccaccatcta ccacctgaga aagaaactgg tggacagcac cgacaaggcc   3120
gacctgcggc tgatctatct ggccctggcc cacatgatca agttccgggg ccacttcctg   3180
atcgagggcg acctgaaccc cgacaacagc gacgtggaca agctgttcat ccagctggtg   3240
cagacctaca accagctgtt cgaggaaaac cccatcaacg ccagcggcgt ggacgccaag   3300
gccatcctgt ctgccagact gagcaagagc agacggctgg aaaatctgat cgcccagctg   3360
cccggcgaga agaagaatgg cctgttcggc aacctgattg ccctgagcct gggcctgacc   3420
cccaacttca agagcaactt cgacctggcc gaggatgcca aactgcagct gagcaaggac   3480
acctacgacg acgacctgga caacctgctg gcccagatcg gcgaccagta cgccgacctg   3540
tttctggccg ccaagaacct gtccgacgcc atcctgctga gcgacatcct gagagtgaac   3600
accgagatca ccaaggcccc cctgagcgcc tctatgatca agagatacga cgagcaccac   3660
caggacctga ccctgctgaa agctctcgtg cggcagcagc tgcctgagaa gtacaaagag   3720
attttcttcg accagagcaa gaacggctac gccggctaca ttgacggcgg agccagccag   3780
gaagagttct acaagttcat caagcccatc ctggaaaaga tggacggcac cgaggaactg   3840
ctcgtgaagc tgaacagaga ggacctgctg cggaagcagc ggaccttcga caacggcagc   3900
atcccccacc agatccacct gggagagctg cacgccattc tgcggcggca ggaagatttt   3960
tacccattcc tgaaggacaa ccgggaaaag atcgagaaga tcctgacctt ccgcatcccc   4020
tactacgtgg gccctctggc caggggaaac agcagattcg cctggatgac cagaaagagc   4080
gaggaaacca tcaccccctg gaacttcgag gaagtggtgg acaagggcgc ttccgcccag   4140
agcttcatcg agcggatgac caacttcgat aagaacctgc ccaacgagaa ggtgctgccc   4200
aagcacagcc tgctgtacga gtacttcacc gtgtataacg agctgaccaa agtgaaatac   4260
gtgaccgagg gaatgagaaa gcccgccttc ctgagcggcg agcagaaaaa ggccatcgtg   4320
gacctgctgt tcaagaccaa ccggaaagtg accgtgaagc agctgaaaga ggactacttc   4380
aagaaaatcg agtgcttcga ctccgtggaa atctccggcg tggaagatcg gttcaacgcc   4440
tccctgggca cataccacga tctgctgaaa attatcaagg acaaggactt cctggacaat   4500
```

-continued

gaggaaaacg aggacattct ggaagatatc gtgctgaccc tgacactgtt tgaggacaga 4560

gagatgatcg aggaacggct gaaaacctat gcccacctgt tcgacgacaa agtgatgaag 4620

cagctgaagc ggcggagata caccggctgg ggcaggctga gccggaagct gatcaacggc 4680

atccgggaca agcagtccgg caagacaatc ctggatttcc tgaagtccga cggcttcgcc 4740

aacagaaact tcatgcagct gatccacgac gacagcctga cctttaaaga ggacatccag 4800

aaagcccagg tgtccggcca gggcgatagc ctgcacgagc acattgccaa tctggccggc 4860

agccccgcca ttaagaaggg catcctgcag acagtgaagg tggtggacga gctcgtgaaa 4920

gtgatgggcc ggcacaagcc cgagaacatc gtgatcgaaa tggccagaga gaaccagacc 4980

acccagaagg gacagaagaa cagccgcgag agaatgaagc ggatcgaaga gggcatcaaa 5040

gagctgggca gccagatcct gaaagaacac cccgtggaaa acacccagct gcagaacgag 5100

aagctgtacc tgtactacct gcagaatggg cgggatatgt acgtggacca ggaactggac 5160

atcaaccggc tgtccgacta cgatgtggac catatcgtgc ctcagagctt tctgaaggac 5220

gactccatcg acaacaaggt gctgaccaga agcgacaaga accggggcaa gagcgacaac 5280

gtgccctccg aagaggtcgt gaagaagatg aagaactact ggcggcagct gctgaacgcc 5340

aagctgatta cccagagaaa gttcgacaat ctgaccaagg ccgagagagg cggcctgagc 5400

gaactggata aggccggctt catcaagaga cagctggtgg aaacccggca gatcacaaag 5460

cacgtggcac agatcctgga ctcccggatg aacactaagt acgacgagaa tgacaagctg 5520

atccgggaag tgaaagtgat caccctgaag tccaagctgg tgtccgattt ccggaaggat 5580

ttccagtttt acaaagtgcg cgagatcaac aactaccacc acgcccacga cgcctacctg 5640

aacgccgtcg tgggaaccgc cctgatcaaa aagtacccta agctggaaag cgagttcgtg 5700

tacggcgact acaaggtgta cgacgtgcgg aagatgatcg ccaagagcga gcaggaaatc 5760

ggcaaggcta ccgccaagta cttcttctac agcaacatca tgaacttttt caagaccgag 5820

attaccctgg ccaacggcga gatccggaag cggcctctga tcgagacaaa cggcgaaacc 5880

ggggagatcg tgtgggataa gggccgggat tttgccaccg tgcggaaagt gctgagcatg 5940

ccccaagtga atatcgtgaa aaagaccgag gtgcagacag gcggcttcag caaagagtct 6000

atcctgccca agaggaacag cgataagctg atcgccagaa agaaggactg ggaccctaag 6060

aagtacggcg gcttcgacag ccccaccgtg gcctattctg tgctggtggt ggccaaagtg 6120

gaaaagggca agtccaagaa actgaagagt gtgaaagagc tgctggggat caccatcatg 6180

gaaagaagca gcttcgagaa gaatcccatc gactttctgg aagccaaggg ctacaaagaa 6240

gtgaaaaagg acctgatcat caagctgcct aagtactccc tgttcgagct ggaaaacggc 6300

cggaagagaa tgctggcctc tgccggcgaa ctgcagaagg gaaacgaact ggccctgccc 6360

tccaaatatg tgaacttcct gtacctggcc agccactatg agaagctgaa gggctccccc 6420

gaggataatg agcagaaaca gctgtttgtg gaacagcaca agcactacct ggacgagatc 6480

atcgagcaga tcagcgagtt ctccaagaga gtgatcctgg ccgacgctaa tctggacaaa 6540

gtgctgtccg cctacaacaa gcaccgggat aagcccatca gagagcaggc cgagaatatc 6600

atccacctgt ttaccctgac caatctggga gcccctgccg ccttcaagta ctttgacacc 6660

accatcgacc ggaagaggta caccagcacc aaagaggtgc tggacgccac cctgatccac 6720

cagagcatca ccggcctgta cgagacacgg atcgacctgt ctcagctggg aggcgacagc 6780

-continued

```
cccaagaaga agagaaaggt ggaggccagc taacatatga ttcgaatgtc tttcttgcgc   6840

tatgacactt ccagcaaaag gtagggcggg ctgcgagacg gcttcccggc gctgcatgca   6900

acaccgatga tgcttcgacc ccccgaagct ccttcggggc tgcatgggcg ctccgatgcc   6960

gctccagggc gagcgctgtt taaatagcca ggcccccgat tgcaaagaca ttatagcgag   7020

ctaccaaagc catattcaaa cacctagatc actaccactt ctacacaggc cactcgagct   7080

tgtgatcgca ctccgctaag gggcgcctc ttcctcttcg tttcagtcac aacccgcaaa   7140

catgacacaa gaatccctgt tacttctcga ccgtattgat tcggatgatt cctacgcgag   7200

cctgcggaac gaccaggaat tctgggaggt gagtcgacga gcaagcccgg cggatcaggc   7260

agcgtgcttg cagatttgac ttgcaacgcc cgcattgtgt cgacgaaggc ttttggctcc   7320

tctgtcgctg tctcaagcag catctaaccc tgcgtcgccg tttccatttg cagccgctgg   7380

cccgccgagc cctggaggag ctcgggctgc cggtgccgcc ggtgctgcgg gtgcccggcg   7440

agagcaccaa ccccgtactg gtcggcgagc ccggcccggt gatcaagctg ttcggcgagc   7500

actggtgcgg tccggagagc ctcgcgtcgg agtcggaggc gtacgcggtc ctggcggacg   7560

ccccggtgcc ggtgccccgc ctcctcggcc gcggcgagct gcggcccggc accggagcct   7620

ggccgtggcc ctacctggtg atgagccgga tgaccggcac cacctggcgg tccgcgatgg   7680

acggcacgac cgaccggaac gcgctgctcg ccctggcccg cgaactcggc cgggtgctcg   7740

gccggctgca cagggtgccg ctgaccggga acaccgtgct cacccccccat tccgaggtct   7800

tcccggaact gctgcgggaa cgccgcgcgg cgaccgtcga ggaccaccgc gggtggggct   7860

acctctcgcc ccggctgctg gaccgcctgg aggactggct gccggacgtg gacacgctgc   7920

tggccggccg cgaaccccgg ttcgtccacg gcgacctgca cgggaccaac atcttcgtgg   7980

acctggccgc gaccgaggtc accgggatcg tcgacttcac cgacgtctat gcgggagact   8040

cccgctacag cctggtgcaa ctgcatctca acgccttccg ggcgaccgc gagatcctgg   8100

ccgcgctgct cgacggggcg cagtggaagc ggaccgagga cttcgcccgc gaactgctcg   8160

ccttcacctt cctgcacgac ttcgaggtgt tcgaggagac cccgctggat ctctccggct   8220

tcaccgatcc ggaggaactg gcgcagttcc tctgggggcc gccggacacc gcccccggcg   8280

cctgataagg atccggcaag actggccccg cttggcaacg caacagtgag cccctcccta   8340

gtgtgtttgg ggatgtgact atgtattcgt gtgttggcca acgggtcaac ccgaacagat   8400

tgatacccgc cttggcattt cctgtcagaa tgtaacgtca gttgatggta ct           8452


<210> SEQ ID NO 129
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 129

ccgtgccggg cggggagacc gccatgg                                         27


<210> SEQ ID NO 130
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 130

ggcccggctg tggctgagga gc                                              22
```

-continued

```
<210> SEQ ID NO 131
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 131

cggtctcccg cccggcacgg                                                   20


<210> SEQ ID NO 132
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 132

gctcctcagc cacagccggg ccgggt                                            26


<210> SEQ ID NO 133
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 133

cgaccctgga aa                                                           12


<210> SEQ ID NO 134
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 134

ccccgccgcc accc                                                         14


<210> SEQ ID NO 135
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 135

tttccagggt cgccatgg                                                     18


<210> SEQ ID NO 136
<211> LENGTH: 10
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 136

ggcggcgggg                                                              10


<210> SEQ ID NO 137
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 137

acccttgtta gccacctccc                                                   20


<210> SEQ ID NO 138
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
```

```
<400> SEQUENCE: 138

gaacgcagtg ctcttcgaag                                                  20


<210> SEQ ID NO 139
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 139

ctcacgccct gctccgtgta                                                  20


<210> SEQ ID NO 140
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 140

ggcgacaact acttcctggt                                                  20


<210> SEQ ID NO 141
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 141

ctcacgccct gctccgtgta                                                  20


<210> SEQ ID NO 142
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 142

gggcgacaac tacttcctgg                                                  20


<210> SEQ ID NO 143
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 143

cctcttcagg gccggggtgg                                                  20


<210> SEQ ID NO 144
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 144

gaggacccag gtggaactgc                                                  20


<210> SEQ ID NO 145
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 145

tcagctccag gcggtcctgg                                                  20
```

-continued

```
<210> SEQ ID NO 146
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 146

agcagcagca gcagtggcag                                                20


<210> SEQ ID NO 147
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 147

tgggcaccgt cagctccagg                                                20


<210> SEQ ID NO 148
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 148

cagcagtggc agcggccacc                                                20


<210> SEQ ID NO 149
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 149

acctctcccc tggccctcat                                                20


<210> SEQ ID NO 150
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 150

ccaggaccgc ctggagctga                                                20


<210> SEQ ID NO 151
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 151

ccgtcagctc caggcggtcc                                                20


<210> SEQ ID NO 152
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 152

agcagcagca gcagtggcag                                                20


<210> SEQ ID NO 153
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 153
```

-continued

```
atgtgccaag caaagcctca                                                20


<210> SEQ ID NO 154
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 154

ttcggtcatg cccgtggatg                                                20


<210> SEQ ID NO 155
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 155

gtcgttgaaa ttcatcgtac                                                20


<210> SEQ ID NO 156
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 156

accacctgtg aagagtttcc                                                20


<210> SEQ ID NO 157
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 157

cgtcgttgaa attcatcgta                                                20


<210> SEQ ID NO 158
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 158

accacctgtg aagagtttcc                                                20


<210> SEQ ID NO 159
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 159

gaacgcagtg cttttcgagg                                                20


<210> SEQ ID NO 160
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 160

acccttgttg gccacctccc                                                20


<210> SEQ ID NO 161
<211> LENGTH: 20
```

-continued

```
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 161

ggtgacaact actatctggt                                             20


<210> SEQ ID NO 162
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 162

ctcacaccct gctccgtgta                                             20


<210> SEQ ID NO 163
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 163

gggtgacaac tactatctgg                                             20


<210> SEQ ID NO 164
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 164

ctcacaccct gctccgtgta                                             20


<210> SEQ ID NO 165
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 165

cgagaacgca gtgcttttcg                                             20


<210> SEQ ID NO 166
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 166

acccttgttg gccacctccc                                             20


<210> SEQ ID NO 167
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 167

atgagccaag caaatcctca                                             20


<210> SEQ ID NO 168
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 168

ttccgtcatg cccgtggaca                                             20
```

-continued

<210> SEQ ID NO 169
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 169

cttcgttgaa aaccattgta 20

<210> SEQ ID NO 170
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 170

ccacctctga agagtttcct 20

<210> SEQ ID NO 171
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 171

cttcgttgaa aaccattgta 20

<210> SEQ ID NO 172
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 172

accacctctg aagagtttcc 20

<210> SEQ ID NO 173
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 173

cttccactca ctctgcgatt 20

<210> SEQ ID NO 174
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 174

accatgtctc agtgtcaagc 20

<210> SEQ ID NO 175
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 175

ggcggcaaca gcggcaacag 20

<210> SEQ ID NO 176
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

```
<400> SEQUENCE: 176

actgctctgc gtggctgcgg                                              20


<210> SEQ ID NO 177
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 177

ccgcagccac gcagagcagt                                              20


<210> SEQ ID NO 178
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 178

gcacctctcc tcgccccgat                                              20


<210> SEQ ID NO 179
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 179

gagggcctat ttcccatgat tcc                                          23


<210> SEQ ID NO 180
<211> LENGTH: 126
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (84)..(102)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 180

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacnnnnnnn nnnnnnnnnn nnccggtgtt tcgtcctttc    120

cacaag                                                               126


<210> SEQ ID NO 181
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (6)..(24)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 181
```

-continued

caccgnnnnn nnnnnnnnnn nnnn 24

<210> SEQ ID NO 182
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(23)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 182

aaacnnnnnn nnnnnnnnnn nnnc 24

<210> SEQ ID NO 183
<211> LENGTH: 126
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 183

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacccctagt cattggaggt gaccggtgtt tcgtcctttc 120

cacaag 126

<210> SEQ ID NO 184
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 184

caccgtcacc tccaatgact aggg 24

<210> SEQ ID NO 185
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 185

aaacccctag tcattggagg tgac 24

<210> SEQ ID NO 186
<211> LENGTH: 192
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

-continued

<400> SEQUENCE: 186

cagaagaaga agggctccca tcacatcaac cggtggcgca ttgccacgaa gcaggccaat 60

ggggaggaca tcgatgtcac ctccaatgac aagcttgcta gcggtgggca accacaaacc 120

cacgagggca gagtgctgct tgctgctggc caggcccctg cgtgggccca agctggactc 180

tggccactcc ct 192

<210> SEQ ID NO 187
<211> LENGTH: 192
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 187

agggagtggc cagagtccag cttgggccca cgcaggggcc tggccagcag caagcagcac 60

tctgccctcg tgggtttgtg gttgcccacc gctagcaagc ttgtcattgg aggtgacatc 120

gatgtcctcc ccattggcct gcttcgtggc aatgcgccac cggttgatgt gatgggagcc 180

cttcttcttc tg 192

<210> SEQ ID NO 188
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 188

ccatcccctt ctgtgaatgt 20

<210> SEQ ID NO 189
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 189

ggagattgga gacacggaga 20

<210> SEQ ID NO 190
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 190

ggctccctgg gttcaaagta 20

<210> SEQ ID NO 191
<211> LENGTH: 21

-continued

```
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 191

agaggggtct ggatgtcgta a                                              21


<210> SEQ ID NO 192
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 192

cgccagggtt ttcccagtca cgac                                           24


<210> SEQ ID NO 193
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 193

gagggtctcg tccttgcggc cgcgctagcg agggcctatt tcccatgatt c             51


<210> SEQ ID NO 194
<211> LENGTH: 133
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (95)..(114)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 194

ctcggtctcg gtaaaaaagc accgactcgg tgccactttt tcaagttgat aacggactag     60

ccttatttta acttgctatt tctagctcta aaacnnnnnn nnnnnnnnnn nnnnggtgtt    120

tcgtcctttc cac                                                       133


<210> SEQ ID NO 195
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 195

gagggtctct ttaccggtga gggcctattt cccatgattc c                        41
```

-continued

```
<210> SEQ ID NO 196
<211> LENGTH: 133
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (95)..(114)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 196

ctcggtctcc tcaaaaaagc accgactcgg tgccactttt tcaagttgat aacggactag     60

ccttatttta acttgctatt tctagctcta aaacnnnnnn nnnnnnnnnn nnnnggtgtt    120

tcgtcctttc cac                                                       133


<210> SEQ ID NO 197
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 197

gagggtctct ttgagctcga gggcctattt cccatgattc                           40


<210> SEQ ID NO 198
<211> LENGTH: 133
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (96)..(115)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 198

ctcggtctcg cgtaaaaaag caccgactcg gtgccacttt ttcaagttga taacggacta     60

gccttatttt aacttgctat ttctagctct aaaacnnnnn nnnnnnnnnn nnnnnggtgt    120

ttcgtccttt cca                                                       133


<210> SEQ ID NO 199
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 199

gagggtctct tacgcgtgtg tctagac                                         27


<210> SEQ ID NO 200
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
```

-continued

```
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 200

ctcggtctca aggacaggga agggagcagt ggttcacgcc tgtaatccca gcaatttggg      60

aggccaaggt gggtagatca cctgagatta ggagttgc                              98


<210> SEQ ID NO 201
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 201

cctgtccttg cggccgcgct agcgagggcc                                       30


<210> SEQ ID NO 202
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 202

cacgcggccg caaggacagg gaagggagca g                                     31


<210> SEQ ID NO 203
<211> LENGTH: 327
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic polypeptide"

<400> SEQUENCE: 203

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
1               5                   10                  15

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
            20                  25                  30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
        35                  40                  45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
    50                  55                  60

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
65                  70                  75                  80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
                85                  90                  95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
            100                 105                 110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
        115                 120                 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
```

-continued

```
    130                 135                 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145                 150                 155                 160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
                165                 170                 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
            180                 185                 190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
        195                 200                 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
    210                 215                 220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Ser
225                 230                 235                 240

Gly Leu Arg Ser Arg Glu Glu Glu Glu Glu Thr Asp Ser Arg Met Pro
                245                 250                 255

His Leu Asp Ser Pro Gly Ser Ser Gln Pro Arg Arg Ser Phe Leu Ser
            260                 265                 270

Arg Val Ile Arg Ala Ala Leu Pro Leu Gln Leu Leu Leu Leu Leu Leu
        275                 280                 285

Leu Leu Leu Ala Cys Leu Leu Pro Ala Ser Glu Asp Asp Tyr Ser Cys
    290                 295                 300

Thr Gln Ala Asn Asn Phe Ala Arg Ser Phe Tyr Pro Met Leu Arg Tyr
305                 310                 315                 320

Thr Asn Gly Pro Pro Pro Thr
                325


<210> SEQ ID NO 204
<211> LENGTH: 3243
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic polynucleotide"

<400> SEQUENCE: 204

accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc     60

aaggtcgaag cgtccatgaa aggaactac attctggggc tggacatcgg gattacaagc    120

gtggggtatg ggattattga ctatgaaaca agggacgtga tcgacgcagg cgtcagactg    180

ttcaaggagg ccaacgtgga aaacaatgag ggacggagaa gcaagagggg agccaggcgc    240

ctgaaacgac ggagaaggca cagaatccag agggtgaaga aactgctgtt cgattacaac    300

ctgctgaccg accattctga gctgagtgga attaatcctt atgaagccag ggtgaaaggc    360

ctgagtcaga agctgtcaga ggaagagttt tccgcagctc tgctgcacct ggctaagcgc    420

cgaggagtgc ataacgtcaa tgaggtggaa gaggacaccg gcaacgagct gtctacaaag    480

gaacagatct cacgcaatag caaagctctg gaagagaagt atgtcgcaga gctgcagctg    540

gaacggctga agaaagatgg cgaggtgaga gggtcaatta ataggttcaa gacaagcgac    600

tacgtcaaag aagccaagca gctgctgaaa gtgcagaagg cttaccacca gctggatcag    660

agcttcatcg atacttatat cgacctgctg gagactcgga gaacctacta tgagggacca    720

ggagaaggga gccccttcgg atggaaagac atcaaggaat ggtacgagat gctgatggga    780
```

-continued

```
cattgcacct attttccaga agagctgaga agcgtcaagt acgcttataa cgcagatctg    840
tacaacgccc tgaatgacct gaacaacctg gtcatcacca gggatgaaaa cgagaaactg    900
gaatactatg agaagttcca gatcatcgaa aacgtgttta agcagaagaa aaagcctaca    960
ctgaaacaga ttgctaagga gatcctggtc aacgaagagg acatcaaggg ctaccgggtg   1020
acaagcactg gaaaaccaga gttcaccaat ctgaaagtgt atcacgatat taaggacatc   1080
acagcacgga aagaaatcat tgagaacgcc gaactgctgg atcagattgc taagatcctg   1140
actatctacc agagctccga ggacatccag gaagagctga ctaacctgaa cagcgagctg   1200
acccaggaag agatcgaaca gattagtaat ctgaaggggt acaccggaac acacaacctg   1260
tccctgaaag ctatcaatct gattctggat gagctgtggc atacaaacga caatcagatt   1320
gcaatcttta accggctgaa gctggtccca aaaaaggtgg acctgagtca gcagaaagag   1380
atcccaacca cactggtgga cgatttcatt ctgtcacccg tggtcaagcg gagcttcatc   1440
cagagcatca aagtgatcaa cgccatcatc aagaagtacg gcctgcccaa tgatatcatt   1500
atcgagctgg ctagggagaa gaacagcaag gacgcacaga agatgatcaa tgagatgcag   1560
aaacgaaacc ggcagaccaa tgaacgcatt gaagagatta tccgaactac cgggaaagag   1620
aacgcaaagt acctgattga aaaaatcaag ctgcacgata tgcaggaggg aaagtgtctg   1680
tattctctgg aggccatccc cctggaggac ctgctgaaca atccattcaa ctacgaggtc   1740
gatcatatta tccccagaag cgtgtccttc gacaattcct ttaacaacaa ggtgctggtc   1800
aagcaggaag agaactctaa aaagggcaat aggactcctt tccagtacct gtctagttca   1860
gattccaaga tctcttacga aacctttaaa aagcacattc tgaatctggc caaaggaaag   1920
ggccgcatca gcaagaccaa aaaggagtac ctgctggaag agcgggacat caacagattc   1980
tccgtccaga aggattttat taaccggaat ctggtggaca caagatacgc tactcgcggc   2040
ctgatgaatc tgctgcgatc ctatttccgg gtgaacaatc tggatgtgaa agtcaagtcc   2100
atcaacggcg ggttcacatc ttttctgagg cgcaaatgga agtttaaaaa ggagcgcaac   2160
aaagggtaca agcaccatgc cgaagatgct ctgattatcg caaatgccga cttcatcttt   2220
aaggagtgga aaaagctgga caaagccaag aaagtgatgg agaaccagat gttcgaagag   2280
aagcaggccg aatctatgcc cgaaatcgag acagaacagg agtacaagga gattttcatc   2340
actcctcacc agatcaagca tatcaaggat ttcaaggact acaagtactc tcaccgggtg   2400
gataaaaagc ccaacagaga gctgatcaat gacaccctgt atagtacaag aaaagacgat   2460
aaggggaata ccctgattgt gaacaatctg aacggactgt acgacaaaga taatgacaag   2520
ctgaaaaagc tgatcaacaa aagtcccgag aagctgctga tgtaccacca tgatcctcag   2580
acatatcaga aactgaagct gattatggag cagtacggcg acgagaagaa cccactgtat   2640
aagtactatg aagagactgg gaactacctg accaagtata gcaaaaagga taatggcccc   2700
gtgatcaaga agatcaagta ctatgggaac aagctgaatg cccatctgga catcacagac   2760
gattacccta acagtcgcaa caaggtggtc aagctgtcac tgaagccata cagattcgat   2820
gtctatctgg acaacggcgt gtataaattt gtgactgtca agaatctgga tgtcatcaaa   2880
aaggagaact actatgaagt gaatagcaag tgctacgaag aggctaaaaa gctgaaaaag   2940
attagcaacc aggcagagtt catcgcctcc ttttacaaca acgacctgat taagatcaat   3000
ggcgaactgt atagggtcat cggggtgaac aatgatctgc tgaaccgcat tgaagtgaat   3060
```

-continued

atgattgaca tcacttaccg agagtatctg gaaaacatga atgataagcg ccccctcga 3120

attatcaaaa caattgcctc taagactcag agtatcaaaa agtactcaac cgacattctg 3180

ggaaacctgt atgaggtgaa gagcaaaaag caccctcaga ttatcaaaaa gggctaagaa 3240

ttc 3243

<210> SEQ ID NO 205
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note=“Description of Artificial Sequence:
Synthetic polynucleotide”

<400> SEQUENCE: 205

gttttagagc tatgctgttt tgaatggtcc caaaacggaa gggcctgagt ccgagcagaa 60

gaagaagttt tagagctatg ctgttttgaa tggtcccaaa ac 102

<210> SEQ ID NO 206
<211> LENGTH: 100
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 206

cggaggacaa agtacaaacg gcagaagctg gaggaggaag ggcctgagtc cgagcagaag 60

aagaagggct cccatcacat caaccggtgg cgcattgcca 100

<210> SEQ ID NO 207
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 207

agctggagga ggaagggcct gagtccgagc agaagaagaa gggctcccac 50

<210> SEQ ID NO 208
<211> LENGTH: 30
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note=“Description of Artificial Sequence:
Synthetic oligonucleotide”

<400> SEQUENCE: 208

gaguccgagc agaagaagaa guuuuagagc 30

<210> SEQ ID NO 209
<211> LENGTH: 49
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note=“Description of Artificial Sequence:
Synthetic oligonucleotide”

<400> SEQUENCE: 209

agctggagga ggaagggcct gagtccgagc agaagagaag ggctcccac 49

<210> SEQ ID NO 210

-continued

<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 210

ctggaggagg aagggcctga gtccgagcag aagaagaagg gctcccatca cat 53

<210> SEQ ID NO 211
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 211

ctggaggagg aagggcctga gtccgagcag aagagaaggg ctcccatcac at 52

<210> SEQ ID NO 212
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 212

ctggaggagg aagggcctga gtccgagcag aagaaagaag ggctcccatc acat 54

<210> SEQ ID NO 213
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 213

ctggaggagg aagggcctga gtccgagcag aagaagggct cccatcacat 50

<210> SEQ ID NO 214
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 214

ctggaggagg aagggcctga gcccgagcag aagggctccc atcacat 47

<210> SEQ ID NO 215
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 215

ctggaggagg aagggcctga gtccgagcag aagaagaagg gctcccat 48

<210> SEQ ID NO 216
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 216

gaguccgagc agaagaagau 20

<210> SEQ ID NO 217
<211> LENGTH: 20
<212> TYPE: RNA

-continued

```
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 217

gaguccgagc agaagaagua                                             20


<210> SEQ ID NO 218
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 218

gaguccgagc agaagaacaa                                             20


<210> SEQ ID NO 219
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 219

gaguccgagc agaagaugaa                                             20


<210> SEQ ID NO 220
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 220

gaguccgagc agaaguagaa                                             20


<210> SEQ ID NO 221
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 221

gaguccgagc agaugaagaa                                             20


<210> SEQ ID NO 222
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
```

-continued

```
<400> SEQUENCE: 222

gaguccgagc acaagaagaa                                                   20


<210> SEQ ID NO 223
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 223

gaguccgagg agaagaagaa                                                   20


<210> SEQ ID NO 224
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 224

gaguccgugc agaagaagaa                                                   20


<210> SEQ ID NO 225
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 225

gagucggagc agaagaagaa                                                   20


<210> SEQ ID NO 226
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 226

gagaccgagc agaagaagaa                                                   20


<210> SEQ ID NO 227
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 227

aatgacaagc ttgctagcgg tggg                                              24


<210> SEQ ID NO 228
```

-continued

```
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 228

aaaacggaag ggcctgagtc cgagcagaag aagaagttt                         39


<210> SEQ ID NO 229
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 229

aaacaggggc cgagattggg tgttcagggc agaggtttt                         39


<210> SEQ ID NO 230
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 230

aaaacggaag ggcctgagtc cgagcagaag aagaagtt                          38


<210> SEQ ID NO 231
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 231

aacggaggga ggggcacaga tgagaaactc agggttttag                        40


<210> SEQ ID NO 232
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 232

agcccttctt cttctgctcg gactcaggcc cttcctcc                          38


<210> SEQ ID NO 233
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 233

cagggaggga ggggcacaga tgagaaactc aggaggcccc                        40


<210> SEQ ID NO 234
```

```
<211> LENGTH: 80
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 234

ggcaatgcgc caccggttga tgtgatggga gcccttctag gaggccccca gagcagccac     60

tggggcctca acactcaggc                                                 80


<210> SEQ ID NO 235
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 235

catcgatgtc ctccccattg gcctgcttcg tgg                                  33


<210> SEQ ID NO 236
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 236

ttcgtggcaa tgcgccaccg gttgatgtga tgg                                  33


<210> SEQ ID NO 237
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 237

tcgtggcaat gcgccaccgg ttgatgtgat ggg                                  33


<210> SEQ ID NO 238
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 238

tccagcttct gccgtttgta ctttgtcctc cgg                                  33


<210> SEQ ID NO 239
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 239

ggagggaggg gcacagatga gaaactcagg agg                                  33


<210> SEQ ID NO 240
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 240

aggggccgag attgggtgtt cagggcagag agg                                  33


<210> SEQ ID NO 241
```

-continued

```
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 241

caagcactga gtgccattag ctaaatgcat agg                                  33


<210> SEQ ID NO 242
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 242

aatgcatagg gtaccaccca caggtgccag ggg                                  33


<210> SEQ ID NO 243
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 243

acacacatgg gaaagcctct gggccaggaa agg                                  33


<210> SEQ ID NO 244
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 244

ggaggaggta gtatacagaa acacagagaa gtagaat                              37


<210> SEQ ID NO 245
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 245

agaatgtaga ggagtcacag aaactcagca ctagaaa                              37


<210> SEQ ID NO 246
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 246

ggacgaaaca ccggaaccat tcaaaacagc atagcaagtt aaaataaggc tagtccgtta     60

tcaacttgaa aaagtggcac cgagtcggtg ctttttt                              98


<210> SEQ ID NO 247
<211> LENGTH: 186
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic polynucleotide"

<400> SEQUENCE: 247
```

-continued

ggacgaaaca ccggtagtat taagtattgt tttatggctg ataaatttct ttgaatttct 60

ccttgattat ttgttataaa agttataaaa taatcttgtt ggaaccattc aaaacagcat 120

agcaagttaa aataaggcta gtccgttatc aacttgaaaa agtggcaccg agtcggtgct 180

tttttt 186

<210> SEQ ID NO 248
<211> LENGTH: 95
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 248

gggttttaga gctatgctgt tttgaatggt cccaaaacgg gtcttcgaga agacgtttta 60

gagctatgct gttttgaatg gtcccaaaac ttttt 95

<210> SEQ ID NO 249
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(34)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 249

aaacnnnnnn nnnnnnnnnn nnnnnnnnnn nnnngt 36

<210> SEQ ID NO 250
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (7)..(36)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 250

taaaacnnnn nnnnnnnnnn nnnnnnnnnn nnnnnn 36

<210> SEQ ID NO 251
<211> LENGTH: 84
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

<400> SEQUENCE: 251

gtggaaagga cgaaacaccg ggtcttcgag aagacctgtt ttagagctag aaatagcaag 60

ttaaaataag gctagtccgt tttt 84

-continued

```
<210> SEQ ID NO 252
<211> LENGTH: 46
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(19)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 252

nnnnnnnnnn nnnnnnnnng uuauuguacu cucaagauuu auuuuu                    46


<210> SEQ ID NO 253
<211> LENGTH: 91
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 253

guuacuuaaa ucuugcagaa gcuacaaaga uaaggcuuca ugccgaaauc aacacccugu     60

cauuuuaugg cagggguguuu ucguuauuua a                                   91


<210> SEQ ID NO 254
<211> LENGTH: 70
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 254

ttttctagtg ctgagtttct gtgactcctc tacattctac ttctctgtgt ttctgtatac     60

tacctcctcc                                                            70


<210> SEQ ID NO 255
<211> LENGTH: 122
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 255

ggaggaaggg cctgagtccg agcagaagaa gaagggctcc catcacatca accggtggcg     60

cattgccacg aagcaggcca atggggagga catcgatgtc acctccaatg actagggtgg    120

gc                                                                   122


<210> SEQ ID NO 256
<211> LENGTH: 48
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(32)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 256
```

-continued

```
acnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnguuuuaga gcuaugcu                48


<210> SEQ ID NO 257
<211> LENGTH: 67
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Combined DNA/RNA
      Molecule: Synthetic oligonucleotide"

<400> SEQUENCE: 257

agcauagcaa guuaaaauaa ggctaguccg uuaucaacuu gaaaaagugg caccgagucg     60

gugcuuu                                                               67


<210> SEQ ID NO 258
<211> LENGTH: 62
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 258

nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc     60

cg                                                                    62


<210> SEQ ID NO 259
<211> LENGTH: 73
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 259

tgaatggtcc caaaacggaa gggcctgagt ccgagcagaa gaagaagttt tagagctatg     60

ctgttttgaa tgg                                                        73


<210> SEQ ID NO 260
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 260

ctggtcttcc acctctctgc cctgaacacc caatctcggc ccctctcgcc accctcctgc     60

atttctgtt                                                             69


<210> SEQ ID NO 261
<211> LENGTH: 138
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
```

```
<400> SEQUENCE: 261

acccaagcac tgagtgccat tagctaaatg catagggtac cacccacagg tgccaggggc      60

ctttcccaaa gttcccagcc ccttctccaa cctttcctgg cccagaggct ttcccatgtg     120

tgtggctgga ccctttga                                                   138


<210> SEQ ID NO 262
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 262

aaaaccaccc ttctctctgg c                                                21


<210> SEQ ID NO 263
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 263

ggagattgga gacacggaga g                                                21


<210> SEQ ID NO 264
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 264

ctggaaagcc aatgcctgac                                                  20


<210> SEQ ID NO 265
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 265

ggcagcaaac tccttgtcct                                                  20


<210> SEQ ID NO 266
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 266
```

-continued

```
gtgctttgca gaggcctacc                                                20


<210> SEQ ID NO 267
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 267

cctggagcgc atgcagtagt                                                20


<210> SEQ ID NO 268
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 268

accttctgtg tttccaccat tc                                             22


<210> SEQ ID NO 269
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 269

ttggggagtg cacagacttc                                                20


<210> SEQ ID NO 270
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic probe"

<400> SEQUENCE: 270

tagctctaaa acttcttctt ctgctcggac                                     30


<210> SEQ ID NO 271
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic probe"

<400> SEQUENCE: 271

ctagccttat tttaacttgc tatgctgttt                                     30


<210> SEQ ID NO 272
<211> LENGTH: 99
<212> TYPE: RNA
```

-continued

```
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: a, c, u, g, unknown or other

<400> SEQUENCE: 272

nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc     60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuu                            99


<210> SEQ ID NO 273
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 273

tagcgggtaa gc                                                         12


<210> SEQ ID NO 274
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 274

tcggtgacat gt                                                         12


<210> SEQ ID NO 275
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 275

actccccgta gg                                                         12


<210> SEQ ID NO 276
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 276

actgcgtgtt aa                                                         12


<210> SEQ ID NO 277
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 277

acgtcgcctg at                                                         12


<210> SEQ ID NO 278
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 278

taggtcgacc ag                                                         12
```

-continued

<210> SEQ ID NO 279
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 279

ggcgttaatg at 12

<210> SEQ ID NO 280
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 280

tgtcgcatgt ta 12

<210> SEQ ID NO 281
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 281

atggaaacgc at 12

<210> SEQ ID NO 282
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 282

gccgaattcc tc 12

<210> SEQ ID NO 283
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 283

gcatggtacg ga 12

<210> SEQ ID NO 284
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 284

cggtactctt ac 12

<210> SEQ ID NO 285
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 285

gcctgtgccg ta 12

<210> SEQ ID NO 286
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 286

tacggtaagt cg 12

<210> SEQ ID NO 287
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 287

cacgaaatta cc 12

<210> SEQ ID NO 288
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 288

aaccaagata cg 12

<210> SEQ ID NO 289
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 289

gagtcgatac gc 12

<210> SEQ ID NO 290
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 290

gtctcacgat cg 12

<210> SEQ ID NO 291
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 291

tcgtcgggtg ca 12

<210> SEQ ID NO 292
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 292

actccgtagt ga 12

<210> SEQ ID NO 293
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 293

caggacgtcc gt 12

<210> SEQ ID NO 294

-continued

<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 294

tcgtatccct ac   12

<210> SEQ ID NO 295
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 295

tttcaaggcc gg   12

<210> SEQ ID NO 296
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 296

cgccggtgga at   12

<210> SEQ ID NO 297
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 297

gaacccgtcc ta   12

<210> SEQ ID NO 298
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 298

gattcatcag cg   12

<210> SEQ ID NO 299
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 299

acaccggtct tc   12

<210> SEQ ID NO 300
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 300

atcgtgccct aa   12

<210> SEQ ID NO 301
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 301

-continued

```
gcgtcaatgt tc                                                          12


<210> SEQ ID NO 302
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 302

ctccgtatct cg                                                          12


<210> SEQ ID NO 303
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 303

ccgattcctt cg                                                          12


<210> SEQ ID NO 304
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 304

tgcgcctcca gt                                                          12


<210> SEQ ID NO 305
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 305

taacgtcgga gc                                                          12


<210> SEQ ID NO 306
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 306

aaggtcgccc at                                                          12


<210> SEQ ID NO 307
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 307

gtcggggact at                                                          12


<210> SEQ ID NO 308
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 308

ttcgagcgat tt                                                          12


<210> SEQ ID NO 309
<211> LENGTH: 12
<212> TYPE: DNA
```

-continued

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 309

tgagtcgtcg ag 12

<210> SEQ ID NO 310
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 310

tttacgcaga gg 12

<210> SEQ ID NO 311
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 311

aggaagtatc gc 12

<210> SEQ ID NO 312
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 312

actcgatacc at 12

<210> SEQ ID NO 313
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 313

cgctacatag ca 12

<210> SEQ ID NO 314
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 314

ttcataaccg gc 12

<210> SEQ ID NO 315
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 315

ccaaacggtt aa 12

<210> SEQ ID NO 316
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 316

cgattccttc gt 12

-continued

```
<210> SEQ ID NO 317
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 317

cgtcatgaat aa                                                          12


<210> SEQ ID NO 318
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 318

agtggcgatg ac                                                          12


<210> SEQ ID NO 319
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 319

cccctacggc ac                                                          12


<210> SEQ ID NO 320
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 320

gccaacccgc ac                                                          12


<210> SEQ ID NO 321
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 321

tgggacaccg gt                                                          12


<210> SEQ ID NO 322
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 322

ttgactgcgg cg                                                          12


<210> SEQ ID NO 323
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 323

actatgcgta gg                                                          12


<210> SEQ ID NO 324
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
```

-continued

```
<400> SEQUENCE: 324

tcacccaaag cg                                                          12


<210> SEQ ID NO 325
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 325

gcaggacgtc cg                                                          12


<210> SEQ ID NO 326
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 326

acaccgaaaa cg                                                          12


<210> SEQ ID NO 327
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 327

cggtgtattg ag                                                          12


<210> SEQ ID NO 328
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 328

cacgaggtat gc                                                          12


<210> SEQ ID NO 329
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 329

taaagcgacc cg                                                          12


<210> SEQ ID NO 330
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 330

cttagtcggc ca                                                          12


<210> SEQ ID NO 331
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 331

cgaaaacgtg gc                                                          12


<210> SEQ ID NO 332
```

-continued

```
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 332

cgtgccctga ac                                                          12


<210> SEQ ID NO 333
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 333

tttaccatcg aa                                                          12


<210> SEQ ID NO 334
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 334

cgtagccatg tt                                                          12


<210> SEQ ID NO 335
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 335

cccaaacggt ta                                                          12


<210> SEQ ID NO 336
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 336

gcgttatcag aa                                                          12


<210> SEQ ID NO 337
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 337

tcgatggtaa ac                                                          12


<210> SEQ ID NO 338
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 338

cgactttttg ca                                                          12


<210> SEQ ID NO 339
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 339
```

-continued

```
tcgacgactc ac                                                          12


<210> SEQ ID NO 340
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 340

acgcgtcaga ta                                                          12


<210> SEQ ID NO 341
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 341

cgtacggcac ag                                                          12


<210> SEQ ID NO 342
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 342

ctatgccgtg ca                                                          12


<210> SEQ ID NO 343
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 343

cgcgtcagat at                                                          12


<210> SEQ ID NO 344
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 344

aagatcggta gc                                                          12


<210> SEQ ID NO 345
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 345

cttcgcaagg ag                                                          12


<210> SEQ ID NO 346
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 346

gtcgtggact ac                                                          12


<210> SEQ ID NO 347
<211> LENGTH: 12
<212> TYPE: DNA
```

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 347

ggtcgtcatc aa 12

<210> SEQ ID NO 348
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 348

gttaacagcg tg 12

<210> SEQ ID NO 349
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 349

tagctaaccg tt 12

<210> SEQ ID NO 350
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 350

agtaaaggcg ct 12

<210> SEQ ID NO 351
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 351

ggtaatttcg tg 12

<210> SEQ ID NO 352
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 352

cagaagaaga agggc 15

<210> SEQ ID NO 353
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 353

ccaatgggga ggacatcgat gtcacctcca atgactaggg tggtgggcaa c 51

<210> SEQ ID NO 354
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 354

ctctggccac tccct 15

-continued

```
<210> SEQ ID NO 355
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 355

acatcgatgt cacctccaat gacaagcttg ctagcggtgg gcaaccacaa ac            52


<210> SEQ ID NO 356
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (6)..(25)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 356

caccgnnnnn nnnnnnnnnn nnnnn                                          25


<210> SEQ ID NO 357
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(24)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 357

aaacnnnnnn nnnnnnnnnn nnnnc                                          25


<210> SEQ ID NO 358
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 358

aacaccgggt cttcgagaag acctgtttta gagctagaaa tagcaagtta aaat          54


<210> SEQ ID NO 359
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"

<400> SEQUENCE: 359

caaaacgggt cttcgagaag acgttttaga gctatgctgt tttgaatggt ccca          54


<210> SEQ ID NO 360
```

-continued

```
<211> LENGTH: 4104
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(4104)

<400> SEQUENCE: 360

atg gac aag aag tac agc atc ggc ctg gac atc ggc acc aac tct gtg       48
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1               5                   10                  15

ggc tgg gcc gtg atc acc gac gag tac aag gtg ccc agc aag aaa ttc       96
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

aag gtg ctg ggc aac acc gac cgg cac agc atc aag aag aac ctg atc      144
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

gga gcc ctg ctg ttc gac agc ggc gaa aca gcc gag gcc acc cgg ctg      192
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

aag aga acc gcc aga aga aga tac acc aga cgg aag aac cgg atc tgc      240
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

tat ctg caa gag atc ttc agc aac gag atg gcc aag gtg gac gac agc      288
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

ttc ttc cac aga ctg gaa gag tcc ttc ctg gtg gaa gag gat aag aag      336
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

cac gag cgg cac ccc atc ttc ggc aac atc gtg gac gag gtg gcc tac      384
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

cac gag aag tac ccc acc atc tac cac ctg aga aag aaa ctg gtg gac      432
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

agc acc gac aag gcc gac ctg cgg ctg atc tat ctg gcc ctg gcc cac      480
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

atg atc aag ttc cgg ggc cac ttc ctg atc gag ggc gac ctg aac ccc      528
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

gac aac agc gac gtg gac aag ctg ttc atc cag ctg gtg cag acc tac      576
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

aac cag ctg ttc gag gaa aac ccc atc aac gcc agc ggc gtg gac gcc      624
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

aag gcc atc ctg tct gcc aga ctg agc aag agc aga cgg ctg gaa aat      672
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

ctg atc gcc cag ctg ccc ggc gag aag aag aat ggc ctg ttc ggc aac      720
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

ctg att gcc ctg agc ctg ggc ctg acc ccc aac ttc aag agc aac ttc      768
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

gac ctg gcc gag gat gcc aaa ctg cag ctg agc aag gac acc tac gac      816
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270
```

-continued

```
gac gac ctg gac aac ctg ctg gcc cag atc ggc gac cag tac gcc gac       864
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

ctg ttt ctg gcc gcc aag aac ctg tcc gac gcc atc ctg ctg agc gac       912
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

atc ctg aga gtg aac acc gag atc acc aag gcc ccc ctg agc gcc tct       960
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

atg atc aag aga tac gac gag cac cac cag gac ctg acc ctg ctg aaa      1008
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

gct ctc gtg cgg cag cag ctg cct gag aag tac aaa gag att ttc ttc      1056
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

gac cag agc aag aac ggc tac gcc ggc tac att gac ggc gga gcc agc      1104
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

cag gaa gag ttc tac aag ttc atc aag ccc atc ctg gaa aag atg gac      1152
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

ggc acc gag gaa ctg ctc gtg aag ctg aac aga gag gac ctg ctg cgg      1200
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

aag cag cgg acc ttc gac aac ggc agc atc ccc cac cag atc cac ctg      1248
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

gga gag ctg cac gcc att ctg cgg cgg cag gaa gat ttt tac cca ttc      1296
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

ctg aag gac aac cgg gaa aag atc gag aag atc ctg acc ttc cgc atc      1344
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

ccc tac tac gtg ggc cct ctg gcc agg gga aac agc aga ttc gcc tgg      1392
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

atg acc aga aag agc gag gaa acc atc acc ccc tgg aac ttc gag gaa      1440
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

gtg gtg gac aag ggc gct tcc gcc cag agc ttc atc gag cgg atg acc      1488
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

aac ttc gat aag aac ctg ccc aac gag aag gtg ctg ccc aag cac agc      1536
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

ctg ctg tac gag tac ttc acc gtg tat aac gag ctg acc aaa gtg aaa      1584
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

tac gtg acc gag gga atg aga aag ccc gcc ttc ctg agc ggc gag cag      1632
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

aaa aag gcc atc gtg gac ctg ctg ttc aag acc aac cgg aaa gtg acc      1680
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

gtg aag cag ctg aaa gag gac tac ttc aag aaa atc gag tgc ttc gac      1728
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575
```

-continued

```
tcc gtg gaa atc tcc ggc gtg gaa gat cgg ttc aac gcc tcc ctg ggc     1776
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

aca tac cac gat ctg ctg aaa att atc aag gac aag gac ttc ctg gac     1824
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

aat gag gaa aac gag gac att ctg gaa gat atc gtg ctg acc ctg aca     1872
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

ctg ttt gag gac aga gag atg atc gag gaa cgg ctg aaa acc tat gcc     1920
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

cac ctg ttc gac gac aaa gtg atg aag cag ctg aag cgg cgg aga tac     1968
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

acc ggc tgg ggc agg ctg agc cgg aag ctg atc aac ggc atc cgg gac     2016
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

aag cag tcc ggc aag aca atc ctg gat ttc ctg aag tcc gac ggc ttc     2064
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

gcc aac aga aac ttc atg cag ctg atc cac gac gac agc ctg acc ttt     2112
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

aaa gag gac atc cag aaa gcc cag gtg tcc ggc cag ggc gat agc ctg     2160
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

cac gag cac att gcc aat ctg gcc ggc agc ccc gcc att aag aag ggc     2208
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

atc ctg cag aca gtg aag gtg gtg gac gag ctc gtg aaa gtg atg ggc     2256
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

cgg cac aag ccc gag aac atc gtg atc gcc atg gcc aga gag aac cag     2304
Arg His Lys Pro Glu Asn Ile Val Ile Ala Met Ala Arg Glu Asn Gln
        755                 760                 765

acc acc cag aag gga cag aag aac agc cgc gag aga atg aag cgg atc     2352
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

gaa gag ggc atc aaa gag ctg ggc agc cag atc ctg aaa gaa cac ccc     2400
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

gtg gaa aac acc cag ctg cag aac gag aag ctg tac ctg tac tac ctg     2448
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

cag aat ggg cgg gat atg tac gtg gac cag gaa ctg gac atc aac cgg     2496
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

ctg tcc gac tac gat gtg gac gcc atc gtg cct cag agc ttt ctg aag     2544
Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

gac gac tcc atc gac gcc aag gtg ctg acc aga agc gac aag gcc cgg     2592
Asp Asp Ser Ile Asp Ala Lys Val Leu Thr Arg Ser Asp Lys Ala Arg
    850                 855                 860

ggc aag agc gac aac gtg ccc tcc gaa gag gtc gtg aag aag atg aag     2640
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880
```

-continued

```
aac tac tgg cgg cag ctg ctg aac gcc aag ctg att acc cag aga aag     2688
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

ttc gac aat ctg acc aag gcc gag aga ggc ggc ctg agc gaa ctg gat     2736
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

aag gcc ggc ttc atc aag aga cag ctg gtg gaa acc cgg cag atc aca     2784
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

aag cac gtg gca cag atc ctg gac tcc cgg atg aac act aag tac gac     2832
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

gag aat gac aag ctg atc cgg gaa gtg aaa gtg atc acc ctg aag tcc     2880
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

aag ctg gtg tcc gat ttc cgg aag gat ttc cag ttt tac aaa gtg cgc     2928
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

gag atc aac aac tac cac cac gcc cac gcc gcc tac ctg aac gcc gtc     2976
Glu Ile Asn Asn Tyr His His Ala His Ala Ala Tyr Leu Asn Ala Val
            980                 985                 990

gtg gga acc gcc ctg atc aaa aag  tac cct aag ctg gaa  agc gag ttc   3024
Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                 1005

gtg tac  ggc gac tac aag gtg  tac gac gtg cgg aag  atg atc gcc      3069
Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                 1015                 1020

aag agc  gag cag gaa atc ggc  aag gct acc gcc aag  tac ttc ttc      3114
Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                 1030                 1035

tac agc  aac atc atg aac ttt  ttc aag acc gag att  acc ctg gcc      3159
Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                 1045                 1050

aac ggc  gag atc cgg aag cgg  cct ctg atc gag aca  aac ggc gaa      3204
Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                 1060                 1065

acc ggg  gag atc gtg tgg gat  aag ggc cgg gat ttt  gcc acc gtg      3249
Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                 1075                 1080

cgg aaa  gtg ctg agc atg ccc  caa gtg aat atc gtg  aaa aag acc      3294
Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                 1090                 1095

gag gtg  cag aca ggc ggc ttc  agc aaa gag tct atc  ctg ccc aag      3339
Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                 1105                 1110

agg aac  agc gat aag ctg atc  gcc aga aag aag gac  tgg gac cct      3384
Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                 1120                 1125

aag aag  tac ggc ggc ttc gac  agc ccc acc gtg gcc  tat tct gtg      3429
Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                 1135                 1140

ctg gtg  gtg gcc aaa gtg gaa  aag ggc aag tcc aag  aaa ctg aag      3474
Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                 1150                 1155

agt gtg  aaa gag ctg ctg ggg  atc acc atc atg gaa  aga agc agc      3519
Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                 1165                 1170
```

-continued

```
ttc gag  aag aat ccc atc gac  ttt ctg gaa gcc aag  ggc tac aaa      3564
Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu Ala Lys  Gly Tyr Lys
    1175                 1180                 1185

gaa gtg  aaa aag gac ctg atc  atc aag ctg cct aag  tac tcc ctg      3609
Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                 1195                 1200

ttc gag  ctg gaa aac ggc cgg  aag aga atg ctg gcc  tct gcc ggc      3654
Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                 1210                 1215

gaa ctg  cag aag gga aac gaa  ctg gcc ctg ccc tcc  aaa tat gtg      3699
Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                 1225                 1230

aac ttc  ctg tac ctg gcc agc  cac tat gag aag ctg  aag ggc tcc      3744
Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                 1240                 1245

ccc gag  gat aat gag cag aaa  cag ctg ttt gtg gaa  cag cac aag      3789
Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                 1255                 1260

cac tac  ctg gac gag atc atc  gag cag atc agc gag  ttc tcc aag      3834
His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                 1270                 1275

aga gtg  atc ctg gcc gac gct  aat ctg gac aaa gtg  ctg tcc gcc      3879
Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                 1285                 1290

tac aac  aag cac cgg gat aag  ccc atc aga gag cag  gcc gag aat      3924
Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                 1300                 1305

atc atc  cac ctg ttt acc ctg  acc aat ctg gga gcc  cct gcc gcc      3969
Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                 1315                 1320

ttc aag  tac ttt gac acc acc  atc gac cgg aag agg  tac acc agc      4014
Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                 1330                 1335

acc aaa  gag gtg ctg gac gcc  acc ctg atc cac cag  agc atc acc      4059
Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                 1345                 1350

ggc ctg  tac gag aca cgg atc  gac ctg tct cag ctg  gga ggc gac      4104
Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                 1360                 1365
```

```
<210> SEQ ID NO 361
<211> LENGTH: 1368
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 361

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
```

-continued

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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
```

-continued

```
                485                 490                 495

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

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

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

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

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

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

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 Ala 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 Ala Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Ala Lys Val Leu Thr Arg Ser Asp Lys Ala 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
```

-continued

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

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

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

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

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

Glu Ile Asn Asn Tyr His His Ala His Ala 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

-continued

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

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

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

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 362
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(57)

<400> SEQUENCE: 362

ggc acc att aaa gaa aat atc att ggt gtt tcc tat gat gaa tat aga       48
Gly Thr Ile Lys Glu Asn Ile Ile Gly Val Ser Tyr Asp Glu Tyr Arg
1               5                   10                  15

tac aga agc                                                           57
Tyr Arg Ser


<210> SEQ ID NO 363
<211> LENGTH: 19
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 363

Gly Thr Ile Lys Glu Asn Ile Ile Gly Val Ser Tyr Asp Glu Tyr Arg
1               5                   10                  15

Tyr Arg Ser


<210> SEQ ID NO 364
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(48)

<400> SEQUENCE: 364

att aaa gaa aat atc att ggc ttt gtt tcc tat gat gaa tat aga tac       48
Ile Lys Glu Asn Ile Ile Gly Phe Val Ser Tyr Asp Glu Tyr Arg Tyr
1               5                   10                  15


<210> SEQ ID NO 365
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
```

-continued

```
      Synthetic peptide"

<400> SEQUENCE: 365

Ile Lys Glu Asn Ile Ile Gly Phe Val Ser Tyr Asp Glu Tyr Arg Tyr
1               5                   10                  15


<210> SEQ ID NO 366
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(48)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 366

ccnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnngg                50


<210> SEQ ID NO 367
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(44)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 367

ccnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnngg                    46


<210> SEQ ID NO 368
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(40)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 368

ccnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn gg                        42


<210> SEQ ID NO 369
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(36)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 369
```

-continued

```
ccnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnngg                              38


<210> SEQ ID NO 370
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(32)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 370

ccnnnnnnnn nnnnnnnnnn nnnnnnnnnn nngg                                  34


<210> SEQ ID NO 371
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(28)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 371

ccnnnnnnnn nnnnnnnnnn nnnnnnnngg                                        30


<210> SEQ ID NO 372
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(24)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 372

ccnnnnnnnn nnnnnnnnnn nnnngg                                            26


<210> SEQ ID NO 373
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 373

ccnnnnnnnn nnnnnnnnnn gg                                                22
```

-continued

```
<210> SEQ ID NO 374
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 374

ccnnnnnnnn nnnnnngg                                                   18


<210> SEQ ID NO 375
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(14)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 375

ccnnnnnnnn nnnngg                                                     16


<210> SEQ ID NO 376
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(13)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 376

ccnnnnnnnn nnngg                                                      15


<210> SEQ ID NO 377
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 377

ccnnnnnnnn nngg                                                       14


<210> SEQ ID NO 378
<211> LENGTH: 13
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
```

-continued

```
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(11)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 378

ccnnnnnnnn ngg                                                    13


<210> SEQ ID NO 379
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(10)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 379

ccnnnnnnnn gg                                                     12


<210> SEQ ID NO 380
<211> LENGTH: 11
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(9)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 380

ccnnnnnnng g                                                      11


<210> SEQ ID NO 381
<211> LENGTH: 10
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 381

ccnnnnnngg                                                        10


<210> SEQ ID NO 382
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic oligonucleotide"
```

-continued

```
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(10)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 382

ggnnnnnnnn cc                                                          12


<210> SEQ ID NO 383
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 383

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa      60

cttgctattt ctagctctaa aactcacatc aaccggtggc gcaggtgttt cgtcctttcc     120

acaag                                                                 125


<210> SEQ ID NO 384
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 384

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa      60

cttgctattt ctagctctaa aactcacatc aaccggtggc gcaggtgttt cgtcctttcc     120

acaag                                                                 125


<210> SEQ ID NO 385
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 385

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa      60

cttgctattt ctagctctaa aacgaggaca aagtacaaac ggcggtgttt cgtcctttcc     120

acaag                                                                 125


<210> SEQ ID NO 386
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 386

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa      60
```

-continued

cttgctattt ctagctctaa aacgaggaca aagtacaaac ggcggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 387
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 387

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgtggcgc attgccacga agcggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 388
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 388

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aaccgagggc agagtgctgc ttgggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 389
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 389

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgagtccg agcagaagaa gaaggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 390
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 390

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgaggaca aagtacaaac ggcggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 391
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 391

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa   60

cttgctattt ctagctctaa aacagcagaa gaagaagggc tccggtgttt cgtcctttcc   120

acaag   125

<210> SEQ ID NO 392
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 392

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa   60

cttgctattt ctagctctaa aactcacatc aaccggtggc gcaggtgttt cgtcctttcc   120

acaag   125

<210> SEQ ID NO 393
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 393

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa   60

cttgctattt ctagctctaa aacccctggc ccaggtgaag gtgggtgttt cgtcctttcc   120

acaag   125

<210> SEQ ID NO 394
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 394

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa   60

cttgctattt ctagctctaa aactccctcc ctggcccagg tgaggtgttt cgtcctttcc   120

acaag   125

<210> SEQ ID NO 395
<211> LENGTH: 125

-continued

```
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 395

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacgaaccgg aggacaaagt acaggtgttt cgtcctttcc    120

acaag                                                                125


<210> SEQ ID NO 396
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 396

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacaggtgaa ggtgtggttc cagggtgttt cgtcctttcc    120

acaag                                                                125


<210> SEQ ID NO 397
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 397

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacggtgaag gtgtggttcc agaggtgttt cgtcctttcc    120

acaag                                                                125


<210> SEQ ID NO 398
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 398

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacgaaccgg aggacaaagt acaggtgttt cgtcctttcc    120

acaag                                                                125


<210> SEQ ID NO 399
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
```

```
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 399

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacccctggc ccaggtgaag gtgggtgttt cgtcctttcc    120

acaag                                                                125


<210> SEQ ID NO 400
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 400

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacaggtgaa ggtgtggttc cagggtgttt cgtcctttcc    120

acaag                                                                125


<210> SEQ ID NO 401
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 401

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacgaggaca aagtacaaac ggcggtgttt cgtcctttcc    120

acaag                                                                125


<210> SEQ ID NO 402
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 402

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacgggaggg aggggcacag atgggtgttt cgtcctttcc    120

acaag                                                                125


<210> SEQ ID NO 403
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 403
```

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aaccaccttc acctgggcca gggggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 404
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 404

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacaccctag tcattggagg tgaggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 405
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 405

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aaccagagca gccactgggg cctggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 406
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 406

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aaccaccttc acctgggcca gggggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 407
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 407

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacccccatt ggcctgcttc gtgggtgttt cgtcctttcc 120

-continued

acaag 125

<210> SEQ ID NO 408
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 408

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacattggcc tgcttcgtgg caaggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 409
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 409

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aactcctcct ccagcttctg ccgggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 410
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 410

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aaccctccag cttctgccgt ttgggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 411
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 411

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacattggcc tgcttcgtgg caaggtgttt cgtcctttcc 120

acaag 125

-continued

```
<210> SEQ ID NO 412
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 412

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa      60

cttgctattt ctagctctaa aacgcagcaa gcagcactct gccggtgttt cgtcctttcc     120

acaag                                                                 125


<210> SEQ ID NO 413
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 413

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa      60

cttgctattt ctagctctaa aacttcttct tctgctcgga ctcggtgttt cgtcctttcc     120

acaag                                                                 125


<210> SEQ ID NO 414
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 414

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa      60

cttgctattt ctagctctaa aacaccggag gacaaagtac aaaggtgttt cgtcctttcc     120

acaag                                                                 125


<210> SEQ ID NO 415
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 415

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa      60

cttgctattt ctagctctaa aactcttctt ctgctcggac tcaggtgttt cgtcctttcc     120

acaag                                                                 125


<210> SEQ ID NO 416
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
```

-continued

```
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 416

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacgttgatg tgatgggagc cctggtgttt cgtcctttcc    120

acaag                                                                125


<210> SEQ ID NO 417
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 417

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacgggccag ggagggaggg gcaggtgttt cgtcctttcc    120

acaag                                                                125


<210> SEQ ID NO 418
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 418

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacgggaggg aggggcacag atgggtgttt cgtcctttcc    120

acaag                                                                125


<210> SEQ ID NO 419
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 419

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa     60

cttgctattt ctagctctaa aacccggttc tggaaccaca cctggtgttt cgtcctttcc    120

acaag                                                                125


<210> SEQ ID NO 420
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"
```

-continued

<400> SEQUENCE: 420

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aactcacctg ggccagggag ggaggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 421
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 421

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aactcacctg ggccagggag ggaggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 422
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 422

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgttctgg aaccacacct tcaggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 423
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 423

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgggaggg aggggcacag atgggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 424
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 424

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

-continued

cttgctattt ctagctctaa aacgggccag ggagggaggg gcaggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 425
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 425

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgttctgg aaccacacct tcaggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 426
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 426

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacaggtgaa ggtgtggttc cagggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 427
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 427

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgaaccgg aggacaaagt acaggtgttt cgtcctttcc 120

acaag 125

<210> SEQ ID NO 428
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 428

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aaccaaaccc acgagggcag agtggtgttt cgtcctttcc 120

acaag 125

-continued

```
<210> SEQ ID NO 429
<211> LENGTH: 125
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 429

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa      60

cttgctattt ctagctctaa aacgagtttc tcatctgtgc cccggtgttt cgtcctttcc     120

acaag                                                                 125
```

```
<210> SEQ ID NO 430
<211> LENGTH: 684
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(123)
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (127)..(159)
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (163)..(399)
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (403)..(684)

<400> SEQUENCE: 430

aaa acc acc ctt ctc tct ggc cca ctg tgt cct ctt cct gcc ctg cca        48
Lys Thr Thr Leu Leu Ser Gly Pro Leu Cys Pro Leu Pro Ala Leu Pro
1               5                   10                  15

tcc cct tct gtg aat gtt aga ccc atg gga gca gct ggt cag agg gga        96
Ser Pro Ser Val Asn Val Arg Pro Met Gly Ala Ala Gly Gln Arg Gly
            20                  25                  30

ccc cgg cct ggg gcc cct aac cct atg tag cct cag tct tcc cat cag       144
Pro Arg Pro Gly Ala Pro Asn Pro Met     Pro Gln Ser Ser His Gln
        35                  40                      45

gct ctc agc tca gcc tga gtg ttg agg ccc cag tgg ctg ctc tgg ggg       192
Ala Leu Ser Ser Ala     Val Leu Arg Pro Gln Trp Leu Leu Trp Gly
        50                      55                  60

cct cct gag ttt ctc atc tgt gcc cct ccc tcc ctg gcc cag gtg aag       240
Pro Pro Glu Phe Leu Ile Cys Ala Pro Pro Ser Leu Ala Gln Val Lys
        65                  70                  75

gtg tgg ttc cag aac cgg agg aca aag tac aaa cgg cag aag ctg gag       288
Val Trp Phe Gln Asn Arg Arg Thr Lys Tyr Lys Arg Gln Lys Leu Glu
    80                  85                  90

gag gaa ggg cct gag tcc gag cag aag aag aag ggc tcc cat cac atc       336
Glu Glu Gly Pro Glu Ser Glu Gln Lys Lys Lys Gly Ser His His Ile
95                  100                 105                 110

aac cgg tgg cgc att gcc acg aag cag gcc aat ggg gag gac atc gat       384
Asn Arg Trp Arg Ile Ala Thr Lys Gln Ala Asn Gly Glu Asp Ile Asp
                115                 120                 125

gtc acc tcc aat gac tag ggt ggg caa cca caa acc cac gag ggc aga       432
Val Thr Ser Asn Asp     Gly Gly Gln Pro Gln Thr His Glu Gly Arg
            130                     135                 140

gtg ctg ctt gct gct ggc cag gcc cct gcg tgg gcc caa gct gga ctc       480
```

-continued

```
Val Leu Leu Ala Ala Gly Gln Ala Pro Ala Trp Ala Gln Ala Gly Leu
            145                 150                 155

tgg cca ctc cct ggc cag gct ttg ggg agg cct gga gtc atg gcc cca      528
Trp Pro Leu Pro Gly Gln Ala Leu Gly Arg Pro Gly Val Met Ala Pro
        160                 165                 170

cag ggc ttg aag ccc ggg gcc gcc att gac aga ggg aca agc aat ggg      576
Gln Gly Leu Lys Pro Gly Ala Ala Ile Asp Arg Gly Thr Ser Asn Gly
    175                 180                 185

ctg gct gag gcc tgg gac cac ttg gcc ttc tcc tcg gag agc ctg cct      624
Leu Ala Glu Ala Trp Asp His Leu Ala Phe Ser Ser Glu Ser Leu Pro
190                 195                 200                 205

gcc tgg gcg ggc ccg ccc gcc acc gca gcc tcc cag ctg ctc tcc gtg      672
Ala Trp Ala Gly Pro Pro Ala Thr Ala Ala Ser Gln Leu Leu Ser Val
                210                 215                 220

tct cca atc tcc                                                      684
Ser Pro Ile Ser
            225


<210> SEQ ID NO 431
<211> LENGTH: 41
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 431

Lys Thr Thr Leu Leu Ser Gly Pro Leu Cys Pro Leu Pro Ala Leu Pro
1               5                   10                  15

Ser Pro Ser Val Asn Val Arg Pro Met Gly Ala Ala Gly Gln Arg Gly
            20                  25                  30

Pro Arg Pro Gly Ala Pro Asn Pro Met
        35                  40


<210> SEQ ID NO 432
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 432

Pro Gln Ser Ser His Gln Ala Leu Ser Ser Ala
1               5                   10


<210> SEQ ID NO 433
<211> LENGTH: 79
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 433

Val Leu Arg Pro Gln Trp Leu Leu Trp Gly Pro Pro Glu Phe Leu Ile
1               5                   10                  15

Cys Ala Pro Pro Ser Leu Ala Gln Val Lys Val Trp Phe Gln Asn Arg
            20                  25                  30

Arg Thr Lys Tyr Lys Arg Gln Lys Leu Glu Glu Glu Gly Pro Glu Ser
        35                  40                  45

Glu Gln Lys Lys Lys Gly Ser His His Ile Asn Arg Trp Arg Ile Ala
    50                  55                  60

Thr Lys Gln Ala Asn Gly Glu Asp Ile Asp Val Thr Ser Asn Asp
65                  70                  75


<210> SEQ ID NO 434
```

-continued

```
<211> LENGTH: 94
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 434

Gly Gly Gln Pro Gln Thr His Glu Gly Arg Val Leu Leu Ala Ala Gly
1               5                   10                  15

Gln Ala Pro Ala Trp Ala Gln Ala Gly Leu Trp Pro Leu Pro Gly Gln
            20                  25                  30

Ala Leu Gly Arg Pro Gly Val Met Ala Pro Gln Gly Leu Lys Pro Gly
        35                  40                  45

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

His Leu Ala Phe Ser Ser Glu Ser Leu Pro Ala Trp Ala Gly Pro Pro
65                  70                  75                  80

Ala Thr Ala Ala Ser Gln Leu Leu Ser Val Ser Pro Ile Ser
                85                  90
```

What is claimed is:

1. A genome wide library comprising a plurality of CRISPR-Cas system guide RNAs comprising guide sequences that are capable of targeting a plurality of target sequences in a plurality of genomic loci in a population of eukaryotic cells.

2. The library of claim **1**, wherein the population of cells is a population of embryonic stem (ES) cells.

3. The library of claim **1**, wherein the target sequence in the genomic locus is a non-coding sequence.

4. The library of claim **1**, wherein gene function of one or more gene products is altered by said targeting.

5. The library of claim **1**, wherein said targeting results in a knockout of gene function.

6. The library of claim **1**, wherein the targeting is of about 100 or more sequences.

7. The library of claim **1**, wherein the targeting is of about 1000 or more sequences.

8. The library of claim **1**, wherein the targeting is of about 20,000 or more sequences.

9. The library of claim **1**, wherein the targeting is of the entire genome.

10. The library of claim **1**, wherein the targeting is of a panel of target sequences focused on a relevant or desirable pathway.

11. The library of claim **10**, wherein the pathway is an immune pathway.

12. The library of claim **5**, wherein targeting is of about 100 or more sequences.

13. The library of claim **5**, wherein targeting is of about 1000 or more sequences.

14. The library of claim **5**, wherein targeting is of about 20,000 or more sequences.

15. The library of claim **5**, wherein targeting is of the entire genome.

16. The library of claim **5**, wherein the targeting is of a panel of target sequences focused on a relevant or desirable pathway.

17. The library of claim **16**, wherein the pathway is an immune pathway.

18. The library of claim **16**, wherein the pathway is a cell division pathway.

19. The library of claim **5**, wherein the knockout of gene function comprises:

introducing into each cell in the population of cells a vector system of one or more vectors comprising an engineered, non-naturally occurring CRISPR-Cas system comprising

I. a Cas protein, and

II. one or more guide RNAs,

wherein components I and II may be same or on different vectors of the system, integrating components I and II into each cell,

wherein the guide sequence targets a unique gene in each cell,

wherein the Cas protein is operably linked to a regulatory element,

wherein when transcribed, the guide RNA comprising the guide sequence directs sequence-specific binding of a CRISPR-Cas system to a target sequence in the genomic loci of the unique gene,

inducing cleavage of the genomic loci by the Cas protein, and

confirming different knockout mutations in a plurality of unique genes in each cell of the population of cells thereby generating a gene knockout cell library.

20. The library of claim **19**, wherein the one or more vectors are plasmid vectors.

21. The library of claim **19**, wherein the regulatory element is an inducible promoter.

22. The library of claim **19**, wherein the inducible promoter is a doxycycline inducible promoter.

23. The library of claim **19**, wherein the confirming of different knockout mutations is by whole exome sequencing.

24. The library of claim **19**, wherein the knockout mutation is achieved in 100 or more unique genes.

25. The library of claim **19**, wherein the knockout mutation is achieved in 1000 or more unique genes.

26. The library of claim **19**, wherein the knockout mutation is achieved in 20,000 or more unique genes.

27. The library of claim **19**, wherein the knockout mutation is achieved in the entire genome,

28. The library of claim **19**, wherein the knockout of gene function is achieved in a plurality of unique genes which function in a particular physiological pathway or condition.

29. The library of claim **28**, wherein the pathway or condition is an immune pathway or condition.

30. The library of claim **28**, wherein the pathway or condition is a cell division pathway or condition.

* * * * *