US 20250250589A1

(19) United States
(12) Patent Application Publication
ZHANG
(10) Pub. No.: US 2025/0250589 A1
(43) Pub. Date: Aug. 7, 2025

(54) **DELIVERY, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION AND THERAPEUTIC APPLICATIONS**

(71) Applicants: **The Broad Institute, Inc.**, Cambridge, MA (US); **Massachusetts Institute of Technology**, Cambridge, MA (US)

(72) Inventor: **Feng ZHANG**, Cambridge, MA (US)

(73) Assignees: **The Broad Institute, Inc.**, Cambridge, MA (US); **Massachusetts Institute of Technology**, Cambridge, MA (US)

(21) Appl. No.: **19/025,611**

(22) Filed: **Jan. 16, 2025**

Related U.S. Application Data

(63) Continuation of application No. 15/430,260, filed on Feb. 10, 2017, which is a continuation of application No. 14/183,512, filed on Feb. 18, 2014, now abandoned, which is a continuation of application No. 14/104,837, filed on Dec. 12, 2013, now abandoned.

(60) Provisional application No. 61/736,527, filed on Dec. 12, 2012, provisional application No. 61/748,427, filed on Jan. 2, 2013, provisional application No. 61/758,468, filed on Jan. 30, 2013, provisional application No. 61/769,046, filed on Feb. 25, 2013, provisional application No. 61/802,174, filed on Mar. 15, 2013, provisional application No. 61/791,409, filed on Mar. 15, 2013, provisional application No. 61/806,375, filed on Mar. 28, 2013, provisional application No. 61/814,263, filed on Apr. 20, 2013, provisional application No. 61/819,803, filed on May 6, 2013, provisional application No. 61/828,130, filed on May 28, 2013, provisional application No. 61/836,123, filed on Jun. 17, 2013, provisional application No. 61/835,931, filed on Jun. 17, 2013, provisional application No. 61/847,537, filed on Jul. 17, 2013.

Publication Classification

(51) **Int. Cl.**

C12N 15/90	(2006.01)
A01K 67/0275	(2024.01)
A01K 67/0278	(2024.01)
C12N 9/22	(2006.01)
C12N 9/96	(2006.01)
C12N 15/01	(2006.01)
C12N 15/10	(2006.01)
C12N 15/113	(2010.01)
C12N 15/63	(2006.01)
C12N 15/79	(2006.01)
C12N 15/82	(2006.01)
C12N 15/85	(2006.01)
C12N 15/86	(2006.01)
G16B 20/00	(2019.01)
G16B 20/20	(2019.01)
G16B 20/30	(2019.01)
G16B 20/50	(2019.01)
G16B 30/00	(2019.01)
G16B 30/10	(2019.01)

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

CPC ***C12N 15/907*** (2013.01); ***A01K 67/0275*** (2013.01); ***A01K 67/0278*** (2013.01); ***C12N 9/22*** (2013.01); ***C12N 9/96*** (2013.01); ***C12N 15/01*** (2013.01); ***C12N 15/102*** (2013.01); ***C12N 15/1082*** (2013.01); ***C12N 15/113*** (2013.01); ***C12N 15/63*** (2013.01); ***C12N 15/8213*** (2013.01); ***C12N 15/85*** (2013.01); ***C12N 15/8509*** (2013.01); ***C12N 15/86*** (2013.01); ***C12N 15/90*** (2013.01); ***G16B 20/00*** (2019.02); ***G16B 20/20*** (2019.02); ***G16B 20/30*** (2019.02); ***G16B 20/50*** (2019.02); ***G16B 30/10*** (2019.02); *A01K 2217/05* (2013.01); *A01K 2217/052* (2013.01); *A01K 2217/07* (2013.01); *A01K 2217/072* (2013.01); *A01K 2227/105* (2013.01); *A01K 2267/03* (2013.01); *C12N 15/79* (2013.01); *C12N 2310/10* (2013.01); *C12N 2310/20* (2017.05); *C12N 2320/11* (2013.01); *C12N 2320/30* (2013.01); *C12N 2750/14143* (2013.01); *C12N 2800/22* (2013.01); *G16B 30/00* (2019.02)

(57) **ABSTRACT**

The invention provides for delivery, engineering and optimization of systems, methods, and compositions for manipulation of sequences and/or activities of target sequences. Provided are delivery systems and tissues or organ which are targeted as sites for delivery. Also 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 to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition.

Specification includes a Sequence Listing.

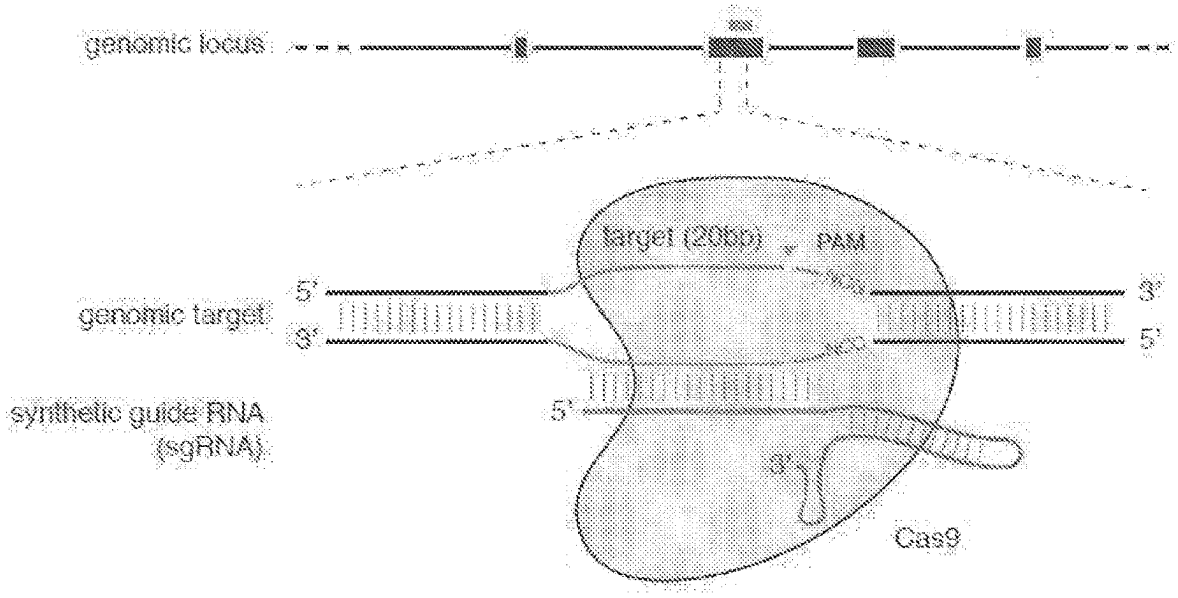

genomic locus
target (20bp)
PAM
genomic target
5'
3'
3'
5'
NGG
synthetic guide RNA
(sgRNA)
5'
3'
Cas9

FIG. 1

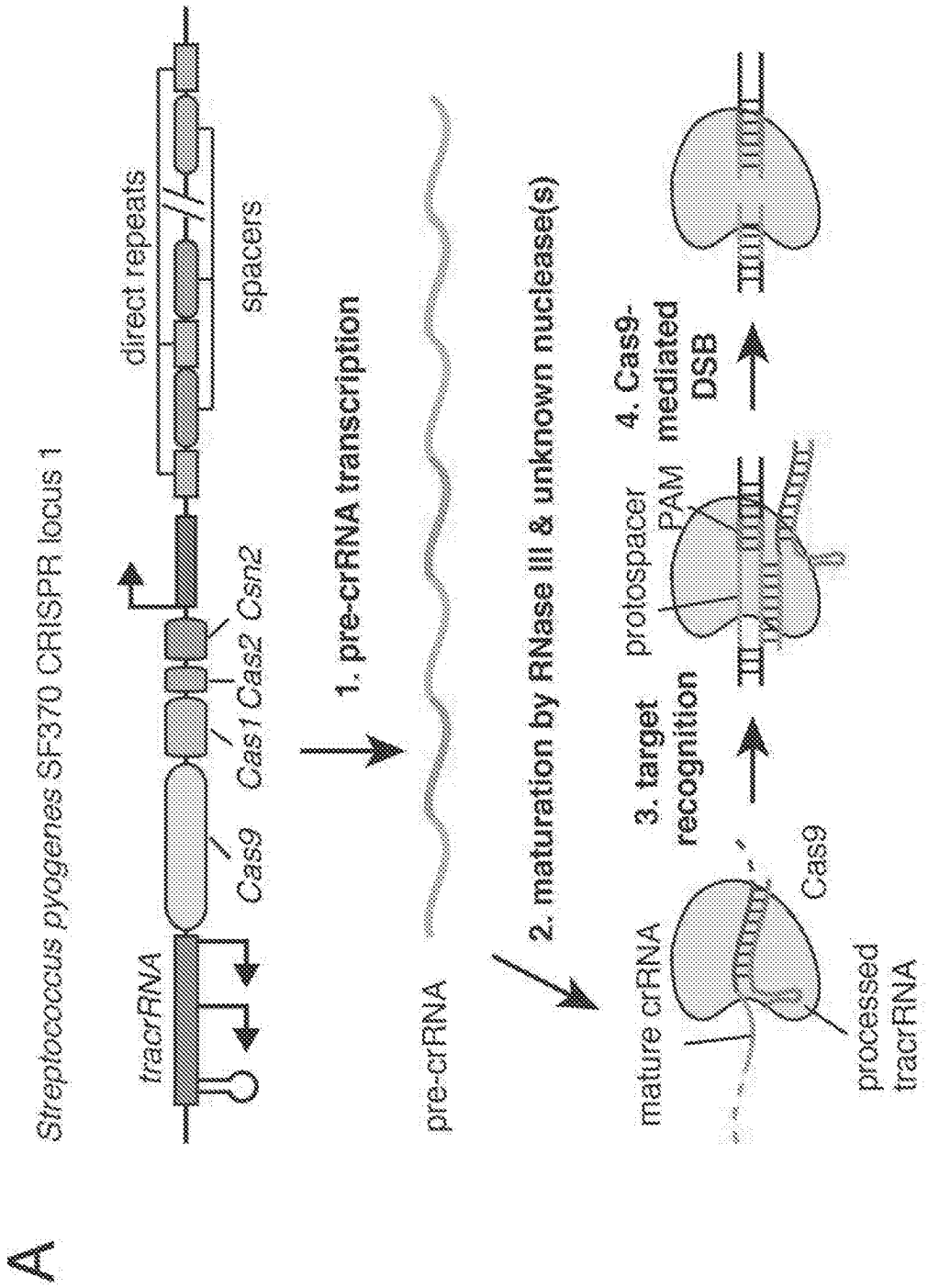

A
Streptococcus pyogenes SF370 CRISPR locus 1
tracrRNA
Cas9
Cas1
Cas2
Csn2
direct repeats
spacers
1. pre-crRNA transcription
pre-crRNA
2. maturation by RNase III & unknown nuclease(s)
mature crRNA
Cas9
processed tracrRNA
3. target recognition
protospacer
PAM
4. Cas9-mediated DSB

FIG. 2A

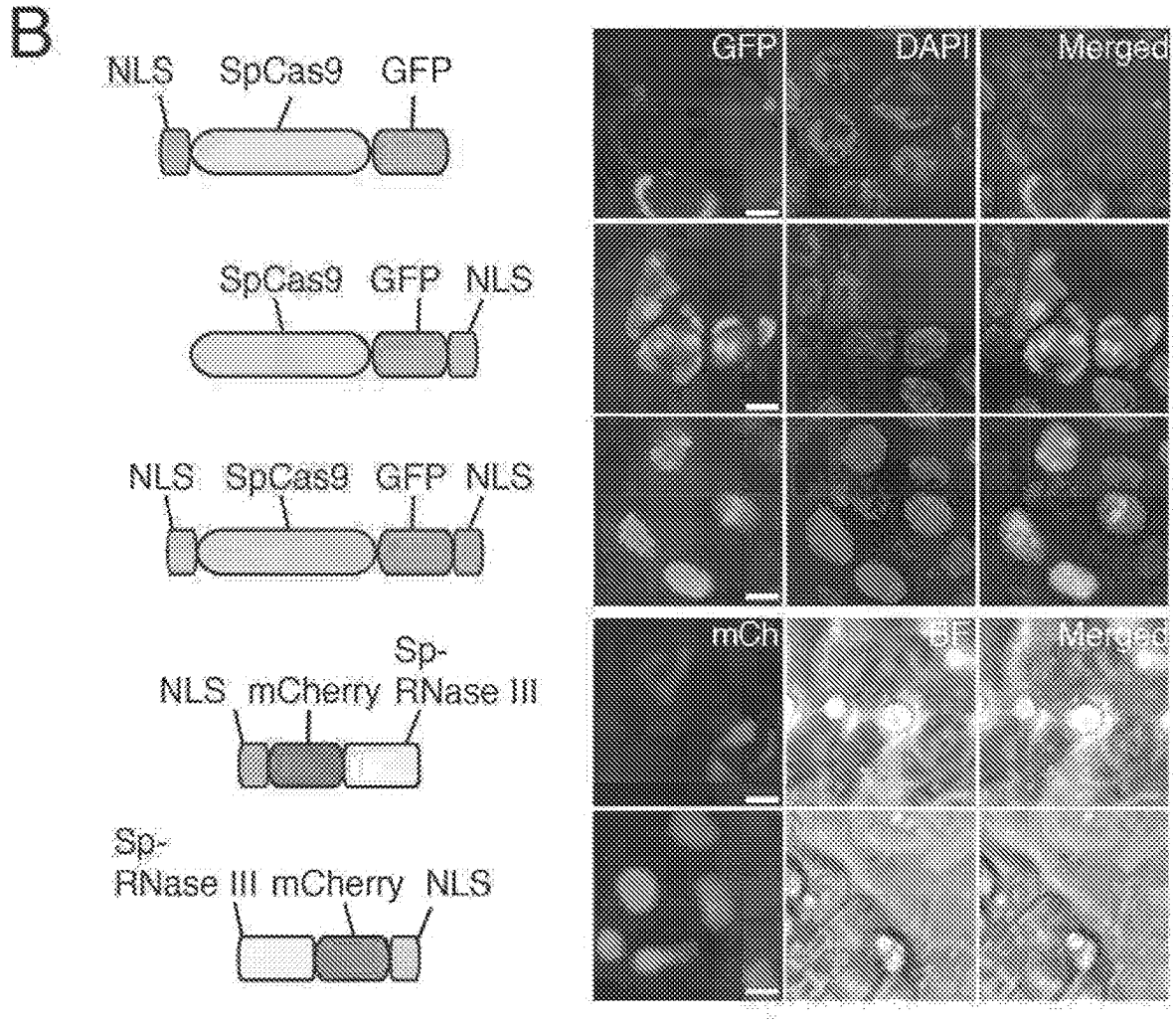

B
NLS
SpCas9
GFP
SpCas9
GFP
NLS
NLS
SpCas9
GFP
NLS
NLS
mCherry
Sp-
RNase III
Sp-
RNase III
mCherry
NLS
GFP
DAPI
Merged
mCh
Merged

FIG. 2B

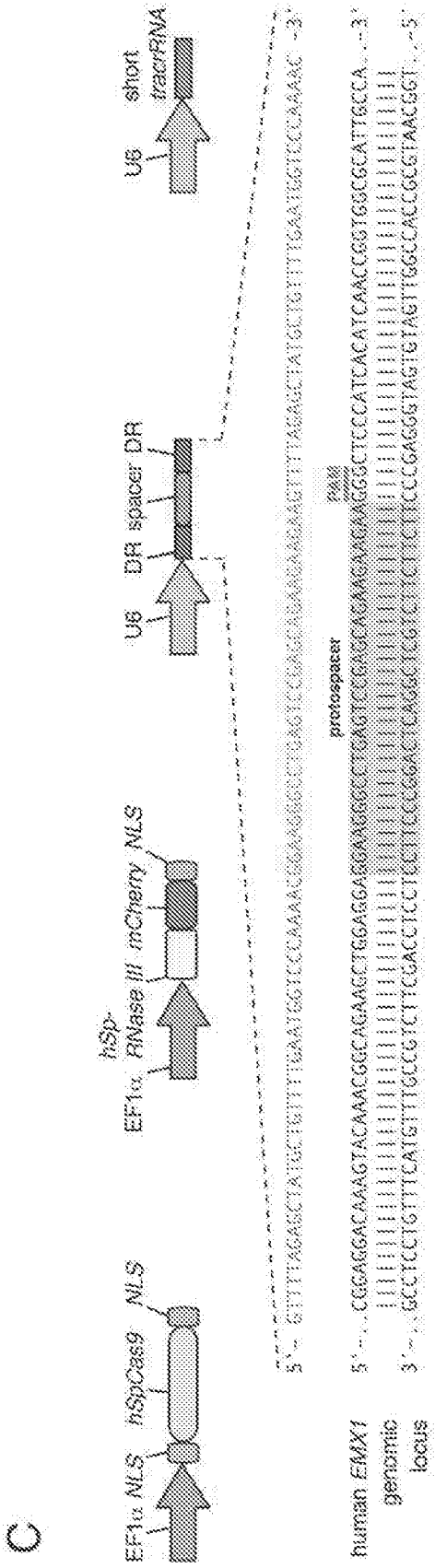

C
EF1α NLS hSpCas9 NLS
hSp-
EF1α RNase III mCherry NLS
U6 DR spacer DR
U6 short tracrRNA
human EMX1 genomic locus
protospacer
PAM

FIG. 2C

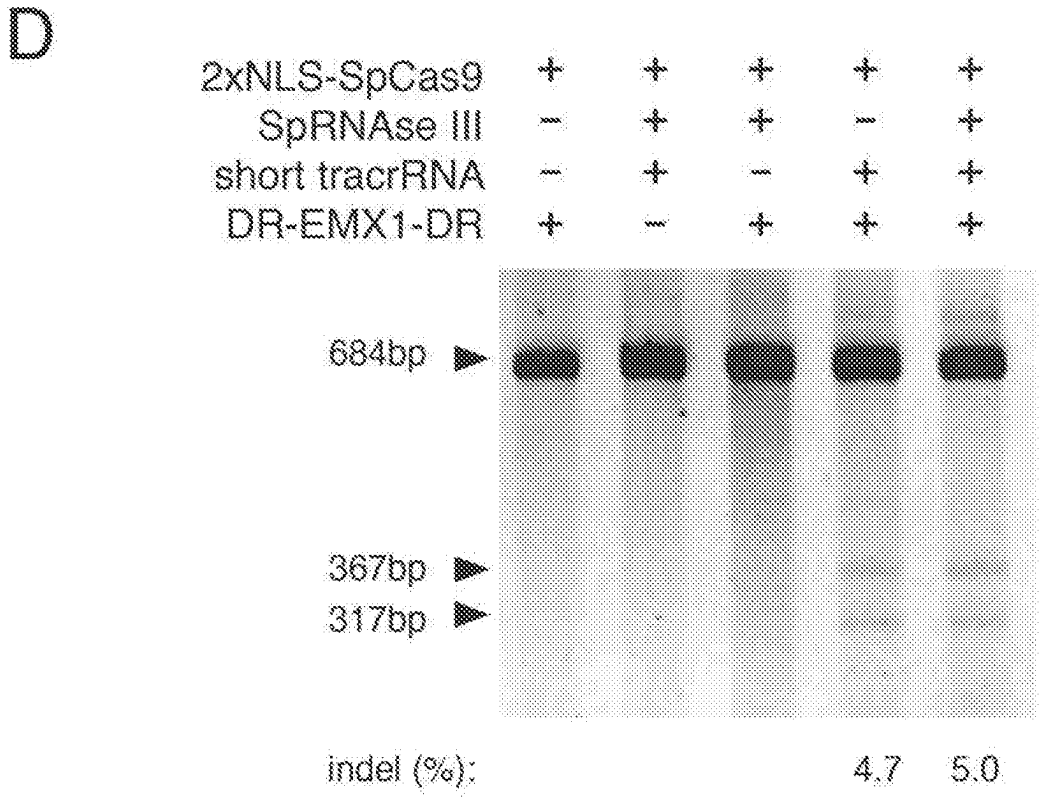

D
2xNLS-SpCas9 + + + + +
SpRNAse III − + + − +
short tracrRNA − + − + +
DR-EMX1-DR + − + + +
684bp
367bp
317bp
indel (%): 4.7 5.0

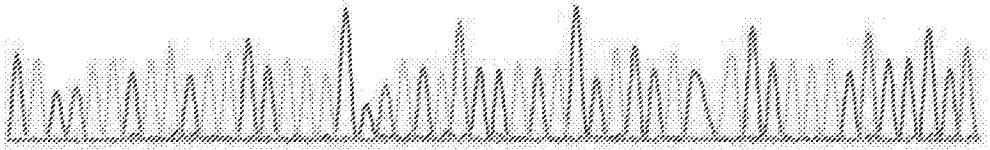
FIG. 2D

E

Target locus 5'-..AGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAC..-3' (PAM)

3'-..TCGACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTG..-5'

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-2F

A

protospacer (1) PAM

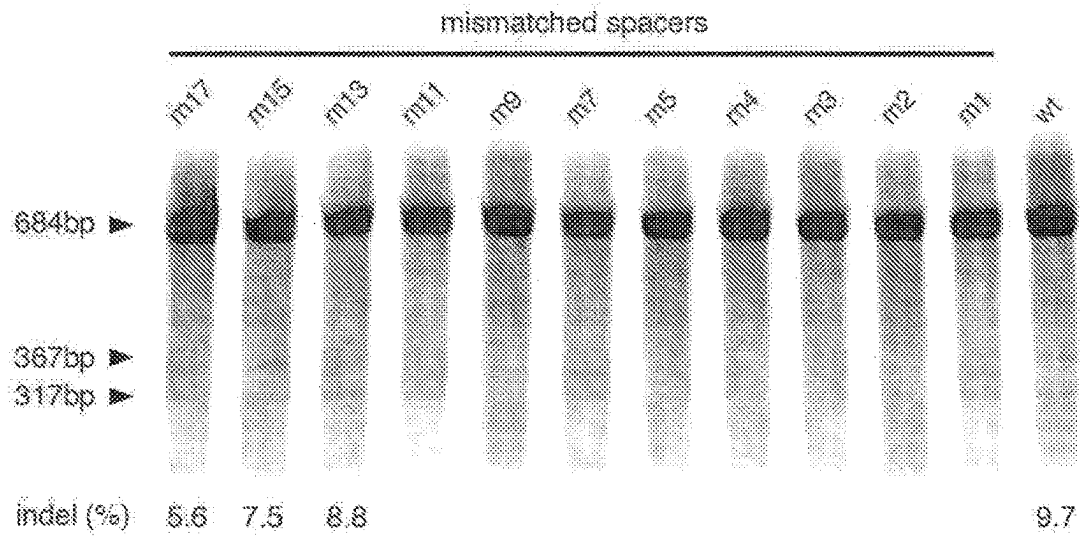
human *EMX1* locus
5'-...CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAT...-3'
3'-...GACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTA...-5'

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



C

left TALEN binding site protospacer (1) PAM

human *EMX1* locus
5'-...CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAT...-3'
3'-...GACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTA...-5'

right TALEN binding site

FIG. 3A-3C

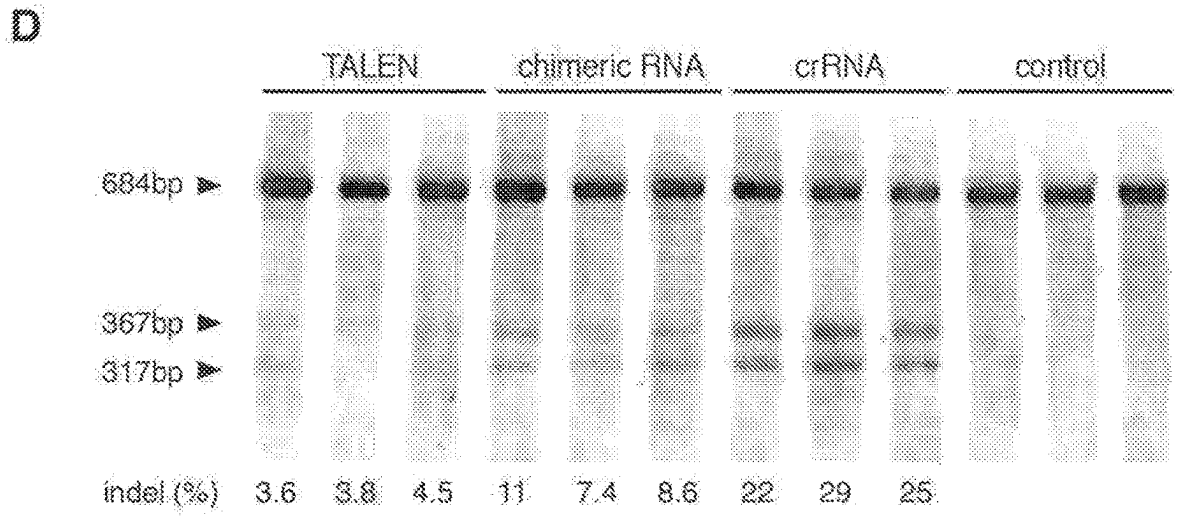

D
TALEN
chimeric RNA
crRNA
control
684bp
367bp
317bp
indel (%)
3.6
3.8
4.5
11
7.4
8.6
22
29
25

FIG. 3D

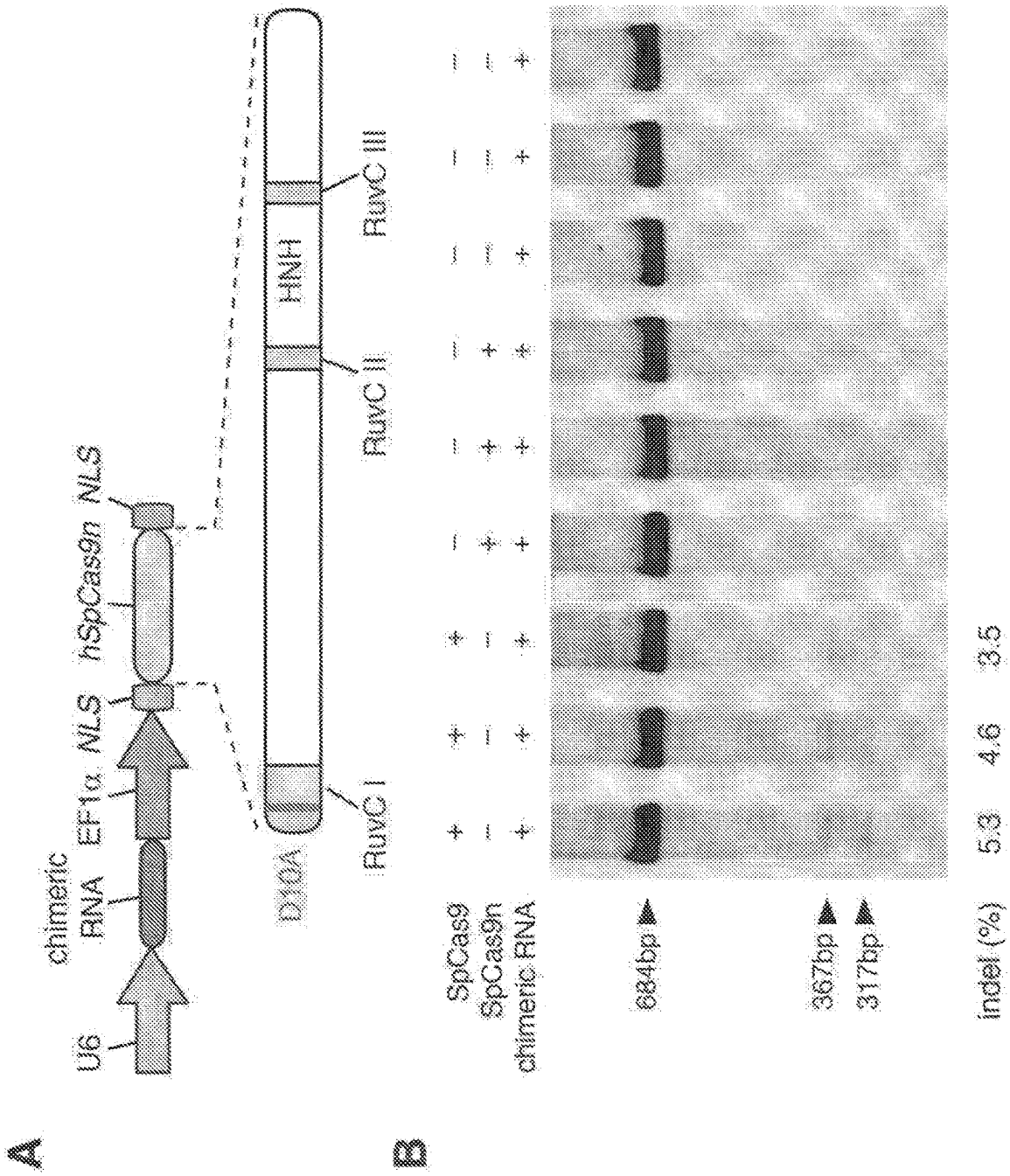

A
U6
chimeric RNA
EF1α
NLS
hSpCas9n
NLS
D10A
RuvC I
RuvC II
HNH
RuvC III
B
SpCas9
SpCas9n
chimeric RNA
684bp
367bp
317bp
indel (%)
5.3
4.6
3.5

FIG. 4A-4B

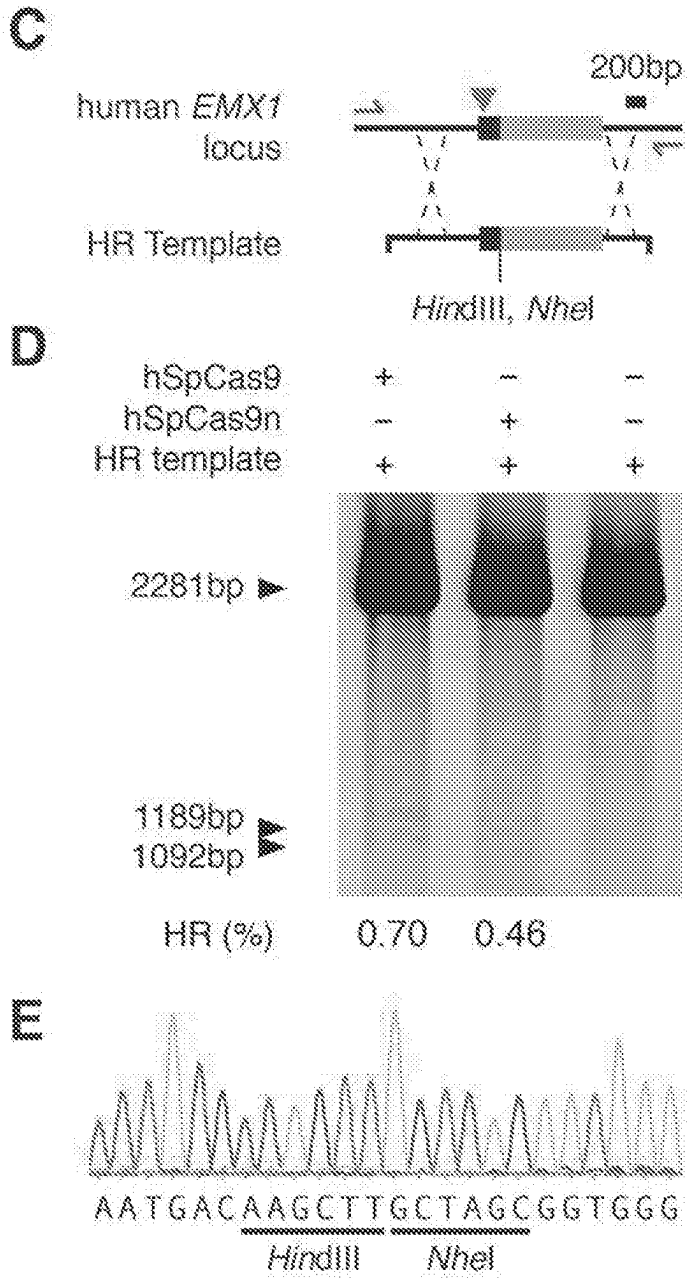

C
200bp
human EMX1 locus
HR Template
HindIII, NheI
D
hSpCas9 + – –
hSpCas9n – + –
HR template + + +
2281bp
1189bp
1092bp
HR (%) 0.70 0.46
E
AATGACAAGCTTGCTAGCGGTGGG
HindIII
NheI

FIG. 4C-4E

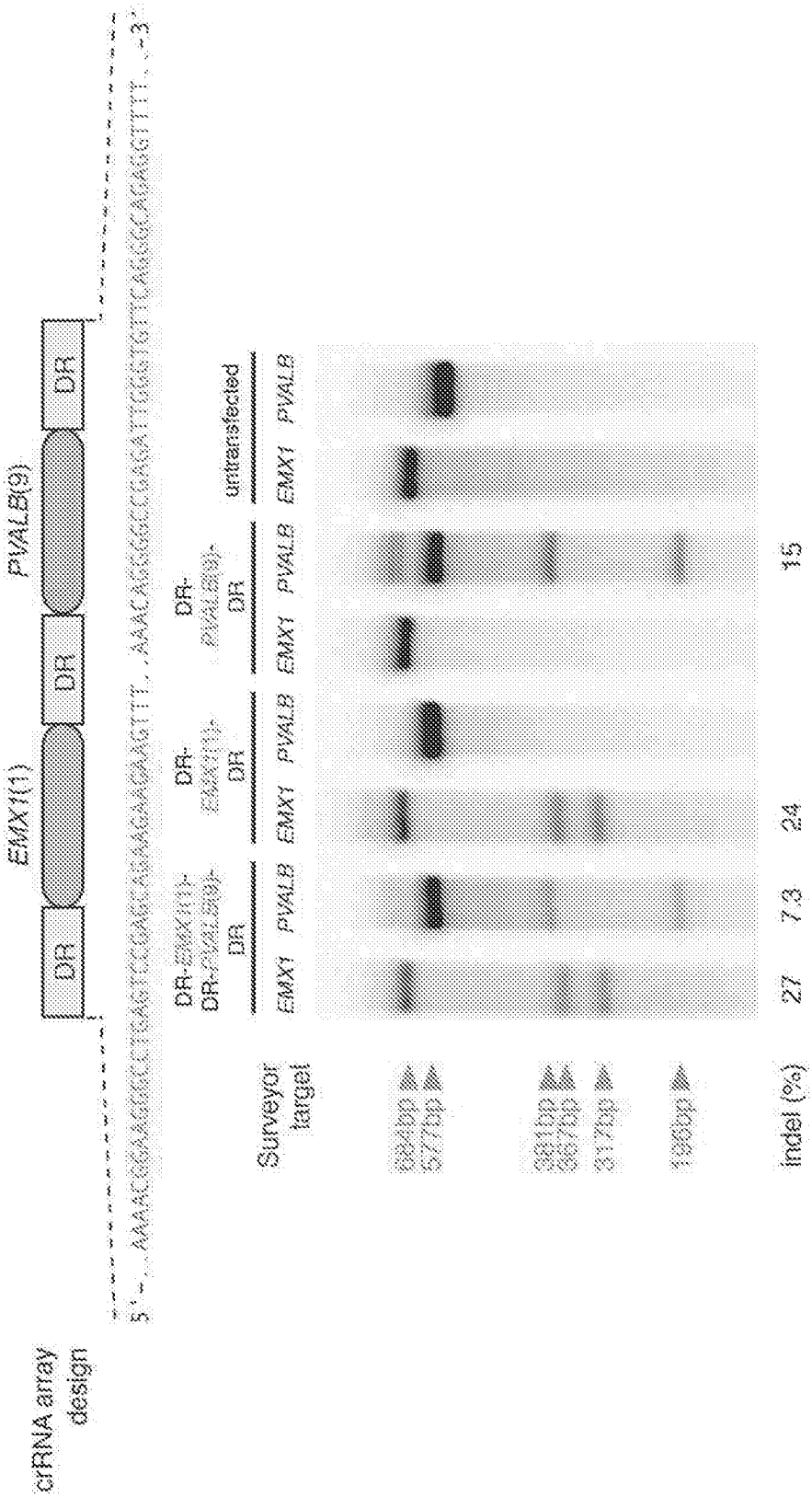

F
crRNA array design
EMX1(1)
PVALB(9)
DR
DR
DR
DR
5'-...AAACAGGGAAGGGCCTTGAGTCCGAGCAGAAGAAGAAGTTT...AAACAGGGGCCGAGATTGGGTGTTCAGGGCAGAGGTTTT...-3'
DR-EMX1(1)-DR-PVALB(9)-DR
DR-EMX1(1)-DR
DR-PVALB(9)-DR
untransfected
EMX1
PVALB
EMX1
PVALB
EMX1
PVALB
EMX1
PVALB
Surveyor target
884bp
577bp
881bp
667bp
317bp
186bp
indel (%)
27
7.3
24
15

FIG. 4F

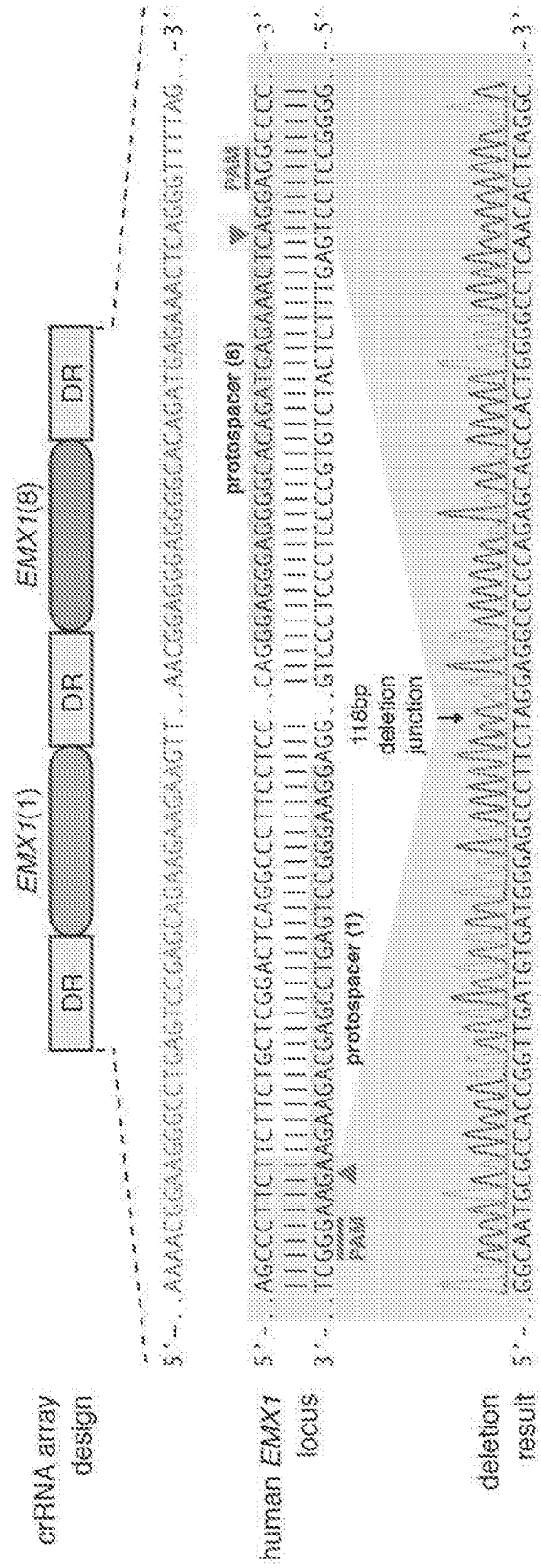
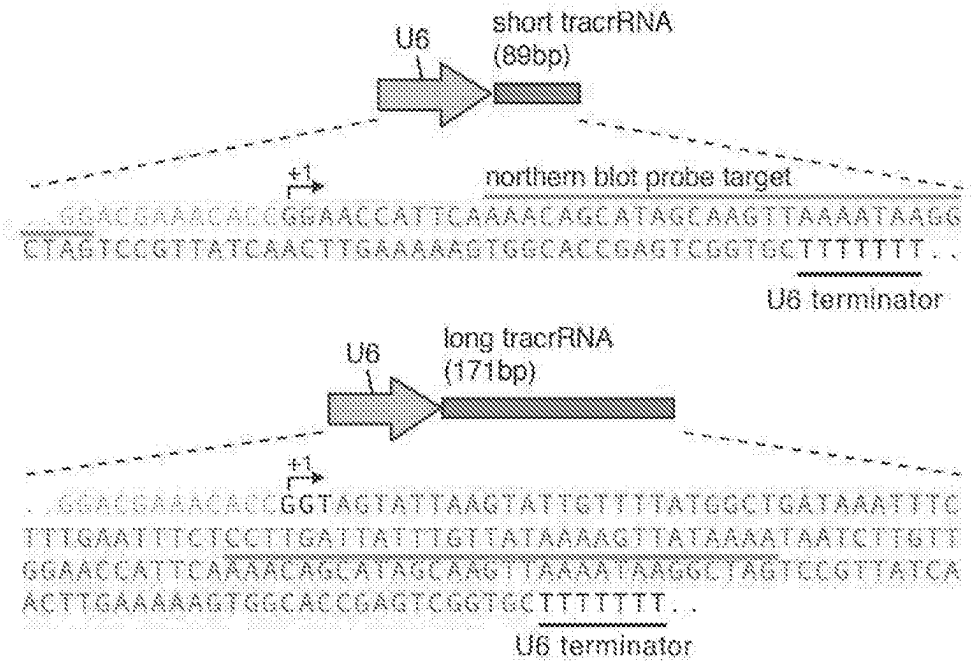

G
crRNA array design
human EMX1 locus
deletion result
DR
EMX1(1)
DR
EMX1(8)
DR
5' - ..AAAACCGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTT..AACGGAGGGAGGGGCACAGATGAGAACTCAGGGTTTTTAG..-3'
5' - ..AGCCCTTCTTCTTCTGCTCGGACTCAGGCCCTTCCTCC..CAGGGAGGGAGGGGCACAGATCAGAACTCAGGAGCCCC..-3'
3' - ..TCGGGAAGAAGACGAGCCTGAGTCCGGGAAGGAGG..GTCCCTCCCTCCCCGTGTCTACTCTTTGAGTCCTCGGGG..-5'
PAM
protospacer (1)
118bp deletion junction
protospacer (8)
PAM
5' - ..GGCAATGCGCCACCGGTTGATGTGATGGGAGCCCTTCTAGGAGGCCCCAGAGCAGCCACTGGGCCTCAACACTCAGGC..-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 ± 0.48	2.1 ± 0.51
		EMX1	6	TCGTGGCAATGCGCCACCGGTTGATGTGAT	GGG	–	293FT	4.0 ± 0.86	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
CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTT..
U6 terminator
U6
long tracrRNA
(171bp)
+1
..GGACGAAACACCGGTAGTATTAAGTATTGTTTTATGGCTGATAAATTTC
TTTGAATTTCTCCTTGATTATTTGTTATAAAAGTTATAAAATAATCTTGTT
GGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
ACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTT..
U6 terminator

B

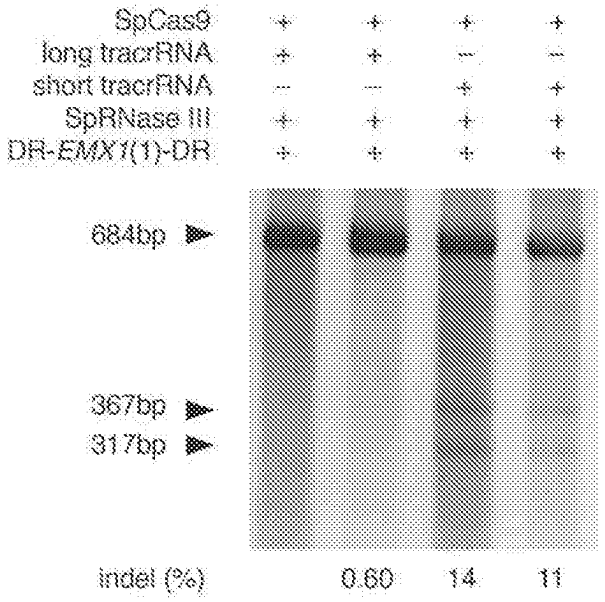

SpCas9 + + + +
long tracrRNA + + − −
short tracrRNA − − + +
SpRNase III + + + +
DR-EMX1(1)-DR + + + +
684bp
367bp
317bp
indel (%) 0.60 14 11

FIG. 6A-6B

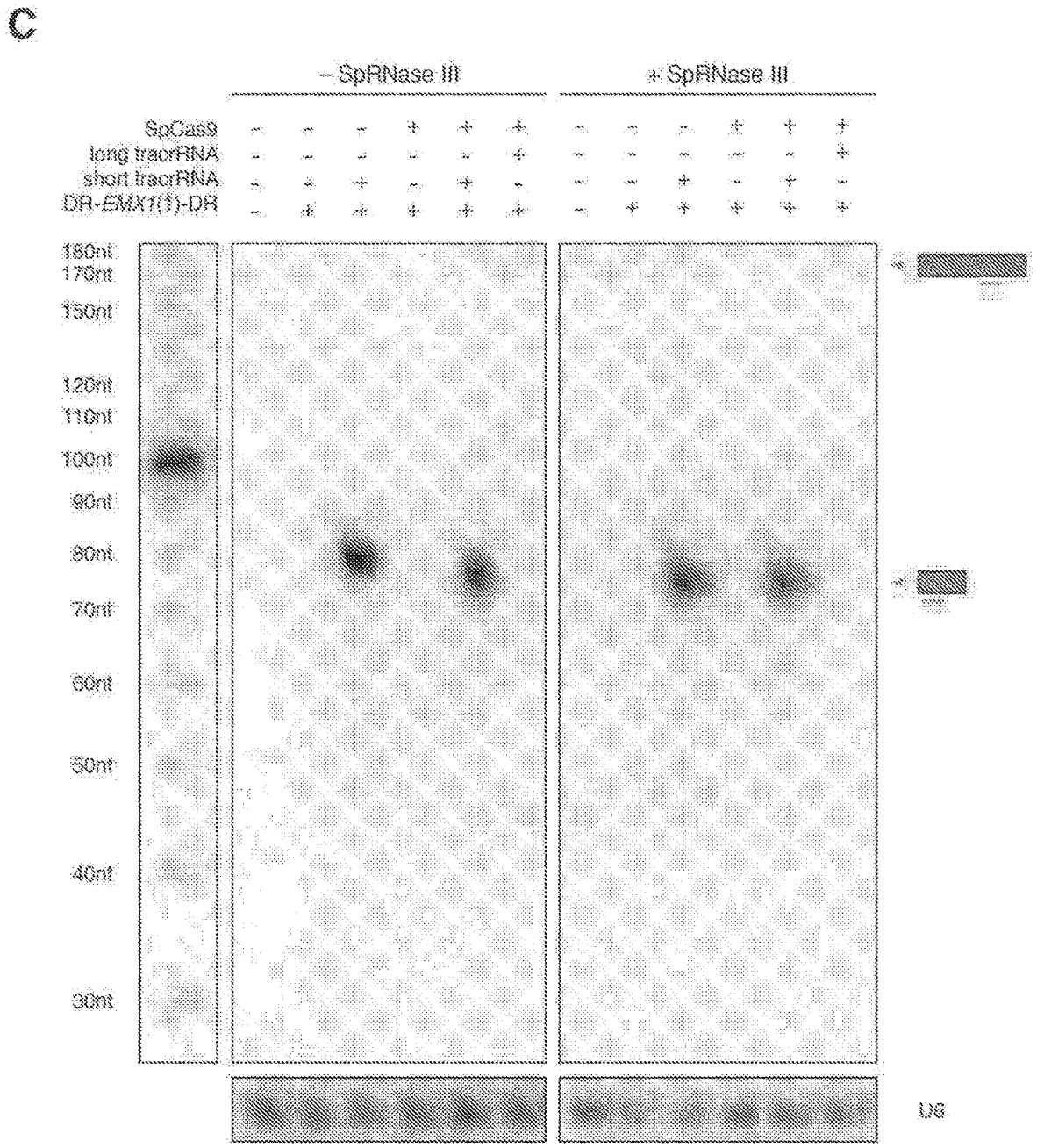

C
– SpRNase III
+ SpRNase III
SpCas9
long tracrRNA
short tracrRNA
DR-EMX1(1)-DR
180nt
170nt
150nt
120nt
110nt
100nt
90nt
80nt
70nt
60nt
50nt
40nt
30nt
U6

FIG. 6C

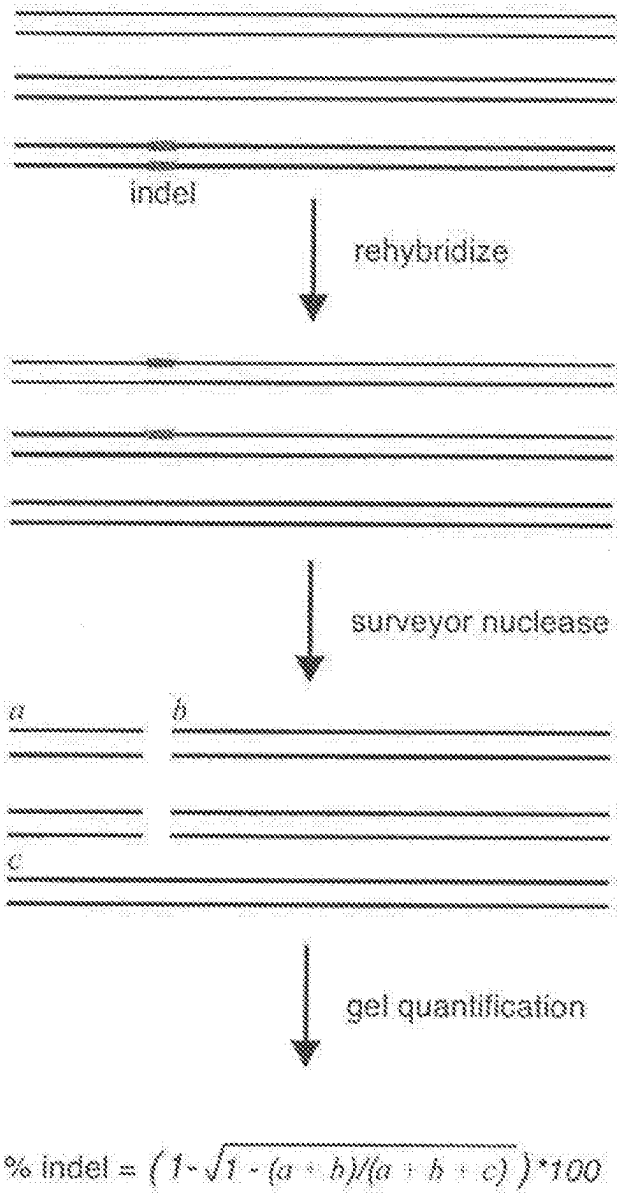

indel
rehybridize
surveyor nuclease
a
b
c
gel quantification
% indel = $\left(1-\sqrt{1-(a+b)/(a+b+c)}\right)*100$

FIG. 7

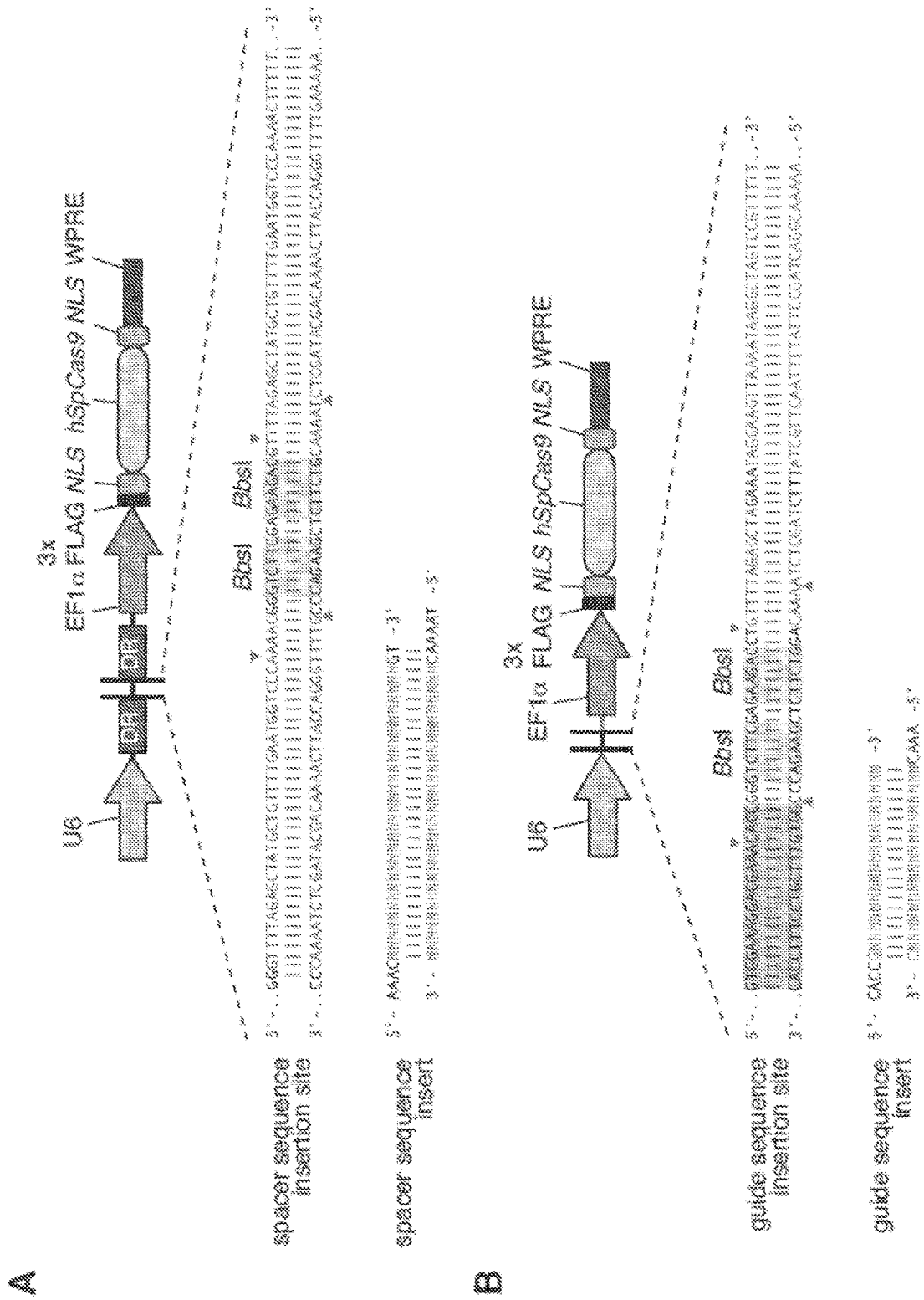
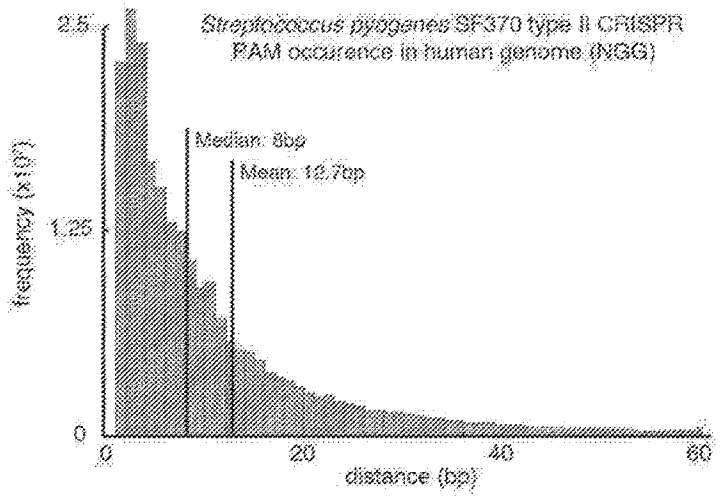
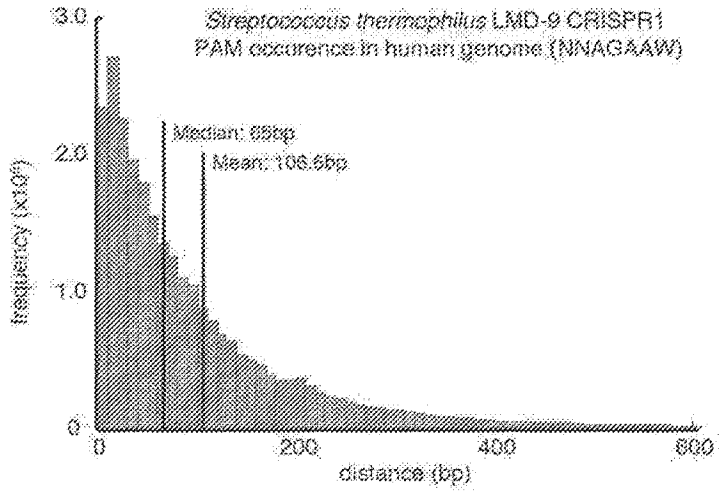

A
U6
3x
EF1α FLAG NLS hSpCas9 NLS WPRE
BtsI
BtsI
spacer sequence insertion site
spacer sequence insert
B
U6
3x
EF1α FLAG NLS hSpCas9 NLS WPRE
BtsI
BtsI
guide sequence insertion site
guide sequence insert

FIG. 8A-8B

A



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-9C

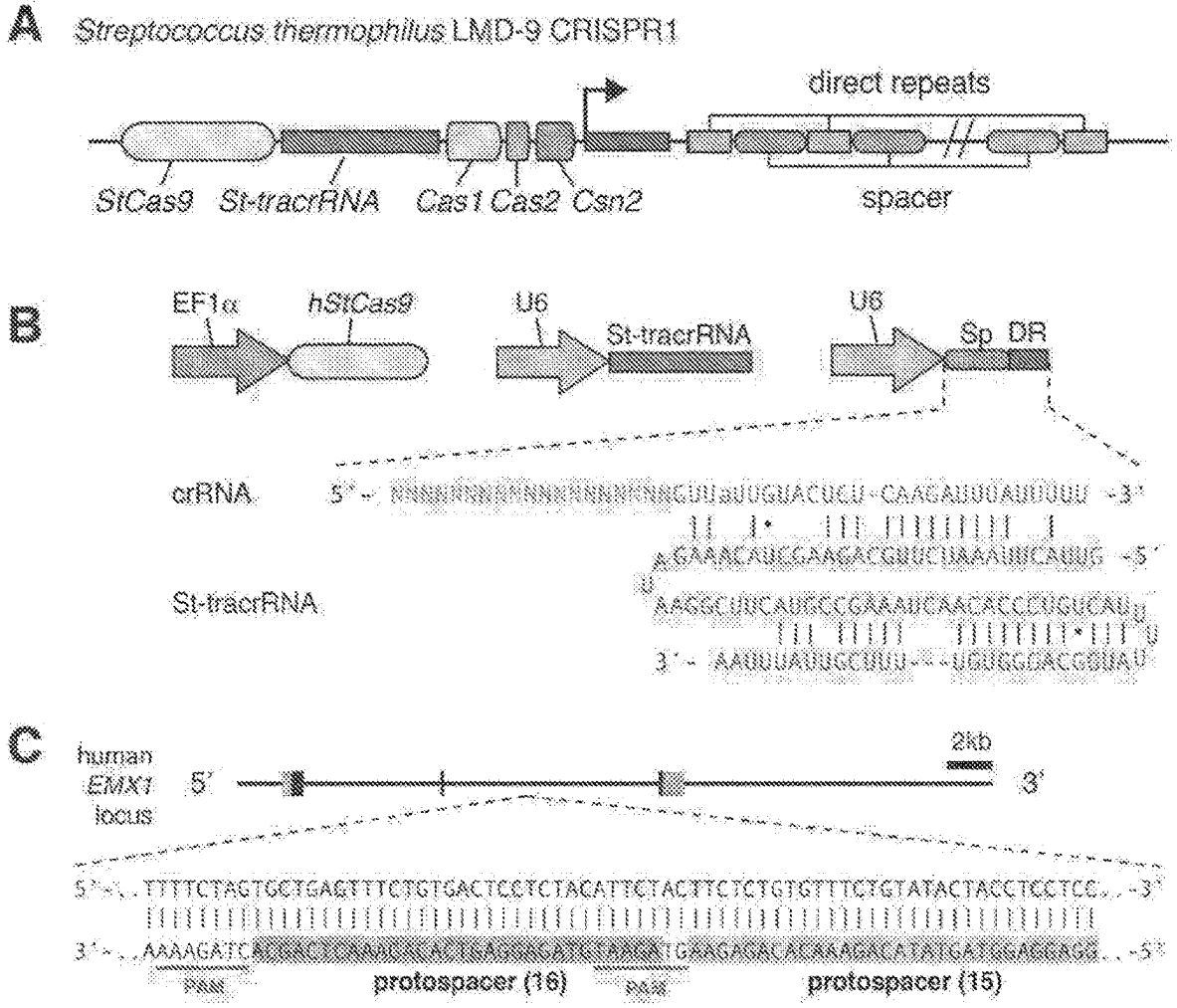

A
Streptococcus thermophilus LMD-9 CRISPR1
direct repeats
StCas9
St-tracrRNA
Cas1
Cas2
Csn2
spacer
B
EF1α
hStCas9
U6
St-tracrRNA
U6
Sp
DR
crRNA
St-tracrRNA
C
human
EMX1
locus
5'
2kb
3'
PAM
protospacer (16)
PAM
protospacer (15)

FIG. 10A-10C

D

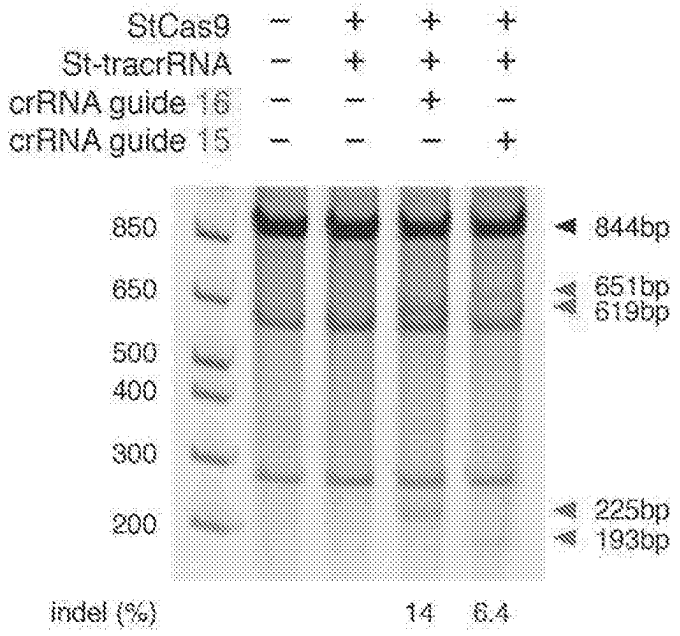

StCas9 – + + +
St-tracrRNA – + + +
crRNA guide 16 – – + –
crRNA guide 15 – – – +
850
650
500
400
300
200
844bp
651bp
619bp
225bp
193bp
indel (%) 14 6.4

FIG. 10D

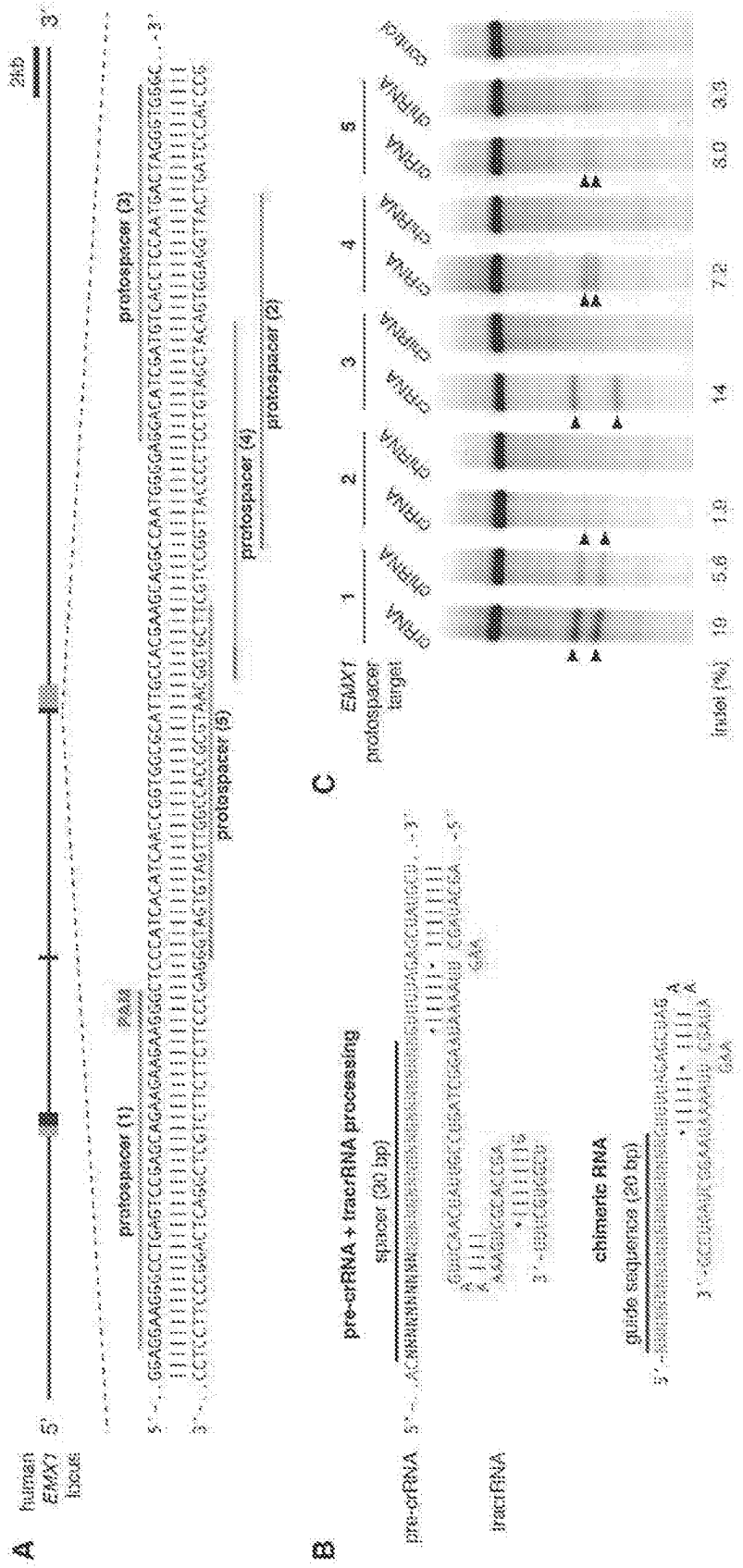

A
human EMX1 locus
2kb
protospacer (1)
PAM
protospacer (3)
protospacer (5)
protospacer (4)
protospacer (2)
B
pre-crRNA + tracrRNA processing
pre-crRNA
spacer (20 bp)
tracrRNA
chimeric RNA
guide sequence (20 bp)
C
EMX1 protospacer target
crRNA
chimeric RNA
indel (%)
19
5.6
1.9
14
7.2
8.0
3.3

FIG. 11A-11C

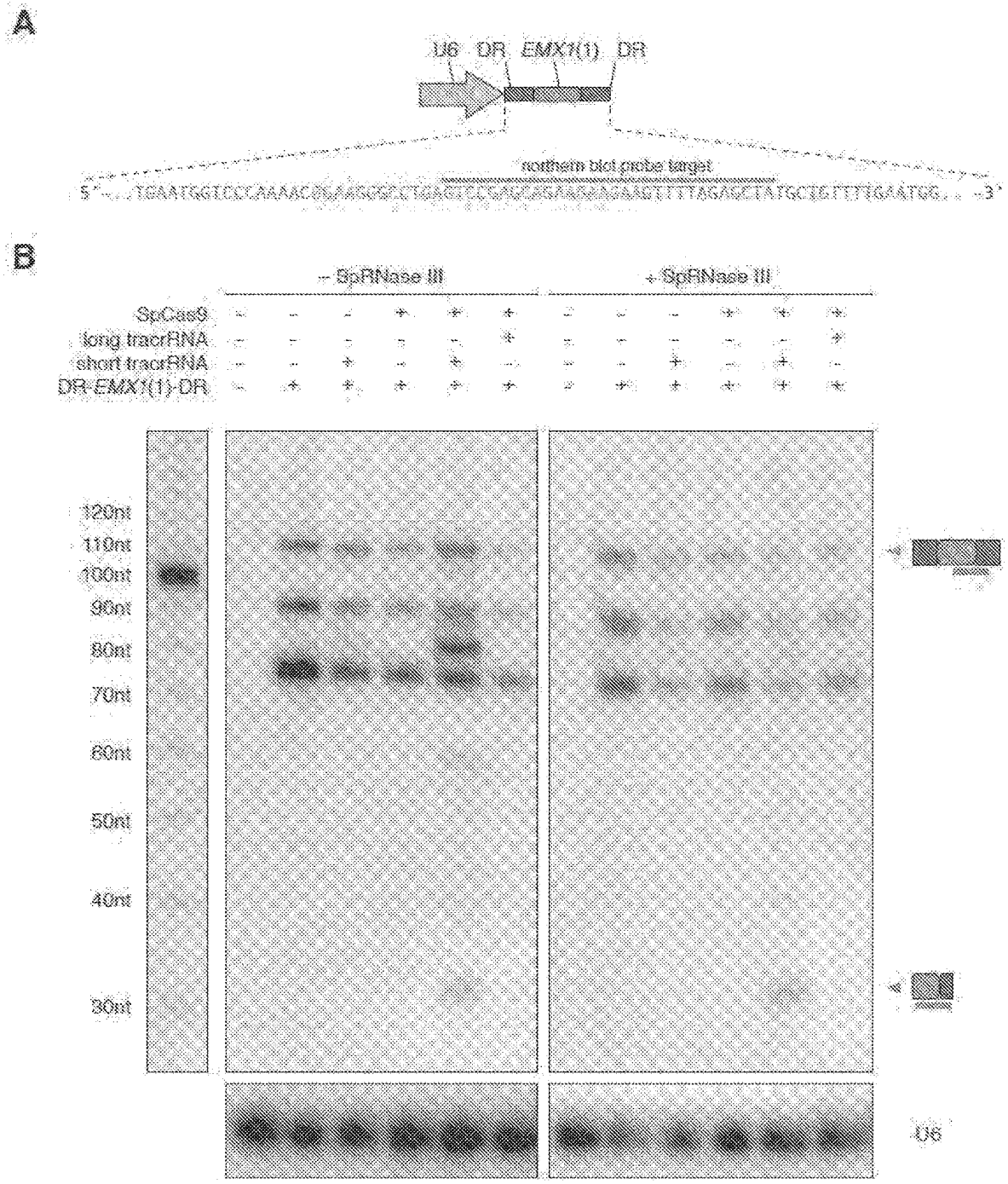

A
U6
DR
EMX1(1)
DR
northern blot probe target
5'
3'
B
– SpRNase III
+ SpRNase III
SpCas9
long tracrRNA
short tracrRNA
DR-EMX1(1)-DR
120nt
110nt
100nt
90nt
80nt
70nt
60nt
50nt
40nt
30nt
U6

FIG. 12A-12B

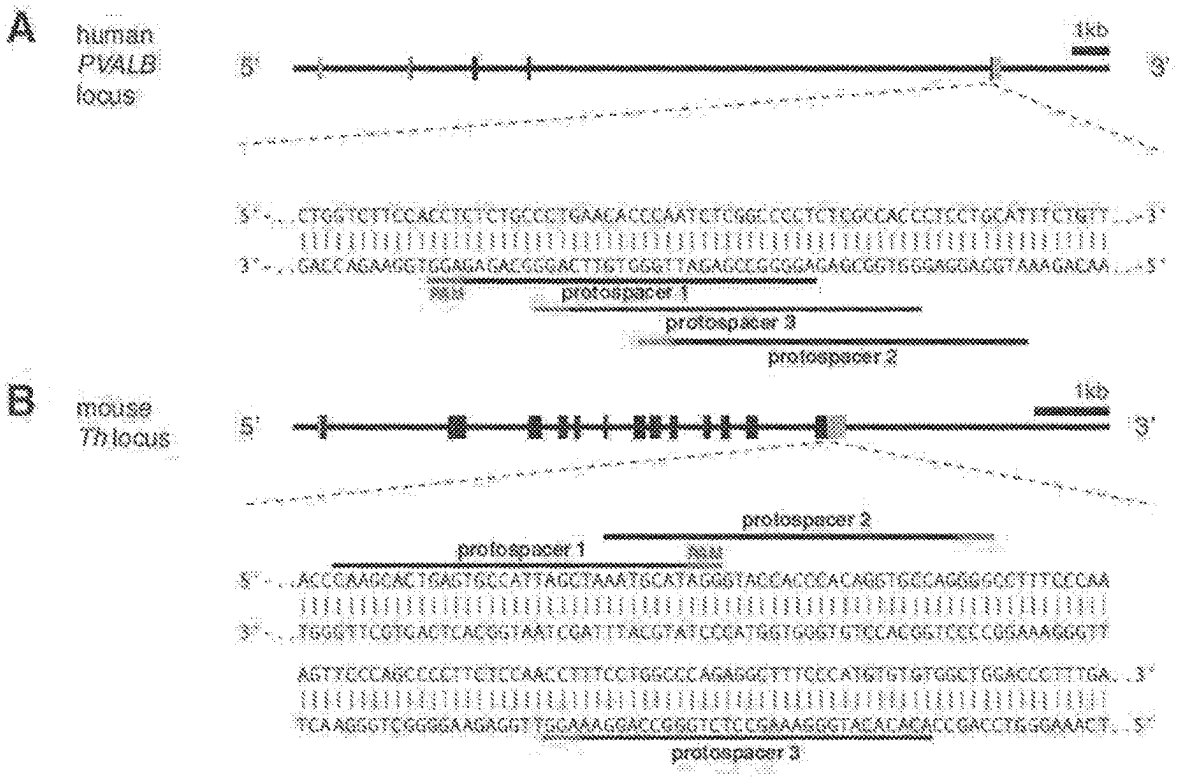

A
human
PVALB
locus
5'
3'
1kb
protospacer 1
protospacer 3
protospacer 2
B
mouse
Th locus
5'
3'
1kb
protospacer 3
protospacer 1
protospacer 3

FIG. 13A-13B

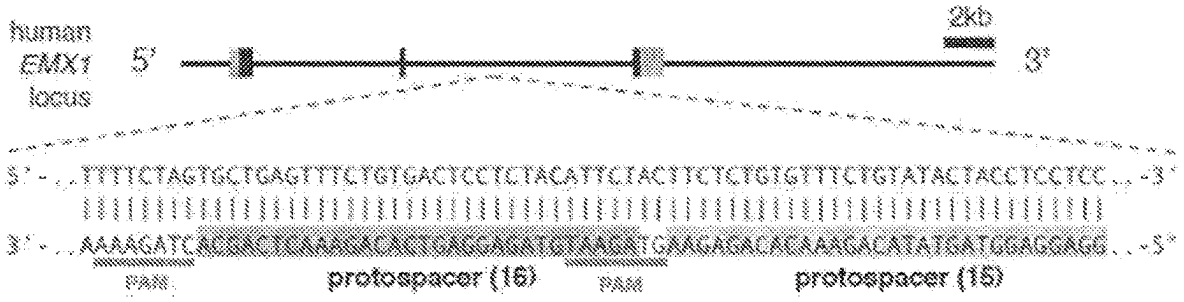

human
EMX1
locus
5'
3'
2kb
5' - . . TTTTCTAGTGCTGAGTTTCTGTGACTCCTCTACATTCTACTTCTCTGTGTTTCTGTATACTACCTCCTCC . . -3'
3' - . . AAAAGATCACGACTCAAACACACTGAGGAGATGTAAGATGAAGAGACACAAACACATATGATGGAGGAGG . . -5'
PAM
protospacer (16)
PAM
protospacer (15)

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

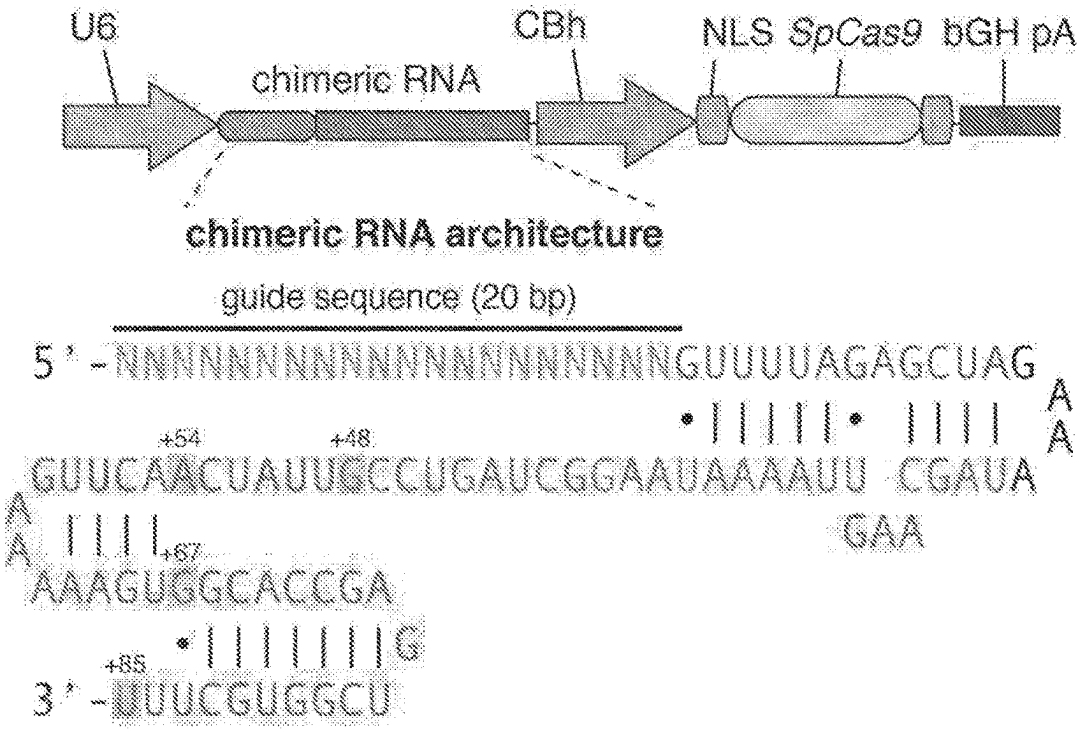

U6
chimeric RNA
CBh
NLS SpCas9 bGH pA
chimeric RNA architecture
guide sequence (20 bp)
5' - NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
A
A
+54
+48
GUUCAACUAUUGCCUGAUCGGAAUAAAAUU CGAUA
A
A
GAA
+67
AAAGUGGCACCGA
G
+85
3' - UUCGUGGCU

FIG. 16A

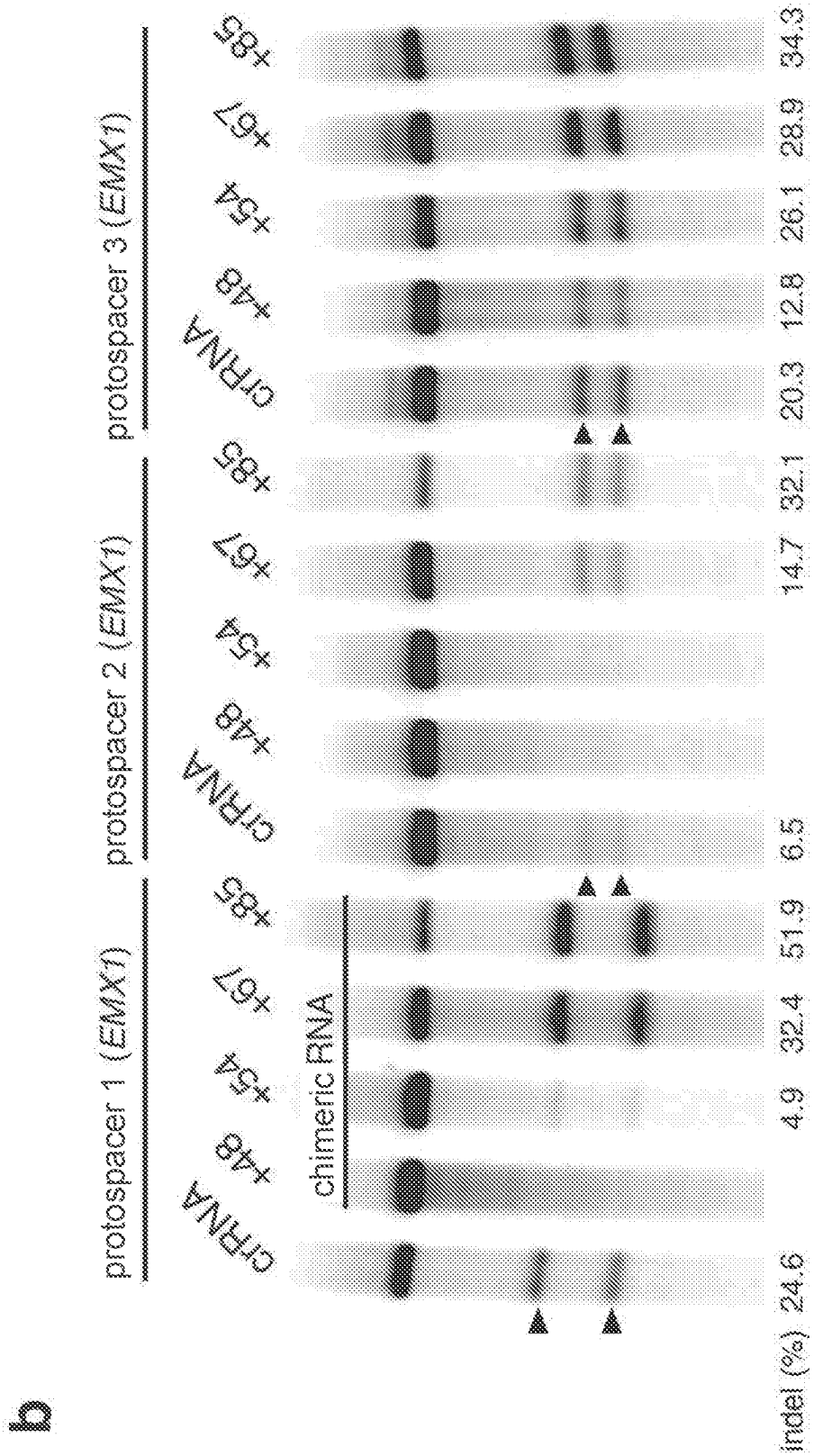

b
protospacer 1 (EMX1)
protospacer 2 (EMX1)
protospacer 3 (EMX1)
crRNA
+48
+54
+67
+85
chimeric RNA
indel (%)
24.6
4.9
32.4
51.9
6.5
14.7
32.1
20.3
12.8
26.1
28.9
34.3

FIG. 16B

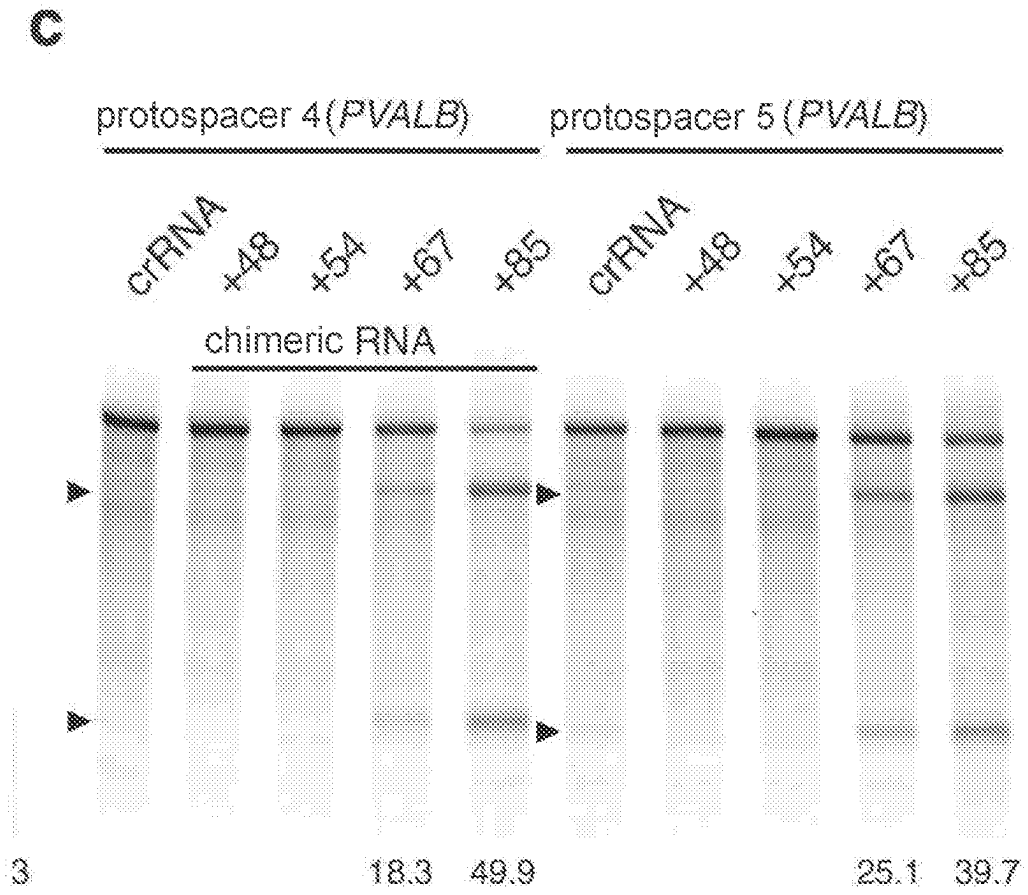
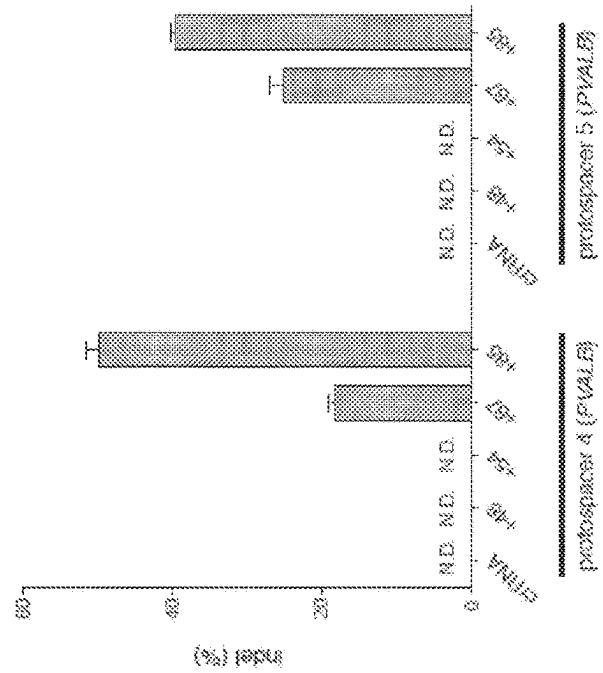

C
protospacer 4 (PVALB)
protospacer 5 (PVALB)
crRNA
+48
+54
+67
+85
crRNA
+48
+54
+67
+85
chimeric RNA
3
18.3
49.9
25.1
39.7

FIG. 16C


b
Indel (%)
60
40
20
0
protospacer 4 (PVALB)
protospacer 5 (PVALB)
N.D.
N.D.
N.D.
N.D.
N.D.
N.D.
CTRNA
6N-1
4N-1
2N-1
0N-1
5N-1

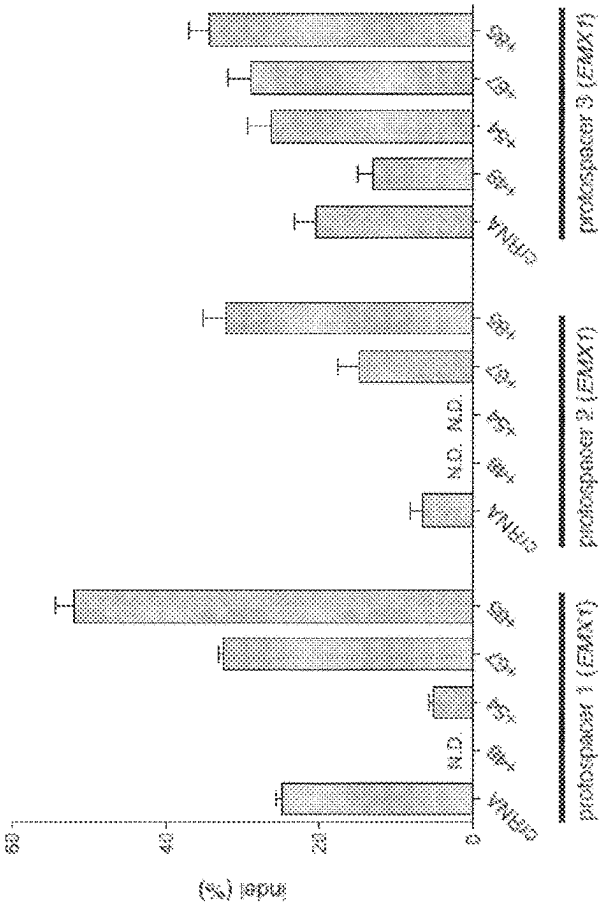

a
Indel (%)
60
40
20
0
protospacer 1 (EMX1)
protospacer 2 (EMX1)
protospacer 3 (EMX1)
N.D.
N.D.
N.D.
CTRNA
6N-1
4N-1
2N-1
0N-1
5N-1

FIG. 17A-17B

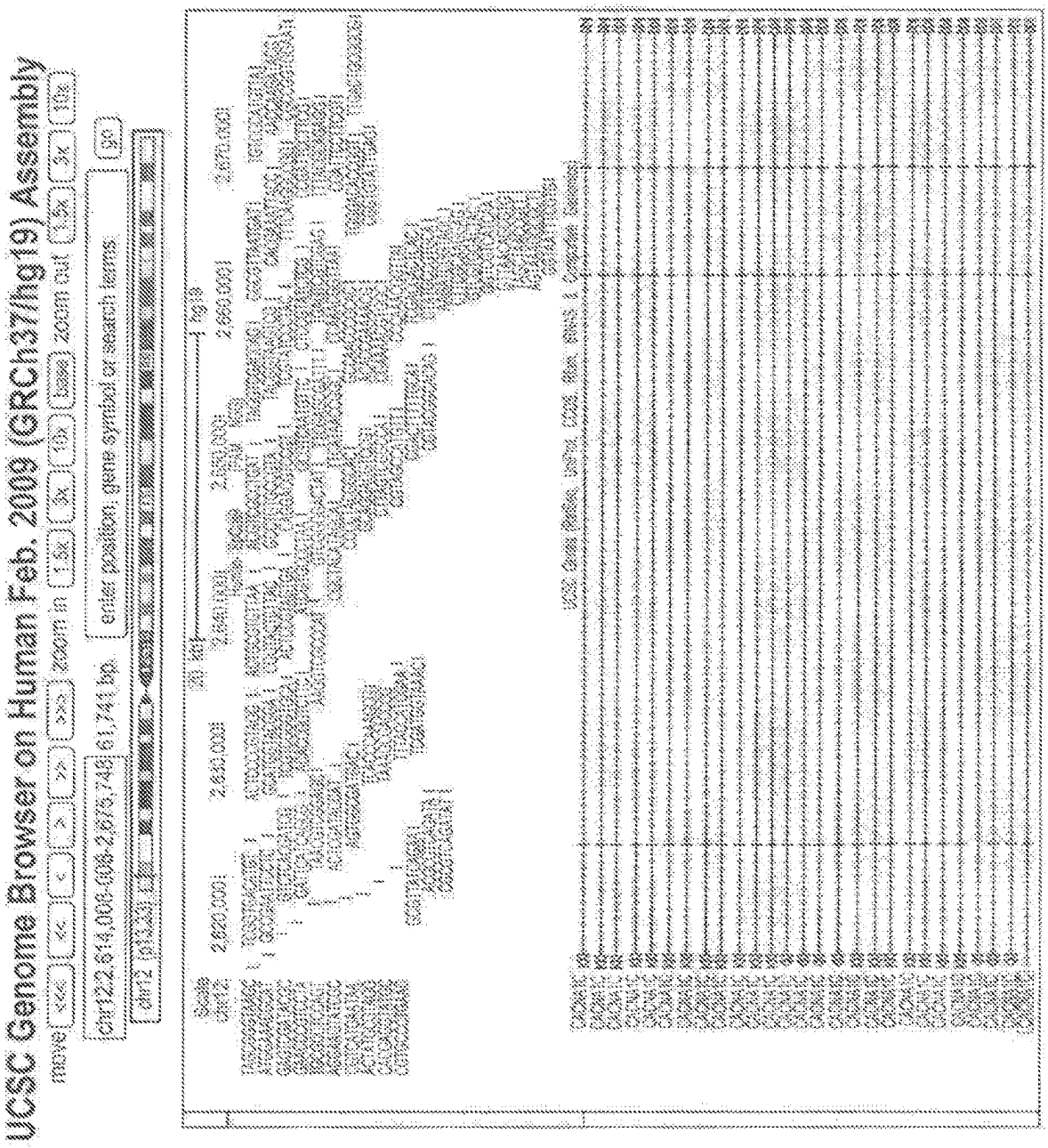

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
move
zoom in
zoom out
enter position, gene symbol or search terms
go
chr12:2,614,006-2,675,748
61,741 bp.
chr12
hg19
2,620,000
2,630,000
2,640,000
2,650,000
2,660,000
2,670,000

FIG. 18

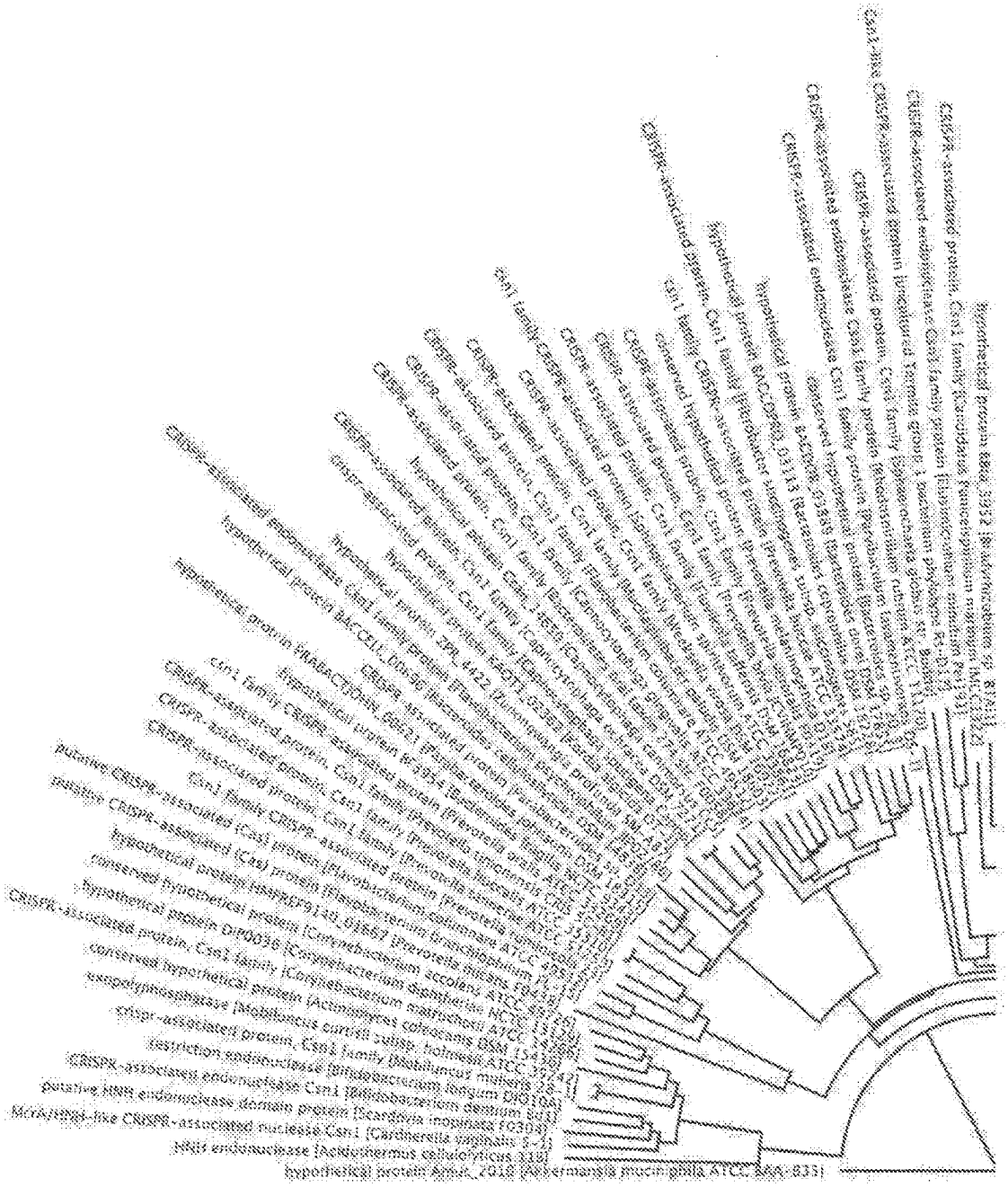


FIG. 19A

hypothetical protein Mbun_4094 [Methanobacter bandungensis JCM]
CRISPR-associated protein, Csn1 family [Azospirillum sp. B510]
hypothetical protein [Dinoroseobacter shibae DFL 12]
hypothetical protein SJA17_3038 [Sphingomonas sp. SJ17]
hypothetical protein [Rhizobium serogroup E24]
CRISPR-associated protein, Csn1 family [Haloferax sp. AL213]
CRISPR-associated endonuclease Csn1 family protein [Haloferax volcanii DS2]
CRISPR-associated protein, Csn1 family [Halorhabdus utahensis DSM 12940

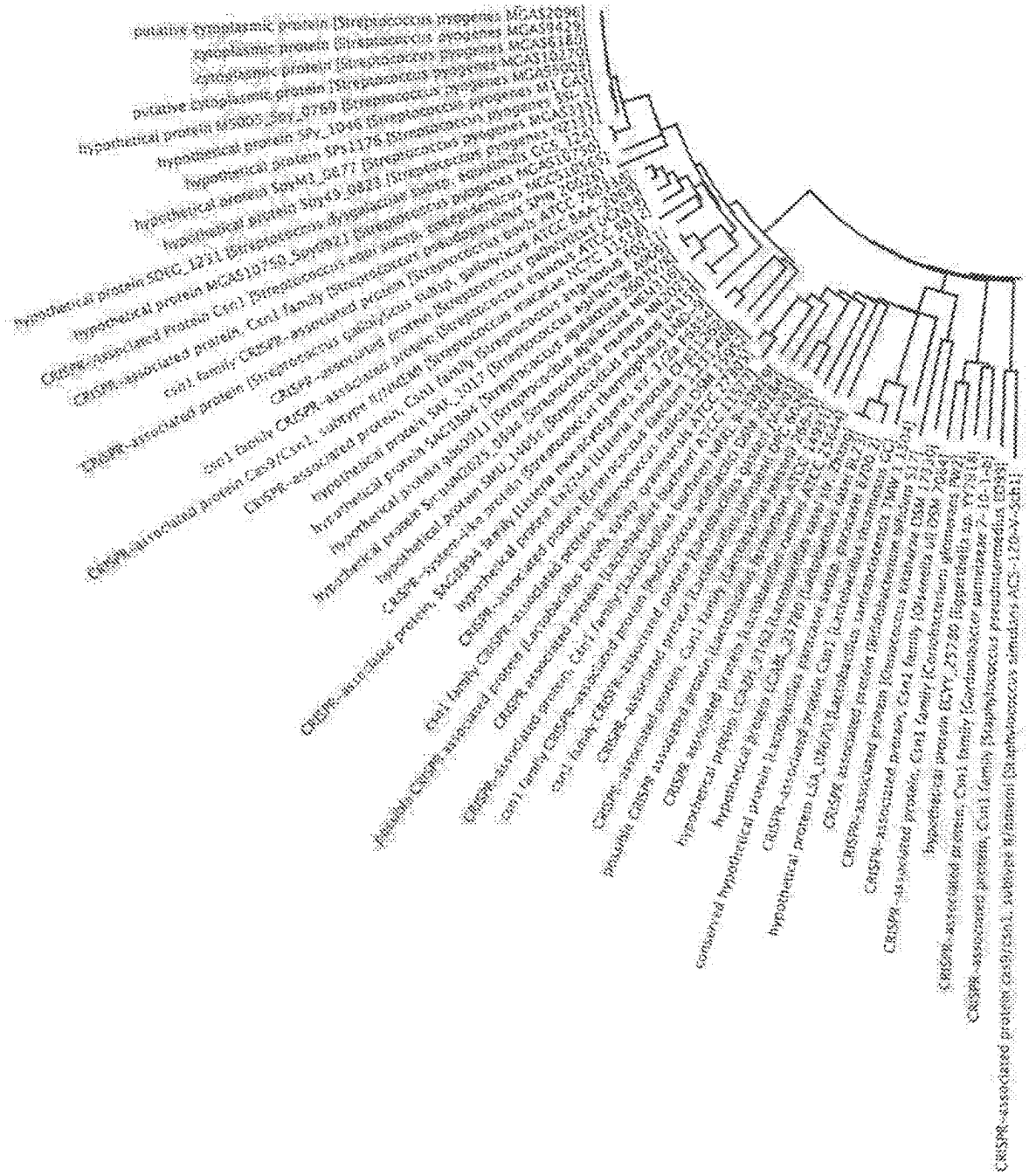


FIG. 19C

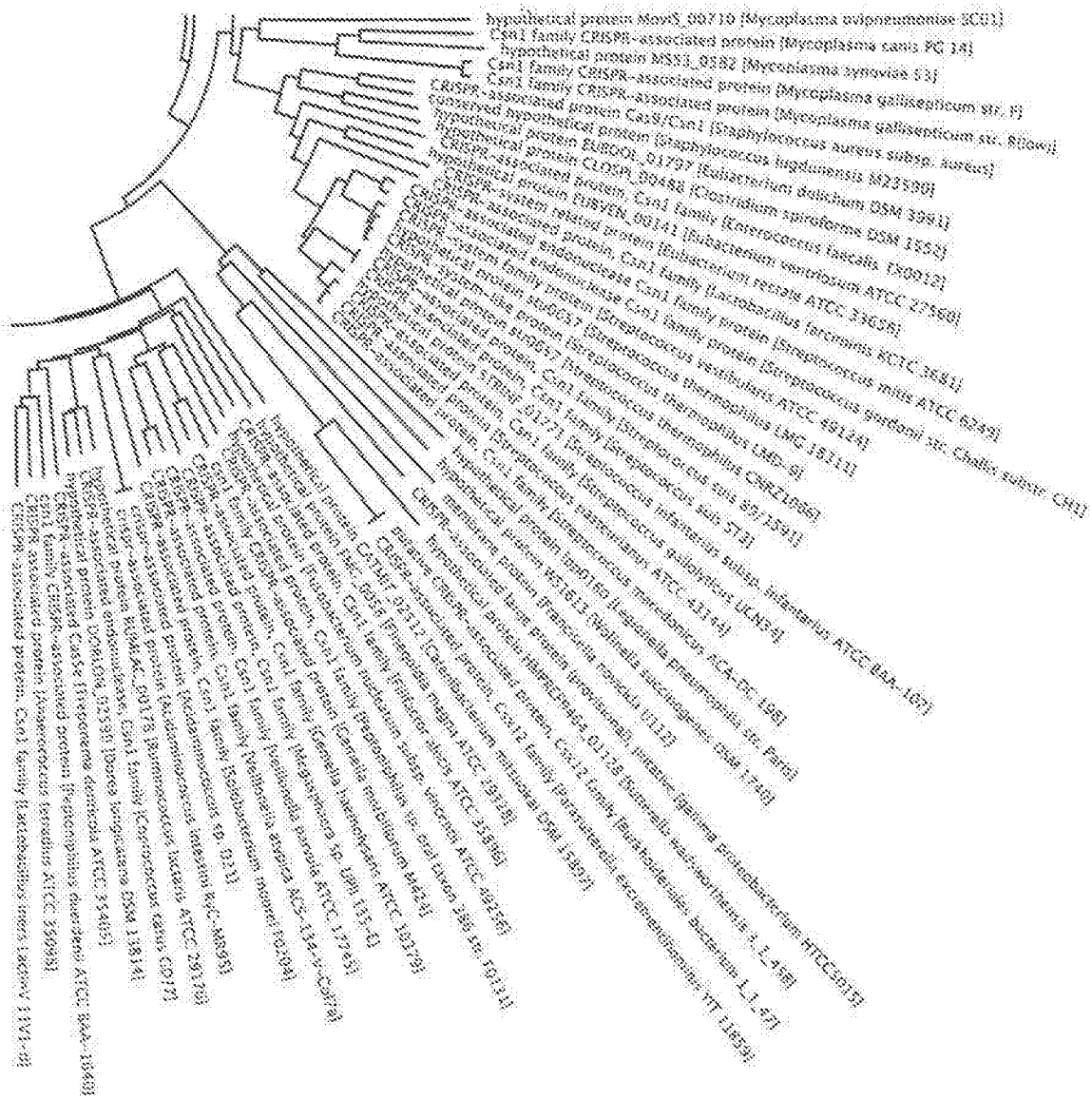

hypothetical protein MovS_00710 [Mycoplasma ovipneumoniae SC01]
Csn1 family CRISPR-associated protein [Mycoplasma canis PG 14]
hypothetical protein MSS3_05882 [Mycoplasma synoviae 53]
Csn1 family CRISPR-associated protein [Mycoplasma gallisepticum str. F]
CRISPR-associated protein Cas9/Csn1 [Mycoplasma gallisepticum str. Rlow]]
conserved hypothetical protein [Staphylococcus aureus subsp. aureus]
hypothetical protein EUBDOL_01797 [Eubacterium dolichum DSM 3991]
hypothetical protein CLOSPI_00488 [Clostridium spiroforme DSM 1552]
CRISPR-associated protein, Csn1 family [Enterococcus faecalis TX0012]
hypothetical protein EUBVEN_00141 [Eubacterium ventriosum ATCC 27560]
CRISPR-system related protein [Lactobacillus brevis ATCC 33656]
CRISPR-associated protein, Csn1 family [Streptococcus mitis ATCC 6249]
CRISPR-associated endonuclease Csn1 family protein [Streptococcus gordonii str. Challis substr. CH1]
CRISPR-associated protein Csn1 family protein [Streptococcus vestibularis ATCC 49124]
hypothetical protein CRISPR-associated protein [Streptococcus thermophilus LMG 18311]
CRISPR-associated protein [Streptococcus thermophilus CNRZ1066]
CRISPR-associated protein [Streptococcus suis 89/1391]
CRISPR-associated protein, Csn1 family [Streptococcus suis 573]
CRISPR-associated protein, Csn1 family [Streptococcus infantarius subsp. infantarius ATCC BAA-102]
CRISPR-associated protein, Csn1 family [Streptococcus gallolyticus UCN34]
hypothetical protein [Streptococcus mutans ATCC 43314]
hypothetical protein [Streptococcus pneumoniae str. Parid]
hypothetical protein [Streptococcus pneumoniae OXC141]
hypothetical protein [Streptococcus pneumoniae 670-6B]
CRISPR-associated protein, Csn1 family [Streptococcus pneumoniae SP195]
hypothetical protein [Streptococcus pneumoniae GA17545]
hypothetical protein [Streptococcus pneumoniae GA47368]
hypothetical protein [Streptococcus pneumoniae GA47901]
hypothetical protein [Streptococcus pneumoniae GA47922]
hypothetical protein [Streptococcus pneumoniae GA47923]
hypothetical protein [Streptococcus pneumoniae GA47928]
hypothetical protein [Streptococcus pneumoniae GA47929]
hypothetical protein [Streptococcus pneumoniae GA47930]
hypothetical protein [Streptococcus pneumoniae GA47931]
hypothetical protein [Streptococcus pneumoniae GA47932]
hypothetical protein [Streptococcus pneumoniae GA47933]
hypothetical protein [Streptococcus pneumoniae GA47934]
hypothetical protein [Streptococcus pneumoniae GA47935]
hypothetical protein [Streptococcus pneumoniae

FIG. 20A

FIG. 20B

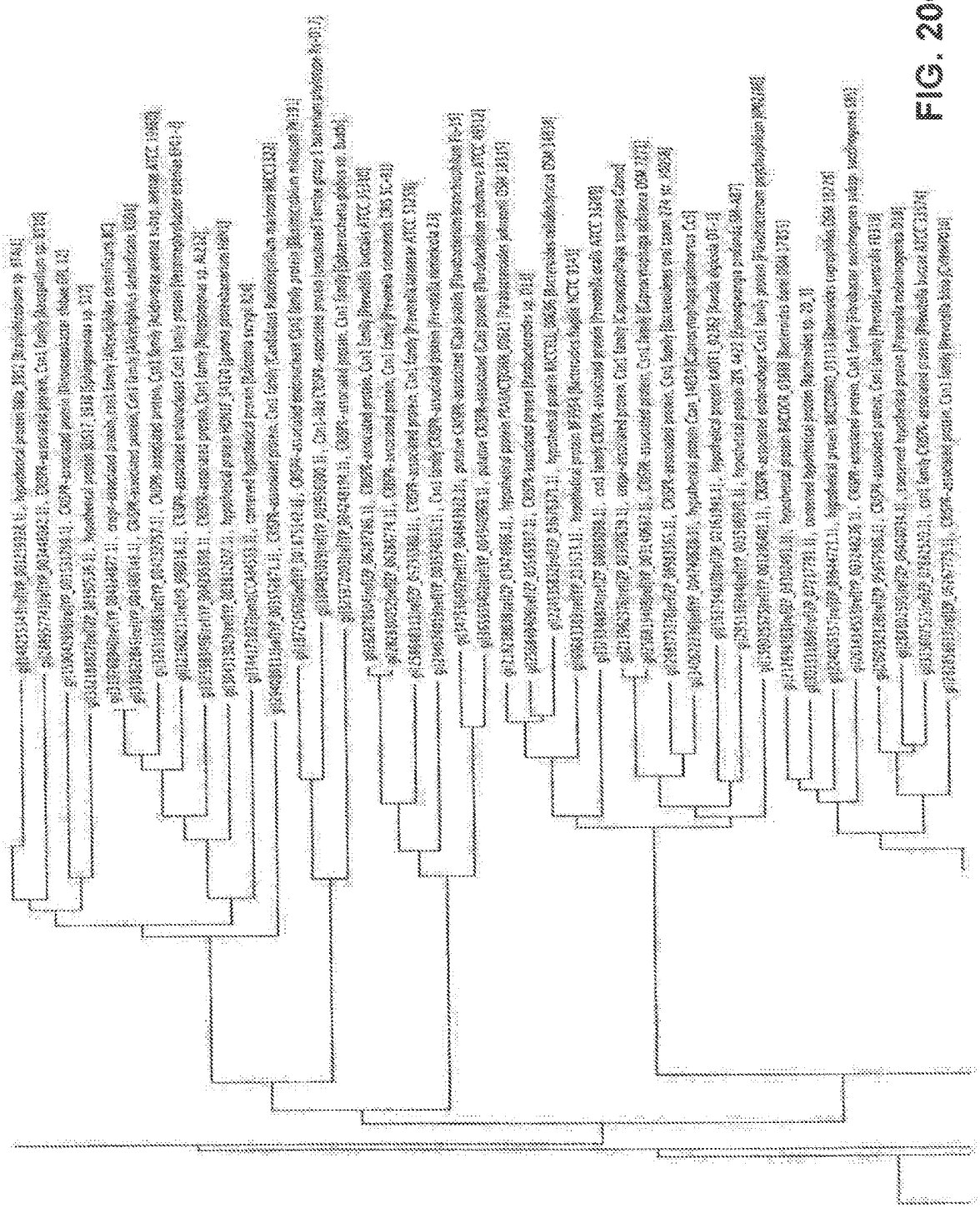


FIG. 20C

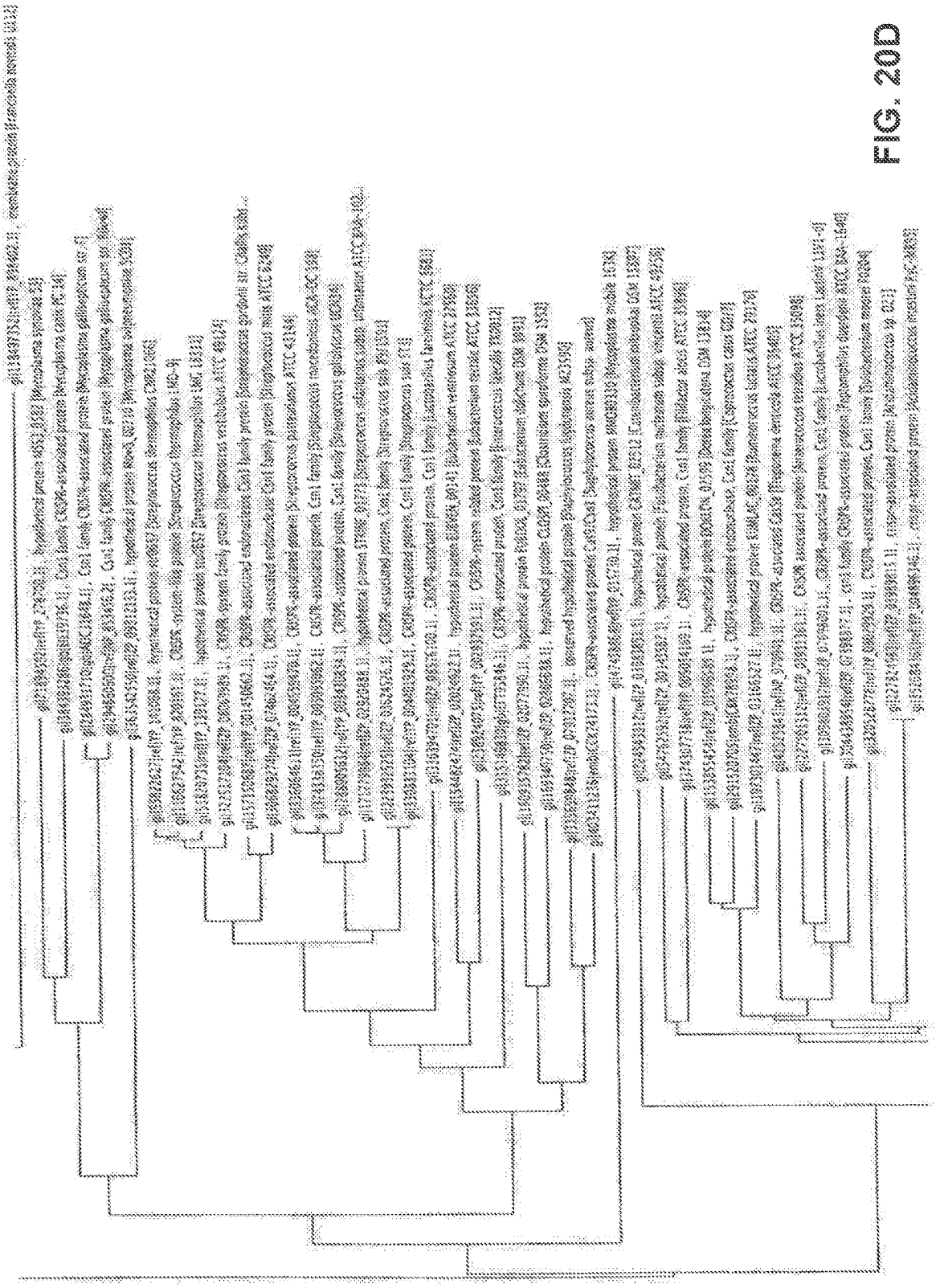


FIG. 20D

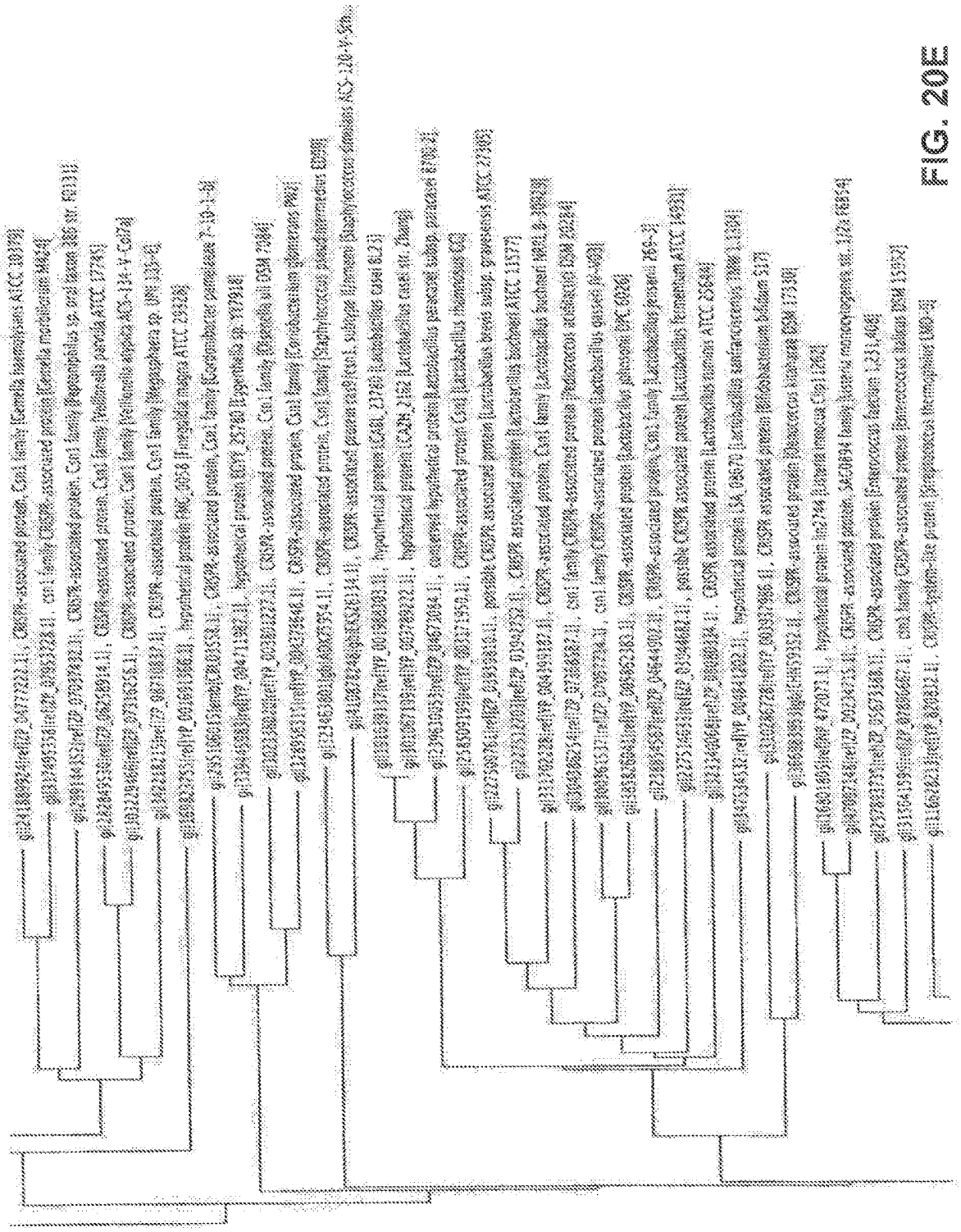


FIG. 20E

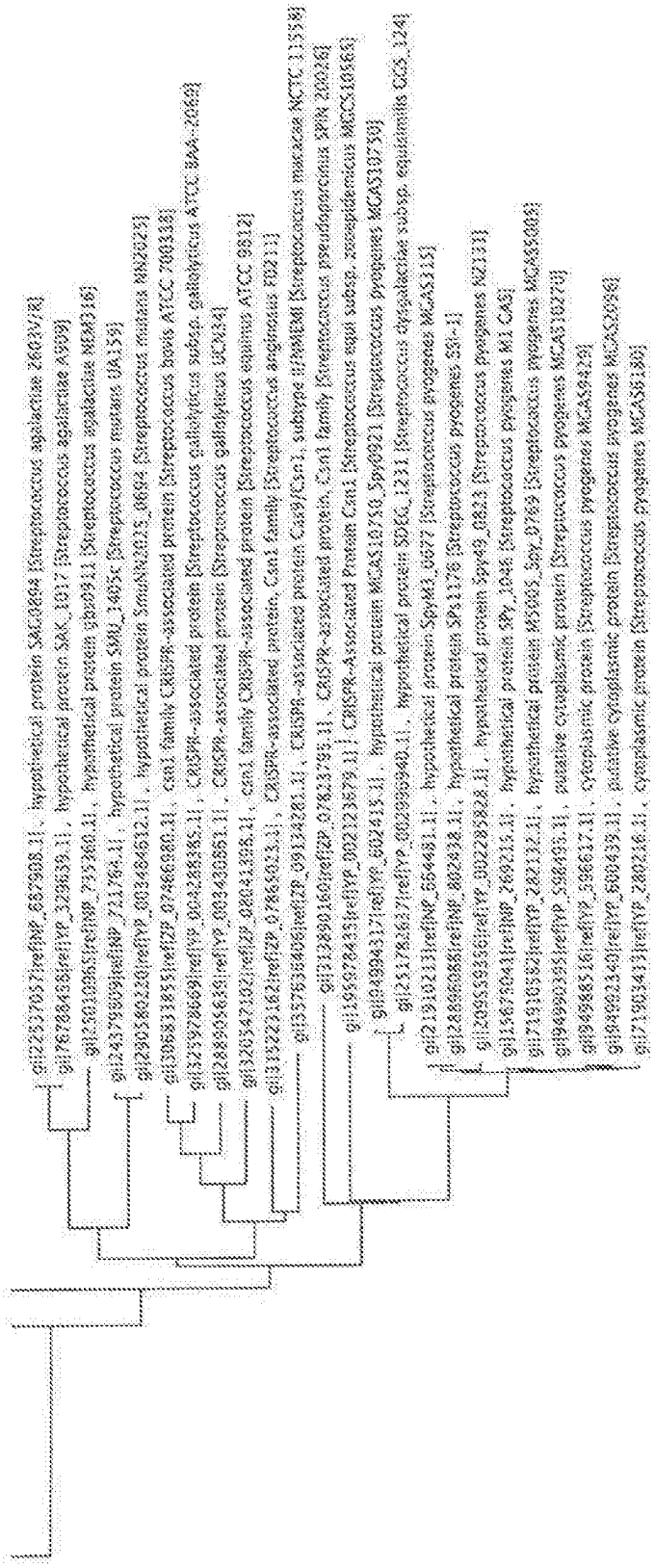

gi|225370517|ref|NP_687908.1| hypothetical protein SAG0864 [Streptococcus agalactiae 2603V/R]
gi|76788439|ref|YP_329839.1| hypothetical protein SAK_1017 [Streptococcus agalactiae A909]
gi|250103095|ref|NP_755360.1| hypothetical protein gbs0911 [Streptococcus agalactiae NEM316]
gi|243779809|ref|NP_711764.1| hypothetical protein SMU_1465c [Streptococcus mutans UA159]
gi|290580220|ref|YP_003484632.1| hypothetical protein SmuNN2025_0694 [Streptococcus mutans NN2025]
gi|306813855|ref|ZP_07466980.1| cas1 family CRISPR-associated protein [Streptococcus bovis ATCC 700338]
gi|323970669|ref|YP_004238385.1| CRISPR-associated protein [Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069]
gi|288905638|ref|YP_003430861.1| CRISPR-associated protein [Streptococcus gallolyticus UCN34]
gi|320547102|ref|ZP_02041398.1| cas1 family CRISPR-associated protein [Streptococcus equinus ATCC 9812]
gi|313223162|ref|ZP_07865013.1| CRISPR-associated protein, Cas1 family [Streptococcus anginosus F0211]
gi|335763840|ref|ZP_09134281.1| CRISPR-associated protein Cas9/Csn1, subtype II/NMEM [Streptococcus massiae NCTC 11558]
gi|312890160|ref|ZP_07823798.1| CRISPR-associated protein, Cas1 family [Streptococcus pseudporcinus SPIN 20026]
gi|169978433|ref|YP_002123879.1| CRISPR-Associated Protein Cas1 [Streptococcus suis subsp. zooepidemicus MGCS10565]
gi|289894317|ref|YP_002415.1| hypothetical protein MGAS10750_Spy0921 [Streptococcus pyogenes MGAS10750]
gi|251782627|ref|YP_002289940.1| hypothetical protein SDEG_1231 [Streptococcus dysgalactiae subsp. equisimilis GGS_124]
gi|219102331|ref|NP_594481.1| hypothetical protein SpyM3_0677 [Streptococcus pyogenes MGAS315]
gi|289890088|ref|NP_802938.1| hypothetical protein SPs1176 [Streptococcus pyogenes SSI-1]
gi|209559336|ref|YP_002285928.1| hypothetical protein Spy49_0821 [Streptococcus pyogenes NZ131]
gi|156786941|ref|NP_269231.1| hypothetical protein SPy_1046 [Streptococcus pyogenes M1 GAS]
gi|71810562|ref|YP_282112.1| hypothetical protein MSM05_Spy_0769 [Streptococcus pyogenes MGAS6018]
gi|94990235|ref|YP_598491.1| putative cytoplasmic protein [Streptococcus pyogenes MGAS10270]
gi|94988516|ref|YP_596127.1| cytoplasmic protein [Streptococcus pyogenes MGAS9429]
gi|94989314|ref|YP_600438.1| putative cytoplasmic protein [Streptococcus pyogenes MGAS2096]
gi|71903413|ref|YP_280216.1| cytoplasmic protein [Streptococcus pyogenes MGAS6180]

FIG. 20F

a

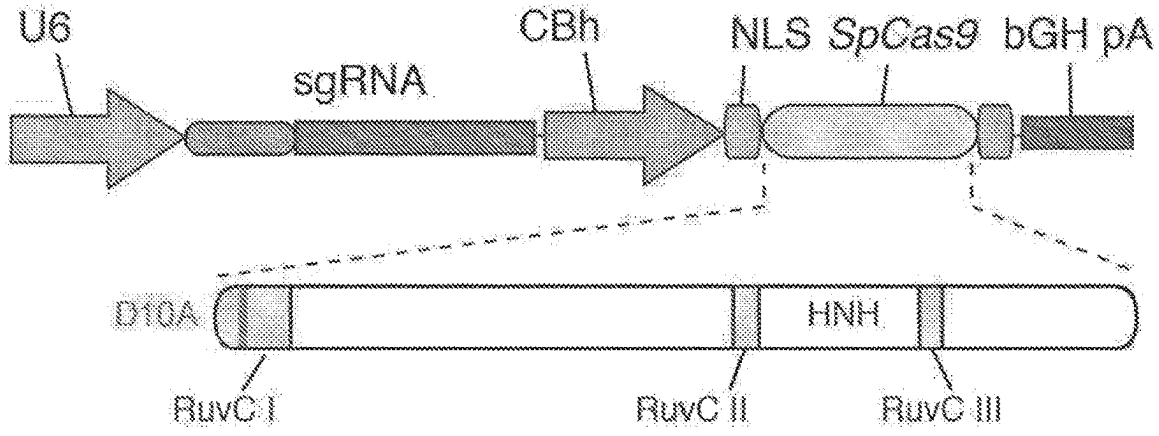


FIG. 21A

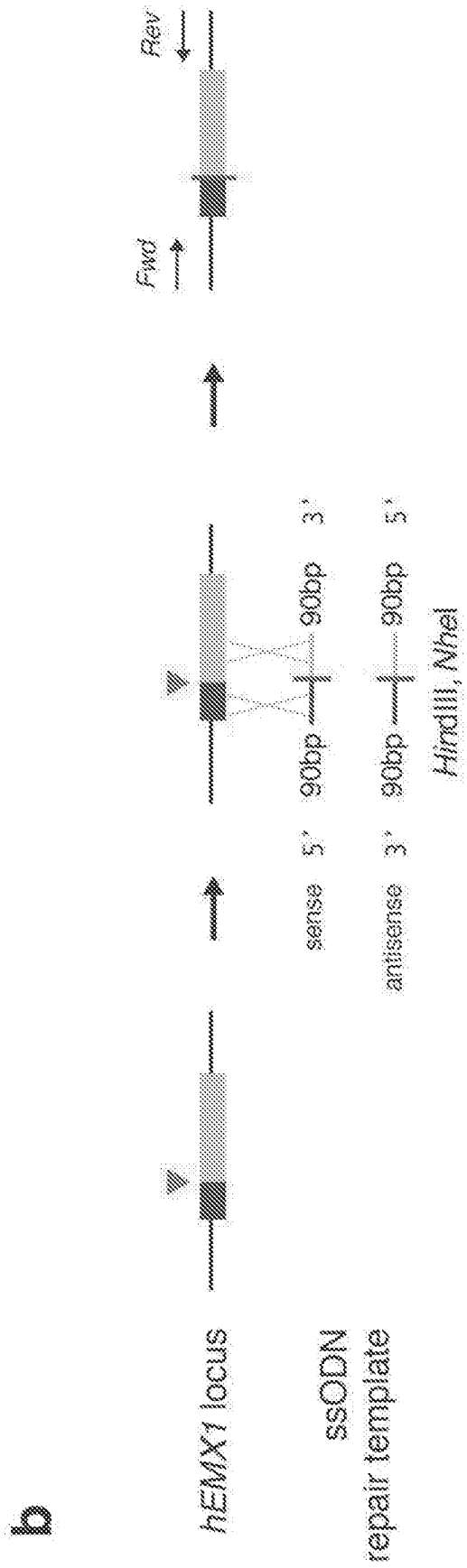

b
hEMX1 locus
ssODN
repair template
sense
5'
90bp
3'
antisense
3'
90bp
5'
HindIII, NheI
Fwd
Rev

FIG. 21B

C

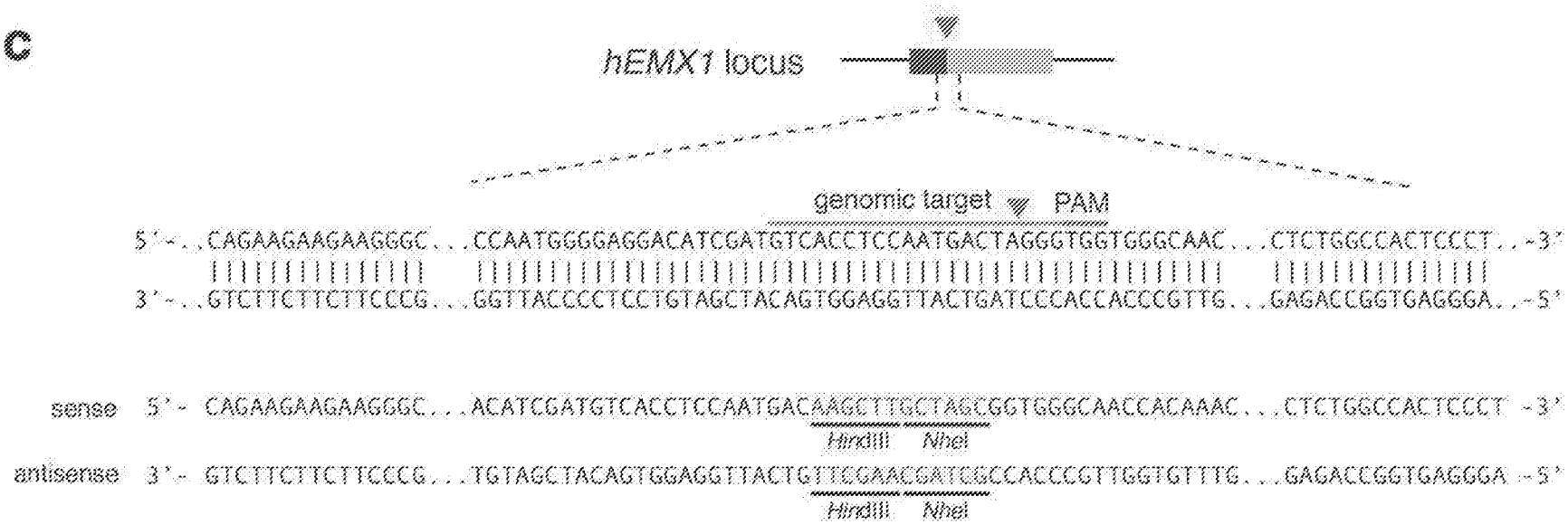

hEMX1 locus
genomic target
PAM
5'-..CAGAAGAAGAAGGGC...CCAATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGTGGGCAAC...CTCTGGCCACTCCCT..-3'
3'-..GTCTTCTTCTTCCCG...GGTTACCCCTCCTGTAGCTACAGTGGAGGTTACTGATCCCACCACCCGTTG...GAGACCGGTGAGGGA..-5'
sense 5'- CAGAAGAAGAAGGGC...ACATCGATGTCACCTCCAATGACAAGCTTGCTAGCGGTGGGCAACCACAAAC...CTCTGGCCACTCCCT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

FIG. 21C

d

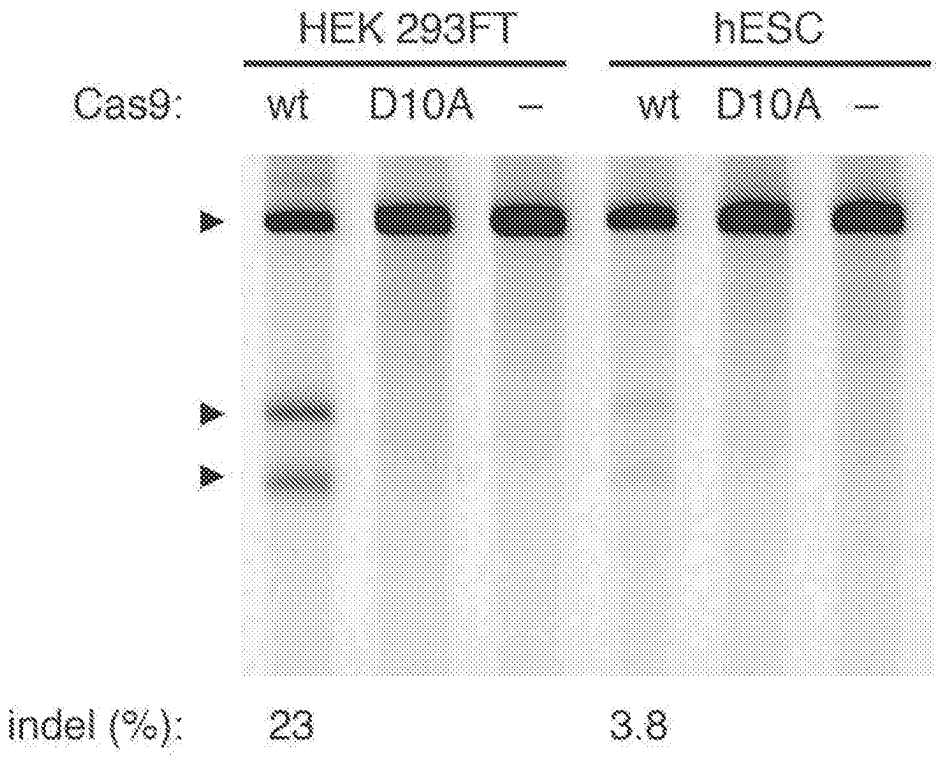


FIG. 21D

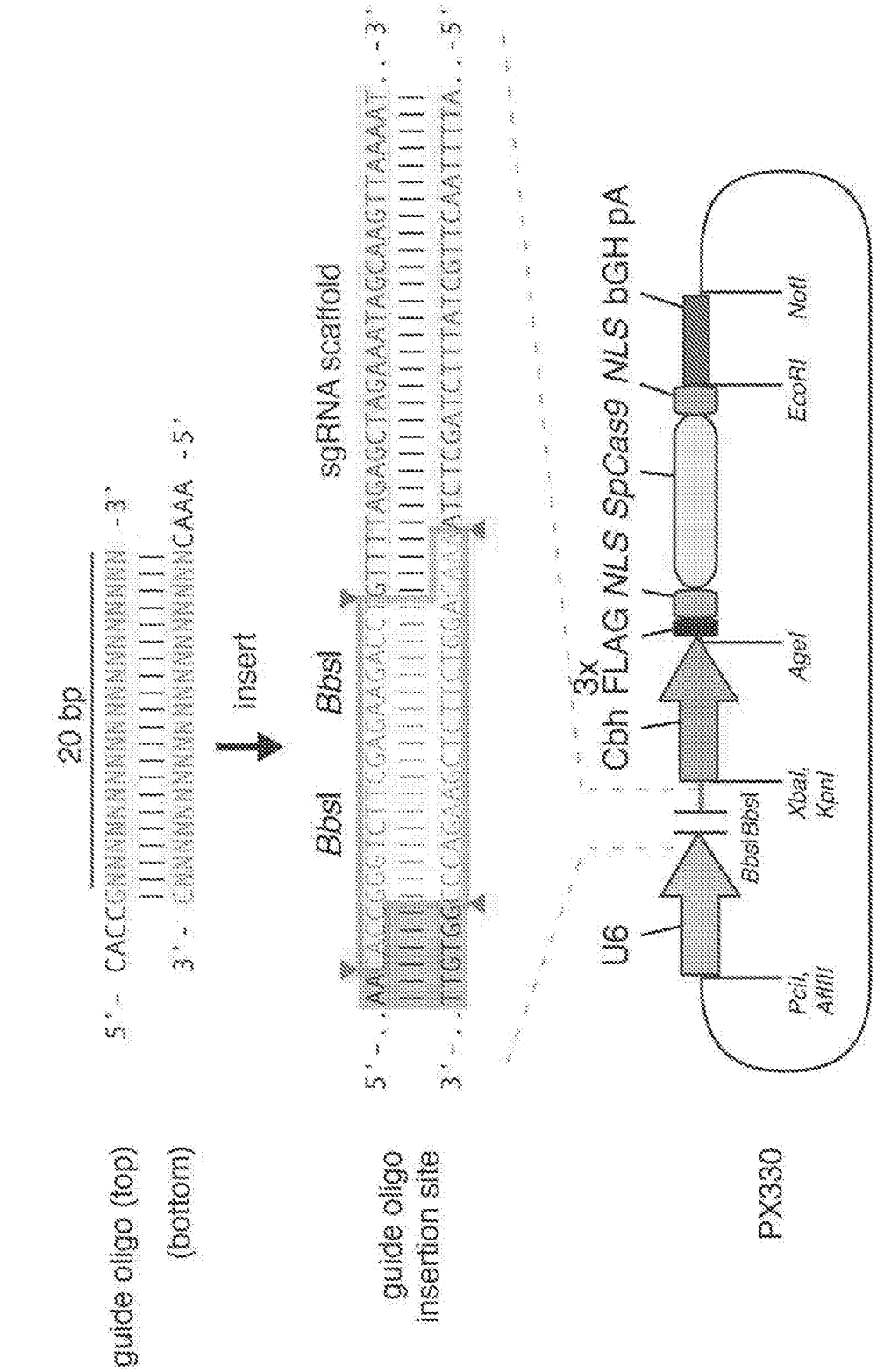

a
guide oligo (top)
(bottom)
5' - CACC
3' - CAAA
20 bp
5' -
3' -
-3'
-5'
insert
guide oligo
insertion site
BbsI
BbsI
sgRNA scaffold
5' -
3' -
-3'
-5'
PX330
U6
3x
Cbh FLAG NLS SpCas9 NLS bGH pA
PciI,
AflIII
BbsI BbsI
XbaI,
KpnI
AgeI
EcoRI
NotI

FIG. 22A

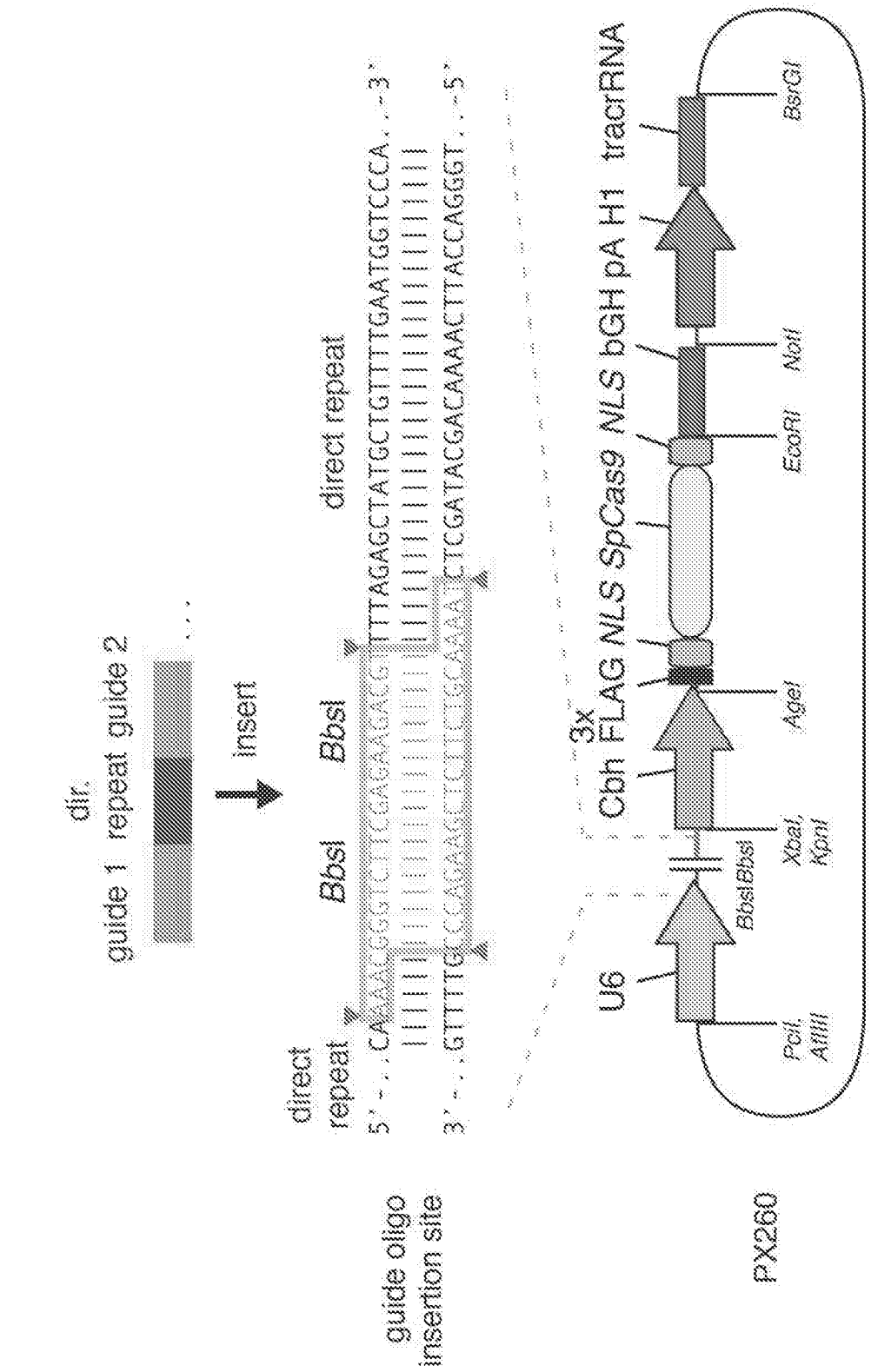

b
guide oligo insertion site
direct repeat
5' - . . CAAACGGTCTTCGACAACG
3' - . . GTTTTGCCAGAAGCTCTTCGCAAAA
BbsI
BbsI
insert
dir. repeat
guide 1
guide 2
. . .
direct repeat
TTTAGAGCTATGCTGTTTTGAATGGTCCCA. . -3'
CTCGATACGACAAAACTTACCAGGGT. . -5'
PX260
U6
Cbh
3x FLAG NLS SpCas9 NLS bGH pA H1 tracrRNA
PciI,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

FIG. 22B

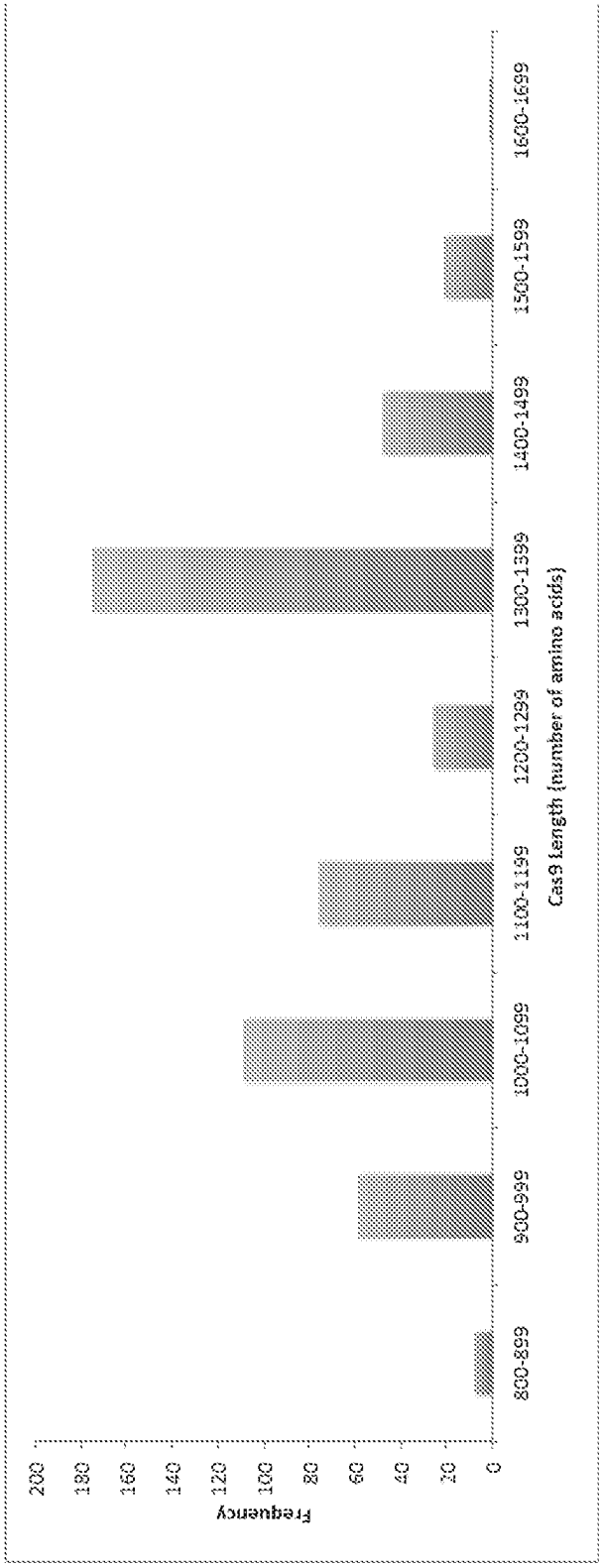

Frequency
200
180
160
140
120
100
80
60
40
20
0
800-899
900-999
1000-1099
1100-1199
1200-1299
1300-1399
1400-1499
1500-1599
1600-1699
Cas9 Length (number of amino acids)

FIG. 23

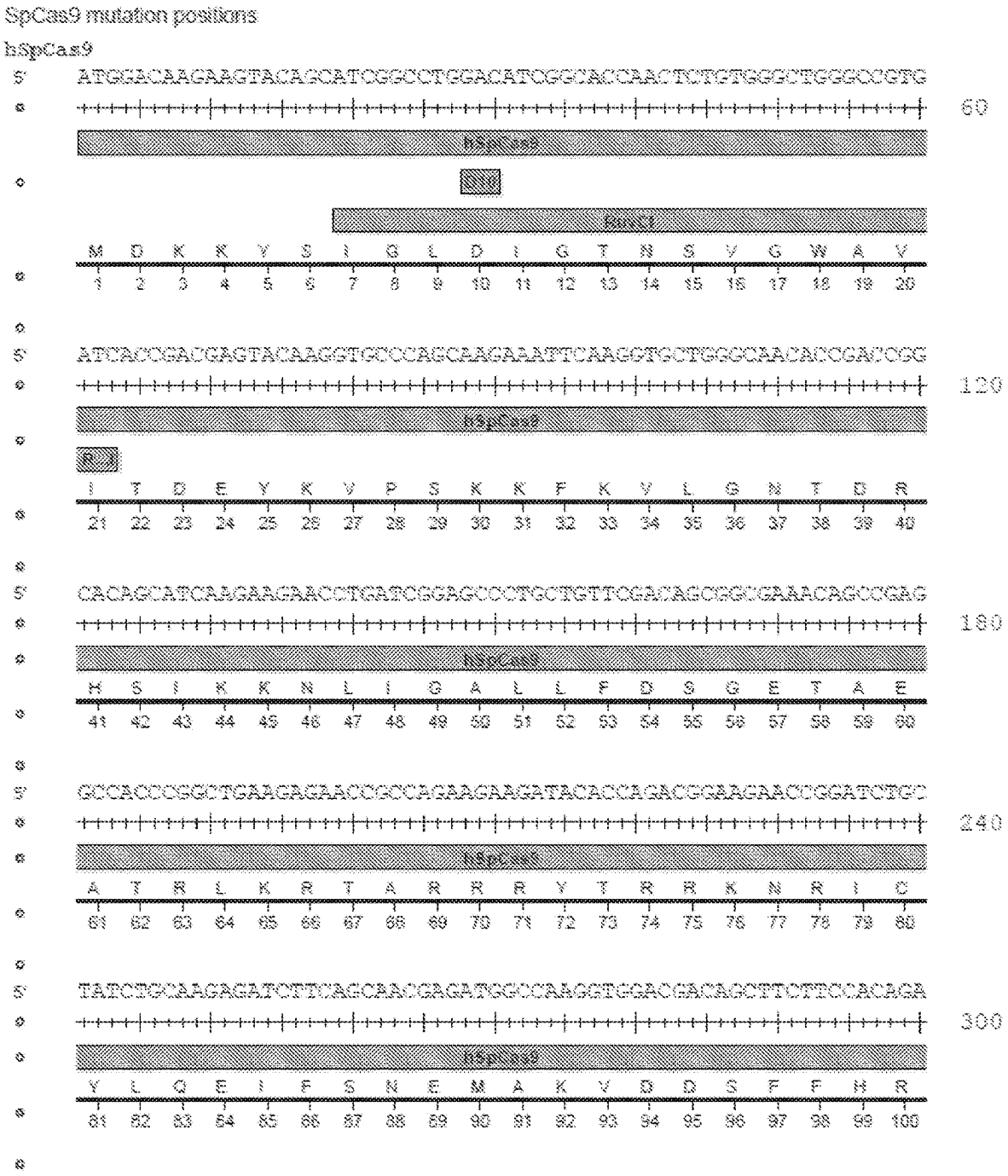

SpCas9 mutation positions
hSpCas9
5' ATGGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
60
hSpCas9
RuvC
M D K K Y S I G L D I G T N S V G W A V
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
5' ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGG
120
hSpCas9
R 3
I T D E Y K V P 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' CGGCTGGAAAATCTGATCGCCCAGCTGCCCGGGGAGAAGAAGAATGGCCTGTTCGGCAAC 720

NSpCas9

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

NSpCas9

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

NSpCas9

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

NSpCas9

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

NSpCas9

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

NSpCas9

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

481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500

5' GTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG 1500

hSpCas9

V V D K G A S A Q S F I E R M T N F D K

501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520

5' AACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTG 1560

hSpCas9

N L P N E K V L P K H S L L Y E Y F T V

521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540

5' TATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTG 1620

hSpCas9

Y N E L T K V K Y V T E G M R K P A F L

541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560

5' AGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACC 1680

hSpCas9

S G E Q K K A I V D L L F K T N R K V T

561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580

5' GTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATC 1740

hSpCas9

V K Q L K E D Y F K K I E C F D S V E I

581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600

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

BsrG I

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

BsrG I

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

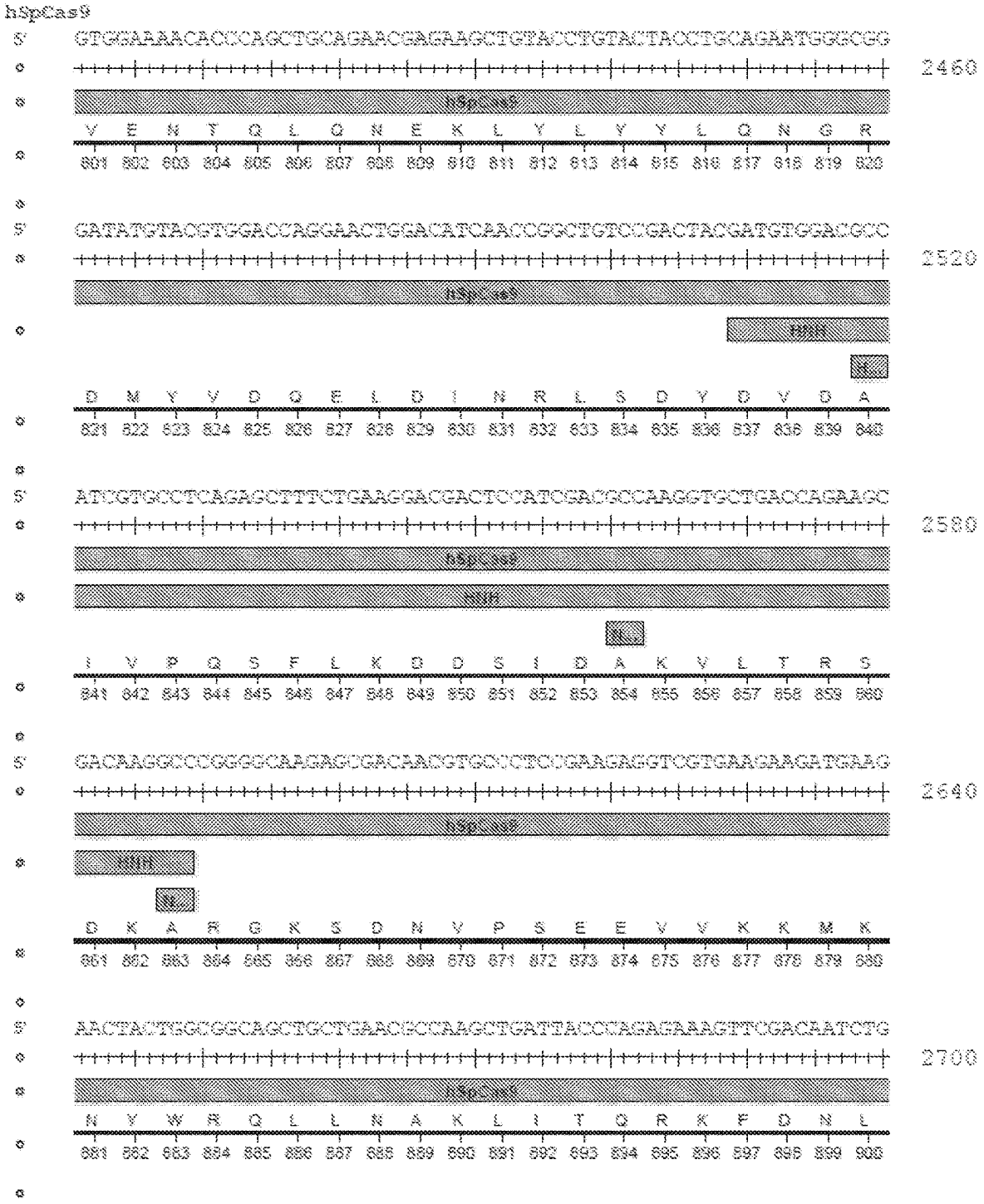

hSpCas9
5' GTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGG
2460
hSpCas9
V E N T Q L Q N E K L Y L Y Y L Q N G R
801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820
5' GATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCC
2520
hSpCas9
HNH
N
D M Y V D Q E 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' GACAAGGCCCCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAG
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

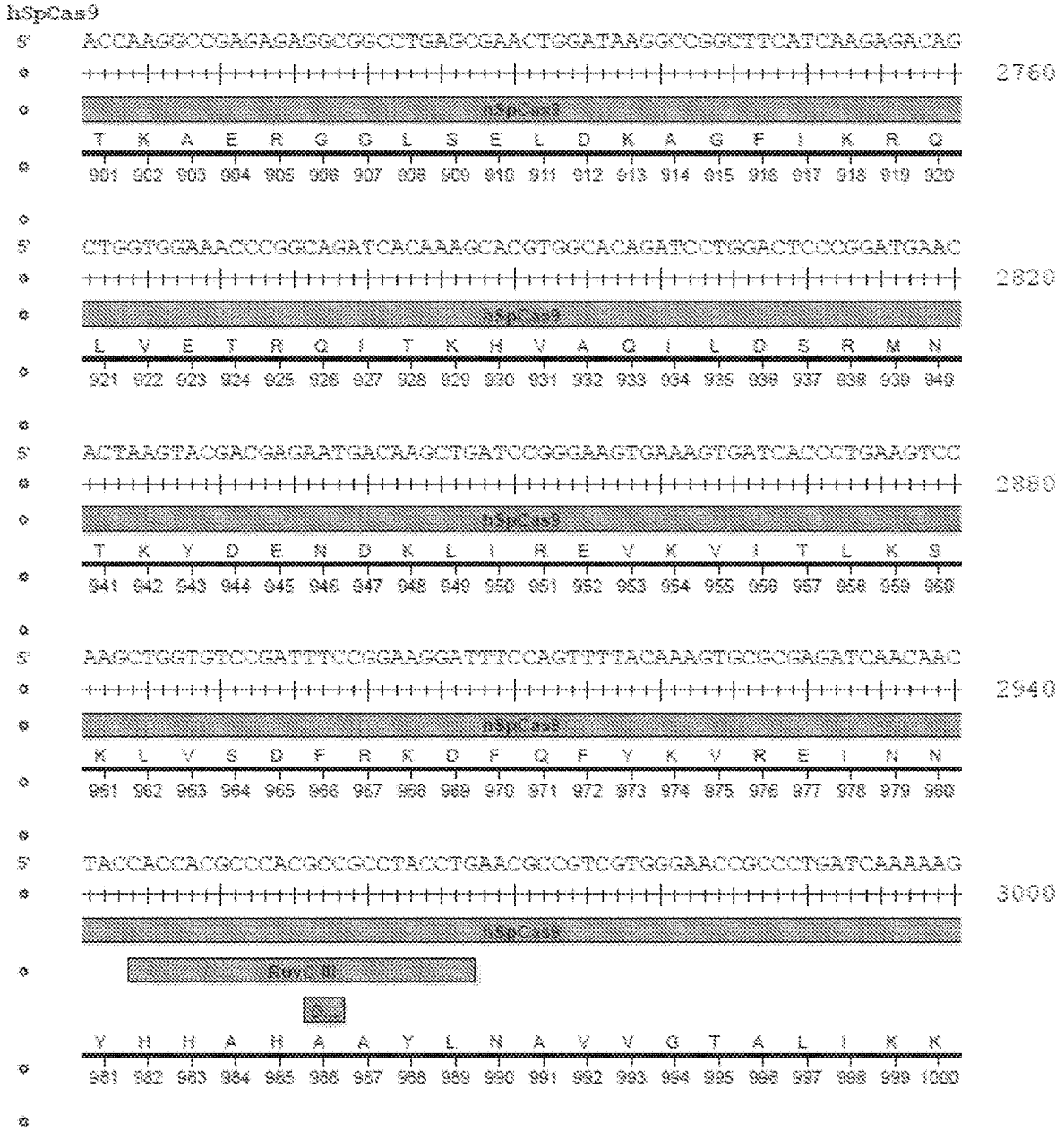

hSpCas9
5' ACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAG
2760
hSpCas9
T K A E R G G L S E L D K A G F I K R Q
901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920
5' CTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAAC
2820
hSpCas9
L V E T R Q I T K H 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' AAGCTGGTGTCCGATTTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAAC
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
Rec1 N
Q
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 G 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' GAGGTGCTGGACGCCACCCTGATCCACCAGAGTATCACCGGCCTGTACGAGACACGGATC 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

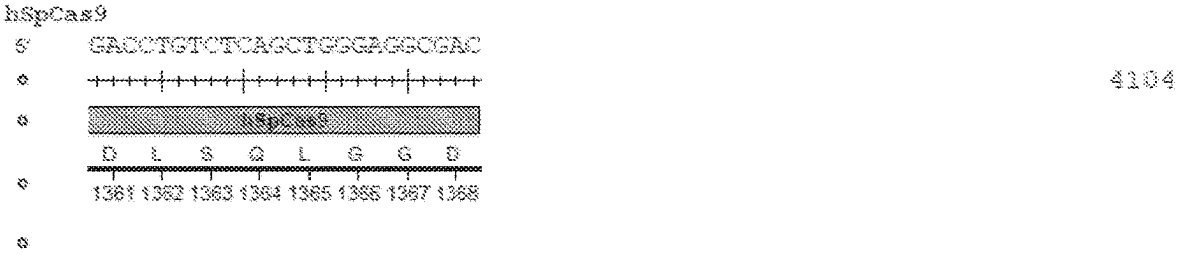

hSpCas9
5'
GACCTGTCTCAGCTGGGAGGCGAC
4104
hSpCas9
D L S Q L G G D
1361 1362 1363 1364 1365 1366 1367 1368

FIG. 24M

Conditional Cas9, Rosa26 targeting vector map

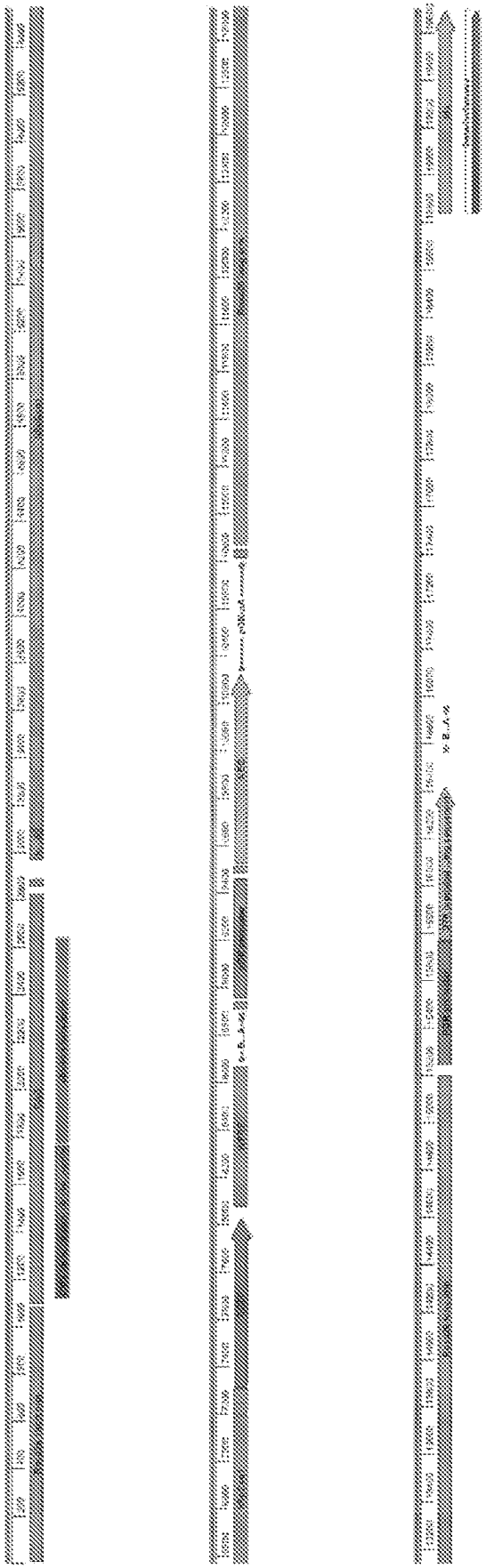
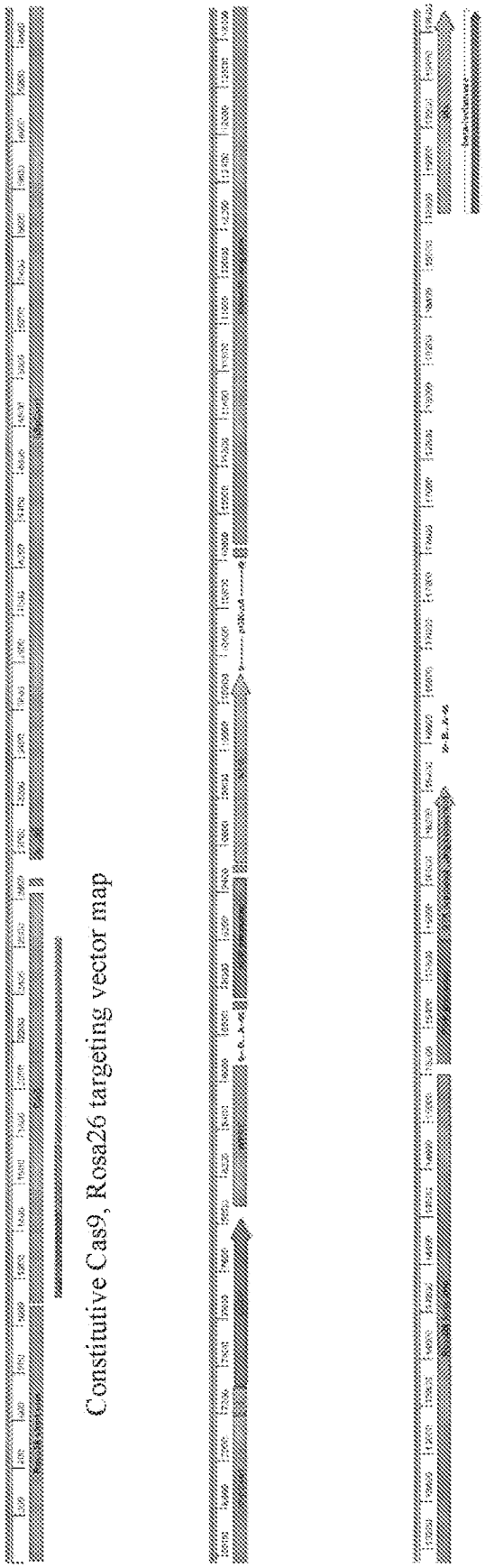


FIG. 25A



Constitutive Cas9, Rosa26 targeting vector map

FIG. 25B

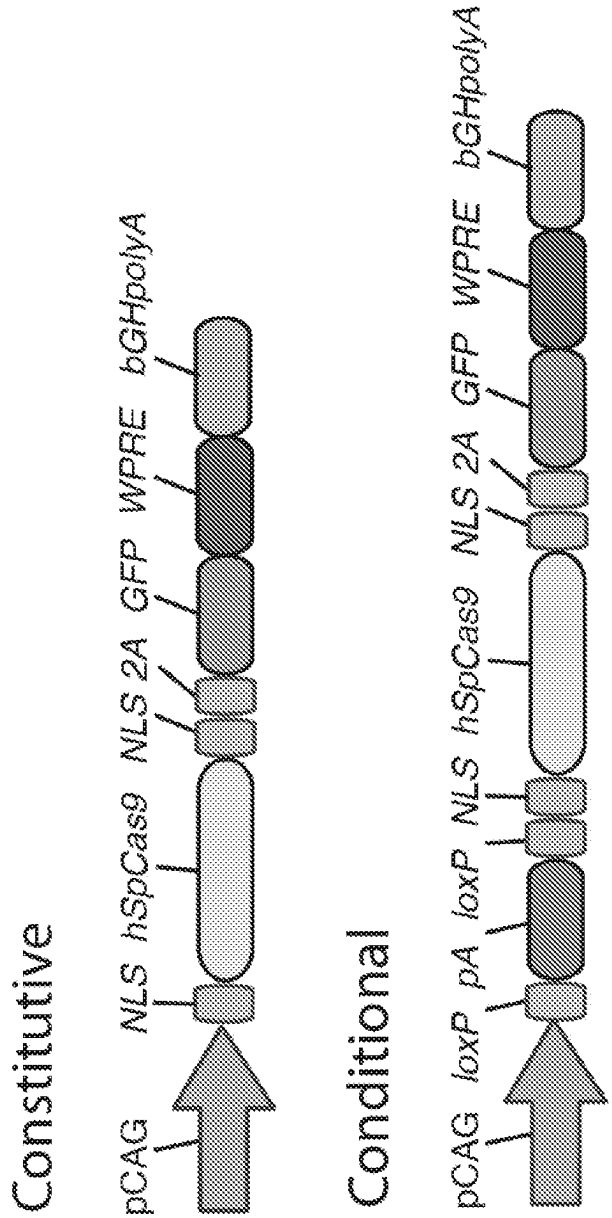
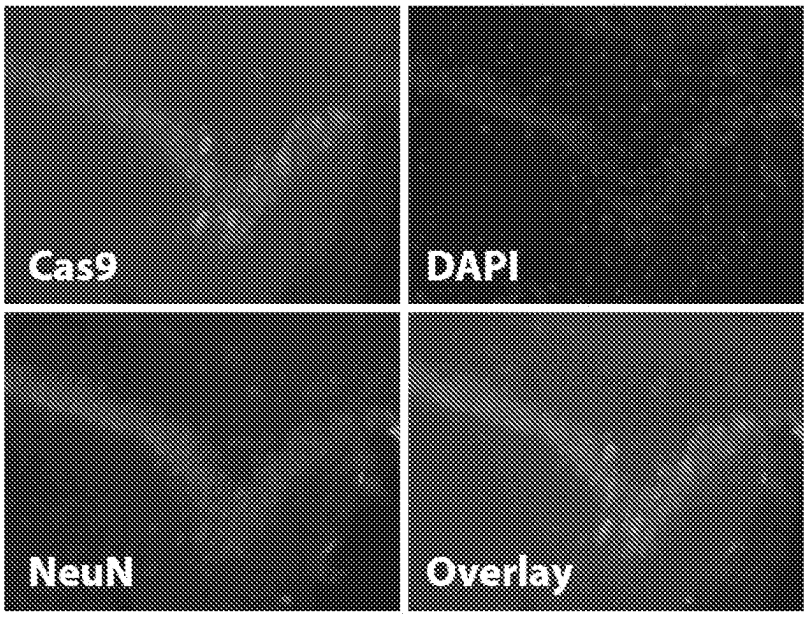

Constitutive
pCAG
NLS
hSpCas9
NLS
2A
GFP
WPRE
bGHpolyA
Conditional
pCAG
loxP
pA
loxP
NLS
hSpCas9
NLS
2A
GFP
WPRE
bGHpolyA

FIG. 26

Cas9 Expression in Mouse Hippocampus (AAV)



Cas9 Expression in Mouse Cortex (AAV)

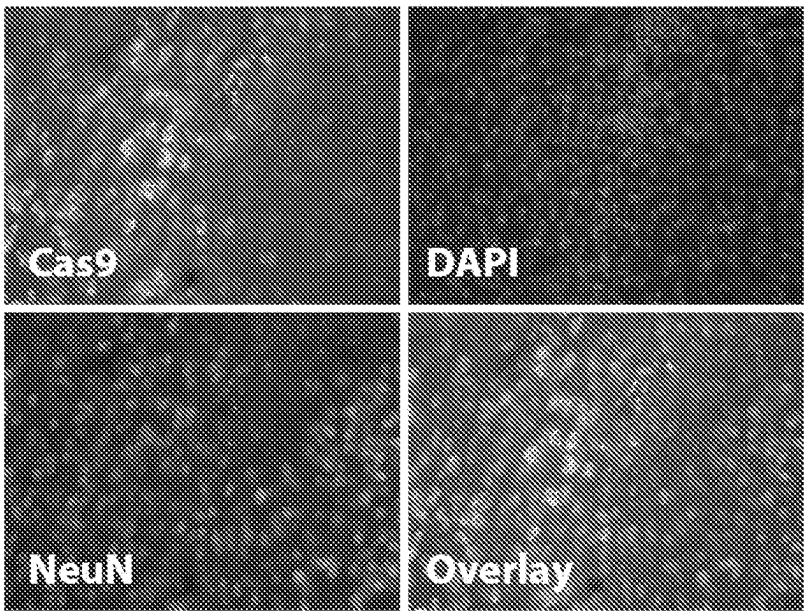


FIG. 27

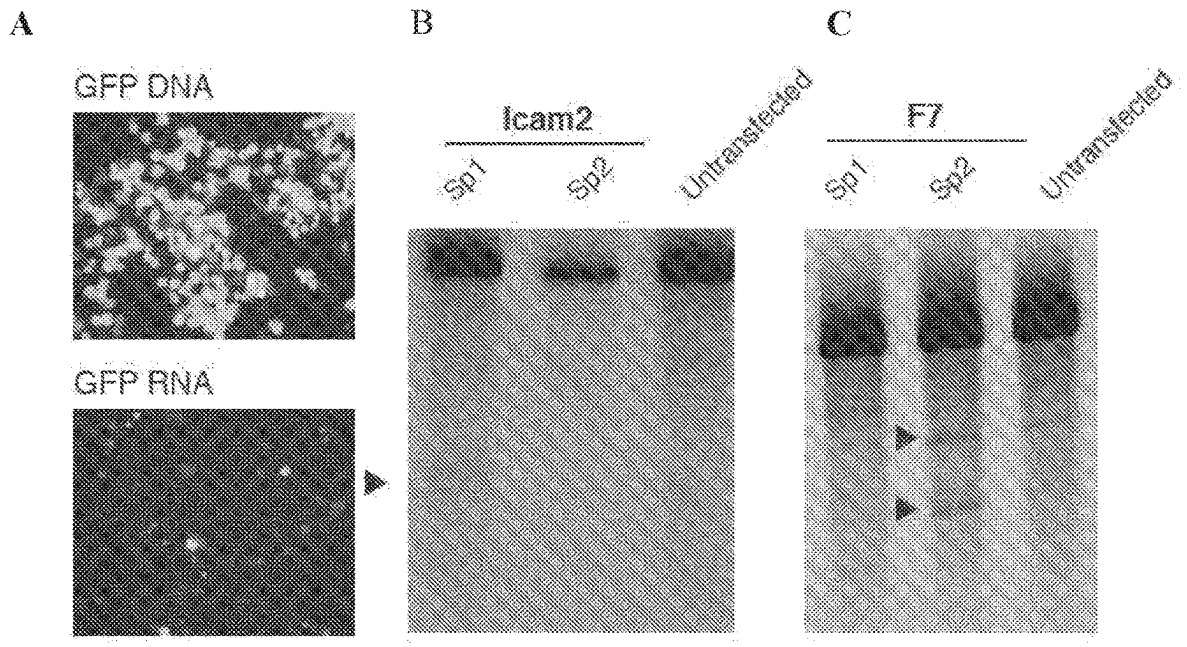

A
B
C
GFP DNA
GFP RNA
Icam2
Sp1
Sp2
Untransfected
F7
Sp1
Sp2
Untransfected

FIG. 28A-28C

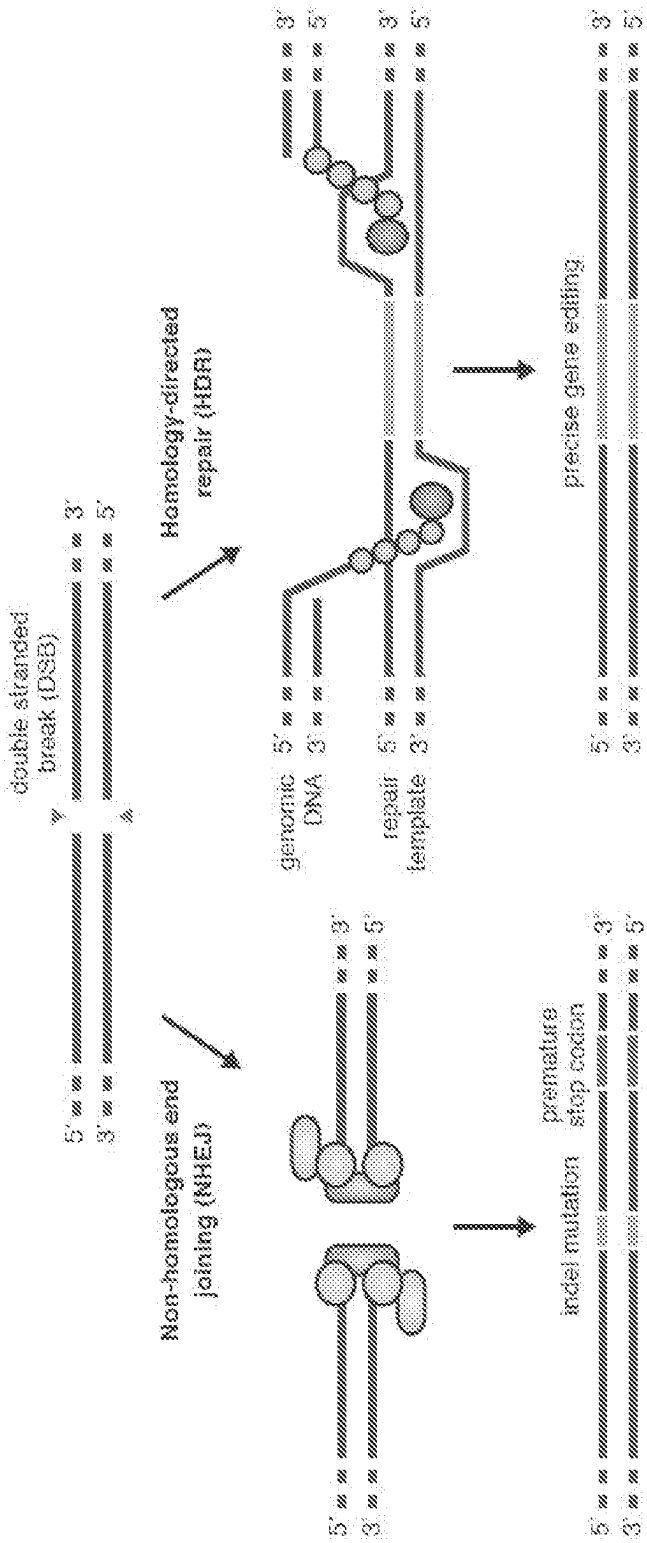

double stranded break (DSB)
5'
3'
3'
5'
Homology-directed repair (HDR)
genomic DNA
repair template
precise gene editing
Non-homologous end joining (NHEJ)
indel mutation
premature stop codon

FIG. 29

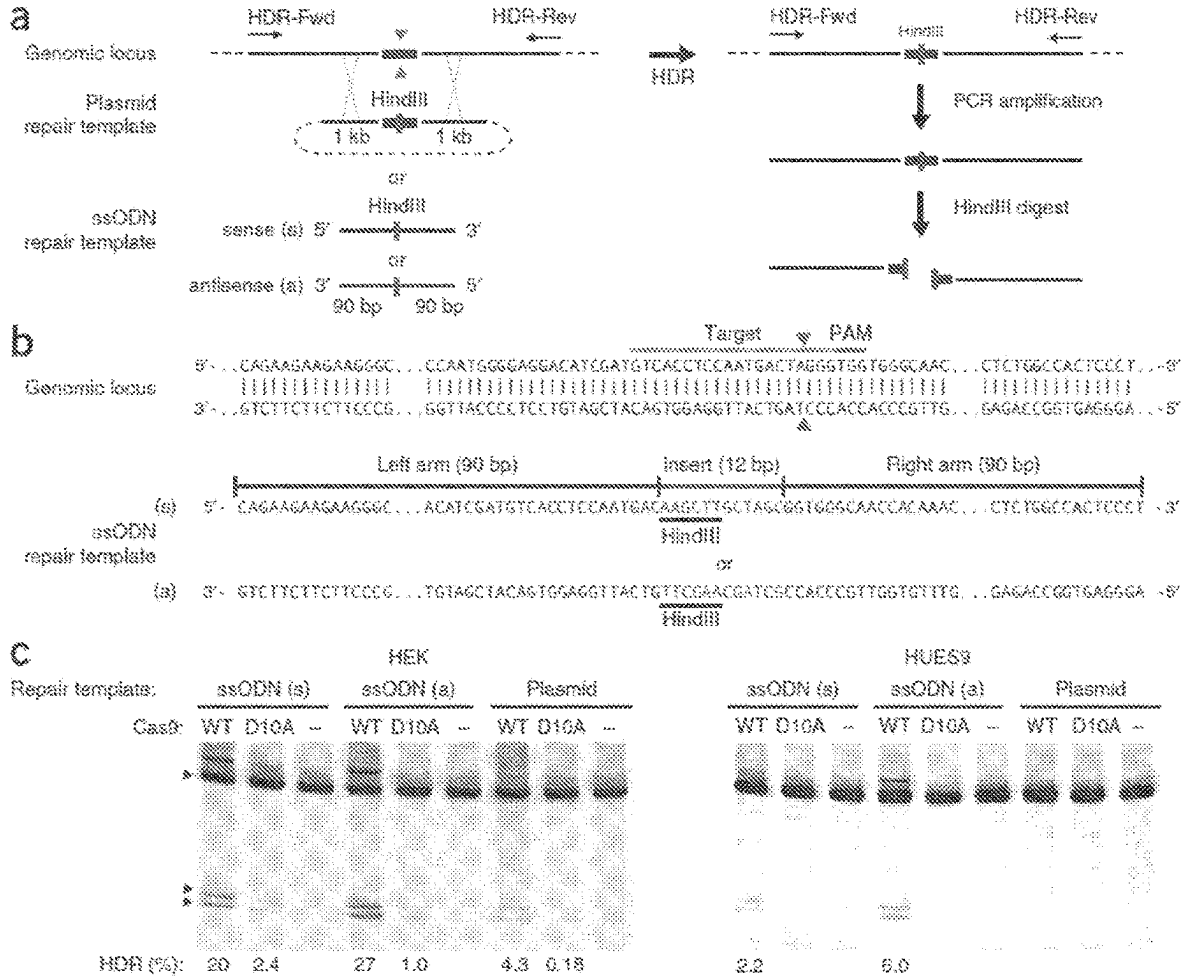

a
HDR-Fwd
HDR-Rev
Genomic locus
Plasmid repair template
HindIII
1 kb
1 kb
or
ssODN repair template
sense (a) 5'
3'
or
antisense (a) 3'
5'
90 bp
90 bp
HDR
HDR-Fwd
HindIII
HDR-Rev
PCR amplification
HindIII digest
b
Target
PAM
Genomic locus
Left arm (90 bp)
Insert (12 bp)
Right arm (90 bp)
(a)
ssODN repair template
HindIII
or
(a)
HindIII
c
HEK
HUES9
Repair template:
ssODN (a)
ssODN (a)
Plasmid
ssODN (a)
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-30C

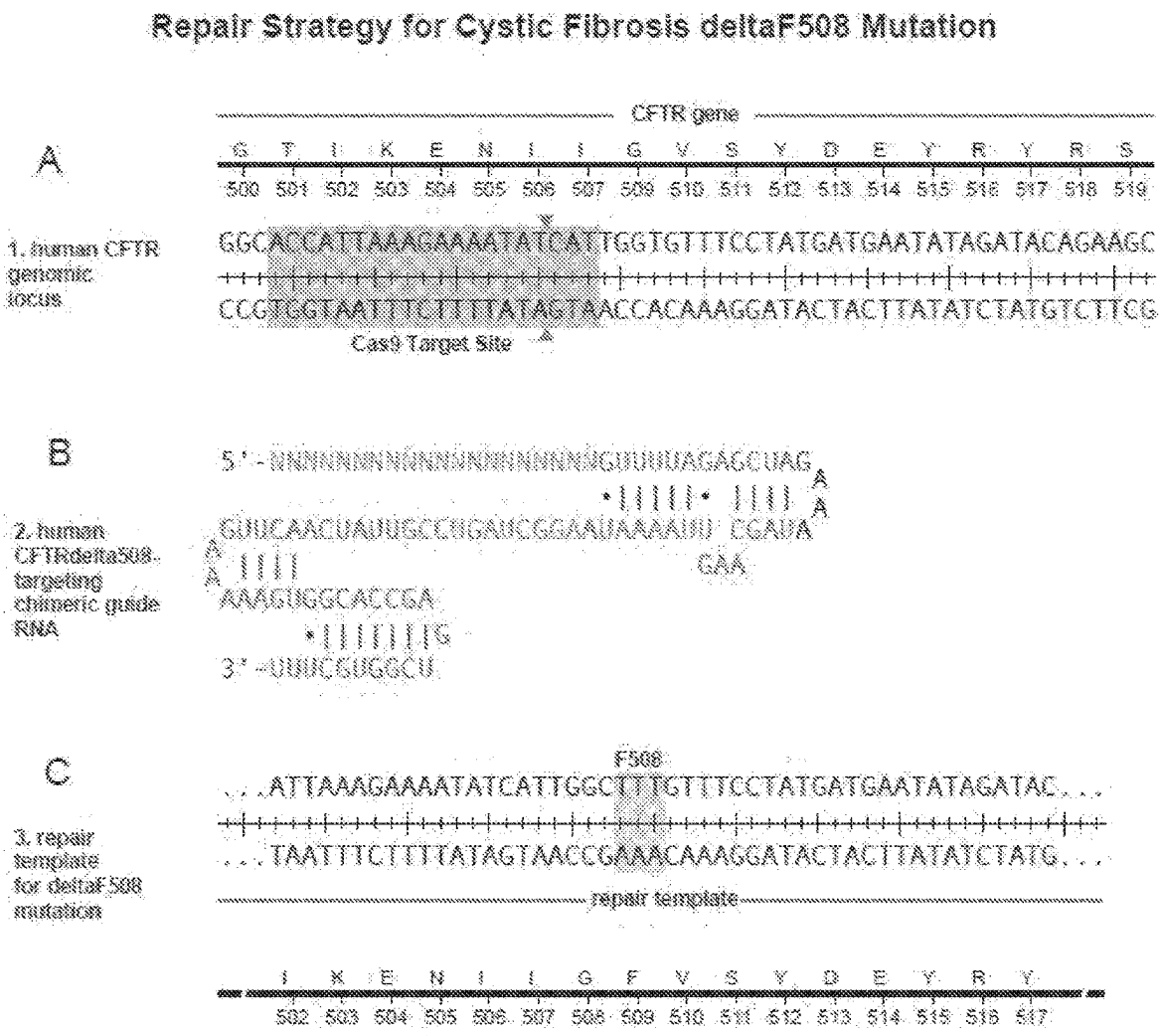
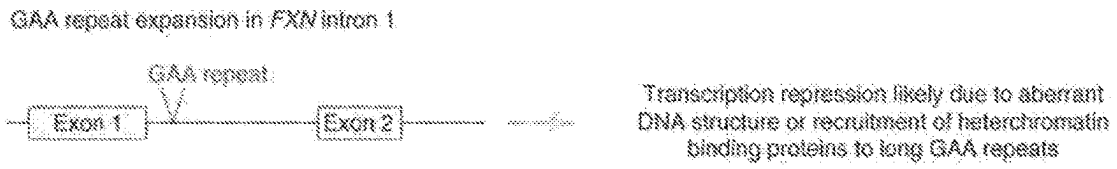

Repair Strategy for Cystic Fibrosis deltaF508 Mutation
A
CFTR gene
G T I K E N I I G V S Y D E Y R Y R S
500 501 502 503 504 505 506 507 509 510 511 512 513 514 515 516 517 518 519
1. human CFTR genomic locus
GGCACCATTAAAGAAAATATCATTGGTGTTTCCTATGATGAATATAGATACAGAAGC
CCGTGGTAATTTCTTTTATAGTAACCACAAAGGATACTACTTATATCTATGTCTTCG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

FIG. 31A-31C

A



B

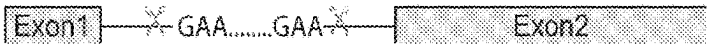


FIG. 32A-32B

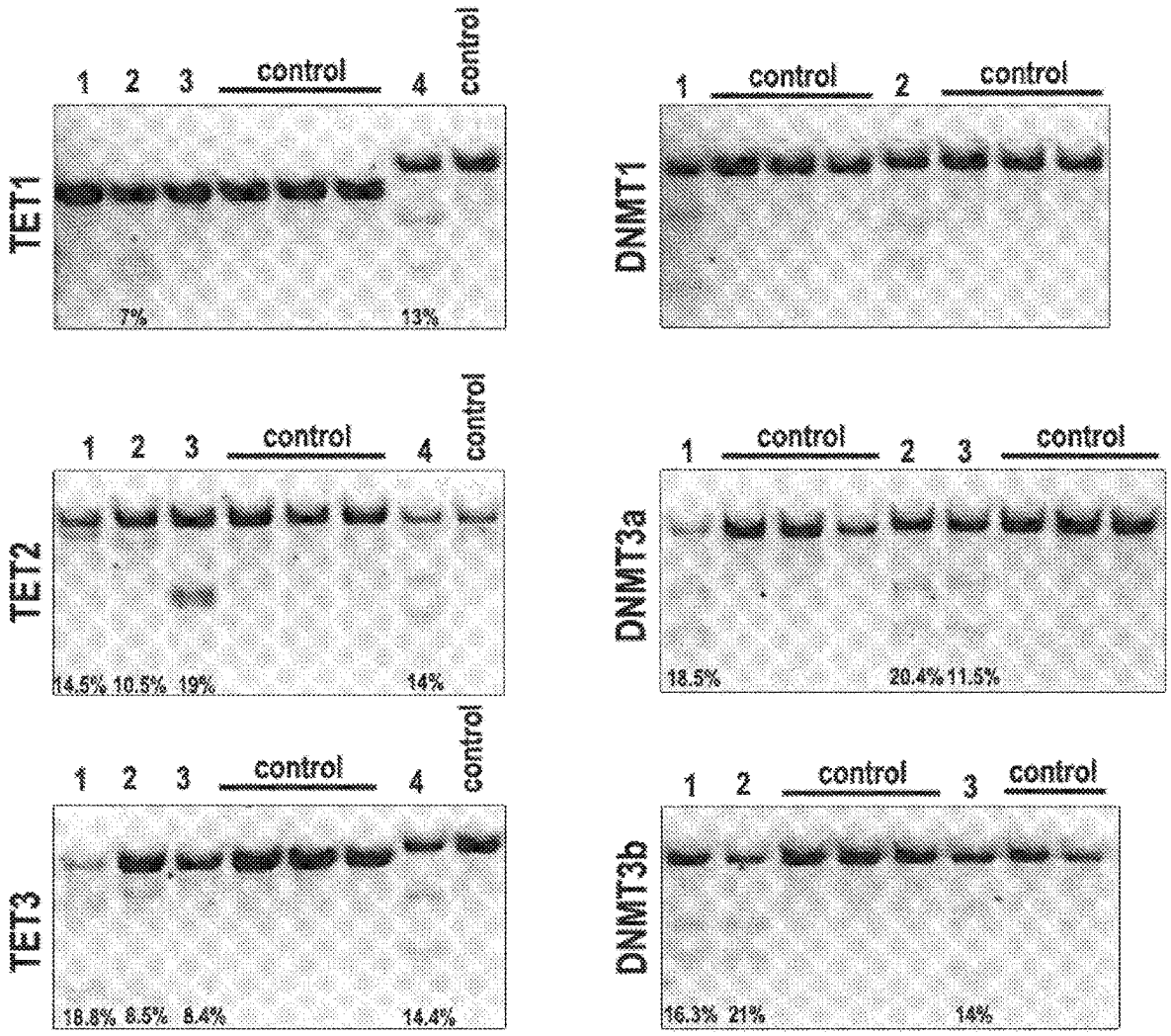

TET1
1 2 3 control 4 control
7%
13%
DNMT1
1 control 2 control
TET2
1 2 3 control 4 control
14.5% 10.5% 19%
14%
DNMT3a
1 control 2 3 control
18.5%
20.4% 11.5%
TET3
1 2 3 control 4 control
18.8% 8.5% 8.4%
14.4%
DNMT3b
1 2 control 3 control
16.3% 21%
14%

FIG. 33

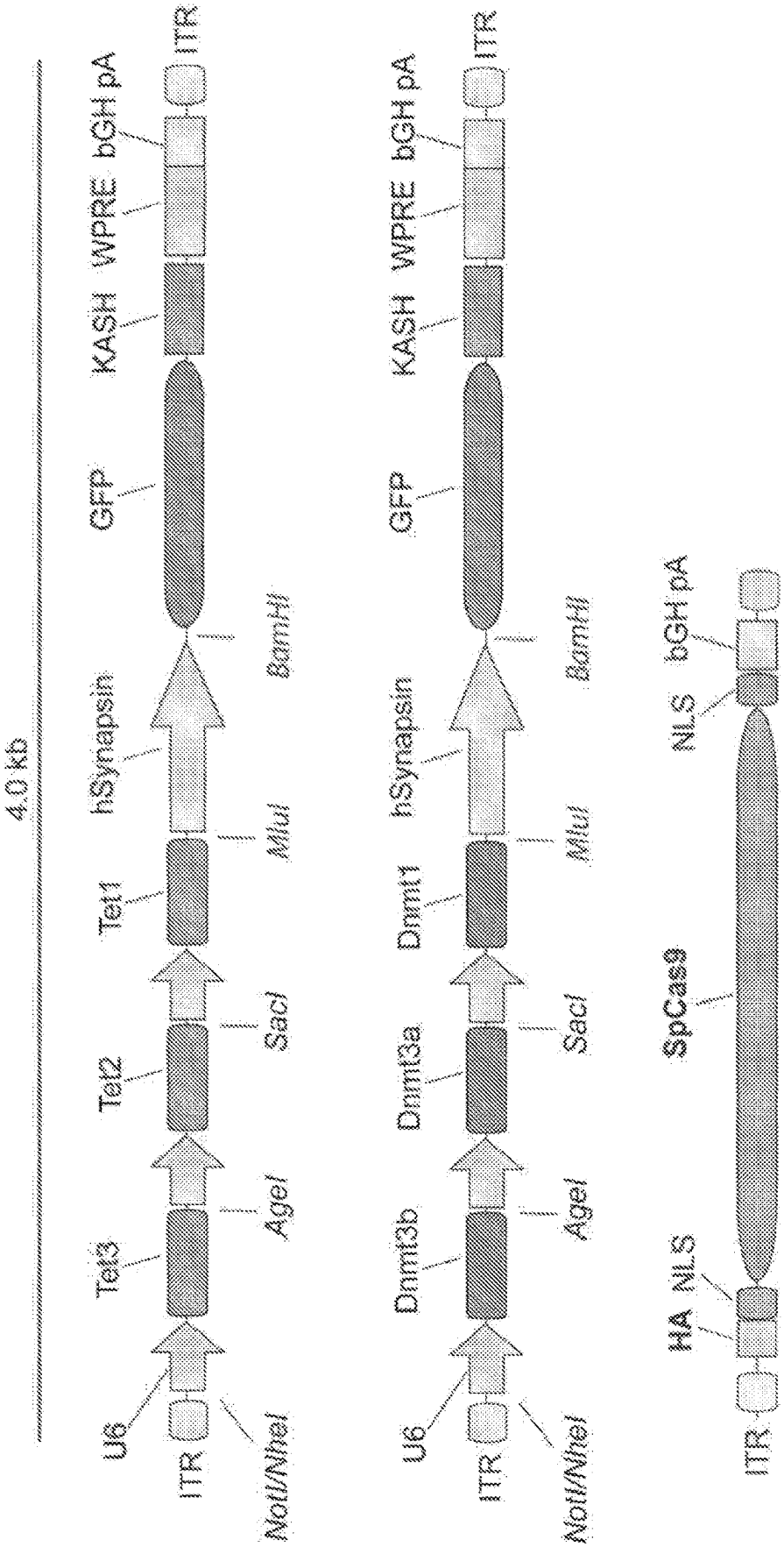

4.0 kb
U6
ITR
NotI/NheI
Tet3
AgeI
Tet2
SacI
Tet1
MluI
hSynapsin
BamHI
GFP
KASH
WPRE
bGH pA
ITR
U6
ITR
NotI/NheI
Dnmt3b
AgeI
Dnmt3a
SacI
Dnmt1
MluI
hSynapsin
BamHI
GFP
KASH
WPRE
bGH pA
ITR
HA
NLS
SpCas9
NLS
bGH pA
ITR

FIG. 34

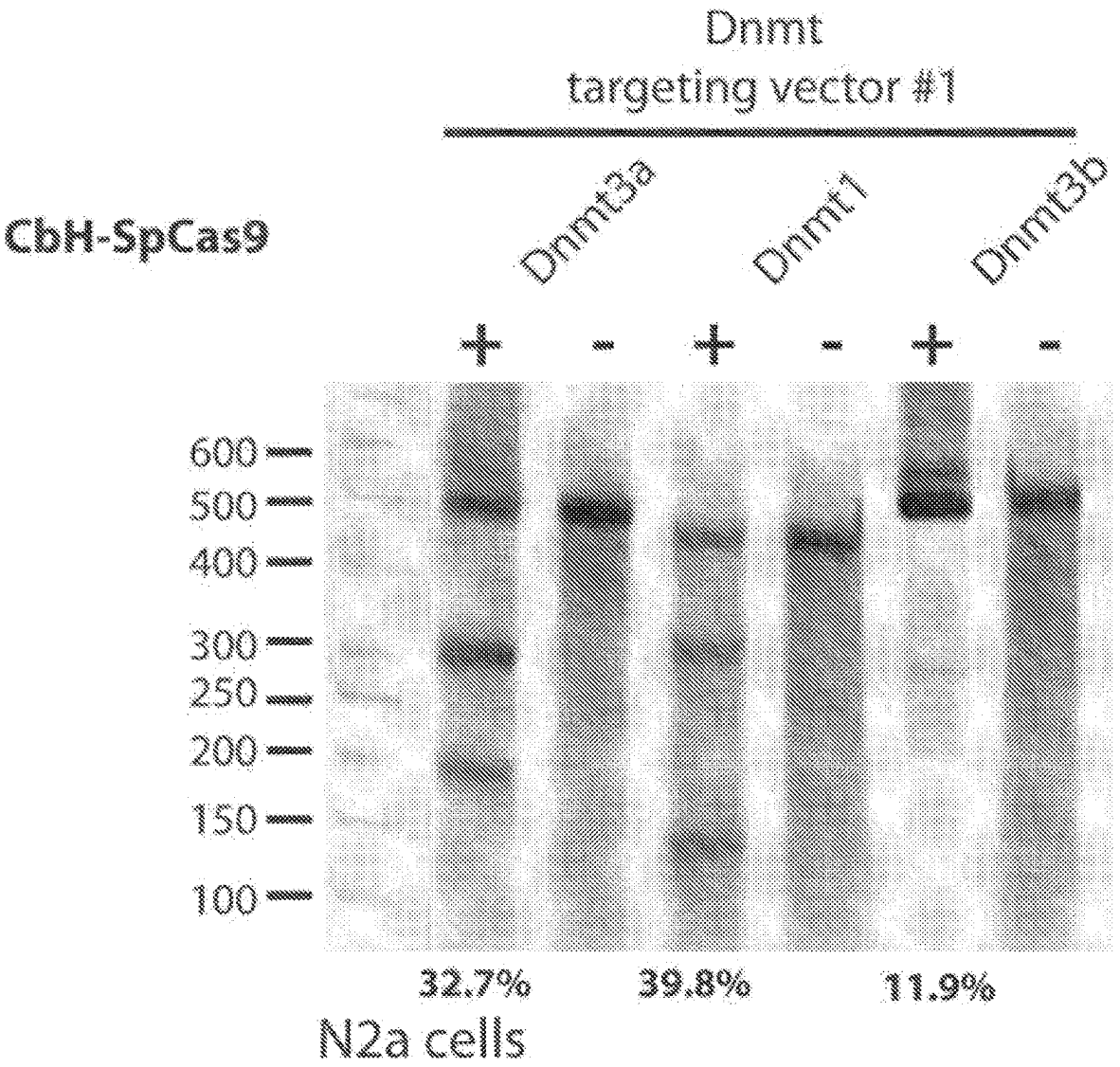

Dnmt
targeting vector #1
CbH-SpCas9
Dnmt3a
Dnmt1
Dnmt3b
+
-
+
-
+
-
600
500
400
300
250
200
150
100
32.7%
39.8%
11.9%
N2a cells

FIG. 35

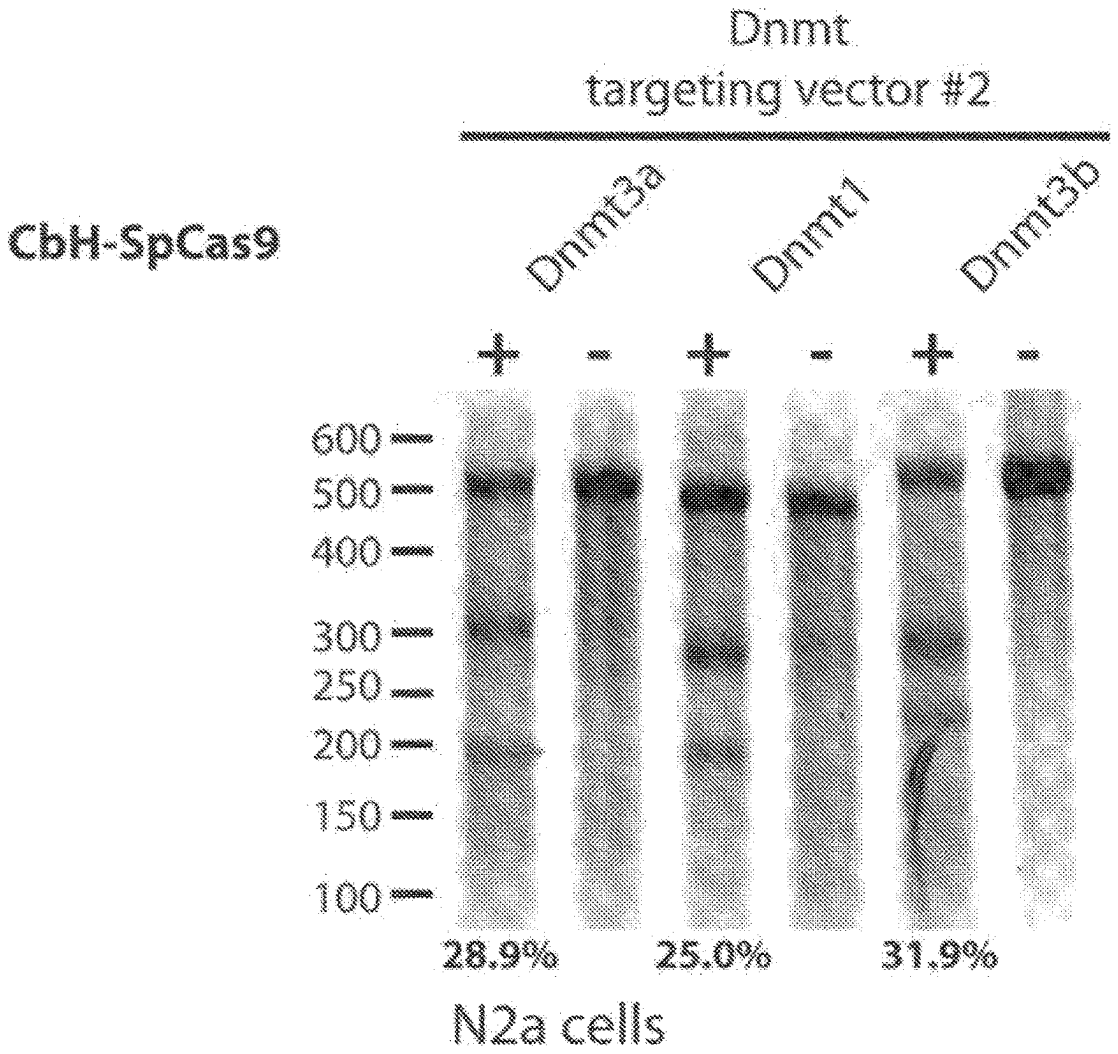

Dnmt
targeting vector #2
CbH-SpCas9
Dnmt3a
Dnmt1
Dnmt3b
+
-
+
-
+
-
600
500
400
300
250
200
150
100
28.9%
25.0%
31.9%
N2a cells

FIG. 36

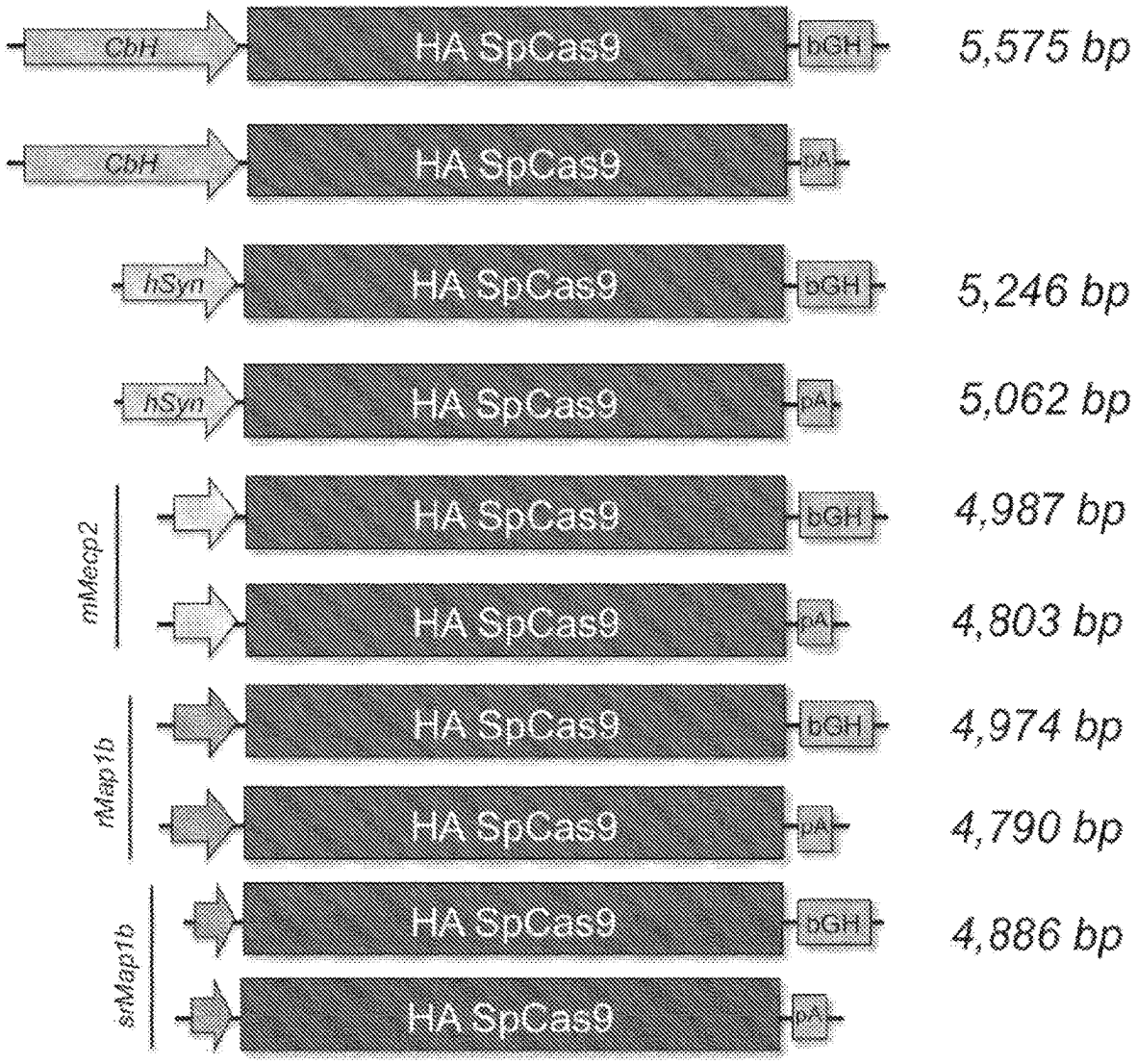

CbH
HA SpCas9
bGH
5,575 bp
CbH
HA SpCas9
bA
hSyn
HA SpCas9
bGH
5,246 bp
hSyn
HA SpCas9
bA
5,062 bp
mMecp2
HA SpCas9
bGH
4,987 bp
HA SpCas9
bA
4,803 bp
rMap1b
HA SpCas9
bGH
4,974 bp
HA SpCas9
bA
4,790 bp
srMap1b
HA SpCas9
bGH
4,886 bp
HA SpCas9
bA

FIG. 37

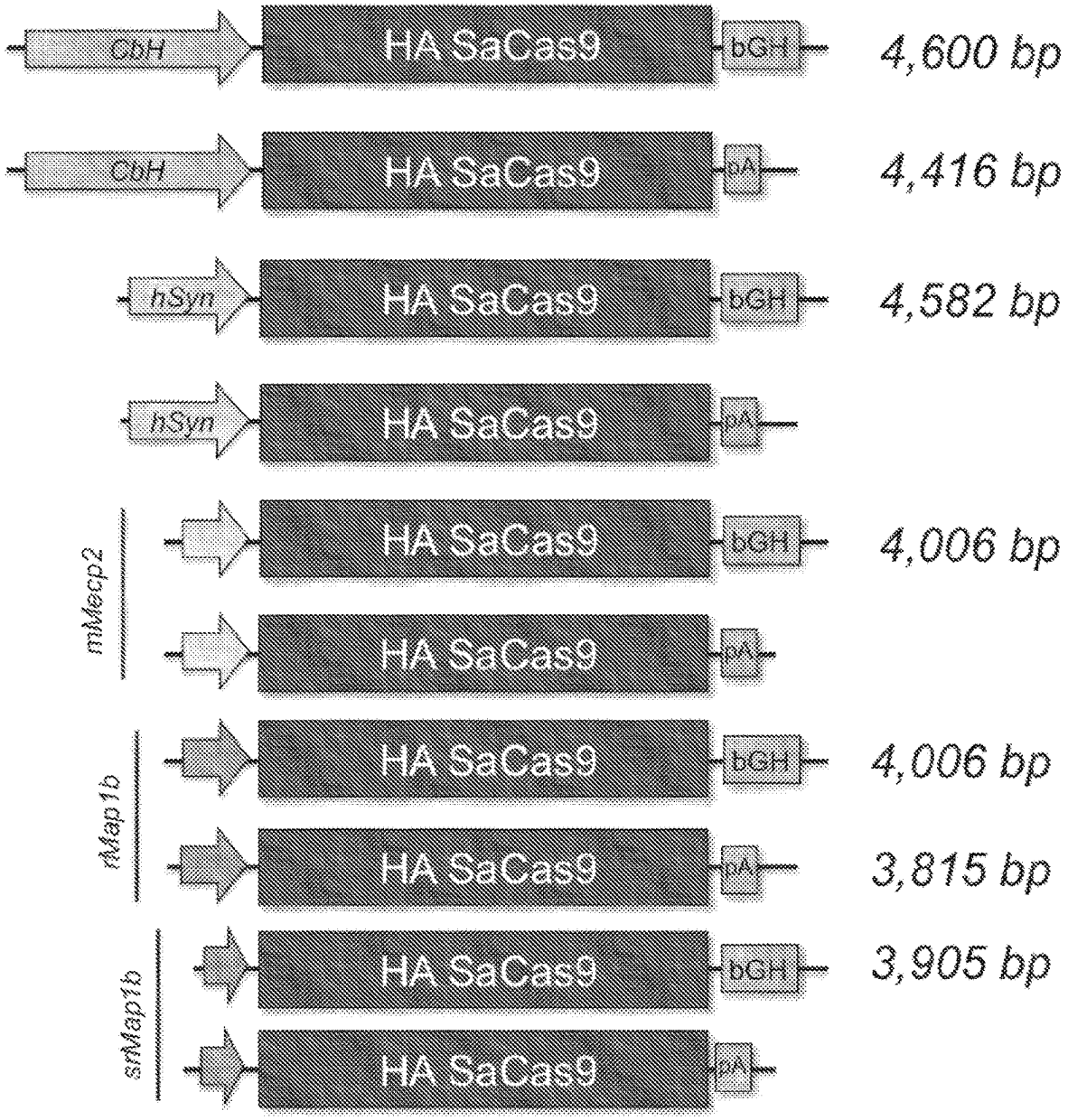

CbH
HA SaCas9
bGH
4,600 bp
CbH
HA SaCas9
bA
4,416 bp
hSyn
HA SaCas9
bGH
4,582 bp
hSyn
HA SaCas9
bA
mMecp2
HA SaCas9
bGH
4,006 bp
HA SaCas9
bA
rMap1b
HA SaCas9
bGH
4,006 bp
HA SaCas9
bA
3,815 bp
srMap1b
HA SaCas9
bGH
3,905 bp
HA SaCas9
bA

FIG. 38

Expression of SpCas9 & SaCas9 in N2a cells

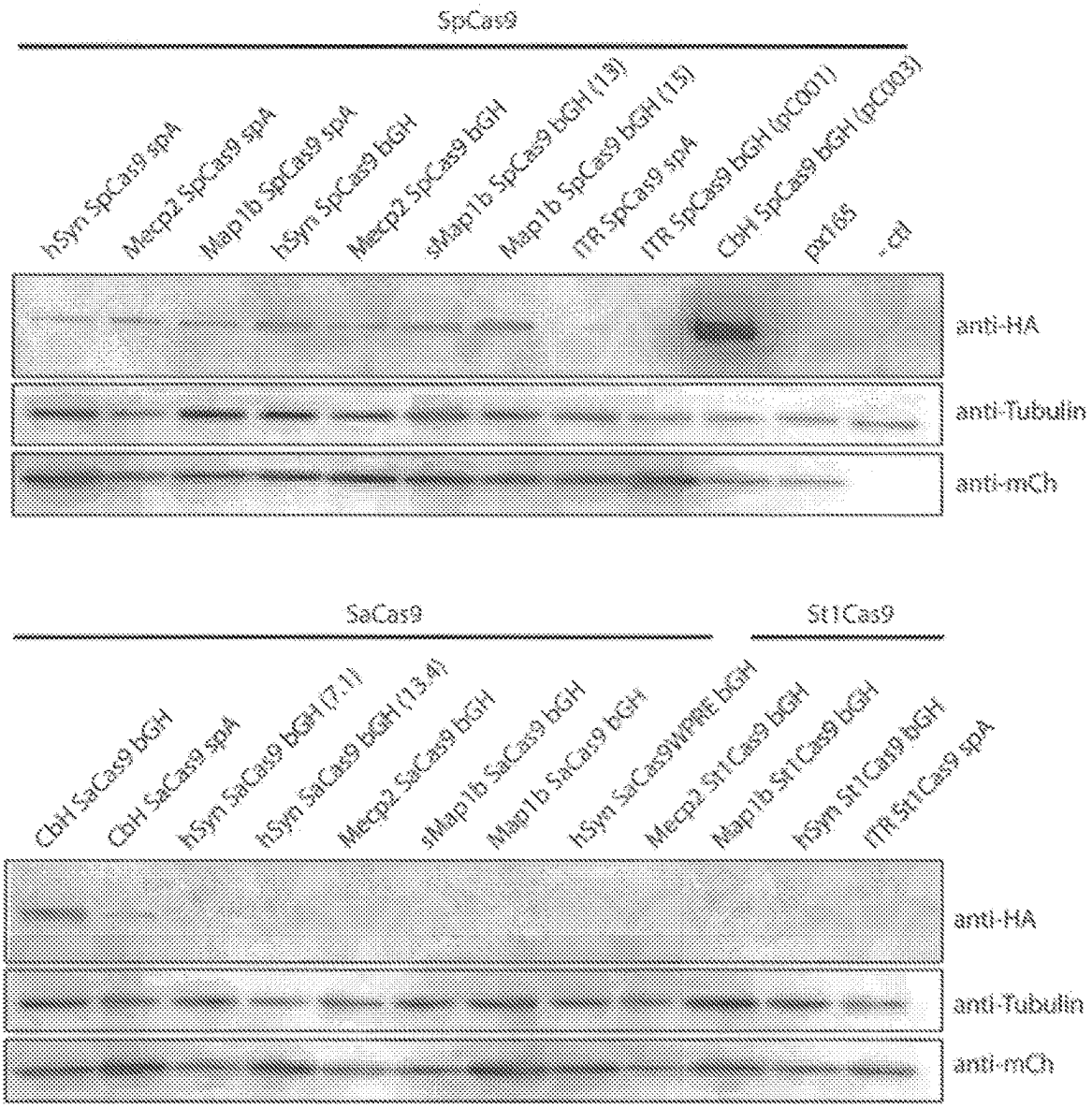


FIG. 39

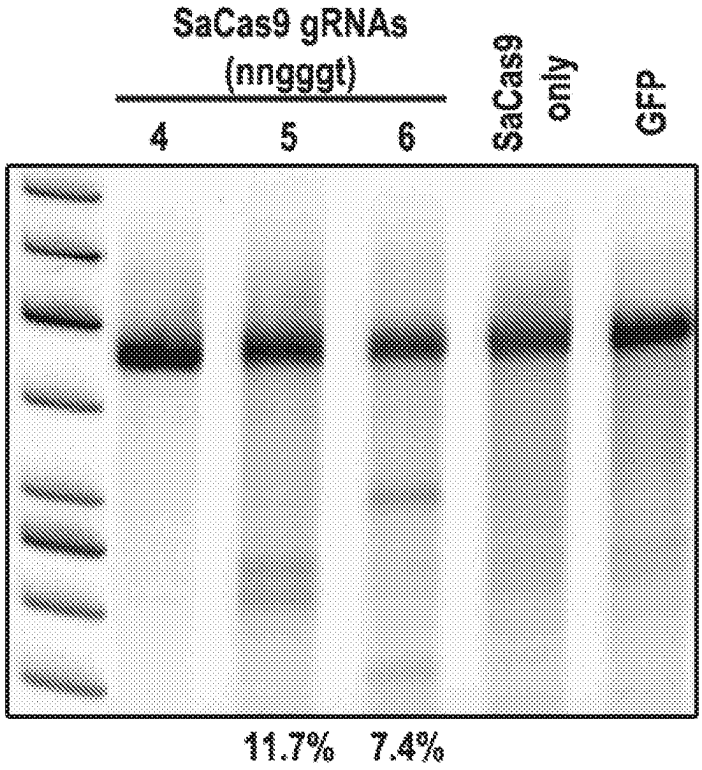


FIG. 40

CA1region of hippocampus

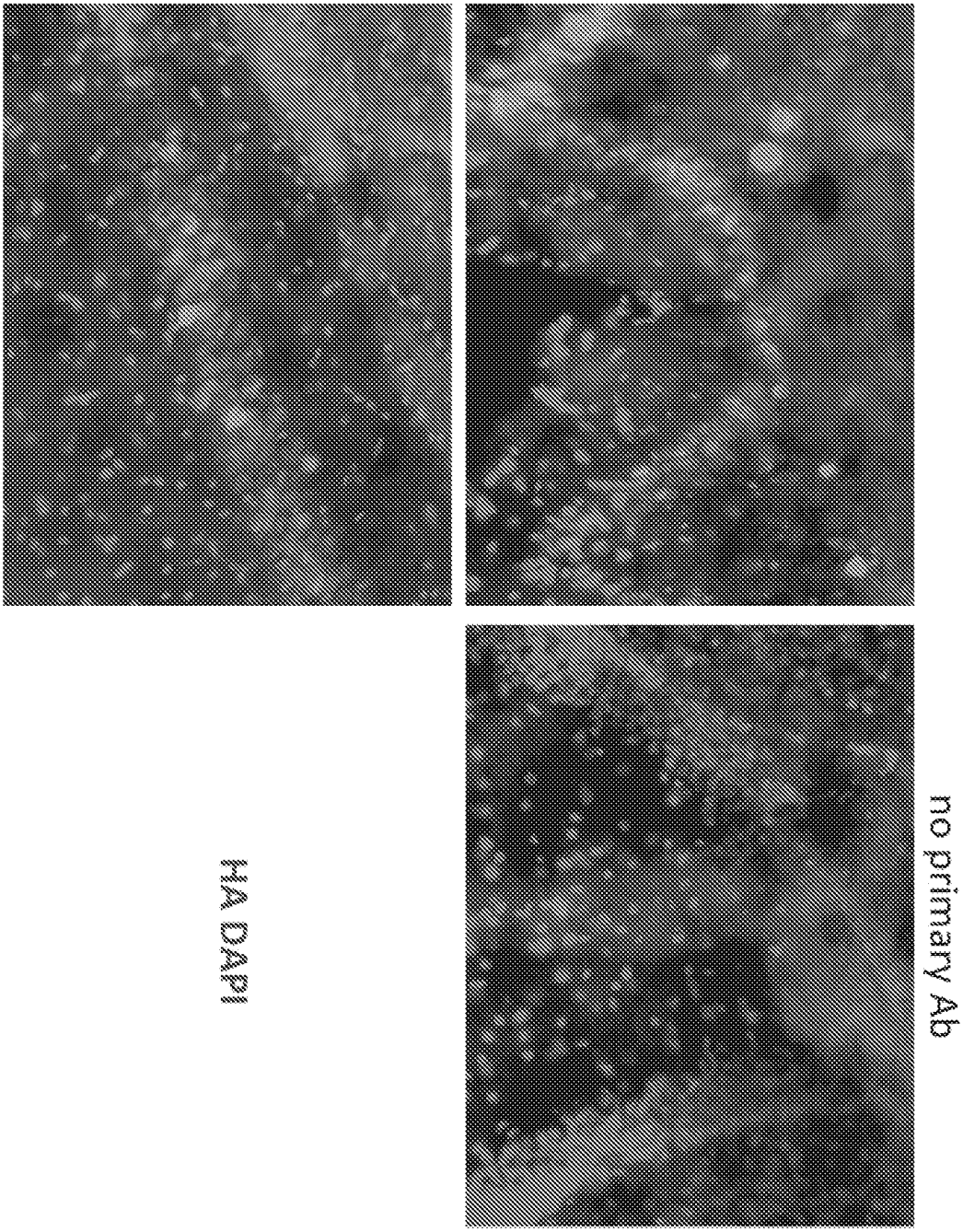


FIG. 41

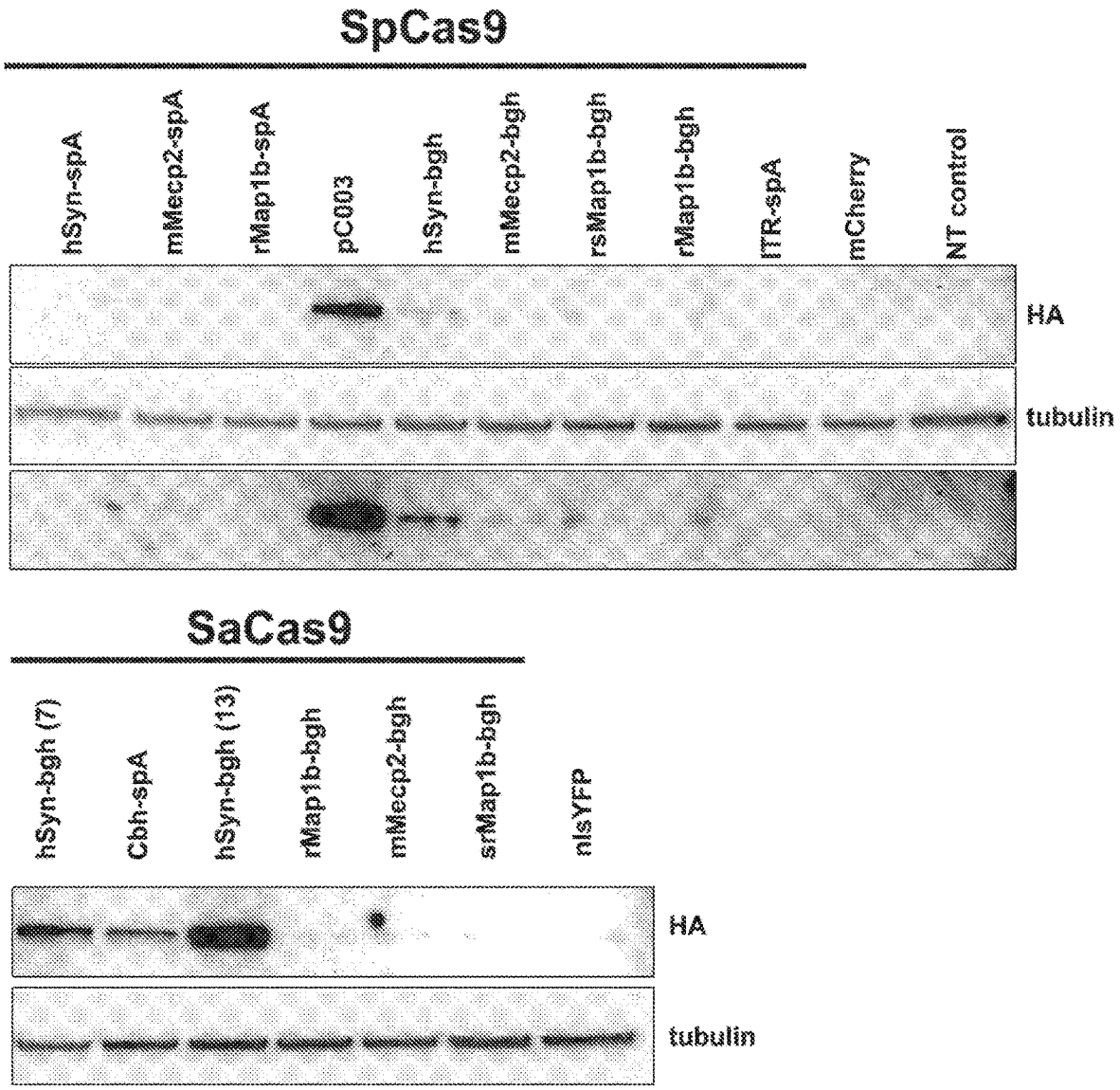

SpCas9
hSyn-spA
mMecp2-spA
rMap1b-spA
pC003
hSyn-bgh
mMecp2-bgh
rsMap1b-bgh
rMap1b-bgh
ITR-spA
mCherry
NT control
HA
tubulin
SaCas9
hSyn-bgh (7)
Cbh-spA
hSyn-bgh (13)
rMap1b-bgh
mMecp2-bgh
srMap1b-bgh
nlsYFP
HA
tubulin

FIG. 42

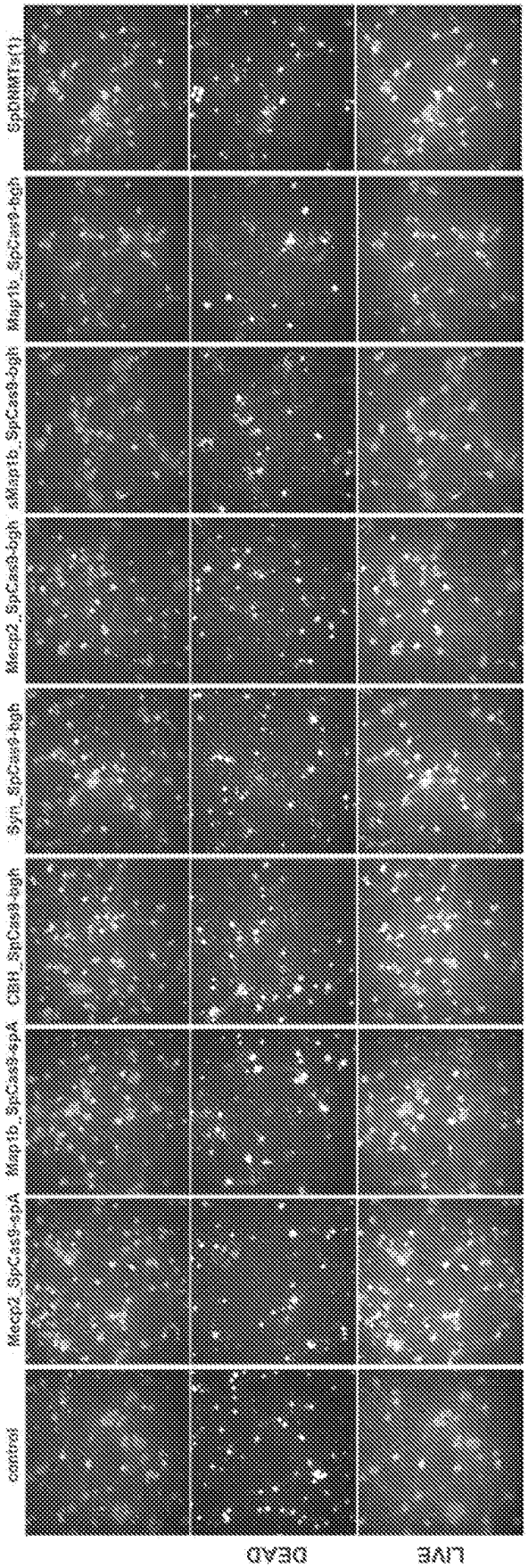

control
Mecp2_SpCas9-spA
Map1b_SpCas9-spA
CBH_SpCas9-bgh
Syn_SpCas9-bgh
Mecp2_SpCas9-bgh
sMap1b_SpCas9-bgh
Map1b_SpCas9-bgh
SpRNMT(sf1)
DEAD
LIVE

FIG. 43

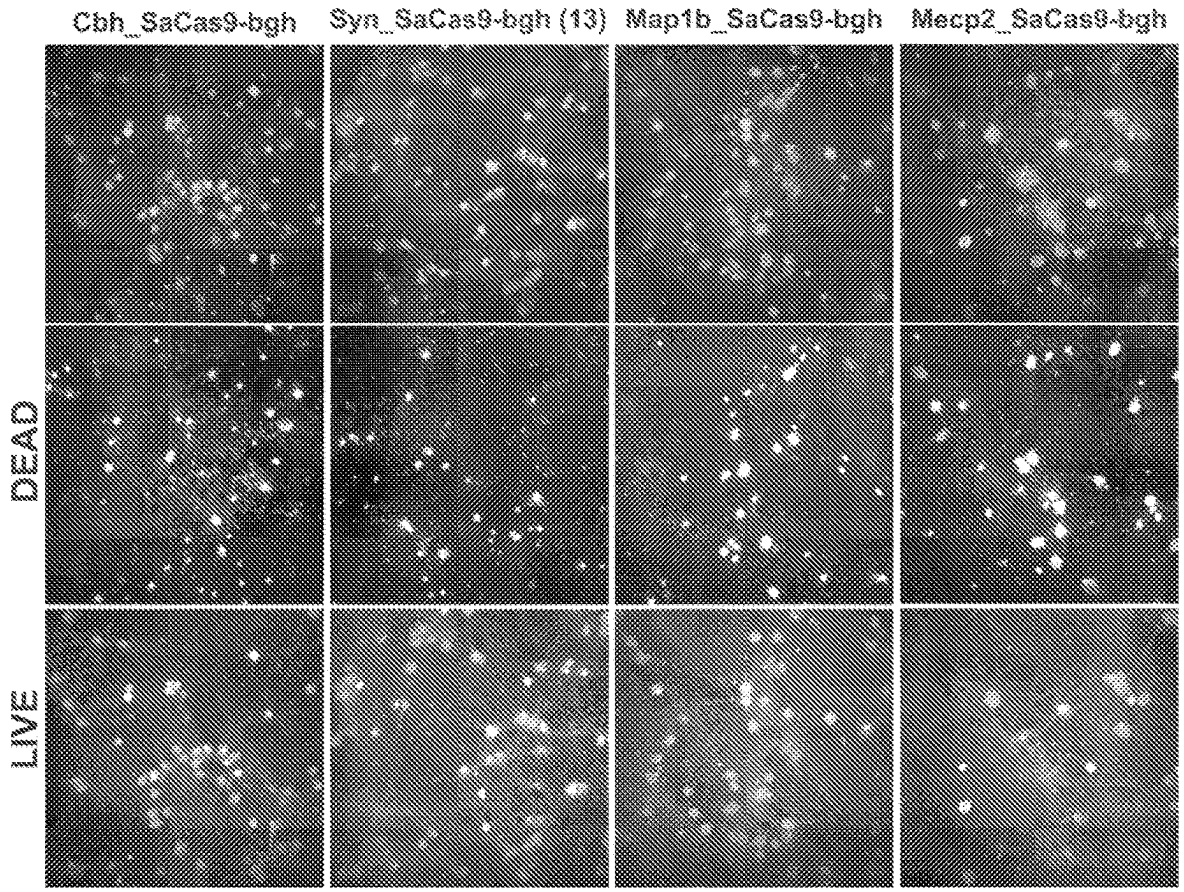

Cbh_SaCas9-bgh
Syn_SaCas9-bgh (13)
Map1b_SaCas9-bgh
Mecp2_SaCas9-bgh
DEAD
LIVE

FIG. 44

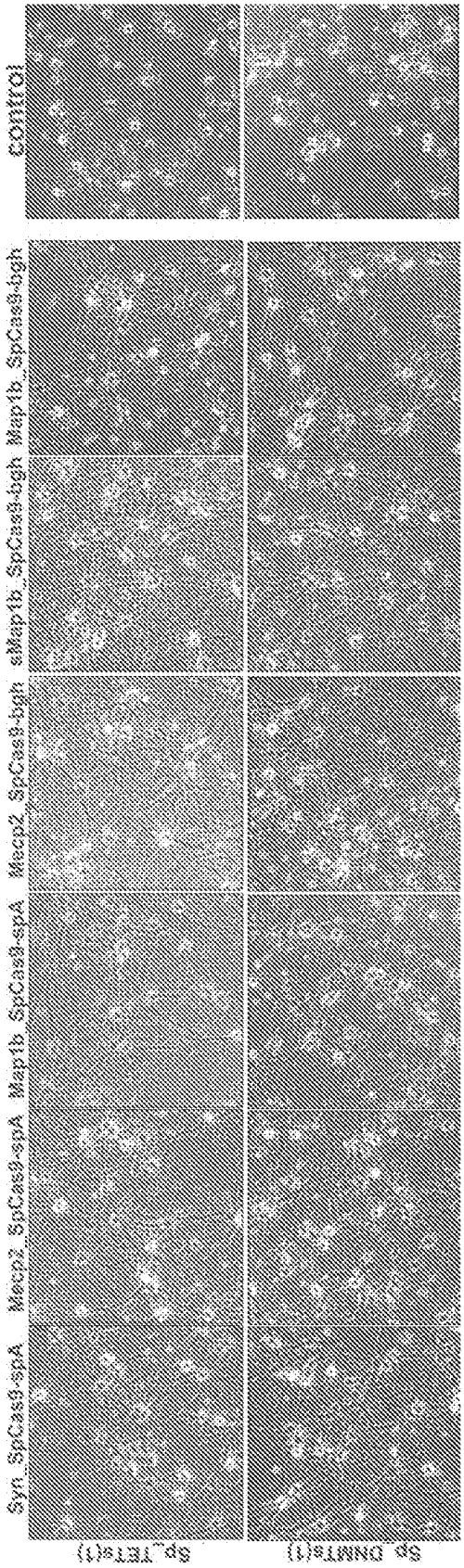

Syn_SpCas9-spA
Mecp2_SpCas9-spA
Map1b_SpCas9-spA
Mecp2_SpCas9-bgh
sMap1b_SpCas9-bgh
Map1b_SpCas9-bgh
control
Sp_TETs(1)
Sp_DNMTs(1)

FIG. 45

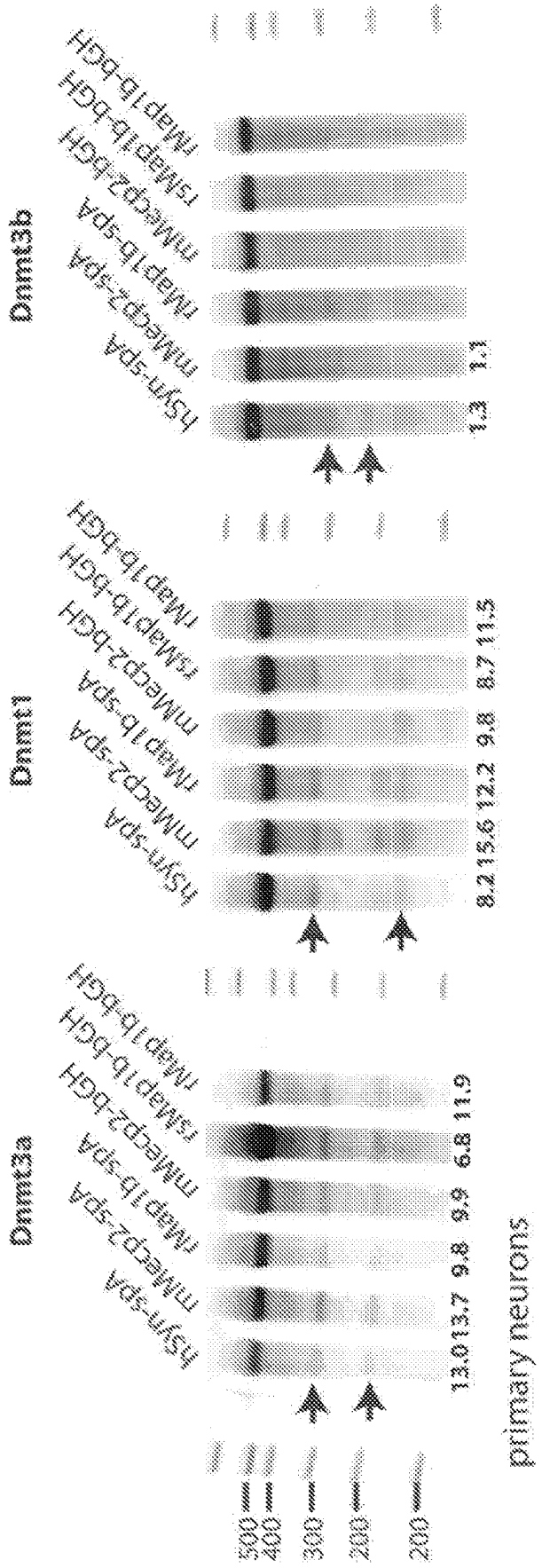

Dnmt3a
Dnmt1
Dnmt3b
hSyn-SpA
mMeCP2-SpA
mMap1b-SpA
rsMeCP2-bGH
rsMap1b-bGH
mMap1b-bGH
500
400
300
200
200
13.0 13.7 9.8 9.9 6.8 11.9
8.2 15.6 12.2 9.8 8.7 11.5
1.3 1.1
primary neurons

FIG. 46

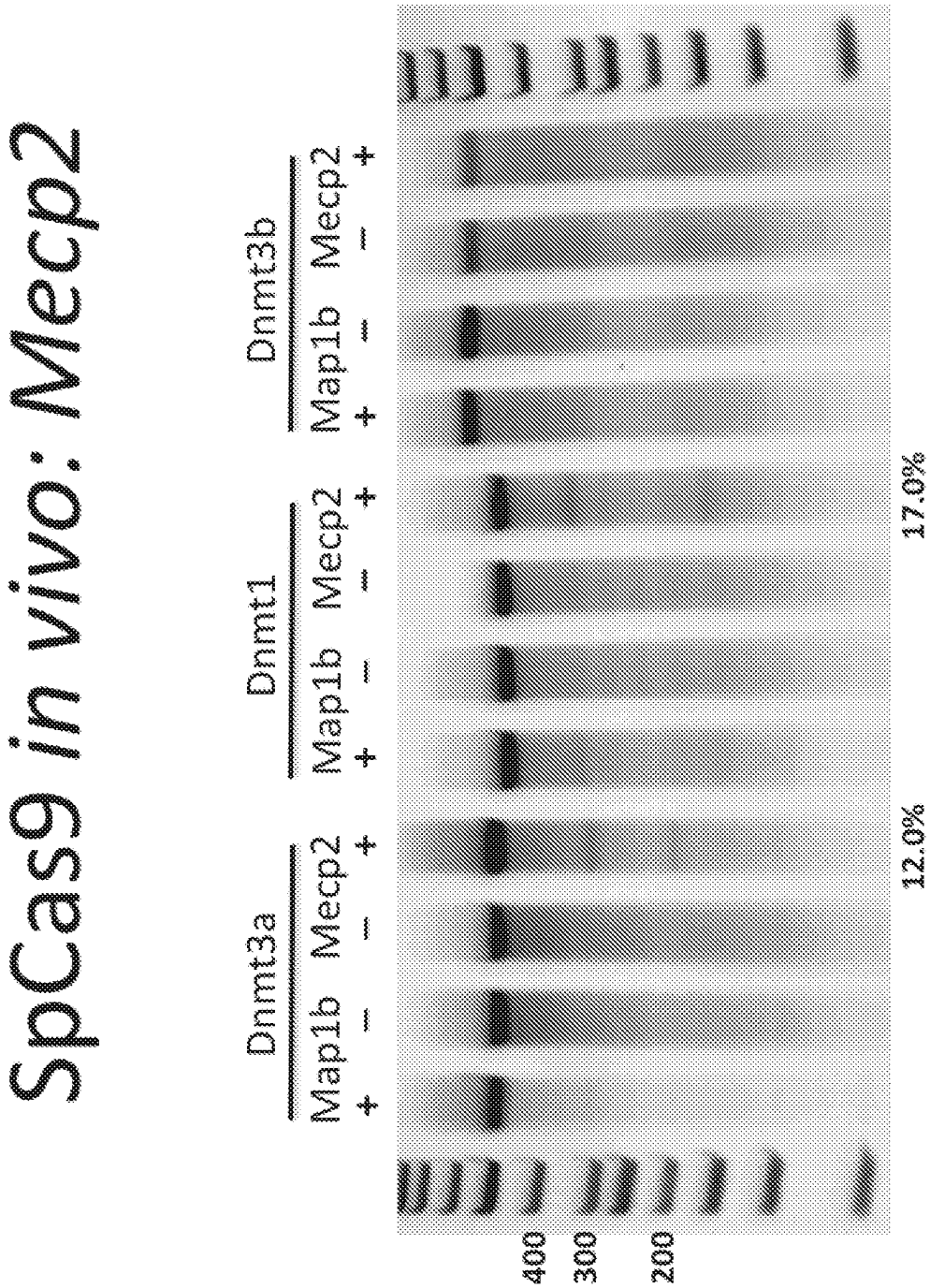

SpCas9 in vivo: Mecp2
Dnmt3a
Dnmt1
Dnmt3b
Map1b
Mecp2
+
−
400
300
200
12.0%
17.0%

FIG. 47

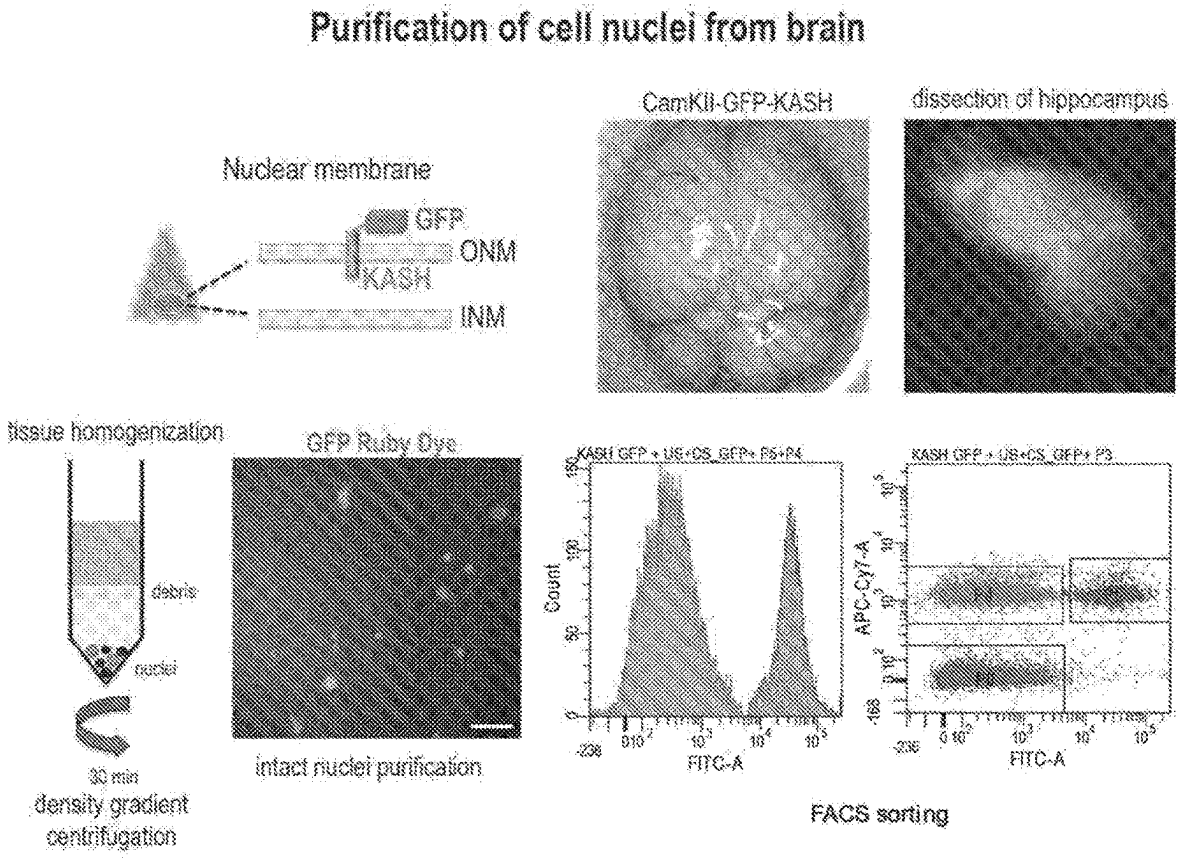

Purification of cell nuclei from brain
Nuclear membrane
GFP
ONM
KASH
INM
CamKII-GFP-KASH
dissection of hippocampus
tissue homogenization
debris
nuclei
30 min
density gradient centrifugation
GFP Ruby Dye
intact nuclei purification
Count
FITC-A
APC-Cy7-A
FACS sorting

FIG. 48

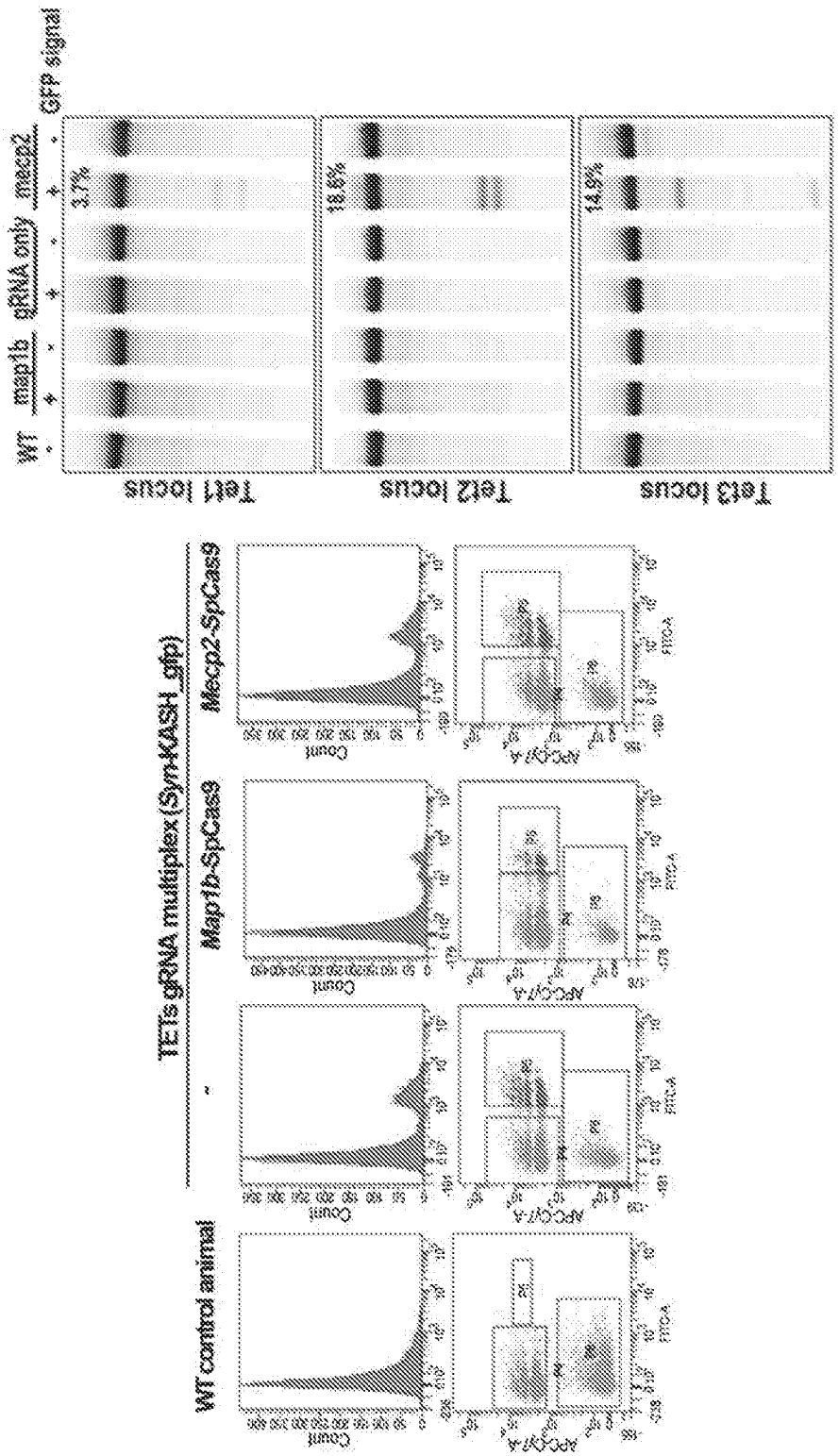

WT control animal
TETs gRNA multiplex (Syn-KASH_gfp)
-
Map1b-SpCas9
Mecp2-SpCas9
Count
APC-CY7-A
FITC-A
Tet1 locus
Tet2 locus
Tet3 locus
WT
map1b
gRNA only
mecp2
GFP signal
3.7%
18.6%
14.9%

FIG. 49

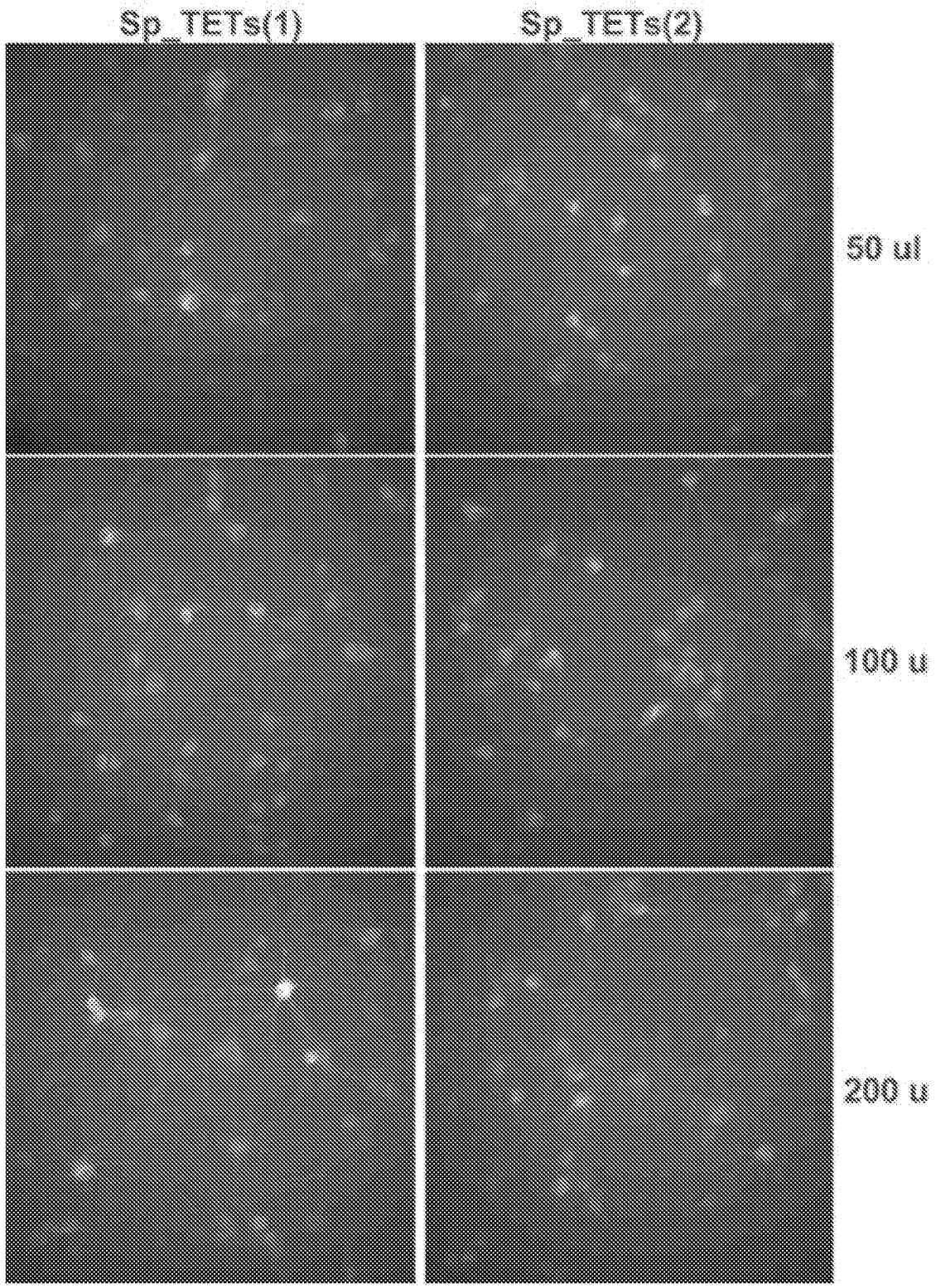


FIG. 50

	pattern	overhang
1	CCnNNNNNNNNNNNNNNNNnnnnnnnnnnNNNNNNNNNNNNNNNNnGG	38
2	CCnNNNNNNNNNNNNNNNNnnnnnnnnNNNNNNNNNNNNNNNNnGG	34
3	CCnNNNNNNNNNNNNNNNNnnnnNNNNNNNNNNNNNNNNnGG	30
4	CCnNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNnGG	26
5	CCnNNNNNNNNNNNNNNNNNNNNNNNNNNNNnGG	22
6	CCnNNNNNNNNNNNNNNNNNNNNNNNNnGG	18
7	CCnNNNNNNNNNNNNNNNNNNNNNNnGG	14
8	CCnNNNNNNNNNNNNNNNNNNnGG	10
9	CCnNNNNNNNNNNNNNNnGG	6
10	CCnNNNNNNNNNNnGG	4
11	CCnNNNNNNNNnGG	3
12	CCnNNNNNNNnGG	2
13	CCnNNNNNNnGG	1
14	CCnNNNNNnGG	0
15	CCnNNNNnGG	1
16	CCnNNNnGG	2
17	CCnNNnGG	3
18	CCnnGG	4
19	CCGG	8
20	GGCC	10
21	GGNNCC	12
22	GGNNNCC	14
23	GGNNNNCC	16
24	GGNNNNNNNCC	20

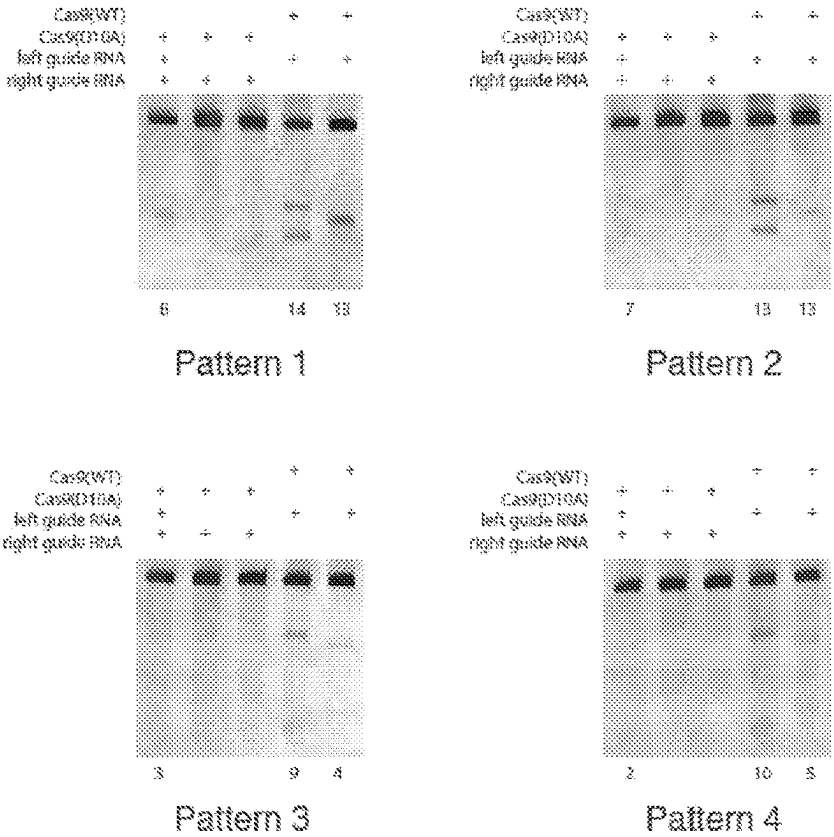


FIG. 51

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

FZ0260	[illegible]	Left guide RNA 1
FZ0261	[illegible]	Left guide RNA 2
FZ0262	[illegible]	Left guide RNA 3
FZ0263	[illegible]	Left guide RNA 4
FZ0264	[illegible]	Left guide RNA 5
FZ0265	[illegible]	Left guide RNA 6
FZ0266	[illegible]	Left guide RNA 7
FZ0267	[illegible]	Left guide RNA 8
FZ0268	[illegible]	Left guide RNA 9
FZ0269	[illegible]	Left guide RNA 10
FZ0270	[illegible]	Left guide RNA 11
FZ0271	[illegible]	Left guide RNA 12
FZ0272	[illegible]	Left guide RNA 13
FZ0273	[illegible]	Left guide RNA 14
FZ0274	[illegible]	Left guide RNA 15
FZ0275	[illegible]	Left guide RNA 16
FZ0276	[illegible]	Left guide RNA 17
FZ0277	[illegible]	Left guide RNA 18
FZ0278	[illegible]	Left guide RNA 19
FZ0279	[illegible]	Left guide RNA 20
FZ0280	[illegible]	Left guide RNA 21
FZ0281	[illegible]	Left guide RNA 22
FZ0282	[illegible]	Left guide RNA 23
FZ0283	[illegible]	Left guide RNA 24
FZ0284	[illegible]	Right guide RNA 1
FZ0285	[illegible]	Right guide RNA 2
FZ0286	[illegible]	Right guide RNA 3
FZ0287	[illegible]	Right guide RNA 4
FZ0288	[illegible]	Right guide RNA 5
FZ0289	[illegible]	Right guide RNA 6
FZ0290	[illegible]	Right guide RNA 7
FZ0291	[illegible]	Right guide RNA 8
FZ0292	[illegible]	Right guide RNA 9
FZ0293	[illegible]	Right guide RNA 10
FZ0294	[illegible]	Right guide RNA 11
FZ0295	[illegible]	Right guide RNA 12
FZ0296	[illegible]	Right guide RNA 13
FZ0297	[illegible]	Right guide RNA 14
FZ0298	[illegible]	Right guide RNA 15
FZ0299	[illegible]	Right guide RNA 16
FZ0300	[illegible]	Right guide RNA 17
FZ0301	[illegible]	Right guide RNA 18
FZ0302	[illegible]	Right guide RNA 19
FZ0303	[illegible]	Right guide RNA 20
FZ0304	[illegible]	Right guide RNA 21
FZ0305	[illegible]	Right guide RNA 22
FZ0306	[illegible]	Right guide RNA 23
FZ0307	[illegible]	Right guide RNA 24

FIG. 52

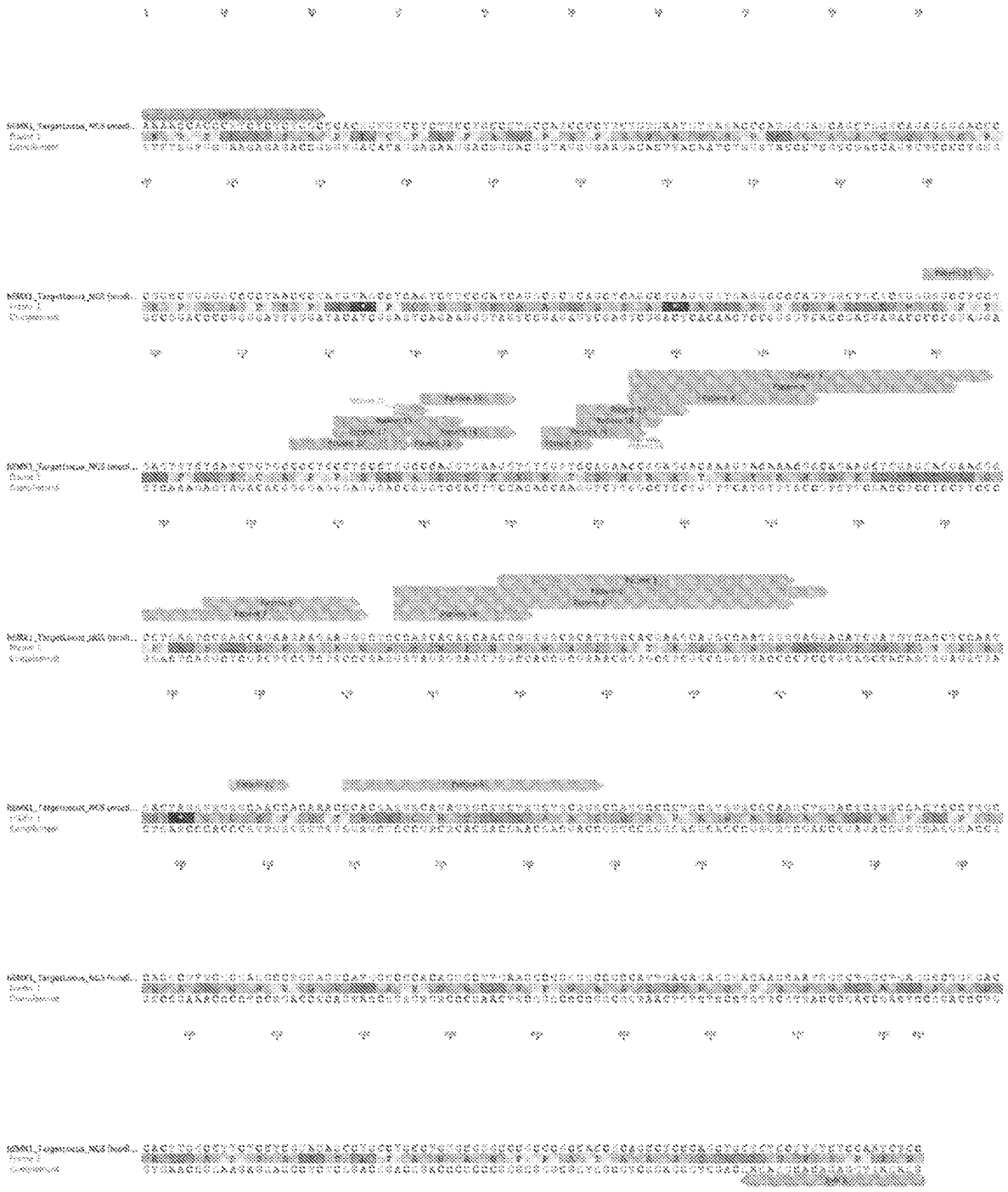


FIG. 53

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

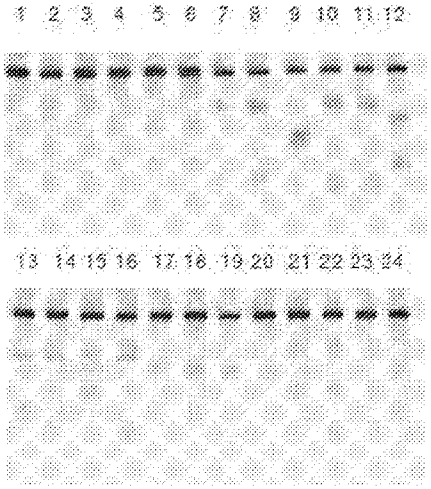


FIG. 54

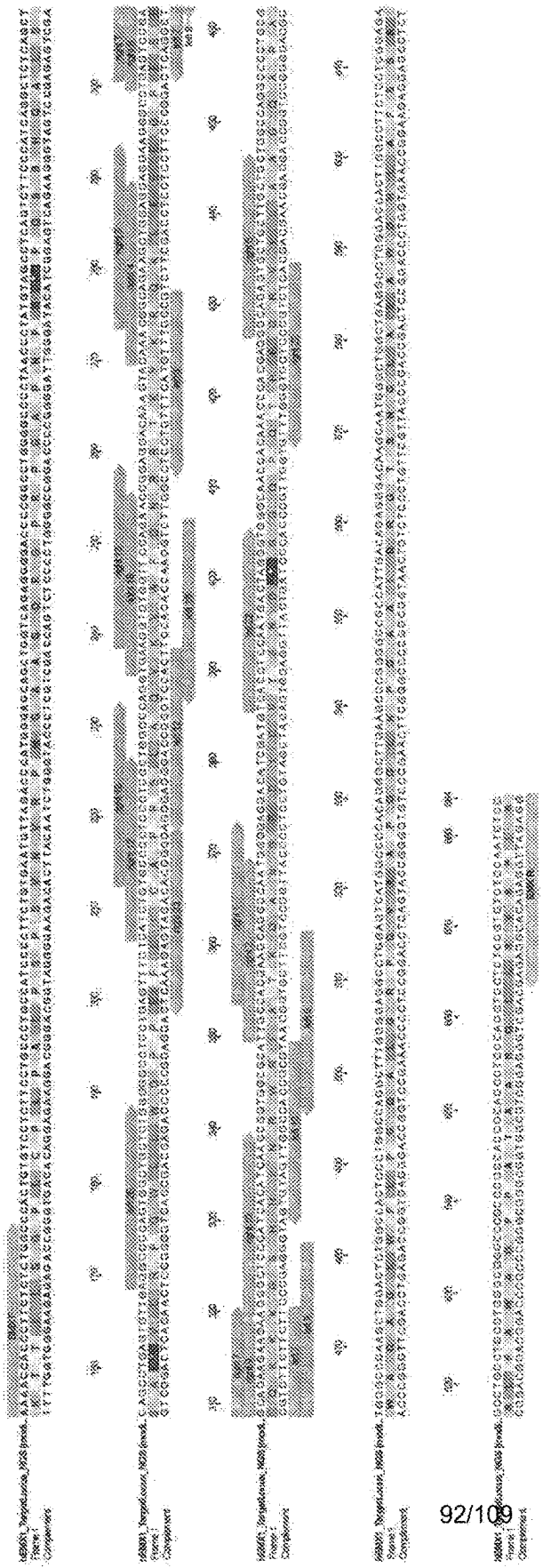


FIG. 55

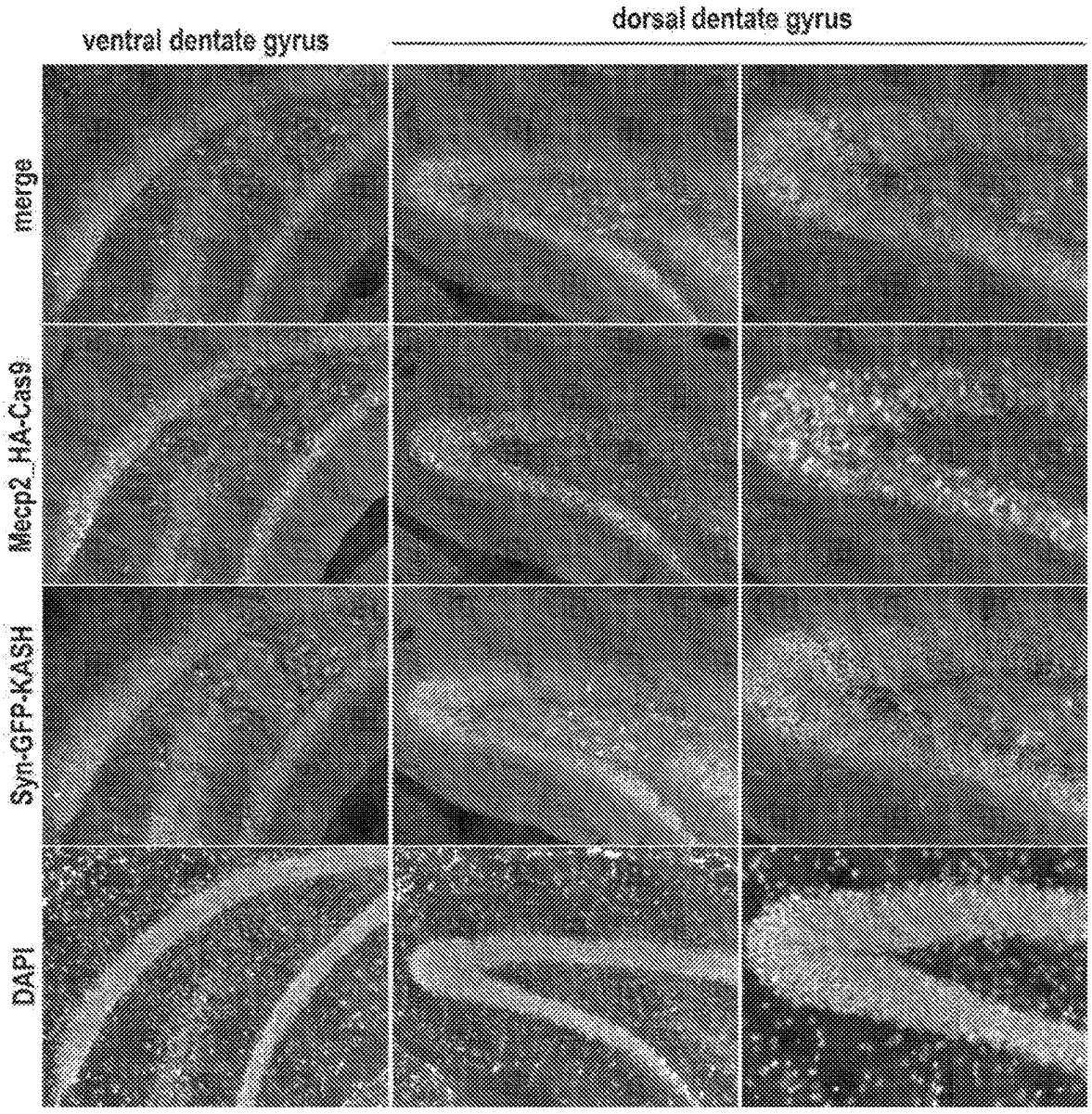

ventral dentate gyrus
dorsal dentate gyrus
merge
Mecp2_HA-Cas9
Syn-GFP-KASH
DAPI

FIG. 56

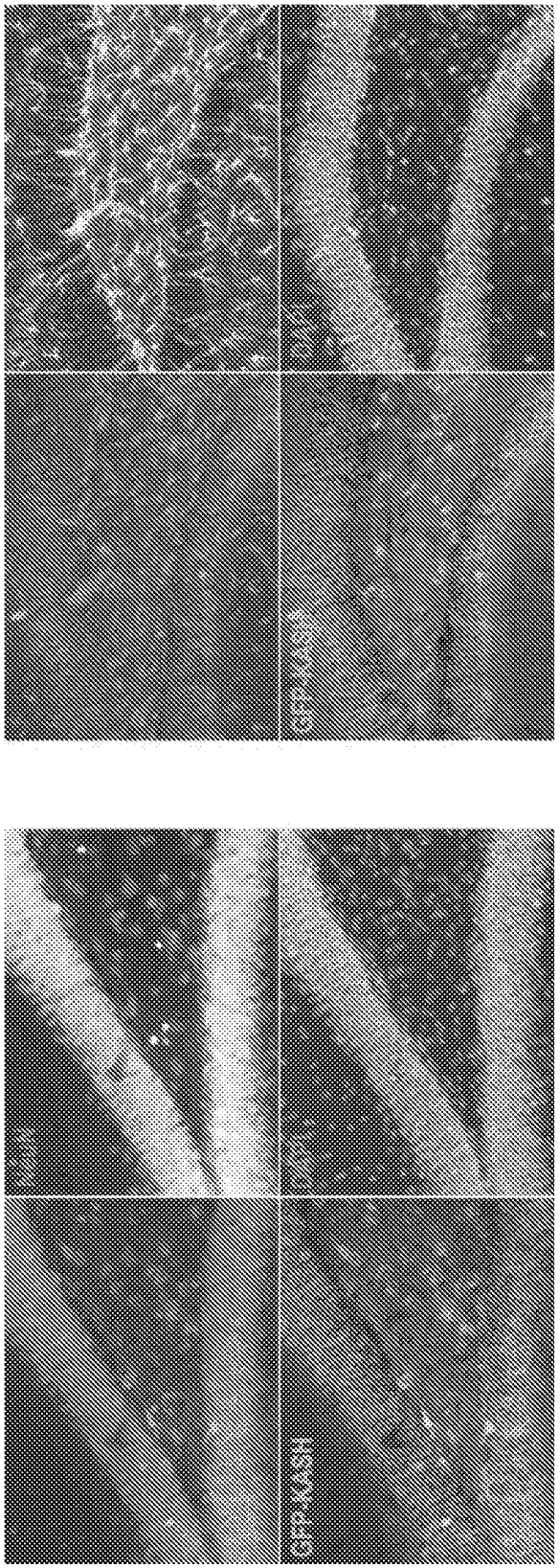


FIG. 57

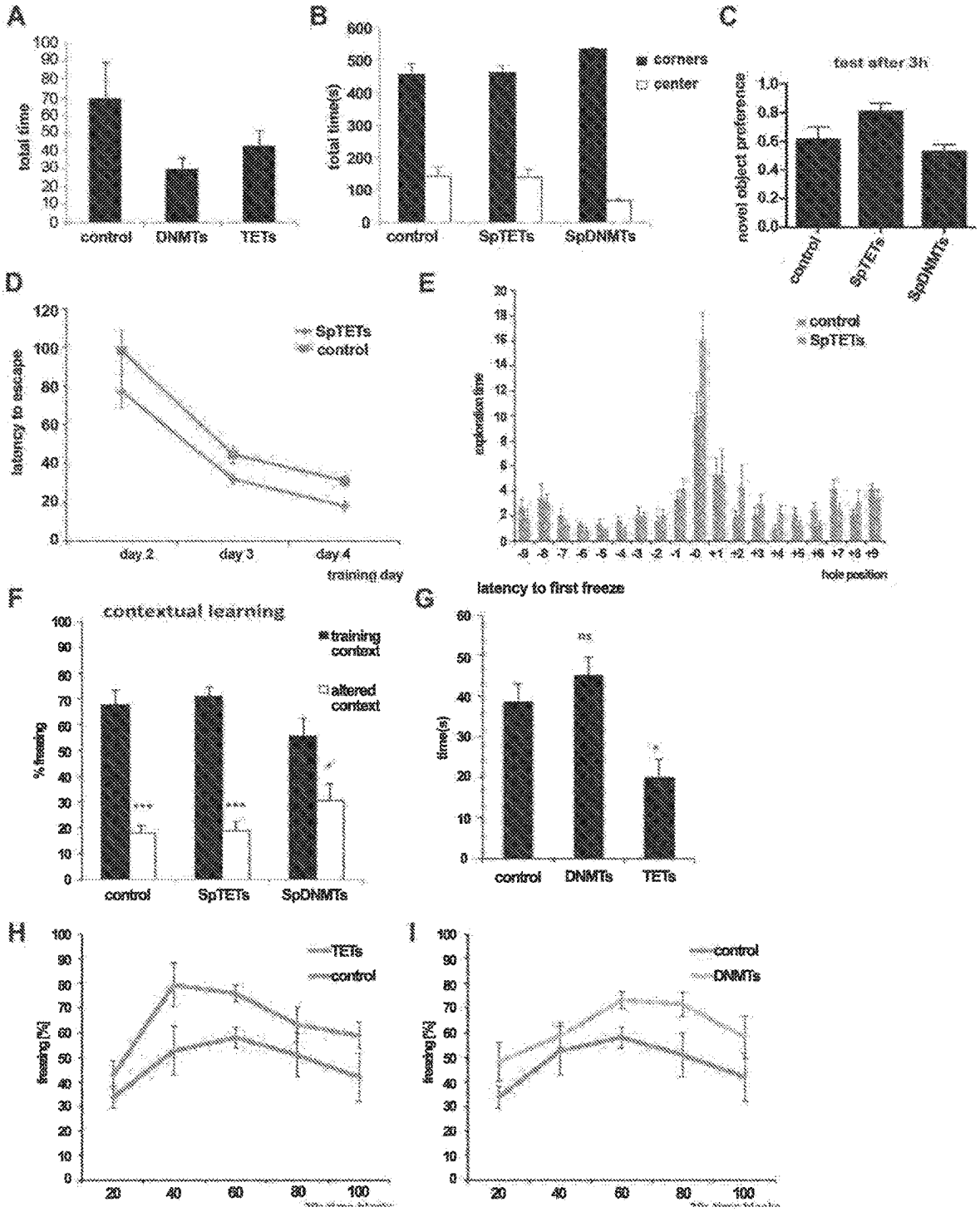

A
total time
control
DNMTs
TETs
B
total time(s)
corners
center
control
SpTETs
SpDNMTs
C
test after 3h
novel object preference
control
SpTETs
SpDNMTs
D
latency to escape
SpTETs
control
day 2
day 3
day 4
training day
E
exploration time
control
SpTETs
hole position
F
contextual learning
% freezing
training context
altered context
control
SpTETs
SpDNMTs
G
latency to first freeze
time(s)
control
DNMTs
TETs
H
freezing [%]
TETs
control
I
freezing [%]
control
DNMTs

FIG. 58

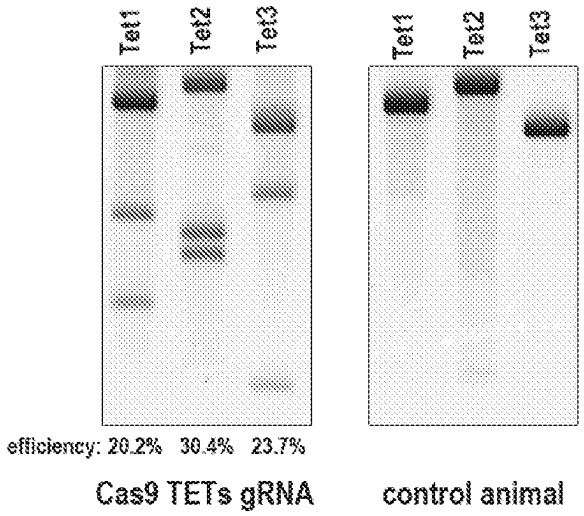

Tet1
Tet2
Tet3
Tet1
Tet2
Tet3
efficiency: 20.2% 30.4% 23.7%
Cas9 TETs gRNA
control animal

FIG. 59

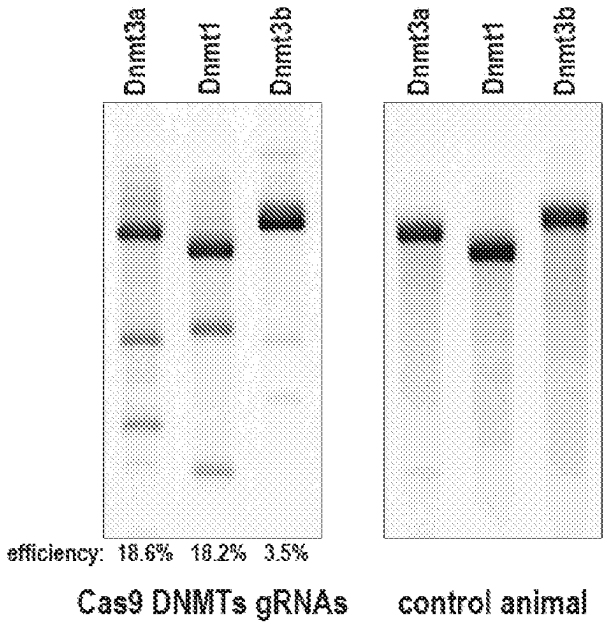

Dnmt3a
Dnmt1
Dnmt3b
Dnmt3a
Dnmt1
Dnmt3b
efficiency: 18.6% 18.2% 3.5%
Cas9 DNMTs gRNAs
control animal

FIG. 60

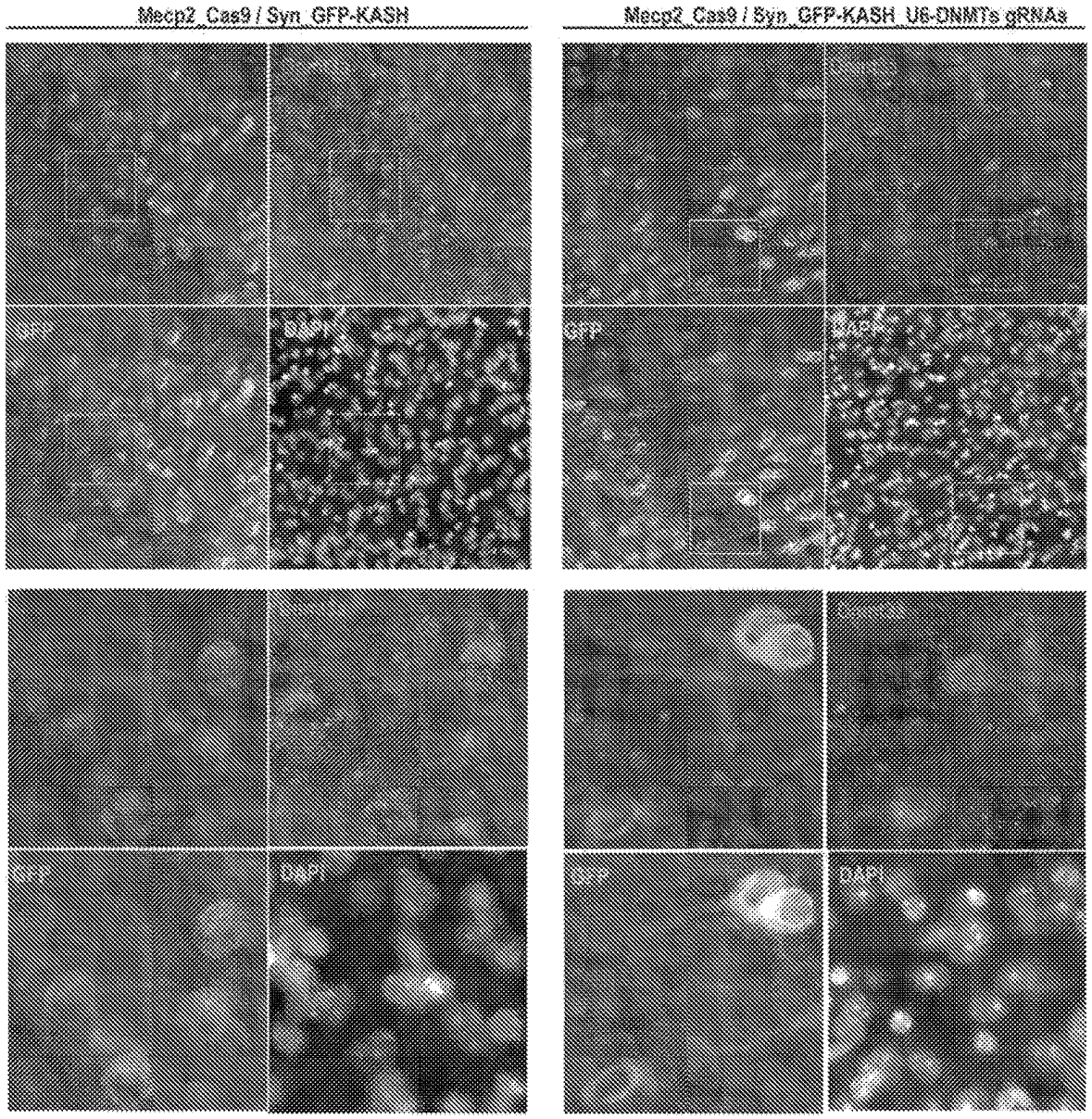

Mecp2_Cas9 / Syn_GFP-KASH
Mecp2_Cas9 / Syn_GFP-KASH_U6-DNMTs gRNAs
GFP
DAPI
GFP
DAPI
GFP
DAPI
GFP
DAPI

FIG. 61

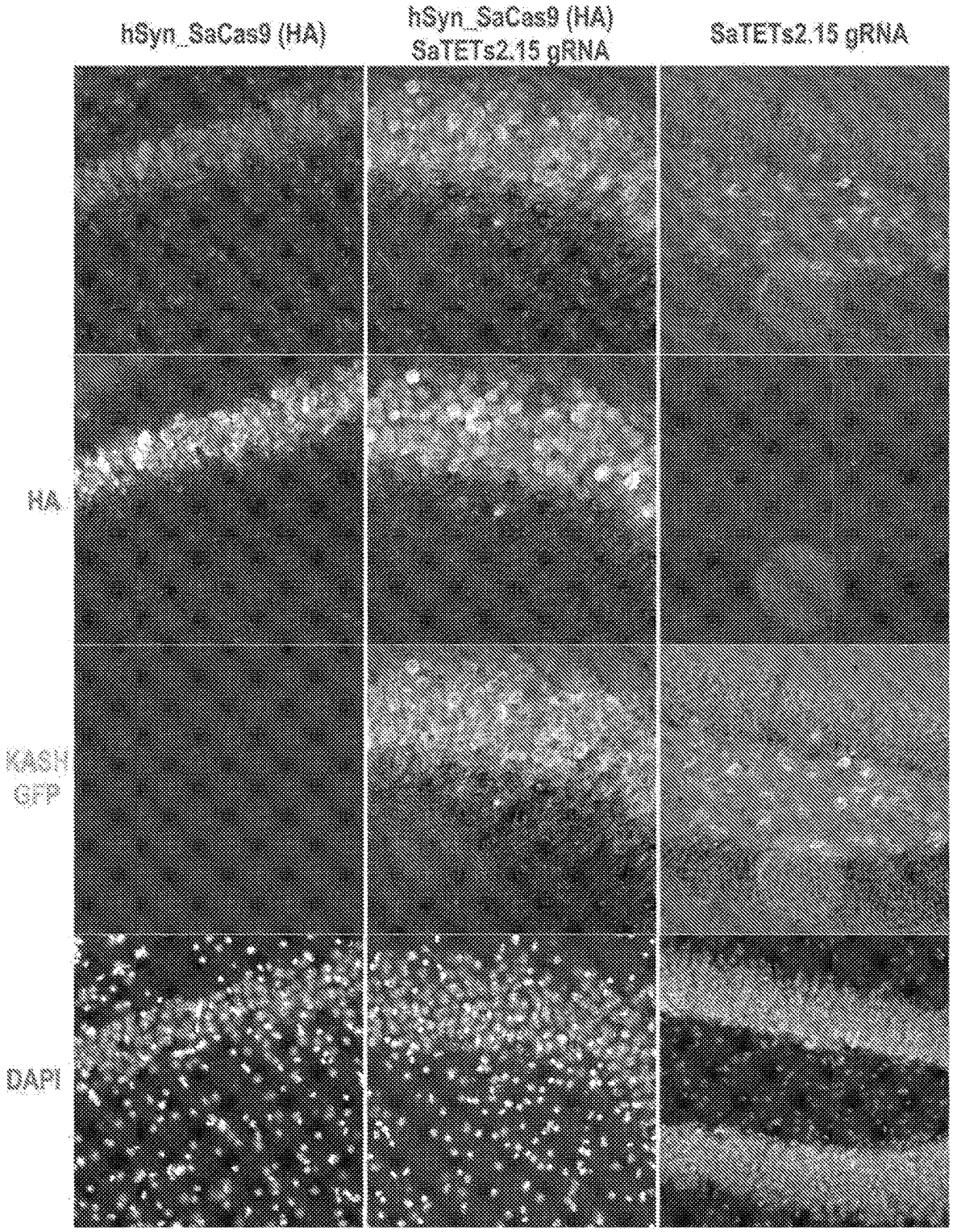

hSyn_SaCas9 (HA)
hSyn_SaCas9 (HA)
SaTETs2.15 gRNA
SaTETs2.15 gRNA
HA
KASH
GFP
DAPI

FIG. 62

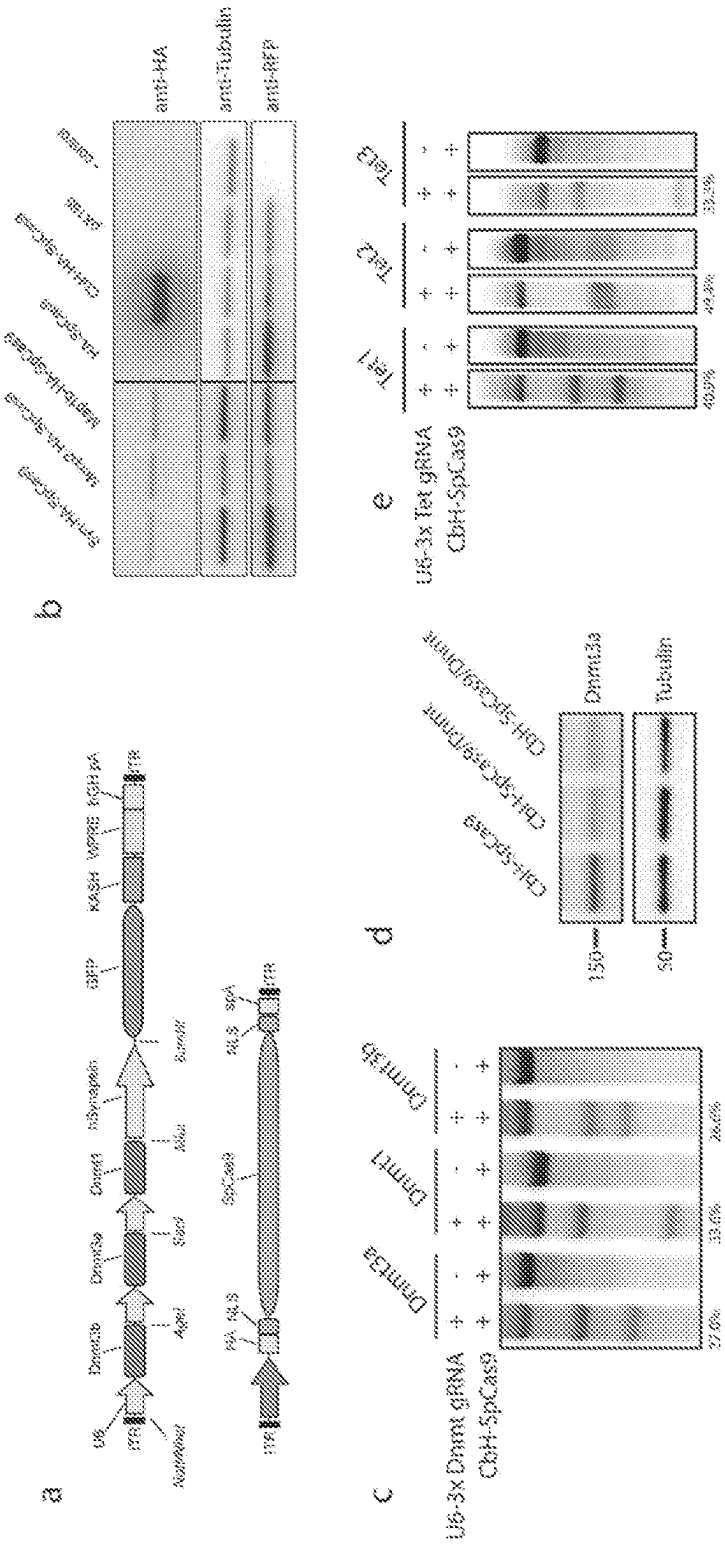

a
b
c
d
e
anti-HA
anti-Tubulin
anti-RFP
U6-3x Dnmt gRNA
CbH-SpCas9
Dnmt3a
Dnmt1
Dnmt3b
Tubulin
U6-3x Tet gRNA
Tet1
Tet2
Tet3

FIG. 63A-63E

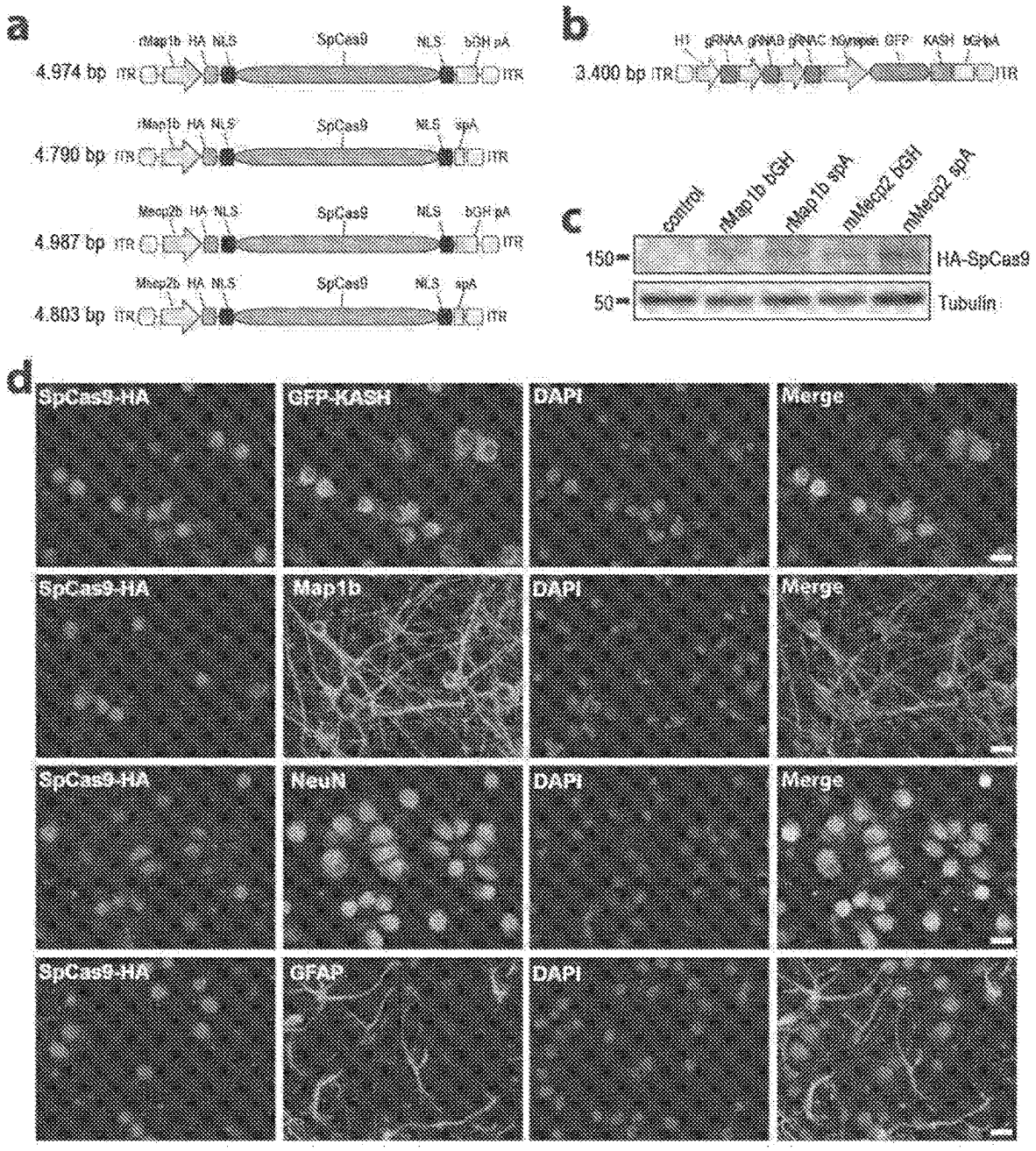

a
rMap1b HA NLS SpCas9 NLS bGH pA
4.974 bp ITR ITR
rMap1b HA NLS SpCas9 NLS spA
4.790 bp ITR ITR
Mecp2b HA NLS SpCas9 NLS bGH pA
4.987 bp ITR ITR
Mecp2b HA NLS SpCas9 NLS spA
4.803 bp ITR ITR
b
H1 gRNA1 gRNA2 gRNA3 hCmpsn GFP KASH bGHpA
3.400 bp ITR ITR
c
control
rMap1b bGH
rMap1b spA
rMecp2 bGH
rMecp2 spA
150
HA-SpCas9
50
Tubulin
d
SpCas9-HA
GFP-KASH
DAPI
Merge
SpCas9-HA
Map1b
DAPI
Merge
SpCas9-HA
NeuN
DAPI
Merge
SpCas9-HA
GFAP
DAPI

FIG. 64A-64D

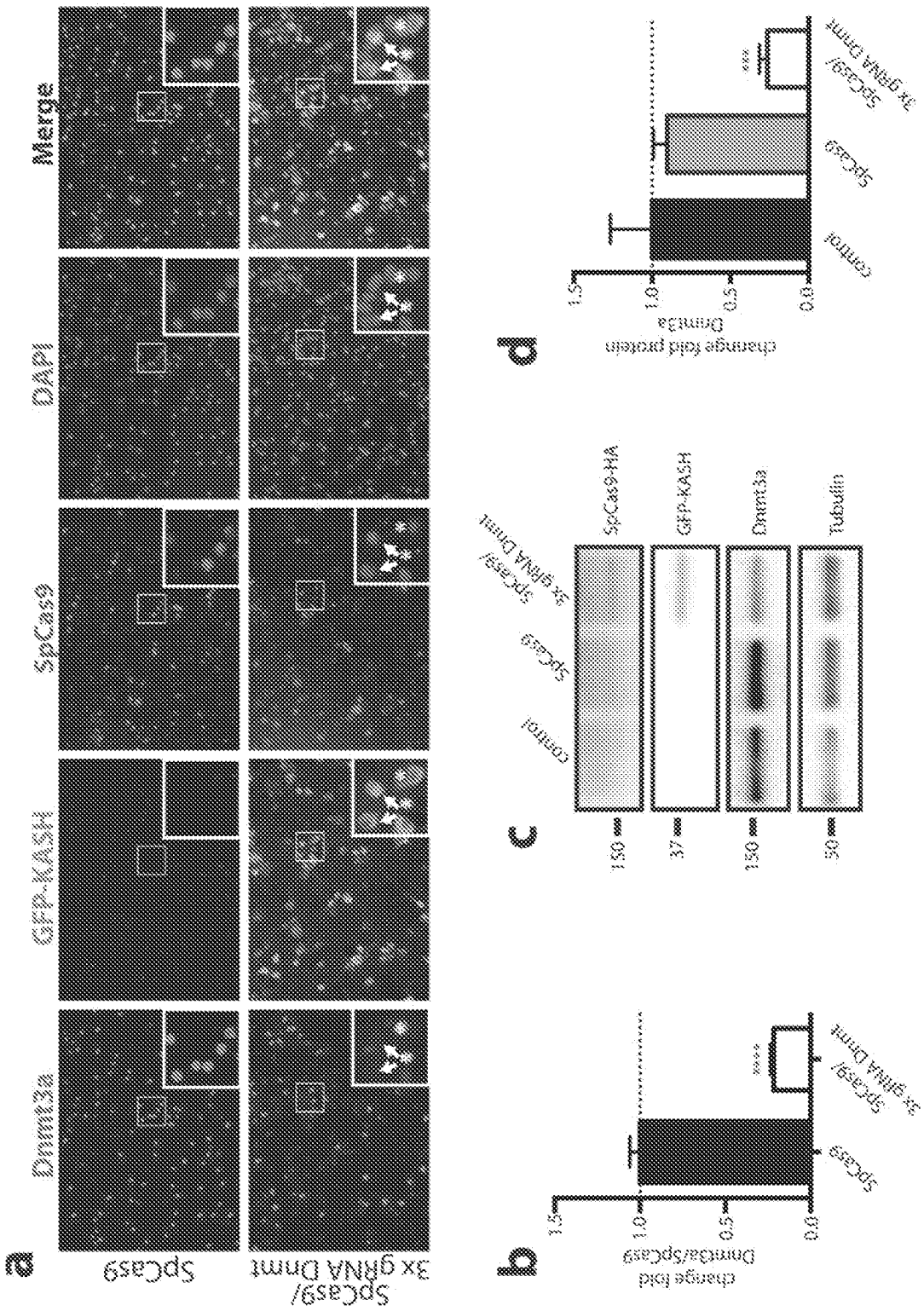

a
Dnmt3a
GFP-KASH
SpCas9
DAPI
Merge
SpCas9
SpCas9/
3x gRNA Dnmt
b
change fold
Dnmt3a/SpCas9
1.5
1.0
0.5
0.0
SpCas9
SpCas9/
3x gRNA Dnmt
c
Control
SpCas9
SpCas9/
3x gRNA Dnmt
SpCas9-HA
GFP-KASH
Dnmt3a
Tubulin
130
37
130
50
d
change fold protein
Dnmt3a
1.5
1.0
0.5
0.0
control
SpCas9
SpCas9/
3x gRNA Dnmt

FIG. 65A-65D

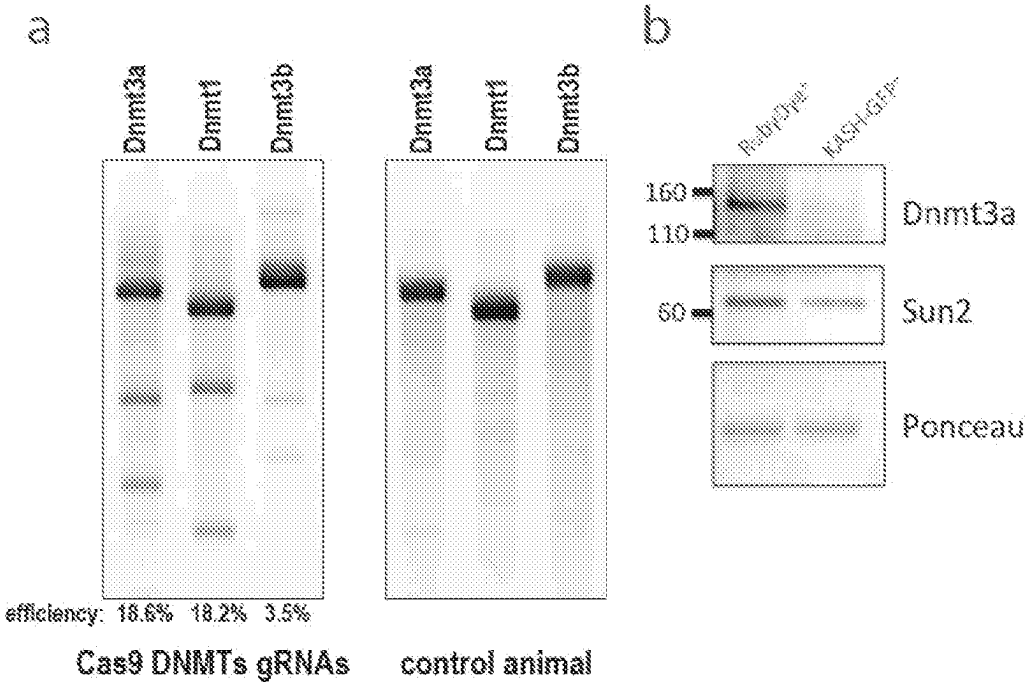

a
Dnmt3a
Dnmt1
Dnmt3b
Dnmt3a
Dnmt1
Dnmt3b
efficiency: 18.6% 18.2% 3.5%
Cas9 DNMTs gRNAs
control animal
b
160
110
60
Dnmt3a
Sun2
Ponceau

FIG. 66A-66B

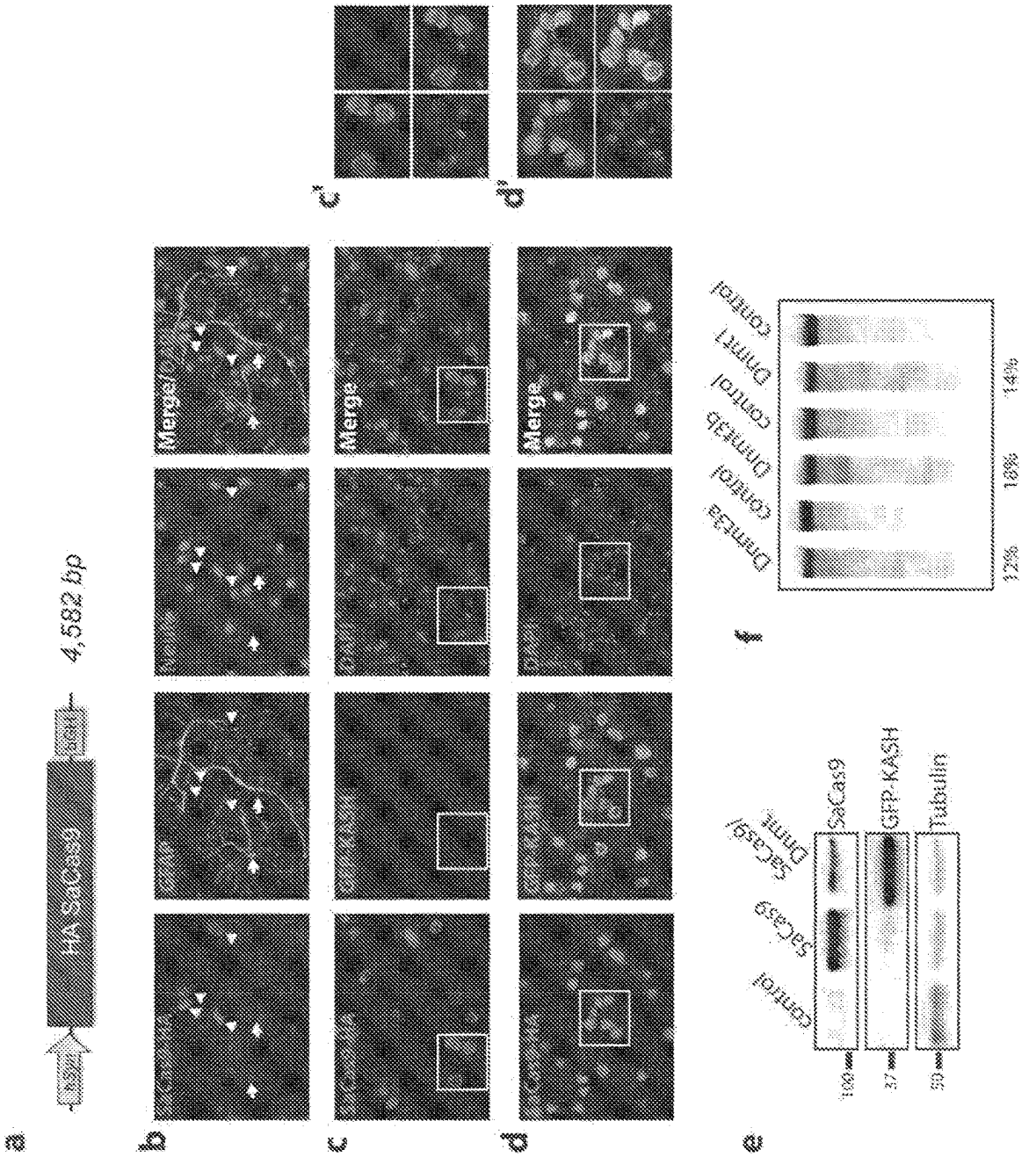

a
HA SaCas9
4,582 bp
b
Merge
c
Merge
d
Merge
c'
d'
e
Control
SaCas9
SaCas9/ Dnmt
SaCas9
GFP-KASH
Tubulin
100
37
50
f
Dnmt3a
Control
Dnmt3b
Control
Dnmt1
Control
12%
18%
14%

FIG. 67A-67F

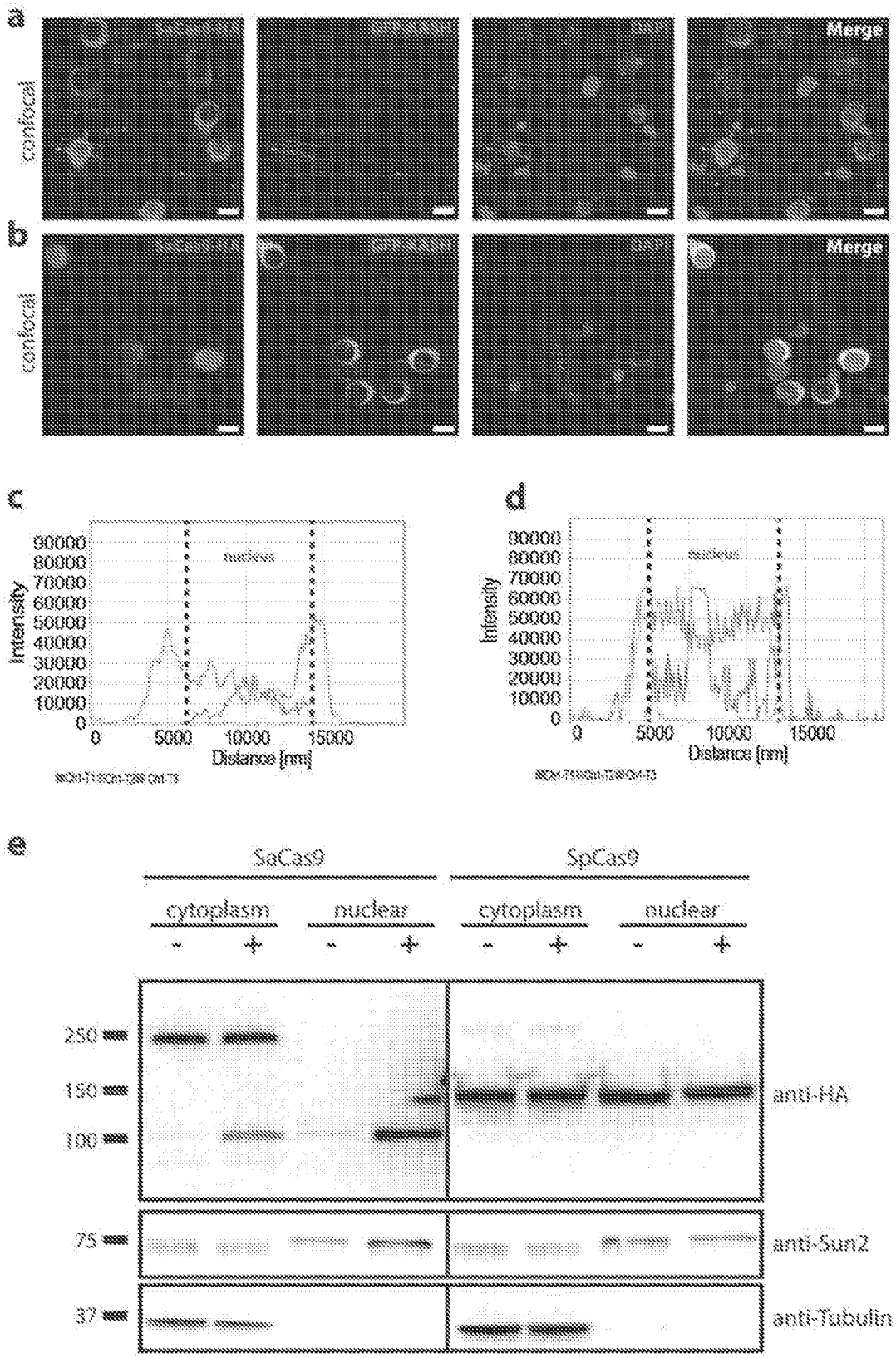
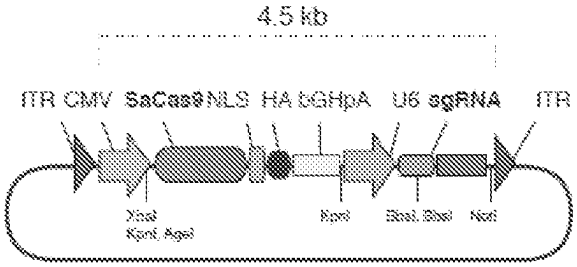
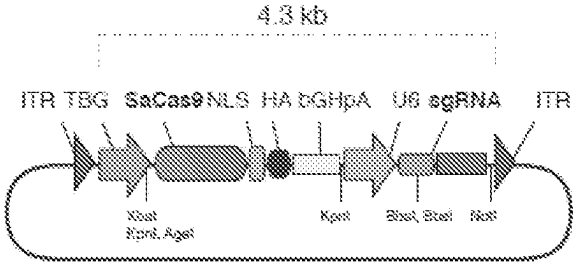

a
confocal
Merge
b
confocal
Merge
c
Intensity
nucleus
Distance [nm]
d
Intensity
nucleus
Distance [nm]
e
SaCas9
SpCas9
cytoplasm
nuclear
cytoplasm
nuclear
-
+
-
+
-
+
-
+
250
150
100
75
37
anti-HA
anti-Sun2
anti-Tubulin

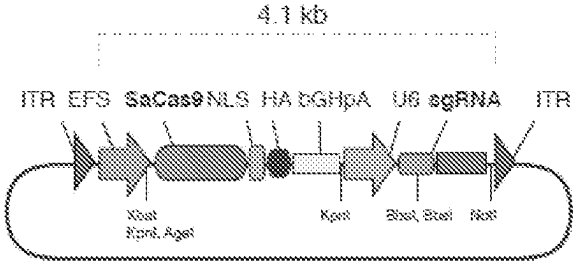
FIG. 68A-68E



AAV-Sa-Cas9 vector:
CMV-SaCas9-NLS-HA-U6-sgRNA



Liver-specific AAV-Sa-Cas9 vector:
TBG-SaCas9-NLS-HA-U6-sgRNA



Alternate AAV-Sa-Cas9 vector:
EFS-SaCas9-NLS-HA-U6-sgRNA

FIG. 69

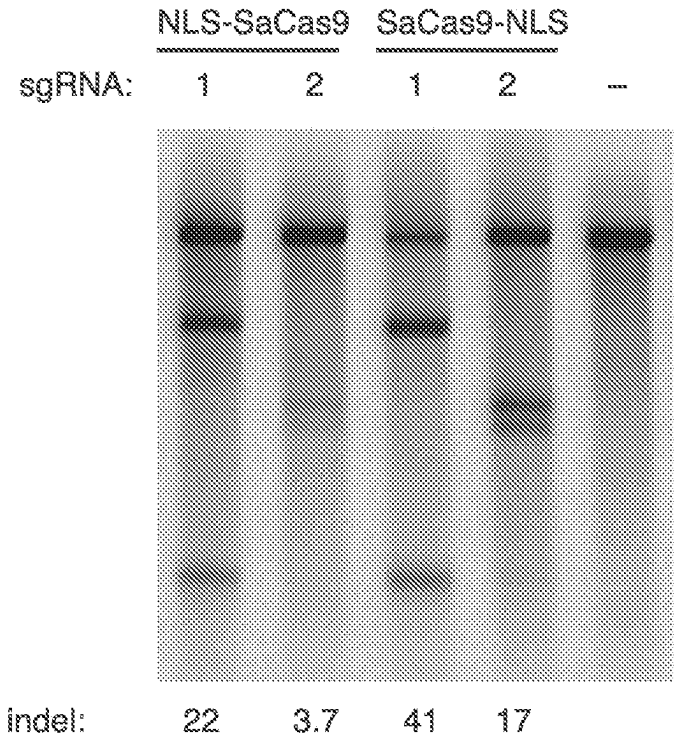


FIG. 70

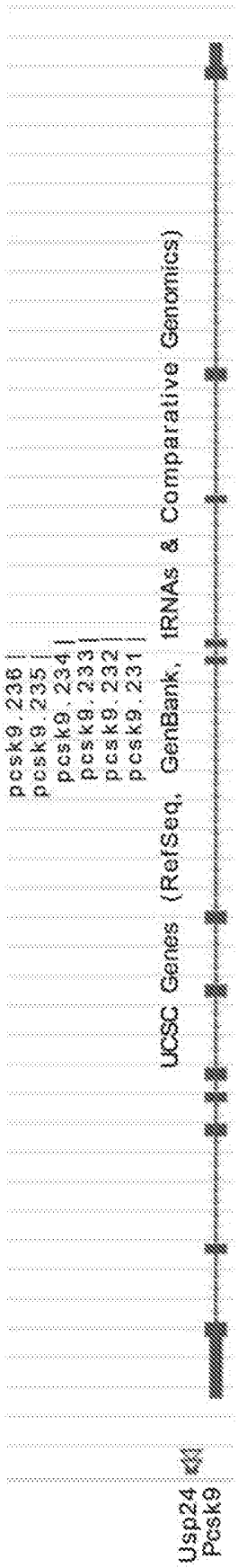

Usp24
Pcsk9
pcsk9.236
pcsk9.235
pcsk9.234
pcsk9.233
pcsk9.232
pcsk9.231
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)

FIG. 71

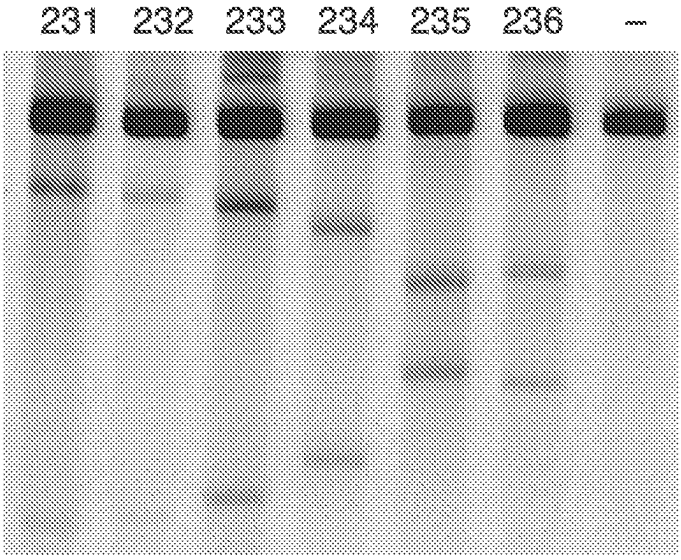


FIG. 72

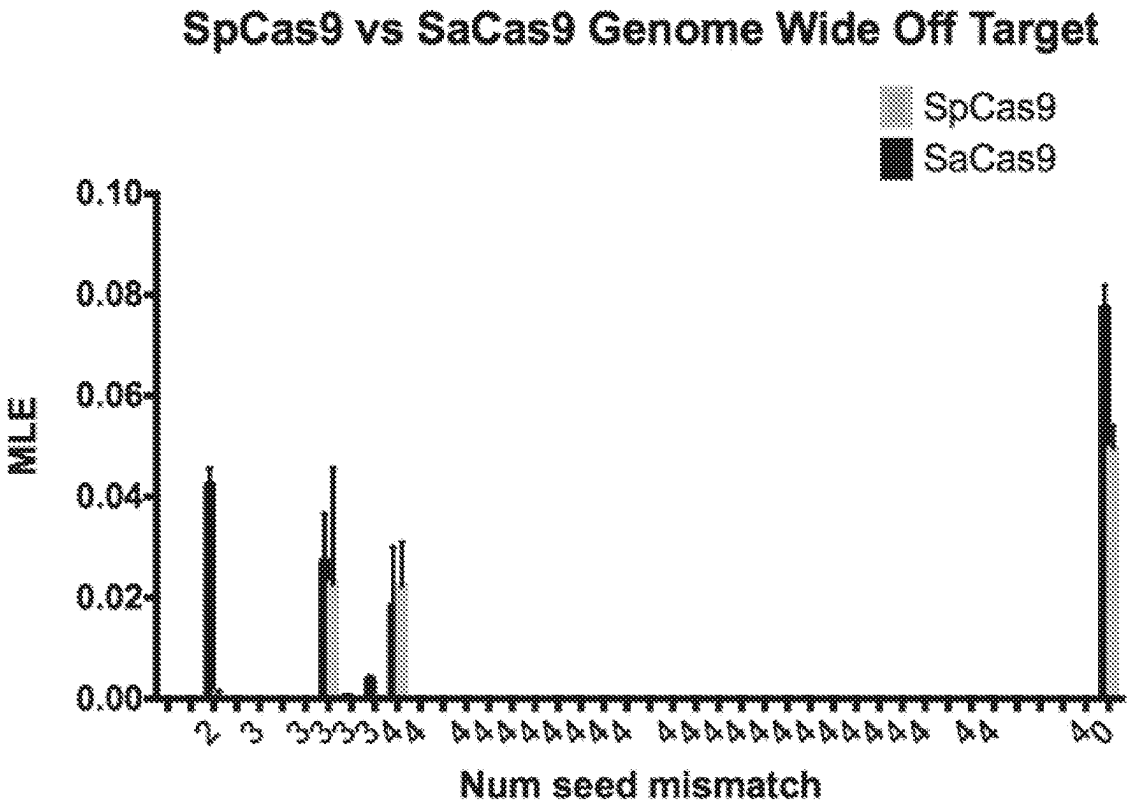

SpCas9 vs SaCas9 Genome Wide Off Target
SpCas9
SaCas9
MLE
0.10
0.08
0.06
0.04
0.02
0.00
Num seed mismatch

FIG. 73

DELIVERY, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION AND THERAPEUTIC APPLICATIONS

RELATED APPLICATIONS AND INCORPORATION BY REFERENCE

[0001] This application is a continuation of U.S. patent application Ser. No. 15/430,260, filed Feb. 10, 2017, which is a continuation of U.S. application Ser. No. 14/183,512 filed Feb. 18, 2014, which is a continuation of U.S. application Ser. No. 14/104,837 filed Dec. 12, 2013 which claims priority to U.S. provisional patent application Nos. 61/736,527 filed Dec. 12, 2012; 61/748,427 filed Jan. 2, 2013; 61/758,468 filed Jan. 30, 2013, 61/769,046 filed Feb. 25, 2013; 61/791,409 and 61/802,174 filed Mar. 15, 2013, 61/806,375 filed Mar. 28, 2013; 61/814,263 filed Apr. 20, 2013; 61/819,803 filed May 6, 2013; 61/828,130 filed May 28, 2013; 61/835,931 and 61/836,123 filed Jun. 17, 2013 and 61/847,537 filed Jul. 17, 2013.

[0002] Reference is also made to U.S. provisional patent application Nos. 61/799,800 filed Mar. 15, 2013; 61/835,931, 61/835,936, 61/836,127, 61/836,101, 61/836,080 and 61/835,973 filed Jun. 17, 2013; 61/862,468 and 61/862,355 filed on Aug. 5, 2013; 61/871,301 filed on Aug. 28, 2013; 61/960,777 filed on Sep. 25, 2013 and 61/961,980 filed on Oct. 28, 2013.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0003] 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.

[0004] 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.

FIELD OF THE INVENTION

[0005] The present invention generally relates to the delivery, engineering, optimization and therapeutic applications 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 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and components thereof.

SEQUENCE LISTING

[0006] The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created on Apr. 16, 2025, is named 114203-6213_SL and is 453,231 bytes in size.

BACKGROUND OF THE INVENTION

[0007] 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. 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 cell-type/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 nucleic 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] 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.

[0011] Aspects of the invention relate to Cas9 enzymes having improved targeting 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 more advanced as there is less coding therefor in the delivery vector than for wild-type Cas9, and/or generating chimeric Cas9 enzymes.

[0012] Also provided are uses of the present sequences, vectors, enzymes or systems, in medicine. Also provided are uses of the same in gene or genome editing.

[0013] 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 transcriptional activation domain may be VP64. In other aspects of the invention, the transcriptional repressor domain 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.

[0014] 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.

[0015] Aspects of the invention also provide for methods of simplifying the cloning and delivery of components of the CRISPR complex. In the 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.

[0016] 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.

[0017] 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.

[0018] 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.

[0019] The catalytic residue in 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.

[0020] 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

[0021] 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, StlCas9 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.

[0022] 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.

[0023] In further embodiments, the invention provides for methods of enhancing the function of Cas9 by generating chimeric Cas9 proteins. Chimeric Cas9 proteins chimeric

Cas9s 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.

[0024] 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.

[0025] 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:

[0026] A)—I. a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises:

[0027] (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell,

[0028] (b) a tracr mate sequence, and

[0029] (c) a tracr sequence, and

[0030] II. a polynucleotide sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences,

[0031] wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

[0032] 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

[0033] 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,

[0034] or

[0035] (B) I. polynucleotides comprising:

[0036] (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, and

[0037] (b) at least one or more tracr mate sequences,

[0038] II. a polynucleotide sequence encoding a CRISPR enzyme, and

[0039] III. a polynucleotide sequence comprising a tracr sequence,

[0040] 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

[0041] 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.

[0042] 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.

[0043] It will be appreciated that where reference is made to a polynucleotide, which is RNA and is said to 'comprise' a feature such a tracr mate sequence, the RNA sequence includes the feature. Where the polynucleotide is DNA and is said to comprise a feature such 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).

[0044] 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.

[0045] 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.

[0046] 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.

[0047] 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.

[0048] 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., and can be part of a vector system as described herein.

[0049] 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.

[0050] 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.

[0051] 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.

[0052] In plants, pathogens are often host-specific. For example, *Fusarium oxysporum* 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.

[0053] 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.

[0054] It will be appreciated that SpCas9 or SaCas9 are those from or derived from *S. pyogenes* or *S. aureus* Cas9.

[0055] 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 by minimizing 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:

[0056] I. a first CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the first polynucleotide sequence comprises:

[0057] (a) a first guide sequence capable of hybridizing to the first target sequence,

[0058] (b) a first tracr mate sequence, and

[0059] (c) a first tracr sequence,

[0060] II. a second CRISPR-Cas system chiRNA polynucleotide sequence, wherein the second polynucleotide sequence comprises:

[0061] (a) a second guide sequence capable of hybridizing to the second target sequence,

[0062] (b) a second tracr mate sequence, and

[0063] (c) a second tracr sequence, and

[0064] 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 tracr 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 by minimizing off-target modifications.

[0065] 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.

[0066] 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 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 5' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs.

[0067] The invention in some embodiments comprehends a method of modifying an organism or a non-human organism by minimizing 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

[0068] I. a first regulatory element operably linked to

[0069] (a) a first guide sequence capable of hybridizing to the first target sequence, and

[0070] (b) at least one or more tracr mate sequences,

[0071] II. a second regulatory element operably linked to

[0072] (a) a second guide sequence capable of hybridizing to the second target sequence, and

[0073] (b) at least one or more tracr mate sequences,

[0074] III. a third regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and

[0075] IV. a fourth regulatory element operably linked to a tracr sequence,

[0076] 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 by minimizing off-target modifications.

[0077] 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, II I and IV may be located on a total of two or three different vectors, with all combinations of locations envisaged, etc.

[0078] 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.

[0079] 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 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 5' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs.

[0080] The invention in some embodiments comprehends a method of modifying a genomic locus of interest by minimizing 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.

[0081] 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 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 5' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs.

[0082] 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.

[0083] 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.

[0084] 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.

[0085] 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.

[0086] 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.

[0087] The invention also comprehends an engineered, non-naturally occurring vector system comprising one or more vectors comprising:

- **[0088]** 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,
- **[0089]** b) a second regulatory element operably linked to a Cas protein,
- **[0090]** 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.

[0091] 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.

[0092] 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.

[0093] 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.

[0094] 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.

[0095] 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.

[0096] 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.

[0097] 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.

[0098] 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.

[0099] 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.

[0100] 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.

[0101] 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.

[0102] 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.

[0103] 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, StlCas9 and so forth.

[0104] 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.

[0105] 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.

[0106] 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.

[0107] For example, a suitable PAM is 5'-NRG or 5'-NNGRR for SpCas9 or SaCas9 enzymes (or derived enzymes), respectively.

[0108] It will be appreciated that SpCas9 or SaCas9 are those from or derived from *S. pyogenes* or *S. aureus* Cas9.

[0109] 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.

[0110] 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.

[0111] These and other embodiments are disclosed or are obvious from and encompassed by, the following Detailed Description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0112] 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:

[0113] 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).

[0114] FIG. **2**A-**2**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. FIG. **2**C discloses SEQ ID NOS 212 and 213, respectively, in order of appearance. FIG. **2**E discloses SEQ ID NOS 214-216, respectively, in order of appearance. FIG. **2**F discloses SEQ ID NOS 217-221, respectively, in order of appearance.

[0115] FIG. **3**A-**3**D shows results of an evaluation of SpCas9 specificity for an example target. FIG. **3**A discloses SEQ ID NOS 222, 215 and 223-233, respectively, in order of appearance. FIG. **3**C discloses SEQ ID NO: 222.

[0116] FIG. **4**A-**4**G show an exemplary vector system and results for its use in directing homologous recombination in eukaryotic cells. FIG. **4**E discloses SEQ ID NO: 234. FIG. **4**F discloses SEQ ID NOS 235 and 236, respectively, in order of appearance. FIG. **4**G discloses SEQ ID NOS 237-241, respectively, in order of appearance.

[0117] FIG. **5** provides a table of protospacer sequences (SEQ ID NOS 93, 92, 91, 242-247, 95, 94 and 248-252, respectively, in order of appearance) 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).

[0118] FIG. **6**A-**6**C shows a comparison of different tracrRNA transcripts for Cas9-mediated gene targeting. FIG. **6**A discloses SEQ ID NOS 253 and 254, respectively, in order of appearance.

[0119] FIG. **7** shows a schematic of a surveyor nuclease assay for detection of double strand break-induced micro-insertions and -deletions.

[0120] FIG. **8**A-**8**B shows exemplary bicistronic expression vectors for expression of CRISPR system elements in eukaryotic cells. FIG. **8**A discloses SEQ ID NOS 255-257, respectively, in order of appearance. FIG. **8**B discloses SEQ ID NOS 258, 182 and 183, respectively, in order of appearance.

[0121] FIG. **9**A-**9**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).

[0122] FIG. **10**A-**10**D shows an exemplary CRISPR system, an example adaptation for expression in eukaryotic cells, and results of tests assessing CRISPR activity. FIG. **10**B discloses SEQ ID NOS 259 and 260, respectively, in order of appearance. FIG. **10**C discloses SEQ ID NO: 261.

[0123] FIG. **11**A-**11**C shows exemplary manipulations of a CRISPR system for targeting of genomic loci in mammalian cells. FIG. **11**A discloses SEQ ID NO: 262. FIG. **11**B discloses SEQ ID NOS 263-265, respectively, in order of appearance.

[0124] FIG. **12**A-**12**B shows the results of a Northern blot analysis of crRNA processing in mammalian cells. FIG. **12**A discloses SEQ ID NO: 266.

[0125] FIG. **13**A-**13**B shows an exemplary selection of protospacers in the human PVALB (SEQ ID NO: 267) and mouse Th loci (SEQ ID NO: 268).

[0126] FIG. **14** shows example protospacer and corresponding PAM sequence targets of the *S. thermophilus* CRISPR system in the human EMX1 locus. FIG. **14** discloses SEQ ID NO: 261.

[0127] FIG. **15** provides a table of sequences (SEQ ID NOS 269-276, 191-192 and 277-278, respectively, in order of appearance) for primers and probes used for Surveyor, RFLP, genomic sequencing, and Northern blot assays.

[0128] FIG. **16**A-**16**C shows exemplary manipulation of a CRISPR system with chimeric RNAs and results of SURVEYOR assays for system activity in eukaryotic cells. FIG. **16**A discloses SEQ ID NO: 279.

[0129] FIG. **17**A-**17**B shows a graphical representation of the results of SURVEYOR assays for CRISPR system activity in eukaryotic cells.

[0130] FIG. **18** shows an exemplary visualization of some *S. pyogenes* Cas9 target sites in the human genome using the UCSC genome browser. FIG. **18** discloses SEQ ID NOS 280-358, respectively, in order of appearance.

[0131] FIG. **19**A-**19**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).

[0132] FIG. **20**A-**20**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).

[0133] FIG. **21**A-**21**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. FIG.

21C discloses SEQ ID NOS 359-361, 359, 362 and 361, respectively, in order of appearance.

[0134] FIG. **22**A-**22**B shows single vector designs for SpCas9. FIG. **22**A discloses SEQ ID NOS 363-365, respectively, in order of appearance. FIG. **22**B discloses SEQ ID NO: 366.

[0135] FIG. **23** shows a graph representing the length distribution of Cas9 orthologs.

[0136] FIG. **24**A-**24**M shows sequences where the mutation points are located within the SpCas9 gene. FIG. **24**A-M discloses the nucleotide sequence as SEQ ID NO: 367 and the amino acid sequence as SEQ ID NO: 368.

[0137] FIG. **25**A shows the Conditional Cas9, Rosa26 targeting vector map.

[0138] FIG. **25**B shows the Constitutive Cas9, Rosa26 targeting vector map.

[0139] FIG. **26** shows a schematic of the important elements in the Constitutive and Conditional Cas9 constructs.

[0140] FIG. **27** shows delivery and in vivo mouse brain Cas9 expression data.

[0141] FIG. **28**A-**28**C 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.

[0142] 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.

[0143] FIG. **30**A-**30**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. FIG. **30**B discloses SEQ ID NOS 359-361, 359, 362 and 361, respectively, in order of appearance.

[0144] FIG. **31**A-**31**C shows the repair strategy for Cystic Fibrosis delta F508 mutation. FIG. **31**A discloses the nucleotide sequence as SEQ ID NO: 371 and the amino acid sequence as 372. FIG. **31**B discloses SEQ ID NO: 279. FIG. **31**C discloses the nucleotide sequence as SEQ ID NO: 373 and the amino acid sequence as SEQ ID NO: 374.

[0145] FIG. **32**A-**32**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.

[0146] 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.

[0147] 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.

[0148] 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.

[0149] 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.

[0150] 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.

[0151] 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.

[0152] 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.

[0153] 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.

[0154] 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.

[0155] 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.

[0156] 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).

[0157] 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).

[0158] 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.

[0159] 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.

[0160] 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.

[0161] 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).

[0162] 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.

[0163] 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.

[0164] FIG. **51** shows (top) a list of spacing (as indicated by the pattern of arrangement for two PAM sequences) between pairs of guide RNAs (SEQ ID NOS 375-391, respectively, in order of appearance). 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.

[0165] FIG. **52** shows a list of U6 reverse primer sequences (SEQ ID NOS 392-438, respectively, in order of appearance) used to generate U6-guide RNA expression casssettes. Each primer needs to be paired with the U6 forward primer "gcactgagggcctatttcccatgattc" (SEQ ID NO: 1) to generate amplicons containing U6 and the desired guide RNA.

[0166] FIG. **53** shows a Genomic sequence map from the human Emx1 locus showing the locations of the **24** patterns listed in FIG. **33**. FIG. **53** discloses the nucleotide sequence as SEQ ID NO: 439 and the amino acid sequences as SEQ ID NOS 440-443, respectively, in order of appearance.

[0167] 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 5' overhang present following cleavage by the Cas9 nickase.

[0168] 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. FIG. **55** discloses the nucleotide sequence as SEQ ID NO: 439 and the amino acid sequences as SEQ ID NOS 440-443, respectively, in order of appearance.

[0169] FIG. **56** shows staining of HA-SpCas9 in dorsal and ventral hippocampus 8 weeks after injection of viruses encoding Mecp2-HA-SpCas9 and 3×gRNA-TETS with Syn-KASH-GFP.

[0170] FIG. **57** shows Syn_GFP-KASH expression 8 weeks after 3×gRNA virus injection is specific for neurons (NeuN positive cells) and not for glia cells (GFAP positive).

[0171] FIG. **58** shows behavior tests conducted 5 weeks after CRISPR-mediated KD of TETs and DNMTs in dentate gyrus (ventral and dorsal part) showed increased level of anxiety and learning deficits. A) time spend in the open arm during elevated plus maze test. B) open field test, time spent in the center of arena vs time in the corners was measured. C) Novel object recognition test, results were measured 3h after familiarization phase. D) Barnes maze; efficiency in finding escape within 3 days of training. E) Barnes maze results. F) Freezing behavior during contextual fear conditioning. G) Latency to first freezing episode during contextual fear conditioning. H) Trace fear conditioning results for TETs KD and DNMTs KD (I). Control-animals injected

with SpCas9 virus and GFP-KASH construct without gRNAs. TETs-animals injected with both SpCas9 and construct encoding gRNAs against Tet1, Tet2 and Tet3. DNMTs-animals injected with both SpCas9 and construct encoding gRNAs against Dnmt1, Dnmt3a and Dnmt3b.

[0172] FIG. **59** shows cutting efficiency of Tet loci in the brain, 8 weeks after Mecp_SpCas9 virus injection in compare to control animals injected with Mecp2_SpCas9 virus only.

[0173] FIG. **60** shows cutting efficiency of Dnmt loci in the brain, 8 weeks after Mecp_SpCas9 virus injection in compare to control animals injected with Mecp2_SpCas9 virus only.

[0174] FIG. **61** shows Dnmt3a staining in the brain, 8 weeks after stereotaxic injection of virus encoding Mecp2_SpCas9 and gRNAs targeting Dnmt loci. Bottom panel shows magnification of ROI indicated on the upper panel.

[0175] FIG. **62** shows staining of Syn_HA-SaCas9 in the dorsal hippocampus, 4 weeks after injection of virus. First column shows animal injected with Sa-Cas9 only, middle column animal injected with both SaCas9 and gRNAs against TETs loci and the right column represents animal injected with only gRNAs encoding virus. SaCas9 nuclear localization depends on the presence of gRNA.

[0176] FIG. **63**A-**63**E shows SpCas9 in N2a cells. A) Targeting- and SpCas9 expression vector. B) Western Blot analysis of N2a cells expressing HA-tagged SpCas9 under the control of different promoters. C) Cutting efficiency of Dnmt loci. D) Western blot analysis demonstrating efficient knock down of Dnmt3a. e) Cutting efficiency of Tet loci.

[0177] FIG. **64**A-**64**D shows SpCas9 in primary neurons. A) Schematic overview of SpCas9 cloning strategies used in this study. Short promoters and short polyA for efficient packaging into AAV delivering system. B) Schematic overview of combined multiplex targeting and nuclear envelope labeling strategy. C) Western blot analysis showing expression of HA-tagged SpCas9 under the control of rMap1b and mMecp2 promoter and bGH and spA signal. D) Immunocytochemistry demonstrating co-expression of SpCas9 and GFP-KASH in primary neurons. SpCas9 under the control of the mMecp2 promoter is expressed in neurons (Map1b, NeuN) but not in astroglia cells (GFAP).

[0178] FIG. **65**A-**65**D shows knock down of Dnmt3a in primary neurons. A) Immunocytochemistry demonstrating efficient knock down of Dnmt3a after targeting with multiplex targeting vector and mMecp2-SpCas9. B) Quantification of Dnmt3a antibody staining in control and targeted neurons. C) Western blot analysis demonstrating reduced Dnmt3a protein level. D) Quantification of Western Blot analysis demonstrating a total knock down of Dnmt3a protein level of approx. 75% in a mixed primary neuron culture (neurons and astroglia).

[0179] FIG. **66**A-**66**B shows knock down of Dnmt3a in vivo. A) Cutting efficiency of Dnmt loci in the brain, 8 weeks after Mecp_SpCas9 virus injection in compare to control animals injected with Mecp2_SpCas9 virus only. B) Western blot analysis showing reduced Dnmt3a protein level in targeted neuronal nuclei (KASH-GFP positive) compared to control nuclei (RubyDye positive) after sorting cell nuclei using FACS.

[0180] FIG. **67**A-**67**F shows expression of SaCas9 in primary neurons. A) Size of SaCas9 expression vector using hSynapsin promoter and bGH signal. B) Expression of SaCas9 in primary neurons (NeuN) but not in astroglia (GFAP). C) Extranuclear localization of SaCas9 in absence of gRNA. C' Higher magnification of SaCas9 positive neurons shown in C). D) Nuclear localization of SaCas9 in presence of gRNA. D') Higher magnification of SaCas9 positive neurons shown in D). E) Western blot analysis demonstrating expression of HA-tagged SaCas9 and GFP-KASH. F) Cutting efficiency of Dnmt loci 1 week after AAV infection.

[0181] FIG. **68**A-**68**E shows gRNA dependent nuclear localization of SaCas9. A) Confocal imaging analysis demonstrating extranuclear localization of SaCas9 in absence of gRNA in primary neurons. B) Nuclear localization of SaCas9 in presence of gRNA. C) Line Scan analysis of confocal picture A) showing extranuclear localization of SaCas9 in absence of gRNA (red, SaCas9 signal; blue, DAPI signal; green, GFP-KASH signal. D) Line Scan analysis of confocal picture B) showing nuclear localization of SaCas9 in presence of gRNA (red, SaCas9 signal; blue, DAPI signal; green, GFP-KASH signal). E) Subcellular localization of SaCas9 and SpCas9 under conditions without (−) and with (+) gRNA in N2a cells. SaCas9 signal at 250 kDa in the cytoplasm fraction (Tubulin positive) indicating dimerization of SaCas9 in the cytoplasm. In the presence of gRNA a shift of SaCas9 protein into the nuclear fraction (Sun2 positive) is visible. SaCas9 signal at 100 kDa indicates a gRNA dependent formation of SaCas9 homomers and transport into the cell nucleus. In contrast, SpCas9 is mainly present as homomer and its nuclear localization is independent of gRNA.

[0182] FIG. **69** shows an AAV-Sa-Cas9 vector, a liver-specific AAV-Sa-Cas9 vector and an alternate AAV-Sa-Cas9 vector.

[0183] FIG. **70** shows data on optimized CMV-SaCas9-NLS-U6-sgRNA vector (submitted vector design last time); new data compares N'-term vs C'-term tagged SaCas9 and shows enhanced cleavage efficiency using C'-term NLS tagging.

[0184] FIG. **71** shows SURVEYOR image showing indels generated by new Pcsk9 targets.

[0185] FIG. **72** shows SaCas9 specificity: genome-wide off target sites (GWOTs) are predicted based on 2 criteria: they contain 4 or fewer mismatched bases to intended SaCas9 target and bear the least restrictive PAM for SaCas9, NNGRR. HEK 293FT cells are transfected with either SpCas9 or SaCas9 with their corresponding sgRNAs at a target site (EMX1: TAGGGTTAGGGGCCCCAGGC (SEQ ID NO: 2)) that has CGGGGT as a PAM (sequence including PAM disclosed as SEQ ID NO: 369) so that it can be cut by either SpCas9 (CGG) or SaCas9 (CGGGGT). DNAs from cells are harvested and analyzed for indels by Illumina sequencing at on-target and 41 predicted off-target loci (following protocols from Hsu et al. Nature Biotech 2013 and data analysis pipeline developed by David Scott and Josh Weinstein).

[0186] FIG. **73** shows that that SaCas9 may have a higher level of off-target activity than SpCas9 at certain loci.

[0187] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE INVENTION

[0188] 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.

[0189] 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.

Cas9

[0190] 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 6. 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.

[0191] The Cas9 may be used as a generic DNA binding protein. For example, and as shown in Example 7, 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.

[0192] 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.

[0193] 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.

Transgenic Animals and Plants

[0194] 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.

[0195] 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.

[0196] 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.

[0197] 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 over-express high levels of oil or alcohols for use in the oil or biofuel industries.

Adeno Associated Virus (AAV)

[0198] In terms of in vivo delivery, AAV is advantageous over other viral vectors for a couple of reasons:

[0199] Low toxicity (this may be due to the purification method not requiring ultra centrifugation of cell particles that can activate the immune response)

[0200] Low probability of causing insertional mutagenesis because it doesn't integrate into the host genome.

[0201] 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

[0202] These species are therefore, in general, preferred Cas9 species. Applicants have shown delivery and in vivo mouse brain Cas9 expression data.

[0203] Two ways to package Cas9 coding nucleic acid molecules, e.g., DNA, into viral vectors to mediate genome modification in vivo are preferred:

[0204] To achieve NHEJ-mediated gene knockout:

Single Virus Vector:

[0205] Vector containing two or more expression cassettes:

[0206] Promoter-Cas9 coding nucleic acid molecule-terminator

[0207] Promoter-gRNA1-terminator

[0208] Promoter-gRNA2-terminator

[0209] Promoter-gRNA (N)-terminator (up to size limit of vector)

Double Virus Vector:

[0210] Vector 1 containing one expression cassette for driving the expression of Cas9

[0211] Promoter-Cas9 coding nucleic acid molecule-terminator

[0212] Vector 2 containing one more expression cassettes for driving the expression of one or more guideRNAs

[0213] Promoter-gRNA1-terminator

[0214] Promoter-gRNA (N)-terminator (up to size limit of vector)

[0215] 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.

[0216] Promoter used to drive Cas9 coding nucleic acid molecule expression can include:

[0217] 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.

[0218] For ubiquitous expression, can use promoters: CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains, etc.

[0219] For brain expression, can use promoters: SynapsinI for all neurons, CaMKIIalpha for excitatory neurons, GAD67 or GAD65 or VGAT for GABAergic neurons, etc.

[0220] For liver expression, can use Albumin promoter.

[0221] For lung expression, can use SP-B.

[0222] For endothelial cells, can use ICAM.

[0223] For hematopoietic cells can use IFNbeta or CD45.

[0224] For Osteoblasts can use OG-2.

[0225] Promoter used to drive guide RNA can include:

[0226] Pol III promoters such as U6 or H1

[0227] Use of Pol II promoter and intronic cassettes to express gRNA

[0228] 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.

[0229] 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.

[0230] 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.

[0231] Various means of delivery are described herein, and further discussed in this section.

[0232] 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.

[0233] 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.

[0234] 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 $1^*\times10^9$-1×10^{10} particles or about 1×10^9-1×10^{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.

[0235] 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.

[0236] 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.

[0237] 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.

Lentivirus

[0238] 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.

[0239] 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 ug 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.

[0240] 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.

[0241] 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.

[0242] 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.).

[0243] 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.

RNA Delivery

[0244] 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-poly A 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.

[0245] To enhance expression and reduce toxicity, the CRISPR enzyme and/or guide RNA can be modified using pseudo-U or 5-Methyl-C.

[0246] mRNA delivery methods are especially promising for liver delivery currently. In particular, for AAV8 is particularly preferred for delivery to the liver.

Nanoparticles

[0247] CRISPR enzyme mRNA and guide RNA may be delivered simultaneously using nanoparticles or lipid envelopes.

[0248] For example, Su X, Fricke J, Kavanagh D G, Irvine D J ("In vitro and in vivo mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles" Mol Pharm. 2011 Jun. 6; 8(3):774-87. doi: 10.1021/mp100390w. Epub 2011 Apr. 1) describes biodegradable core-shell structured nanoparticles with a poly(β-amino ester) (PBAE) core enveloped by a phospholipid bilayer shell. These were developed for in vivo mRNA delivery. The pH-responsive PBAE component was chosen to promote endosome disruption, while the lipid surface layer was selected to minimize toxicity of the polycation core. Such are, therefore, preferred for delivering RNA of the present invention.

[0249] In one embodiment, nanoparticles based on self assembling bioadhesive polymers are contemplated, which may be applied to oral delivery of peptides, intravenous delivery of peptides and nasal delivery of peptides, all to the brain. Other embodiments, such as oral absorption and ocular deliver of hydrophobic drugs are also contemplated. The molecular envelope technology involves an engineered polymer envelope which is protected and delivered to the site of the disease (see, e.g., Mazza, M. et al. ACSNano, 2013. 7 (2): 1016-1026; Siew, A., et al. Mol Pharm, 2012. 9(1):14-28; Lalatsa, A., et al. J Contr Rel, 2012. 161(2):523-36; Lalatsa, A., et al., Mol Pharm, 2012. 9(6):1665-80; Lalatsa, A., et al. Mol Pharm, 2012. 9(6):1764-74; Garrett, N. L., et al. J Biophotonics, 2012. 5(5-6):458-68; Garrett, N. L., et al. J Raman Spect, 2012. 43(5):681-688; Ahmad, S., et al. J Royal Soc Interface 2010. 7:S423-33; Uchegbu, I. F. Expert Opin Drug Deliv, 2006. 3(5):629-40; Qu, X., et al. Biomacromolecules, 2006. 7(12):3452-9 and Uchegbu, I. F., et al. Int J Pharm, 2001. 224:185-199). Doses of about 5 mg/kg are contemplated, with single or multiple doses, depending on the target tissue.

[0250] In one embodiment, nanoparticles that can deliver RNA to a cancer cell to stop tumor growth developed by Dan

Anderson's lab at MIT may be used/and or adapted to the CRISPR Cas system of the present invention. In particular, the Anderson lab developed fully automated, combinatorial systems for the synthesis, purification, characterization, and formulation of new biomaterials and nanoformulations. See, e.g., Alabi et al., Proc Natl Acad Sci USA. 2013 Aug. 6; 110(32):12881-6; Zhang et al., Adv Mater. 2013 Sep. 6; 25(33):4641-5; Jiang et al., Nano Lett. 2013 Mar. 13; 13(3):1059-64; Karagiannis et al., ACS Nano. 2012 Oct. 23; 6(10):8484-7; Whitehead et al., ACS Nano. 2012 Aug. 28; 6(8):6922-9 and Lee et al., Nat Nanotechnol. 2012 Jun. 3; 7(6):389-93.

[0251] U.S. patent application No. 20110293703 relates to lipidoid compounds are also particularly useful in the administration of polynucleotides, which may be applied to deliver the CRISPR Cas system of the present invention. In one aspect, the aminoalcohol lipidoid compounds are combined with an agent to be delivered to a cell or a subject to form microparticles, nanoparticles, liposomes, or micelles. The agent to be delivered by the particles, liposomes, or micelles may be in the form of a gas, liquid, or solid, and the agent may be a polynucleotide, protein, peptide, or small molecule. The minoalcohol lipidoid compounds may be combined with other aminoalcohol lipidoid compounds, polymers (synthetic or natural), surfactants, cholesterol, carbohydrates, proteins, lipids, etc. to form the particles. These particles may then optionally be combined with a pharmaceutical excipient to form a pharmaceutical composition.

[0252] US Patent Publication No. 0110293703 also provides methods of preparing the aminoalcohol lipidoid compounds. One or more equivalents of an amine are allowed to react with one or more equivalents of an epoxide-terminated compound under suitable conditions to form an aminoalcohol lipidoid compound of the present invention. In certain embodiments, all the amino groups of the amine are fully reacted with the epoxide-terminated compound to form tertiary amines. In other embodiments, all the amino groups of the amine are not fully reacted with the epoxide-terminated compound to form tertiary amines thereby resulting in primary or secondary amines in the aminoalcohol lipidoid compound. These primary or secondary amines are left as is or may be reacted with another electrophile such as a different epoxide-terminated compound. As will be appreciated by one skilled in the art, reacting an amine with less than excess of epoxide-terminated compound will result in a plurality of different aminoalcohol lipidoid compounds with various numbers of tails. Certain amines may be fully functionalized with two epoxide-derived compound tails while other molecules will not be completely functionalized with epoxide-derived compound tails. For example, a diamine or polyamine may include one, two, three, or four epoxide-derived compound tails off the various amino moieties of the molecule resulting in primary, secondary, and tertiary amines. In certain embodiments, all the amino groups are not fully functionalized. In certain embodiments, two of the same types of epoxide-terminated compounds are used. In other embodiments, two or more different epoxide-terminated compounds are used. The synthesis of the aminoalcohol lipidoid compounds is performed with or without solvent, and the synthesis may be performed at higher temperatures ranging from 30.-100 C., preferably at approximately 50.-90 C. The prepared aminoalcohol lipidoid compounds may be optionally purified. For example, the mixture of aminoalcohol lipidoid compounds may be purified to yield an aminoalcohol lipidoid compound with a particular number of epoxide-derived compound tails. Or the mixture may be purified to yield a particular stereo- or regioisomer. The aminoalcohol lipidoid compounds may also be alkylated using an alkyl halide (e.g., methyl iodide) or other alkylating agent, and/or they may be acylated.

[0253] US Patent Publication No. 0110293703 also provides libraries of aminoalcohol lipidoid compounds prepared by the inventive methods. These aminoalcohol lipidoid compounds may be prepared and/or screened using high-throughput techniques involving liquid handlers, robots, microtiter plates, computers, etc. In certain embodiments, the aminoalcohol lipidoid compounds are screened for their ability to transfect polynucleotides or other agents (e.g., proteins, peptides, small molecules) into the cell.

[0254] US Patent Publication No. 20130302401 relates to a class of poly (beta-amino alcohols) (PBAAs) has been prepared using combinatorial polymerization. The inventive PBAAs may be used in biotechnology and biomedical applications as coatings (such as coatings of films or multilayer films for medical devices or implants), additives, materials, excipients, non-biofouling agents, micropatterning agents, and cellular encapsulation agents. When used as surface coatings, these PBAAs elicited different levels of inflammation, both in vitro and in vivo, depending on their chemical structures. The large chemical diversity of this class of materials allowed us to identify polymer coatings that inhibit macrophage activation in vitro. Furthermore, these coatings reduce the recruitment of inflammatory cells, and reduce fibrosis, following the subcutaneous implantation of carboxylated polystyrene microparticles. These polymers may be used to form polyelectrolyte complex capsules for cell encapsulation. The invention may also have many other biological applications such as antimicrobial coatings, DNA or siRNA delivery, and stem cell tissue engineering. The teachings of US Patent Publication No. 20130302401 may be applied to the CRISPR Cas system of the present invention.

[0255] In another embodiment, lipid nanoparticles (LNPs) are contemplated. In particular, an antitransthyretin small interfering RNA encapsulated in lipid nanoparticles (see, e.g., Coelho et al., N Engl J Med 2013; 369:819-29) may be applied to the CRISPR Cas system of the present invention. Doses of about 0.01 to about 1 mg per kg of body weight administered intravenously are contemplated. Medications to reduce the risk of infusion-related reactions are contemplated, such as dexamethasone, acetampinophen, diphenhydramine or cetirizine, and ranitidine are contemplated. Multiple doses of about 0.3 mg per kilogram every 4 weeks for five doses are also contemplated.

[0256] LNPs have been shown to be highly effective in delivering siRNAs to the liver (see, e.g., Tabernero et al., Cancer Discovery, April 2013, Vol. 3, No. 4, pages 363-470) and are therefore contemplated for delivering CRISPR Cas to the liver. A dosage of about four doses of 6 mg/kg of the LNP every two weeks may be contemplated. Tabernero et al. demonstrated that tumor regression was observed after the first 2 cycles of LNPs dosed at 0.7 mg/kg, and by the end of 6 cycles the patient had achieved a partial response with complete regression of the lymph node metastasis and substantial shrinkage of the liver tumors. A complete response was obtained after 40 doses in this patient, who has remained in remission and completed treatment after receiv-

ing doses over 26 months. Two patients with RCC and extrahepatic sites of disease including kidney, lung, and lymph nodes that were progressing following prior therapy with VEGF pathway inhibitors had stable disease at all sites for approximately 8 to 12 months, and a patient with PNET and liver metastases continued on the extension study for 18 months (36 doses) with stable disease.

[0257] However, the charge of the LNP must be taken into consideration. As cationic lipids combined with negatively charged lipids to induce nonbilayer structures that facilitate intracellular delivery. Because charged LNPs are rapidly cleared from circulation following intravenous injection, ionizable cationic lipids with pKa values below 7 were developed (see, e.g., Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-220 December 2011). Negatively charged polymers such as siRNA oligonucleotides may be loaded into LNPs at low pH values (e.g., pH 4) where the ionizable lipids display a positive charge. However, at physiological pH values, the LNPs exhibit a low surface charge compatible with longer circulation times. Four species of ionizable cationic lipids have been focused upon, namely 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxy-keto-N,N-dimethyl-3-aminopropane (DLinKDMA), and 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA). It has been shown that LNP siRNA systems containing these lipids exhibit remarkably different gene silencing properties in hepatocytes in vivo, with potencies varying according to the series DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP employing a Factor VII gene silencing model (see, e.g., Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-220 December 2011). A dosage of 1 μg/ml levels may be contemplated, especially for a formulation containing DLinKC2-DMA.

[0258] Preparation of LNPs and CRISPR Cas encapsulation may be used/and or adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-220 December 2011). The cationic lipids 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA), (3-o-[2"-(methoxypolyethyleneglycol 2000) succinoyl]-1,2-dimyristoyl-sn-glycol (PEG-S-DMG), and R-3-[(ω-methoxy-poly(ethylene glycol)2000) carbamoyl]-1,2-dimyristyloxlpropyl-3-amine (PEG-C-DOMG) may be provided by Tekmira Pharmaceuticals (Vancouver, Canada) or synthesized. Cholesterol may be purchased from Sigma (St Louis, MO). The specific CRISPR Cas RNA may be encapsulated in LNPs containing DLinDAP, DLinDMA, DLinK-DMA, and DLinKC2-DMA (cationic lipid: DSPC: CHOL:PEGS-DMG or PEG-C-DOMG at 40:10:40:10 molar ratios). When required, 0.2% SP-DiOC18 (Invitrogen, Burlington, Canada) may be incorporated to assess cellular uptake, intracellular delivery, and biodistribution. Encapsulation may be performed by dissolving lipid mixtures comprised of cationic lipid: DSPC: cholesterol: PEG-c-DOMG (40:10:40:10 molar ratio) in ethanol to a final lipid concentration of 10 mmol/l. This ethanol solution of lipid may be added drop-wise to 50 mmol/l citrate, pH 4.0 to form multilamellar vesicles to produce a final concentration of 30% ethanol vol/vol. Large unilamellar vesicles may be formed following extrusion of multilamellar vesicles through two stacked 80 nm Nuclepore polycarbonate filters using the Extruder (Northern Lipids, Vancouver, Canada). Encapsulation may be achieved by adding RNA dissolved at 2 mg/ml in 50 mmol/l citrate, pH 4.0 containing 30% ethanol vol/vol drop-wise to extruded preformed large unilamellar vesicles and incubation at 31° C. for 30 minutes with constant mixing to a final RNA/lipid weight ratio of 0.06/1 wt/wt. Removal of ethanol and neutralization of formulation buffer were performed by dialysis against phosphate-buffered saline (PBS), pH 7.4 for 16 hours using Spectra/Por 2 regenerated cellulose dialysis membranes. Nanoparticle size distribution may be determined by dynamic light scattering using a NICOMP 370 particle sizer, the vesicle/intensity modes, and Gaussian fitting (Nicomp Particle Sizing, Santa Barbara, CA). The particle size for all three LNP systems may be ~70 nm in diameter. siRNA encapsulation efficiency may be determined by removal of free siRNA using VivaPureD MiniH columns (Sartorius Stedim Biotech) from samples collected before and after dialysis. The encapsulated RNA may be extracted from the eluted nanoparticles and quantified at 260 nm. siRNA to lipid ratio was determined by measurement of cholesterol content in vesicles using the Cholesterol E enzymatic assay from Wako Chemicals USA (Richmond, VA).

[0259] Preparation of large LNPs may be used/and or adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-220 December 2011. A lipid premix solution (20.4 mg/ml total lipid concentration) may be prepared in ethanol containing DLinKC2-DMA, DSPC, and cholesterol at 50:10:38.5 molar ratios. Sodium acetate may be added to the lipid premix at a molar ratio of 0.75:1 (sodium acetate: DLinKC2-DMA). The lipids may be subsequently hydrated by combining the mixture with 1.85 volumes of citrate buffer (10 mmol/l, pH 3.0) with vigorous stirring, resulting in spontaneous liposome formation in aqueous buffer containing 35% ethanol. The liposome solution may be incubated at 37° C. to allow for time-dependent increase in particle size. Aliquots may be removed at various times during incubation to investigate changes in liposome size by dynamic light scattering (Zetasizer Nano Z S, Malvern Instruments, Worcestershire, UK). Once the desired particle size is achieved, an aqueous PEG lipid solution (stock=10 mg/ml PEG-DMG in 35% (vol/vol) ethanol) may be added to the liposome mixture to yield a final PEG molar concentration of 3.5% of total lipid. Upon addition of PEG-lipids, the liposomes should their size, effectively quenching further growth. RNA may then be added to the empty liposomes at an siRNA to total lipid ratio of approximately 1:10 (wt: wt), followed by incubation for 30 minutes at 37° C. to form loaded LNPs. The mixture may be subsequently dialyzed overnight in PBS and filtered with a 0.45-μm syringe filter.

[0260] Spherical Nucleic Acid (SNA™) constructs and other nanoparticles (particularly gold nanoparticles) are also contemplate as a means to delivery CRISPR/Cas system to intended targets. Significant data show that AuraSense Therapeutics' Spherical Nucleic Acid (SNA™) constructs, based upon nucleic acid-functionalized gold nanoparticles, are superior to alternative platforms based on multiple key success factors, such as:

[0261] High in vivo stability. Due to their dense loading, a majority of cargo (DNA or siRNA) remains bound to the constructs inside cells, conferring nucleic acid stability and resistance to enzymatic degradation.

[0262] Deliverability. For all cell types studied (e.g., neurons, tumor cell lines, etc.) the constructs demonstrate a transfection efficiency of 99% with no need for carriers or transfection agents.

[0263] Therapeutic targeting. The unique target binding affinity and specificity of the constructs allow exquisite specificity for matched target sequences (i.e., limited off-target effects).

[0264] Superior efficacy. The constructs significantly outperform leading conventional transfection reagents (Lipofectamine 2000 and Cytofectin).

[0265] Low toxicity. The constructs can enter a variety of cultured cells, primary cells, and tissues with no apparent toxicity.

[0266] No significant immune response. The constructs elicit minimal changes in global gene expression as measured by whole-genome microarray studies and cytokine-specific protein assays.

[0267] Chemical tailorability. Any number of single or combinatorial agents (e.g., proteins, peptides, small molecules) can be used to tailor the surface of the constructs.

[0268] This platform for nucleic acid-based therapeutics may be applicable to numerous disease states, including inflammation and infectious disease, cancer, skin disorders and cardiovascular disease.

[0269] Citable literature includes: Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang et al., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi et al., Proc. Natl. Acad. Sci. USA. 2013 110(19):7625-7630, Jensen et al., Sci. Transl. Med. 5, 209ra152 (2013) and Mirkin, et al., Small, doi.org/10.1002/smll.201302143.

[0270] Self-assembling nanoparticles with siRNA may be constructed with polyethyleneimine (PEI) that is PEGylated with an Arg-Gly-Asp (RGD) peptide ligand attached at the distal end of the polyethylene glycol (PEG), for example, as a means to target tumor neovasculature expressing integrins and used to deliver siRNA inhibiting vascular endothelial growth factor receptor-2 (VEGF R2) expression and thereby tumor angiogenesis (see, e.g., Schiffelers et al., Nucleic Acids Research, 2004, Vol. 32, No. 19). Nanoplexes may be prepared by mixing equal volumes of aqueous solutions of cationic polymer and nucleic acid to give a net molar excess of ionizable nitrogen (polymer) to phosphate (nucleic acid) over the range of 2 to 6. The electrostatic interactions between cationic polymers and nucleic acid resulted in the formation of polyplexes with average particle size distribution of about 100 nm, hence referred to here as nanoplexes. A dosage of about 100 to 200 mg of CRISPR Cas is envisioned for delivery in the self-assembling nanoparticles of Schiffelers et al.

[0271] The nanoplexes of Bartlett et al. (PNAS, Sep. 25, 2007, vol. 104, no. 39) may also be applied to the present invention. The nanoplexes of Bartlett et al. are prepared by mixing equal volumes of aqueous solutions of cationic polymer and nucleic acid to give a net molar excess of ionizable nitrogen (polymer) to phosphate (nucleic acid) over the range of 2 to 6. The electrostatic interactions between cationic polymers and nucleic acid resulted in the formation of polyplexes with average particle size distribution of about 100 nm, hence referred to here as nanoplexes. The DOTA-siRNA of Bartlett et al. was synthesized as follows: 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid mono (N-hydroxysuccinimide ester) (DOTA-NHSester) was ordered from Macrocyclics (Dallas, TX). The amine modified RNA sense strand with a 100-fold molar excess of DOTA-NHS-ester in carbonate buffer (pH 9) was added to a microcentrifuge tube. The contents were reacted by stirring for 4 h at room temperature. The DOTA-RNAsense conjugate was ethanol-precipitated, resuspended in water, and annealed to the unmodified antisense strand to yield DOTA-siRNA. All liquids were pretreated with Chelex-100 (Bio-Rad, Hercules, CA) to remove trace metal contaminants. Tf-targeted and nontargeted siRNA nanoparticles may be formed by using cyclodextrin-containing polycations. Typically, nanoparticles were formed in water at a charge ratio of 3 (+/−) and an siRNA concentration of 0.5 g/liter. One percent of the adamantane-PEG molecules on the surface of the targeted nanoparticles were modified with Tf (adamantane-PEG-Tf). The nanoparticles were suspended in a 5% (wt/vol) glucose carrier solution for injection.

[0272] Davis et al. (Nature, Vol 464, 15 Apr. 2010) conducts a siRNA clinical trial that uses a targeted nanoparticle-delivery system (clinical trial registration number NCT00689065). Patients with solid cancers refractory to standard-of-care therapies are administered doses of targeted nanoparticles on days 1, 3, 8 and 10 of a 21-day cycle by a 30-min intravenous infusion. The nanoparticles consist of a synthetic delivery system containing: (1) a linear, cyclodextrin-based polymer (CDP), (2) a human transferrin protein (TF) targeting ligand displayed on the exterior of the nanoparticle to engage TF receptors (TFR) on the surface of the cancer cells, (3) a hydrophilic polymer (polyethylene glycol (PEG) used to promote nanoparticle stability in biological fluids), and (4) siRNA designed to reduce the expression of the RRM2 (sequence used in the clinic was previously denoted siR2B+5). The TFR has long been known to be upregulated in malignant cells, and RRM2 is an established anti-cancer target. These nanoparticles (clinical version denoted as CALAA-01) have been shown to be well tolerated in multi-dosing studies in non-human primates. Although a single patient with chronic myeloid leukaemia has been administered siRNA by liposomal delivery, Davis et al.'s clinical trial is the initial human trial to systemically deliver siRNA with a targeted delivery system and to treat patients with solid cancer. To ascertain whether the targeted delivery system can provide effective delivery of functional siRNA to human tumours, Davis et al. investigated biopsies from three patients from three different dosing cohorts; patients A, B and C, all of whom had metastatic melanoma and received CALAA-01 doses of 18, 24 and 30 mg m **2** siRNA, respectively. Similar doses may also be contemplated for the CRISPR Cas system of the present invention. The delivery of the invention may be achieved with nanoparticles containing a linear, cyclodextrin-based polymer (CDP), a human transferrin protein (TF) targeting ligand displayed on the exterior of the nanoparticle to engage TF receptors (TFR) on the surface of the cancer cells and/or a hydrophilic polymer (for example, polyethylene glycol (PEG) used to promote nanoparticle stability in biological fluids).

Exosomes

[0273] Exosomes are endogenous nano-vesicles that transport RNAs and proteins which can deliver short interfering (si) RNA to the brain in mice. To reduce immunogenicity, Alvarez-Erviti et al. (2011, Nat Biotechnol 29:341) used self-derived dendritic cells for exosome production. Targeting was achieved by engineering the dendritic cells to express Lamp2b, an exosomal membrane protein, fused to the neuron-specific RVG peptide3. Purified exosomes were loaded with exogenous siRNA by electroporation. Intravenously injected RVG-targeted exosomes delivered GAPDH siRNA specifically to neurons, microglia, oligodendrocytes in the brain, resulting in a specific gene knockdown. Pre-exposure to RVG exosomes did not attenuate knockdown, and non-specific uptake in other tissues was not observed. The therapeutic potential of exosome-mediated siRNA delivery was demonstrated by the strong mRNA (60%) and protein (62%) knockdown of BACE1, a therapeutic target in Alzheimer's disease.

[0274] To obtain a pool of immunologically inert exosomes, Alvarez-Erviti et al. harvested bone marrow from inbred C57BL/6 mice with a homogenous major histocompatibility complex (MHC) haplotype. As immature dendritic cells produce large quantities of exosomes devoid of T-cell activators such as MHC-II and CD86, Alvarez-Erviti et al. selected for dendritic cells with granulocyte/macrophage-colony stimulating factor (GM-CSF) for 7 d. Exosomes were purified from the culture supernatant the following day using well-established ultracentrifugation protocols. The exosomes produced were physically homogenous, with a size distribution peaking at 80 nm in diameter as determined by nanoparticle tracking analysis (NTA) and electron microscopy. Alvarez-Erviti et al. obtained 6-12 μg of exosomes (measured based on protein concentration) per 10^6 cells.

[0275] Next, Alvarez-Erviti et al. investigated the possibility of loading modified exosomes with exogenous cargoes using electroporation protocols adapted for nanoscale applications. As electroporation for membrane particles at the nanometer scale is not well-characterized, nonspecific Cy5-labeled siRNA was used for the empirical optimization of the electroporation protocol. The amount of encapsulated siRNA was assayed after ultracentrifugation and lysis of exosomes. Electroporation at 400 V and 125 μF resulted in the greatest retention of siRNA and was used for all subsequent experiments.

[0276] Alvarez-Erviti et al. administered 150 μg of each BACE1 siRNA encapsulated in 150 μg of RVG exosomes to normal C57BL/6 mice and compared the knockdown efficiency to four controls: untreated mice, mice injected with RVG exosomes only, mice injected with BACE1 siRNA complexed to an in vivo cationic liposome reagent and mice injected with BACE1 siRNA complexed to RVG-9R, the RVG pep tide conjugated to 9 D-arginines that electrostatically binds to the siRNA. Cortical tissue samples were analyzed 3 d after administration and a significant protein knockdown (45%, $P<0.05$, versus 62%, $P<0.01$) in both siRNA-RVG-9R-treated and siRNARVG exosome-treated mice was observed, resulting from a significant decrease in BACE1 mRNA levels (66% [+ or −] 15%, $P<0.001$ and 61% [+ or −] 13% respectively, $P<0.01$). Moreover, Applicants demonstrated a significant decrease (55%, $P<0.05$) in the total [beta]-amyloid 1-42 levels, a main component of the amyloid plaques in Alzheimer's pathology, in the RVG-exosome-treated animals. The decrease observed was greater than the β-amyloid 1-40 decrease demonstrated in normal mice after intraventricular injection of BACE1 inhibitors. Alvarez-Erviti et al. carried out 5'-rapid amplification of cDNA ends (RACE) on BACE1 cleavage product, which provided evidence of RNAi-mediated knockdown by the siRNA.

[0277] Finally, Alvarez-Erviti et al. investigated whether siRNA-RVG exosomes induced immune responses in vivo by assessing IL-6, IP-10, TNFα and IFN-α serum concentrations. Following siRNA-RVG exosome treatment, non-significant changes in all cytokines were registered similar to siRNA-transfection reagent treatment in contrast to siRNA-RVG-9R, which potently stimulated IL-6 secretion, confirming the immunologically inert profile of the exosome treatment. Given that exosomes encapsulate only 20% of siRNA, delivery with RVG-exosome appears to be more efficient than RVG-9R delivery as comparable mRNA knockdown and greater protein knockdown was achieved with fivefold less siRNA without the corresponding level of immune stimulation. This experiment demonstrated the therapeutic potential of RVG-exosome technology, which is potentially suited for long-term silencing of genes related to neurodegenerative diseases. The exosome delivery system of Alvarez-Erviti et al. may be applied to deliver the CRISPR-Cas system of the present invention to therapeutic targets, especially neurodegenerative diseases. A dosage of about 100 to 1000 mg of CRISPR Cas encapsulated in about 100 to 1000 mg of RVG exosomes may be contemplated for the present invention.

[0278] El-Andaloussi et al. (Nature Protocols 7, 2112-2126 (2012)) discloses how exosomes derived from cultured cells can be harnessed for delivery of siRNA in vitro and in vivo. This protocol first describes the generation of targeted exosomes through transfection of an expression vector, comprising an exosomal protein fused with a peptide ligand. Next, El-Andaloussi et al. explain how to purify and characterize exosomes from transfected cell supernatant. Next, El-Andaloussi et al. detail crucial steps for loading siRNA into exosomes. Finally, El-Andaloussi et al. outline how to use exosomes to efficiently deliver siRNA in vitro and in vivo in mouse brain. Examples of anticipated results in which exosome-mediated siRNA delivery is evaluated by functional assays and imaging are also provided. The entire protocol takes ~3 weeks. Delivery or administration according to the invention may be performed using exosomes produced from self-derived dendritic cells.

[0279] In another embodiment, the plasma exosomes of Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130) are contemplated. Exosomes are nano-sized vesicles (30-90 nm in size) produced by many cell types, including dendritic cells (DC), B cells, T cells, mast cells, epithelial cells and tumor cells. These vesicles are formed by inward budding of late endosomes and are then released to the extracellular environment upon fusion with the plasma membrane. Because exosomes naturally carry RNA between cells, this property might be useful in gene therapy.

[0280] Exosomes from plasma are prepared by centrifugation of buffy coat at 900 g for 20 min to isolate the plasma followed by harvesting cell supernatants, centrifuging at 300 g for 10 min to eliminate cells and at 16 500 g for 30 min followed by filtration through a 0.22 mm filter. Exosomes are pelleted by ultracentrifugation at 120 000 g for 70 min. Chemical transfection of siRNA into exosomes is carried out

according to the manufacturer's instructions in RNAi Human/Mouse Starter Kit (Quiagen, Hilden, Germany). siRNA is added to 100 ml PBS at a final concentration of 2 mmol/ml. After adding HiPerFect transfection reagent, the mixture is incubated for 10 min at RT. In order to remove the excess of micelles, the exosomes are re-isolated using aldehyde/sulfate latex beads. The chemical transfection of CRISPR Cas into exosomes may be conducted similarly to siRNA. The exosomes may be co-cultured with monocytes and lymphocytes isolated from the peripheral blood of healthy donors. Therefore, it may be contemplated that exosomes containing CRISPR Cas may be introduced to monocytes and lymphocytes of and autologously reintroduced into a human. Accordingly, delivery or administration according to the invention may beperformed using plasma exosomes.

Liposomes

[0281] Delivery or administration according to the invention can be performed with liposomes. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes have gained considerable attention as drug delivery carriers because they are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB) (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0282] Liposomes can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes as drug carriers. Although liposome formation is spontaneous when a lipid film is mixed with an aqueous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi: 10.1155/2011/469679 for review).

[0283] Several other additives may be added to liposomes in order to modify their structure and properties. For instance, either cholesterol or sphingomyelin may be added to the liposomal mixture in order to help stabilize the liposomal structure and to prevent the leakage of the liposomal inner cargo. Further, liposomes are prepared from hydrogenated egg phosphatidylcholine or egg phosphatidylcholine, cholesterol, and dicetyl phosphate, and their mean vesicle sizes were adjusted to about 50 and 100 nm. (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi: 10.1155/2011/469679 for review).

[0284] Conventional liposome formulation is mainly comprised of natural phospholipids and lipids such as 1,2-distearoryl-sn-glycero-3-phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines and monosialoganglioside. Since this formulation is made up of phospholipids only, liposomal formulations have encountered many challenges, one of the ones being the instability in plasma. Several attempts to overcome these challenges have been made, specifically in the manipulation of the lipid membrane. One of these attempts focused on the manipulation of cholesterol. Addition of cholesterol to conventional formulations reduces rapid release of the encapsulated bioactive compound into the plasma or 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) increases the stability (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi: 10.1155/2011/469679 for review).

[0285] In a particularly advantageous embodiment, Trojan Horse liposomes (also known as Molecular Trojan Horses) are desirable and protocols may be found at http://cshprotocols.cshlp.org/content/2010/4/pdb.protS407 long. These particles allow delivery of a transgene to the entire brain after an intravascular injection. Without being bound by limitation, it is believed that neutral lipid particles with specific antibodies conjugated to surface allow crossing of the blood brain barrier via endocytosis. Applicant postulates utilizing Trojan Horse Liposomes to deliver the CRISPR family of nucleases to the brain via an intravascular injection, which would allow whole brain transgenic animals without the need for embryonic manipulation. About 1-5 g of DNA may be contemplated for in vivo administration in liposomes.

[0286] In another embodiment, the CRISPR Cas system may be administered in liposomes, such as a stable nucleic-acid-lipid particle (SNALP) (see, e.g., Morrissey et al., Nature Biotechnology, Vol. 23, No. 8, August 2005). Daily intravenous injections of about 1, 3 or 5 mg/kg/day of a specific CRISPR Cas targeted in a SNALP are contemplated. The daily treatment may be over about three days and then weekly for about five weeks. In another embodiment, a specific CRISPR Cas encapsulated SNALP) administered by intravenous injection to at doses of abpit 1 or 2.5 mg/kg are also contemplated (see, e.g., Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006). The SNALP formulation may contain the lipids 3-N-[(wmethoxypoly(ethylene glycol) 2000) carbamoyl]-!,2-dimyristyloxy-propylamine (PEG-C-DMA), 1,2-dilinoleyloxy-N,N-dimethyl-3-aminopropane (DLinDMA), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC) and cholesterol, in a 2:40:10:48 molar percent ratio (see, e.g., Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006).

[0287] In another embodiment, stable nucleic-acid-lipid particles (SNALPs) have proven to be effective delivery molecules to highly vascularized HepG2-derived liver tumors but not in poorly vascularized HCT-116 derived liver tumors (see, e.g., Li, Gene Therapy (2012) 19, 775-780). The SNALP liposomes may be prepared by formulating D-Lin-DMA and PEG-C-DMA with distearoylphosphatidylcholine (DSPC), Cholesterol and siRNA using a 25:1 lipid/siRNA ratio and a 48/40/10/2 molar ratio of Cholesterol/D-Lin-DMA/DSPC/PEG-C-DMA. The resulted SNALP liposomes are about 80-100 nm in size.

[0288] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich, St Louis, MO, USA), dipalmitoylphosphatidylcholine (Avanti Polar Lipids, Alabaster, AL, USA), 3-N-[(w-methoxy poly(ethylene glycol)2000)carbamoyl]-1,2-dimyrestyloxypropylamine, and cationic 1,2-dilinoleyloxy-3-N,Ndimethylaminopropane (see, e.g., Geisbert et al., Lancet 2010; 375: 1896-905). A dosage of about 2 mg/kg total CRISPR Cas per dose administered as, for example, a bolus intravenous infusion may be contemplated.

[0289] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich), 1,2-distearoyl-

sn-glycero-3-phosphocholine (DSPC; Avanti Polar Lipids Inc.), PEG-CDMA, and 1,2-dilinoleyloxy-3-(N;N-dimethyl) aminopropane (DLinDMA) (see, e.g., Judge, J. Clin. Invest. 119:661-673 (2009)). Formulations used for in vivo studies may comprise a final lipid/RNA mass ratio of about 9:1.

[0290] The safety profile of RNAi nanomedicines has been reviewed by Barros and Gollob of Alnylam Pharmaceuticals (see, e.g., Advanced Drug Delivery Reviews 64 (2012) 1730-1737). The stable nucleic acid lipid particle (SNALP) is comprised of four different lipids—an ionizable lipid (DLinDMA) that is cationic at low pH, a neutral helper lipid, cholesterol, and a diffusible polyethylene glycol (PEG)-lipid. The particle is approximately 80 nm in diameter and is charge-neutral at physiologic pH. During formulation, the ionizable lipid serves to condense lipid with the anionic siRNA during particle formation. When positively charged under increasingly acidic endosomal conditions, the ionizable lipid also mediates the fusion of SNALP with the endosomal membrane enabling release of siRNA into the cytoplasm. The PEG-lipid stabilizes the particle and reduces aggregation during formulation, and subsequently provides a neutral hydrophilic exterior that improves pharmacokinetic properties.

[0291] To date, two clinical programs have been initiated using SNALPsiRNA formulations. Tekmira Pharmaceuticals recently completed a phase I single-dose study of SNALP-ApoB in adult volunteers with elevated LDL cholesterol. ApoB is predominantly expressed in the liver and jejunum and is essential for the assembly and secretion of VLDL and LDL. Seventeen subjects received a single dose of SNALP-ApoB (dose escalation across 7 dose levels). There was no evidence of liver toxicity (anticipated as the potential dose-limiting toxicity based on preclinical studies). One (of two) subjects at the highest dose experienced flu-like symptoms consistent with immune system stimulation, and the decision was made to conclude the trial.

[0292] Alnylam Pharmaceuticals has similarly advanced ALN-TTR01, which employs the SNALP technology described above and targets hepatocyte production of both mutant and wild-type TTR to treat TTR amyloidosis (ATTR). Three ATTR syndromes have been described: familial amyloidotic polyneuropathy (FAP) and familial amyloidotic cardiomyopathy (FAC)—both caused by autosomal dominant mutations in TTR; and senile systemic amyloidosis (SSA) cause by wildtype TTR. A placebo-controlled, single dose-escalation phase I trial of ALN-TTR01 was recently completed in patients with ATTR. ALN-TTR01 was administered as a 15-minute IV infusion to 31 patients (23 with study drug and 8 with placebo) within a dose range of 0.01 to 1.0 mg/kg (based on siRNA). Treatmentwaswell tolerated with no significant increases in liver function tests. Infusion-related reactions were noted in 3 of 23 patients at ≥0.4 mg/kg; all responded to slowing of the infusion rate and all continued on study. Minimal and transient elevations of serum cytokines IL-6, IP-10 and IL-Ira were noted in two patients at the highest dose of 1 mg/kg (as anticipated from preclinical and NHP studies). Lowering of serum TTR, the expected pharmacodynamics effect of ALN-TTR01, was observed at 1 mg/kg.

[0293] In yet another embodiment, a SNALP may be made by solubilizing a cationic lipid, DSPC, cholesterol and PEG-lipid were solubilized in ethanol at a molar ratio of 40:10:40:10, respectively (see, Semple et al., Nature Niotechnology, Volume 28 Number 2 Feb. 2010, pp. 172-177). The lipid mixture was added to an aqueous buffer (50 mM citrate, pH 4) with mixing to a final ethanol and lipid concentration of 30% (vol/vol) and 6.1 mg/ml, respectively, and allowed to equilibrate at 22° C. for 2 min before extrusion. The hydrated lipids were extruded through two stacked 80 nm pore-sized filters (Nuclepore) at 22° C. using a Lipex Extruder (Northern Lipids) until a vesicle diameter of 70-90 nm, as determined by dynamic light scattering analysis, was obtained. This generally required 1-3 passes. The siRNA (solubilized in a 50 mM citrate, pH 4 aqueous solution containing 30% ethanol) was added to the pre-equilibrated (35° C.) vesicles at a rate of ~5 ml/min with mixing. After a final target siRNA/lipid ratio of 0.06 (wt/wt) was reached, the mixture was incubated for a further 30 min at 35° C. to allow vesicle reorganization and encapsulation of the siRNA. The ethanol was then removed and the external buffer replaced with PBS (155 mM NaCl, 3 mM Na2HPO4, 1 mM KH2PO4, pH 7.5) by either dialysis or tangential flow diafiltration. siRNA were encapsulated in SNALP using a controlled step-wise dilution method process. The lipid constituents of KC2-SNALP were DLin-KC2-DMA (cationic lipid), dipalmitoylphosphatidylcholine (DPPC; Avanti Polar Lipids), synthetic cholesterol (Sigma) and PEG-C-DMA used at a molar ratio of 57.1:7.1:34.3:1.4. Upon formation of the loaded particles, SNALP were dialyzed against PBS and filter sterilized through a 0.2 μm filter before use. Mean particle sizes were 75-85 nm and 90-95% of the siRNA was encapsulated within the lipid particles. The final siRNA/lipid ratio in formulations used for in vivo testing was ~0.15 (wt/wt). LNP-siRNA systems containing Factor VII siRNA were diluted to the appropriate concentrations in sterile PBS immediately before use and the formulations were administered intravenously through the lateral tail vein in a total volume of 10 ml/kg. This method may be extrapolated to the CRISPR Cas system of the present invention.

Other Lipids

[0294] Other cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA) may be utilized to encapsulate CRISPR Cas similar to SiRNA (see, e.g., Jayaraman, Angew. Chem. Int. Ed. 2012, 51, 8529-8533). A preformed vesicle with the following lipid composition may be contemplated: amino lipid, distearoylphosphatidylcholine (DSPC), cholesterol and (R)-2,3-bis (octadecyloxy) propyl-1-(methoxy poly(ethylene glycol)2000)propylcarbamate (PEG-lipid) in the molar ratio 40/10/40/10, respectively, and a FVII siRNA/total lipid ratio of approximately 0.05 (w/w). To ensure a narrow particle size distribution in the range of 70-90 nm and a low polydispersity index of 0.11_0.04 (n=56), the particles may be extruded up to three times through 80 nm membranes prior to adding the CRISPR Cas RNA. Particles containing the highly potent amino lipid 16 may be used, in which the molar ratio of the four lipid components 16, DSPC, cholesterol and PEG-lipid (50/10/38.5/1.5) which may be further optimized to enhance in vivo activity.

[0295] Michael S D Kormann et al. ("Expression of therapeutic proteins after delivery of chemically modified mRNA in mice: Nature Biotechnology, Volume: 29, Pages: 154-157 (2011) Published online 9 Jan. 2011) describes the use of lipid envelopes to deliver RNA. Use of lipid envelopes is also preferred in the present invention.

[0296] In another embodiment, lipids may be formulated with the CRISPR Cas system of the present invention to form lipid nanoparticles (LNPs). Lipids include, but are not limited to, DLin-KC2-DMA4, C12-200 and colipids disteroylphosphatidyl choline, cholesterol, and PEG-DMG may be formulated with CRISPR Cas instead of siRNA (see, e.g., Novobrantseva, Molecular Therapy-Nucleic Acids (2012) 1, e4; doi: 10.1038/mtna.2011.3) using a spontaneous vesicle formation procedure. The component molar ratio may be about 50/10/38.5/1.5 (DLin-KC2-DMA or C12-200/disteroylphosphatidyl choline/cholesterol/PEG-DMG). The final lipid: siRNA weight ratio may be ~12:1 and 9:1 in the case of DLin-KC2-DMA and C12-200 lipid nanoparticles (LNPs), respectively. The formulations may have mean particle diameters of ~80 nm with >90% entrapment efficiency. A 3 mg/kg dose may be contemplated.

[0297] Tekmira has a portfolio of approximately 95 patent families, in the U.S. and abroad, that are directed to various aspects of LNPs and LNP formulations (see, e.g., U.S. Pat. Nos. 7,982,027; 7,799,565; 8,058,069; 8,283,333; 7,901,708; 7,745,651; 7,803,397; 8,101,741; 8,188,263; 7,915,399; 8,236,943 and 7,838,658 and European Pat. Nos. 1766035; 1519714; 1781593 and 1664316), all of which may be used/and or adapted to the present invention.

[0298] The CRISPR Cas system may be delivered encapsulated in PLGA Microspheres such as that further described in US published applications 20130252281 and 20130245107 and 20130244279 (assigned to Moderna Therapeutics) which relate to aspects of formulation of compositions comprising modified nucleic acid molecules which may encode a protein, a protein precursor, or a partially or fully processed form of the protein or a protein precursor. The formulation may have a molar ratio 50:10:38.5:1.5-3.0 (cationic lipid: fusogenic lipid: cholesterol: PEG lipid). The PEG lipid may be selected from, but is not limited to PEG-c-DOMG, PEG-DMG. The fusogenic lipid may be DSPC. See also, Schrum et al., Delivery and Formulation of Engineered Nucleic Acids, US published application No. 20120251618.

[0299] Nanomerics' technology addresses bioavailability challenges for a broad range of therapeutics, including low molecular weight hydrophobic drugs, peptides, and nucleic acid based therapeutics (plasmid, siRNA, miRNA). Specific administration routes for which the technology has demonstrated clear advantages include the oral route, transport across the blood-brain-barrier, delivery to solid tumours, as well as to the eye. See, e.g., Mazza et al., 2013, ACS Nano. 2013 Feb. 26; 7(2):1016-26; Uchegbu and Siew, 2013, J Pharm Sci. 102(2):305-10 and Lalatsa et al., 2012, J Control Release. 2012 Jul. 20; 161(2):523-36.

[0300] US Patent Publication No. 20050019923 describes cationic dendrimers for delivering bioactive molecules, such as polynucleotide molecules, peptides and polypeptides and/or pharmaceutical agents, to a mammalian body. The dendrimers are suitable for targeting the delivery of the bioactive molecules to, for example, the liver, spleen, lung, kidney or heart. Dendrimers are synthetic 3-dimensional macromolecules that are prepared in a step-wise fashion from simple branched monomer units, the nature and functionality of which can be easily controlled and varied. Dendrimers are synthesised from the repeated addition of building blocks to a multifunctional core (divergent approach to synthesis), or towards a multifunctional core (convergent approach to synthesis) and each addition of a 3-dimensional shell of building blocks leads to the formation of a higher generation of the dendrimers. Polypropylenimine dendrimers start from a diaminobutane core to which is added twice the number of amino groups by a double Michael addition of acrylonitrile to the primary amines followed by the hydrogenation of the nitriles. This results in a doubling of the amino groups. Polypropylenimine dendrimers contain 100% protonable nitrogens and up to 64 terminal amino groups (generation 5, DAB 64). Protonable groups are usually amine groups which are able to accept protons at neutral pH. The use of dendrimers as gene delivery agents has largely focused on the use of the polyamidoamine. and phosphorous containing compounds with a mixture of amine/amide or N—P(O_2)S as the conjugating units respectively with no work being reported on the use of the lower generation polypropylenimine dendrimers for gene delivery. Polypropylenimine dendrimers have also been studied as pH sensitive controlled release systems for drug delivery and for their encapsulation of guest molecules when chemically modified by peripheral amino acid groups. The cytotoxicity and interaction of polypropylenimine dendrimers with DNA as well as the transfection efficacy of DAB 64 has also been studied.

[0301] US Patent Publication No. 20050019923 is based upon the observation that, contrary to earlier reports, cationic dendrimers, such as polypropylenimine dendrimers, display suitable properties, such as specific targeting and low toxicity, for use in the targeted delivery of bioactive molecules, such as genetic material. In addition, derivatives of the cationic dendrimer also display suitable properties for the targeted delivery of bioactive molecules. See also, Bioactive Polymers, US published application No. 20080267903, which discloses "Various polymers, including cationic polyamine polymers and dendrimeric polymers, are shown to possess anti-proliferative activity, and may therefore be useful for treatment of disorders characterised by undesirable cellular proliferation such as neoplasms and tumours, inflammatory disorders (including autoimmune disorders), psoriasis and atherosclerosis. The polymers may be used alone as active agents, or as delivery vehicles for other therapeutic agents, such as drug molecules or nucleic acids for gene therapy. In such cases, the polymers' own intrinsic anti-tumour activity may complement the activity of the agent to be delivered."

Supercharged Proteins

[0302] Supercharged proteins are a class of engineered or naturally occurring proteins with unusually high positive or negative net theoretical charge. Both supernegatively and superpositively charged proteins exhibit a remarkable ability to withstand thermally or chemically induced aggregation. Superpositively charged proteins are also able to penetrate mammalian cells. Associating cargo with these proteins, such as plasmid DNA, siRNA, or other proteins, can enable the functional delivery of these macromolecules into mammalian cells both in vitro and in vivo. David Liu's lab reported the creation and characterization of supercharged proteins in 2007 (Lawrence et al., 2007, Journal of the American Chemical Society 129, 10110-10112).

[0303] The nonviral delivery of siRNA and plasmid DNA into mammalian cells are valuable both for research and therapeutic applications (Akinc et al., 2010, Nat. Biotech. 26, 561-569). Purified +36 GFP protein (or other superpositively charged protein) is mixed with siRNAs in the appropriate serum-free media and allowed to complex prior

addition to cells. Inclusion of serum at this stage inhibits formation of the supercharged protein-siRNA complexes and reduces the effectiveness of the treatment. The following protocol has been found to be effective for a variety of cell lines (McNaughton et al., 2009, Proc. Natl. Acad. Sci. USA 106, 6111-6116). However, pilot experiments varying the dose of protein and siRNA should be performed to optimize the procedure for specific cell lines.

[0304] (1) One day before treatment, plate 1×10^5 cells per well in a 48-well plate.

[0305] (2) On the day of treatment, dilute purified +36 GFP protein in serumfree media to a final concentration 200 nM. Add siRNA to a final concentration of 50 nM. Vortex to mix and incubate at room temperature for 10 min.

[0306] (3) During incubation, aspirate media from cells and wash once with PBS.

[0307] (4) Following incubation of +36 GFP and siRNA, add the protein-siRNA complexes to cells.

[0308] (5) Incubate cells with complexes at 37 C for 4 h.

[0309] (6) Following incubation, aspirate the media and wash three times with 20 U/mL heparin PBS. Incubate cells with serum-containing media for a further 48 h or longer depending upon the assay for knockdown.

[0310] (7) Analyze cells by immunoblot, qPCR, phenotypic assay, or other appropriate method.

[0311] David Liu's lab has further found +36 GFP to be an effective plasmid delivery reagent in a range of cells. As plasmid DNA is a larger cargo than siRNA, proportionately more +36 GFP protein is required to effectively complex plasmids. For effective plasmid delivery Applicants have developed a variant of +36 GFP bearing a C-terminal HA2 peptide tag, a known endosome-disrupting peptide derived from the influenza virus hemagglutinin protein. The following protocol has been effective in a variety of cells, but as above it is advised that plasmid DNA and supercharged protein doses be optimized for specific cell lines and delivery applications.

[0312] (1) One day before treatment, plate 1×10^5 per well in a 48-well plate.

[0313] (2) On the day of treatment, dilute purified p36 GFP protein in serumfree media to a final concentration 2 mM. Add 1 mg of plasmid DNA. Vortex to mix and incubate at room temperature for 10 min.

[0314] (3) During incubation, aspirate media from cells and wash once with PBS.

[0315] (4) Following incubation of p36 GFP and plasmid DNA, gently add the protein-DNA complexes to cells.

[0316] (5) Incubate cells with complexes at 37 C for 4 h.

[0317] (6) Following incubation, aspirate the media and wash with PBS. Incubate cells in serum-containing media and incubate for a further 24-48 h.

[0318] (7) Analyze plasmid delivery (e.g., by plasmid-driven gene expression) as appropriate.

[0319] See also, e.g., McNaughton et al., Proc. Natl. Acad. Sci. USA 106, 6111-6116 (2009); Cronican et al., ACS Chemical Biology 5, 747-752 (2010); Cronican et al., Chemistry & Biology 18, 833-838 (2011); Thompson et al., Methods in Enzymology 503, 293-319 (2012); Thompson, D. B., et al., Chemistry & Biology 19 (7), 831-843 (2012). The methods of the super charged proteins may be used and/or adapted for delivery of the CRISPR Cas system of the present invention.

Implantable Devices

[0320] In another embodiment, implantable devices are also contemplated for delivery of the CRISPR Cas system. For example, US Patent Publication 20110195123 discloses an implantable medical device which elutes a drug locally and in prolonged period is provided, including several types of such a device, the treatment modes of implementation and methods of implantation. The device comprising of polymeric substrate, such as a matrix for example, that is used as the device body, and drugs, and in some cases additional scaffolding materials, such as metals or additional polymers, and materials to enhance visibility and imaging. The selection of drug is based on the advantageous of releasing drug locally and in prolonged period, where drug is released directly to the extracellular matrix (ECM) of the diseased area such as tumor, inflammation, degeneration or for symptomatic objectives, or to injured smooth muscle cells, or for prevention. One kind of drug is the gene silencing drugs based on RNA interference (RNAi), including but not limited to si RNA, sh RNA, or antisense RNA/DNA, ribozyme and nucleoside analogs. Therefore, this system may be used/and or adapted to the CRISPR Cas system of the present invention. The modes of implantation in some embodiments are existing implantation procedures that are developed and used today for other treatments, including brachytherapy and needle biopsy. In such cases the dimensions of the new implant described in this invention are similar to the original implant. Typically a few devices are implanted during the same treatment procedure.

[0321] As described in US Patent Publication 20110195123, there is provided a drug delivery implantable or insertable system, including systems applicable to a cavity such as the abdominal cavity and/or any other type of administration in which the drug delivery system is not anchored or attached, comprising a biostable and/or degradable and/or bioabsorbable polymeric substrate, which may for example optionally be a matrix. It should be noted that the term "insertion" also includes implantation. The drug delivery system is preferably implemented as a "Loder" as described in US Patent Publication 20110195123.

[0322] The polymer or plurality of polymers are biocompatible, incorporating an agent and/or plurality of agents, enabling the release of agent at a controlled rate, wherein the total volume of the polymeric substrate, such as a matrix for example, in some embodiments is optionally and preferably no greater than a maximum volume that permits a therapeutic level of the agent to be reached. As a non-limiting example, such a volume is preferably within the range of 0.1 m^3 to 1000 mm^3, as required by the volume for the agent load. The Loder may optionally be larger, for example when incorporated with a device whose size is determined by functionality, for example and without limitation, a knee joint, an intra-uterine or cervical ring and the like.

[0323] The drug delivery system (for delivering the composition) is designed in some embodiments to preferably employ degradable polymers, wherein the main release mechanism is bulk erosion; or in some embodiments, non degradable, or slowly degraded polymers are used, wherein the main release mechanism is diffusion rather than bulk erosion, so that the outer part functions as membrane, and its

internal part functions as a drug reservoir, which practically is not affected by the surroundings for an extended period (for example from about a week to about a few months). Combinations of different polymers with different release mechanisms may also optionally be used. The concentration gradient at the surface is preferably maintained effectively constant during a significant period of the total drug releasing period, and therefore the diffusion rate is effectively constant (termed "zero mode" diffusion). By the term "constant" it is meant a diffusion rate that is preferably maintained above the lower threshold of therapeutic effectiveness, but which may still optionally feature an initial burst and/or fluctuate, for example increasing and decreasing to a certain degree. The diffusion rate is preferably so maintained for a prolonged period, and it can be considered constant to a certain level to optimize the therapeutically effective period, for example the effective silencing period.

[0324] The drug delivery system optionally and preferably is designed to shield the nucleotide based therapeutic agent from degradation, whether chemical in nature or due to attack from enzymes and other factors in the body of the subject.

[0325] The drug delivery system as described in US Patent Publication 20110195123 is optionally associated with sensing and/or activation appliances that are operated at and/or after implantation of the device, by non and/or minimally invasive methods of activation and/or acceleration/deceleration, for example optionally including but not limited to thermal heating and cooling, laser beams, and ultrasonic, including focused ultrasound and/or RF (radiofrequency) methods or devices.

[0326] According to some embodiments of US Patent Publication 20110195123, the site for local delivery may optionally include target sites characterized by high abnormal proliferation of cells, and suppressed apoptosis, including tumors, active and or chronic inflammation and infection including autoimmune diseases states, degenerating tissue including muscle and nervous tissue, chronic pain, degenerative sites, and location of bone fractures and other wound locations for enhancement of regeneration of tissue, and injured cardiac, smooth and striated muscle. The site for local delivery also may optionally include sites enabling performing preventive activities including pregnancy, prevention of infection and aging.

[0327] The site for implantation of the composition, or target site, preferably features a radius, area and/or volume that is sufficiently small for targeted local delivery. For example, the target site optionally has a diameter in a range of from about 0.1 mm to about 5 cm.

[0328] The location of the target site is preferably selected for maximum therapeutic efficacy. For example, the composition of the drug delivery system (optionally with a device for implantation as described above) is optionally and preferably implanted within or in the proximity of a tumor environment, or the blood supply associated thereof.

[0329] For example the composition (optionally with the device) is optionally implanted within or in the proximity to pancreas, prostate, breast, liver, via the nipple, within the vascular system and so forth.

[0330] The target location is optionally selected from the group consisting of (as non-limiting examples only, as optionally any site within the body may be suitable for implanting a Loder): 1. brain at degenerative sites like in Parkinson or Alzheimer disease at the basal ganglia, white and gray matter; 2. spine as in the case of amyotrophic lateral sclerosis (ALS); 3. uterine cervix to prevent HPV infection; 4. active and chronic inflammatory joints; 5. dermis as in the case of psoriasis; 6. sympathetic and sensoric nervous sites for analgesic effect; 7. Intra osseous implantation; 8. acute and chronic infection sites; 9. Intra vaginal; 10. Inner ear—auditory system, labyrinth of the inner ear, vestibular system; 11. Intra tracheal; 12. Intra-cardiac; coronary, epicardiac; 13. urinary bladder; 14. biliary system; 15. parenchymal tissue including and not limited to the kidney, liver, spleen; 16. lymph nodes; 17. salivary glands; 18. dental gums; 19. Intra-articular (into joints); 20. Intra-ocular; 21. Brain tissue; 22. Brain ventricles; 23. Cavities, including abdominal cavity (for example but without limitation, for ovary cancer); 24. Intra esophageal and 25. Intra rectal.

[0331] Optionally insertion of the system (for example a device containing the composition) is associated with injection of material to the ECM at the target site and the vicinity of that site to affect local pH and/or temperature and/or other biological factors affecting the diffusion of the drug and/or drug kinetics in the ECM, of the target site and the vicinity of such a site.

[0332] Optionally, according to some embodiments, the release of said agent could be associated with sensing and/or activation appliances that are operated prior and/or at and/or after insertion, by non and/or minimally invasive and/or else methods of activation and/or acceleration/deceleration, including laser beam, radiation, thermal heating and cooling, and ultrasonic, including focused ultrasound and/or RF (radiofrequency) methods or devices, and chemical activators.

[0333] According to other embodiments of US Patent Publication 20110195123, the drug preferably comprises a gene silencing biological RNAi drug, for example for localized cancer cases in breast, pancreas, brain, kidney, bladder, lung, and prostate as described below. Moreover, many drugs other than siRNA are applicable to be encapsulated in Loder, and can be used in association with this invention, as long as such drugs can be encapsulated with the Loder substrate, such as a matrix for example. Such drugs include approved drugs that are delivered today by methods other than of this invention, including Amphotericin B for fungal infection; antibiotics such as in osteomyelitis; pain killers such as narcotics; anti degenerative such as in Alzheimer or Parkinson diseases in a Loder implanted in the vicinity of the spine in the case of back pain. Such a system may be used and/or adapted to deliver the CRISPR Cas system of the present invention.

[0334] For example, for specific applications such as prevention of growth or regrowth of smooth muscle cells (that are injured during a stenting procedure and as a result tend to proliferate), the drug may optionally be siRNA that silence smooth muscle cells, including H19 silencing, or a drug selected from the group consisting of taxol, rapamycin and rapamycin-analogs. In such cases the Loder is preferably either a Drug Eluting Stent (DES), with prolonged release at constant rate, or a dedicated device that is implanted separately, in association to the stent. All of this may be used/and or adapted to the CRISPR Cas system of the present invention.

[0335] As another example of a specific application, neuro and muscular degenerative diseases develop due to abnormal gene expression. Local delivery of silencing RNAs may

have therapeutic properties for interfering with such abnormal gene expression. Local delivery of anti apoptotic, anti inflammatory and anti degenerative drugs including small drugs and macromolecules may also optionally be therapeutic. In such cases the Loder is applied for prolonged release at constant rate and/or through a dedicated device that is implanted separately. All of this may be used and/or adapted to the CRISPR Cas system of the present invention.

[0336] As yet another example of a specific application, psychiatric and cognitive disorders are treated with gene modifiers. Gene knockdown with silencing RNA is a treatment option. Loders locally delivering nucleotide based agents to central nervous system sites are therapeutic options for psychiatric and cognitive disorders including but not limited to psychosis, bi-polar diseases, neurotic disorders and behavioral maladies. The Loders could also deliver locally drugs including small drugs and macromolecules upon implantation at specific brain sites. All of this may be used and/or adapted to the CRISPR Cas system of the present invention.

[0337] As another example of a specific application, silencing of innate and/or adaptive immune mediators at local sites enables the prevention of organ transplant rejection. Local delivery of silencing RNAs and immunomodulating reagents with the Loder implanted into the transplanted organ and/or the implanted site renders local immune suppression by repelling immune cells such as CD8 activated against the transplanted organ. All of this may be used/and or adapted to the CRISPR Cas system of the present invention.

[0338] As another example of a specific application, vascular growth factors including VEGFs and angiogenin and others are essential for neovascularization. Local delivery of the factors, peptides, peptidomimetics, or suppressing their repressors is an important therapeutic modality; silencing the repressors and local delivery of the factors, peptides, macromolecules and small drugs stimulating angiogenesis with the Loder is therapeutic for peripheral, systemic and cardiac vascular disease.

[0339] The method of insertion, such as implantation, may optionally already be used for other types of tissue implantation and/or for insertions and/or for sampling tissues, optionally without modifications, or alternatively optionally only with non-major modifications in such methods. Such methods optionally include but are not limited to brachytherapy methods, biopsy, endoscopy with and/or without ultrasound, such as ERCP, stereotactic methods into the brain tissue, Laparoscopy, including implantation with a laparoscope into joints, abdominal organs, the bladder wall and body cavities.

CRISPR Enzyme mRNA and Guide RNA

[0340] 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.

[0341] 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.

[0342] Additional administrations of CRISPR enzyme mRNA and/or guide RNA might be useful to achieve the most efficient levels of genome modification.

[0343] 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' (SEQ ID NO: 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' (SEQ ID NO: 4) and 2: 5'-GAGTCTAAGCAGAAGAAGAA-3' (SEQ ID NO: 5). 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.

[0344] 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' (SEQ ID NO: 6) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNN-5' (SEQ ID NO: 7)
13	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNN-3' (SEQ ID NO: 8) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNN-5' (SEQ ID NO: 9)
12	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNN-3' (SEQ ID NO: 10) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNN-5' (SEQ ID NO: 11)
11	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNN-3' (SEQ ID NO: 12) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNN-5' (SEQ ID NO: 13)
10	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNN-3' (SEQ ID NO: 14) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNN-5' (SEQ ID NO: 15)
9	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNN-3' (SEQ ID NO: 16) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNN-5' (SEQ ID NO: 17)
8	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNN-3' (SEQ ID NO: 18) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNN-5' (SEQ ID NO: 19)

-continued

Over-hang length (bp)	Guide RNA design (guide sequence and PAM color coded)
7	5'-NNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNG GNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 20) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNC CNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 21)
6	5'-NNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNGG NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 22) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNCC NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 23)
5	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNGGN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 24) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNCCN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 25)
4	5'-NNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNGGNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 26) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNCCNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 27)
3	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNGGNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 28) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNCCNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 29)
2	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNGGNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 30) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNCCNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 31)
1	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNGGNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 32) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNCCNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 33)
blunt	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNGGNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 34) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNCCNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 35)
1	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNGGNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 36) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNCCNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 37)
2	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNGGNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 38) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNCCNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 39)
3	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNGGNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 40) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNCCNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 41)
4	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNGGNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 42) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNNCCNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 43)
5	5'-NNNNNNNNNNNNNNNNNNNNCCNNNGGNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 44) 3'-NNNNNNNNNNNNNNNNNNNNGGNNNCCNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 45)
6	5'-NNNNNNNNNNNNNNNNNNNNCCNNGGNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 46) 3'-NNNNNNNNNNNNNNNNNNNNGGNNCCNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 47)

-continued

Over-hang length (bp)	Guide RNA design (guide sequence and PAM color coded)
7	5'-NNNNNNNNNNNNNNNNNNNNNCCNGGNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 48) 3'-NNNNNNNNNNNNNNNNNNNNGGNCCNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 49)
8	5'-NNNNNNNNNNNNNNNNNNNNCCGGNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 50) 3'-NNNNNNNNNNNNNNNNNNNNGGCCNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 51)
12	5'-NNNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 52) 3'-NNNNNNNNNNNNNNNNNNNNNNCCGGNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 53)
13	5'-NNNNNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 54) 3'-NNNNNNNNNNNNNNNNNNNNNNCCNGGNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 55)
14	5'-NNNNNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 54) 3'-NNNNNNNNNNNNNNNNNNNNNNNCCNGGNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 56)
15	5'-NNNNNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 54) 3'-NNNNNNNNNNNNNNNNNNNNNNNCCNNNGGNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 57)
16	5'-NNNNNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 54) 3'-NNNNNNNNNNNNNNNNNNNNNNNCCNNNNGGNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 58)
17	5'-NNNNNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNN-3' (SEQ ID NO: 54) 3'-NNNNNNNNNNNNNNNNNNNNNNNCCNNNNNGGNNNNNN NNNNNNNNNNNNNNNNNNNNNN-5' (SEQ ID NO: 59)

[0345] Further interrogation of the system have given Applicants evidence of 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 5' overhang. In embodiments of the invention 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 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 3' overhang is at most 150, 100 or 25 base pairs or at least 15, 10 or 1 base pairs. In preferred embodiments 3' overhang is 1-100 basepairs.

[0346] 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.

[0347] 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.

[0348] 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.

Liver, Proprotein Convertase Subtilisin Kexin 9 (PCSK9)

[0349] Proprotein convertase subtilisin kexin 9 (PCSK9) is a member of the subtilisin serine protease family. PCSK9 is primarily expressed by the liver and is critical for the down regulation of hepatocyte LDL receptor expression. LDL-C levels in plasma are highly elevated in humans with gain of function mutations in PCSK9, classifying them as having severe hypercholesterolemia. Therefore, PCSK9 is an attractive target for CRISPR. PCS9K-targeted CRISPR may be formulated in a lipid particle and for example administered at about 15, 45, 90, 150, 250 and 400 μg/kg intraveneously (see, e.g., http://www.alnylam.com/capella/wp-content/uploads/2013/08/ALN-PCS02-001-Protocol-Lancet.pdf).

[0350] Bailey et al. (J Mol Med (Berl). 1999 January; 77 (1): 244-9) discloses insulin delivery by ex-vivo somatic cell gene therapy involves the removal of non-B-cell somatic cells (e.g. fibroblasts) from a diabetic patient, and genetically altering them in vitro to produce and secrete insulin. The cells can be grown in culture and returned to the donor as a source of insulin replacement. Cells modified in this way could be evaluated before implantation, and reserve stocks could be cryopreserved. By using the patient's own cells, the procedure should obviate the need for immunosuppression and overcome the problem of tissue supply, while avoiding a recurrence of cell destruction. Ex-vivo somatic cell gene therapy requires an accessible and robust cell type that is amenable to multiple transfections and subject to controlled proliferation. Special problems associated with the use of non-B-cell somatic cells include the processing of proinsulin to insulin, and the conferment of sensitivity to glucose-stimulated proinsulin biosynthesis and regulated insulin release. Preliminary studies using fibroblasts, pituitary cells, kidney (COS) cells and ovarian (CHO) cells suggest that these challenges could be met, and that ex-vivo somatic cell gene therapy offers a feasible approach to insulin replacement therapy. The system of Bailey et al. may be used/and or adapted to the CRISPR Cas system of the present invention for delivery to the liver.

[0351] The methods of Sato et al. (Nature Biotechnology Volume 26 Number 4 Apr. 2008, Pages 431-442) may be applied to the CRISPR Cas system of the present invention for delivery to the liver. Sato et al. found that treatments with the siRNA-bearing vitamin A-coupled liposomes almost completely resolved liver fibrosis and prolonged survival in rats with otherwise lethal dimethylnitrosamine-induced liver cirrhosis in a dose- and duration-dependent manner. Cationic liposomes (Lipotrust) containing O,O'-ditetradecanoyl-N-(a-trimethylammonioacetyl) diethanolamine chloride (DC-6-14) as a cationic lipid, cholesterol and dioleoylphosphatidylethanolamine at a molar ratio of 4:3:3 (which has shown high transfection efficiency under serum-containing conditions for in vitro and in vivo gene delivery) were purchased from Hokkaido System Science. The liposomes were manufactured using a freeze-dried empty liposomes method and prepared at a concentration of 1 mM (DC-16-4) by addition of double-distilled water (DDW) to the lyophilized lipid mixture under vortexing before use. To prepare VA-coupled liposomes, 200 nmol of vitamin A (retinol, Sigma) dissolved in DMSO was mixed with the liposome suspensions (100 nmol as DC-16-4) by vortexing in a 1.5 ml tube at 25 1 C. To prepare VA-coupled liposomes carrying siRNAgp46 (VA-lip-siRNAgp46), a solution of siRNAgp46 (580 pmol/ml in DDW) was added to the retinol-coupled liposome solution with stirring at 25 C. The ratio of siRNA to DC-16-4 was 1:11.5 (mol/mol) and the siRNA to liposome ratio (wt/wt) was 1:1. Any free vitamin A or siRNA that was not taken up by liposomes were separated from liposomal preparations using a micropartition system (VIVASPIN 2 concentrator 30,000 MWCO PES, VIVASCIENCE). The liposomal suspension was added to the filters and centrifuged at 1,500 g for 5 min 3 times at 25 1C. Fractions were collected and the material trapped in the filter was reconstituted with PBS to achieve the desired dose for in vitro or in vivo use. Three injections of 0.75 mg/kg siRNA were given every other day to rats. The system of Sato et al. may be used/and or adapted to the CRISPR Cas system of the present invention for delivery to the liver by delivering about 0.5 to 1 mg/kg of CRISPR Cas RNA in the liposomes as described by Sato et al. to humans.

[0352] The methods of Rozema et al. (PNAS, Aug. 7, 2007, vol. 104, no. 32) for a vehicle for the delivery of siRNA to hepatocytes both in vitro and in vivo, which Rozema et al. have named siRNA Dynamic PolyConjugates may also be applied to the present invention. Key features of the Dynamic Poly-Conjugate technology include a membrane-active polymer, the ability to reversibly mask the activity of this polymer until it reaches the acidic environment of endosomes, and the ability to target this modified polymer and its siRNA cargo specifically to hepatocytes in vivo after simple, low-pressure i.v. injection. SATA-modified siRNAs are synthesized by reaction of 5' aminemodified siRNA with 1 weight equivalents (wt eq) of Nsuccinimidyl-S-acetylthioacetate (SATA) reagent (Pierce) and 0.36 wt eq of $NaHCO_3$ in water at 4° C. for 16 h. The modified siRNAs are then precipitated by the addition of 9 vol of ethanol and incubation at 80° C. for 2 h. The precipitate is resuspended in 1× siRNA buffer (Dharmacon) and quantified by measuring absorbance at the 260-nm wavelength. PBAVE (30 mg/ml in 5 mMTAPS, pH 9) is modified by addition of 1.5

wt % SMPT (Pierce). After a 1-h incubation, 0.8 mg of SMPT-PBAVE was added to 400 ul of isotonic glucose solution containing 5 mM TAPS (pH 9). To this solution was added 50 μg of SATA-modified siRNA. For the dose-response experiments where [PBAVE] was constant, different amounts of siRNA are added. The mixture is then incubated for 16 h. To the solution is then added 5.6 mg of Hepes free base followed by a mixture of 3.7 mg of CDM-NAG and 1.9 mg of CDM-PEG. The solution is then incubated for at least 1 h at room temperature before injection. CDM-PEG and CDM-NAG are synthesized from the acid chloride generated by using oxalyl chloride. To the acid chloride is added 1.1 molar equivalents polyethylene glycol monomethyl ether (molecular weight average of 450) to generate CDM-PEG or (aminoethoxy) ethoxy-2-(acetylamino)-2-deoxy-β-D-glucopyranoside to generate CDM-NAG. The final product is purified by using reverse-phase HPLC with a 0.1% TFA water/acetonitrile gradient. About 25 to 50 μg of siRNA was delivered to mice. The system of Rozema et al. may be applied to the CRISPR Cas system of the present invention for delivery to the liver, for example by envisioning a dosage of about 50 to about 200 mg of CRISPR Cas for delivery to a human.

Bone

[0353] Oakes and Lieberman (Clin Orthop Relat Res. 2000 October; (379 Suppl):S101-12) discusses delivery of genes to the bone. By transferring genes into cells at a specific anatomic site, the osteoinductive properties of growth factors can be used at physiologic doses for a sustained period to facilitate a more significant healing response. The specific anatomic site, the quality of the bone, and the soft-tissue envelope, influences the selection of the target cells for regional gene therapy. Gene therapy vectors delivered to a treatment site in osteoconductive carriers have yielded promising results. Several investigators have shown exciting results using ex vivo and in vivo regional gene therapy in animal models. Such a system may be used/and or adapted to the CRISPR Cas system for delivery to the bone.

Brain

[0354] Delivery options for the brain include encapsulation of CRISPR enzyme and guide RNA in the form of either DNA or RNA into liposomes and conjugating to molecular Trojan horses for trans-blood brain barrier (BBB) delivery. Molecular Trojan horses have been shown to be effective for delivery of B-gal expression vectors into the brain of non-human primates. The same approach can be used to delivery vectors containing CRISPR enzyme and guide RNA. For instance, Xia C F and Boado R J, Pardridge W M ("Antibody-mediated targeting of siRNA via the human insulin receptor using avidin-biotin technology." Mol Pharm. 2009 May-June; 6(3):747-51. doi: 10.1021/mp800194) describes how delivery of short interfering RNA (siRNA) to cells in culture, and in vivo, is possible with combined use of a receptor-specific monoclonal antibody (mAb) and avidin-biotin technology. The authors also report that because the bond between the targeting mAb and the siRNA is stable with avidin-biotin technology, and RNAi effects at distant sites such as brain are observed in vivo following an intravenous administration of the targeted siRNA.

[0355] Zhang et al. (Mol Ther. 2003 January; 7(1):11-8.)) describe how expression plasmids encoding reporters such as luciferase were encapsulated in the interior of an "artificial virus" comprised of an 85 nm pegylated immunoliposome, which was targeted to the rhesus monkey brain in vivo with a monoclonal antibody (MAb) to the human insulin receptor (HIR). The HIRMAb enables the liposome carrying the exogenous gene to undergo transcytosis across the blood-brain barrier and endocytosis across the neuronal plasma membrane following intravenous injection. The level of luciferase gene expression in the brain was 50-fold higher in the rhesus monkey as compared to the rat. Widespread neuronal expression of the beta-galactosidase gene in primate brain was demonstrated by both histochemistry and confocal microscopy. The authors indicate that this approach makes feasible reversible adult transgenics in 24 hours. Accordingly, the use of immunoliposome is preferred. These may be used in conjunction with antibodies to target specific tissues or cell surface proteins.

[0356] Other means of delivery or RNA are also preferred, such as via nanoparticles (Cho, S., Goldberg, M., Son, S., Xu, Q., Yang, F., Mei, Y., Bogatyrev, S., Langer, R. and Anderson, D., Lipid-like nanoparticles for small interfering RNA delivery to endothelial cells, Advanced Functional Materials, 19: 3112-3118, 2010) or exosomes (Schroeder, A., Levins, C., Cortez, C., Langer, R., and Anderson, D., Lipid-based nanotherapeutics for siRNA delivery, Journal of Internal Medicine, 267: 9-21, 2010, PMID: 20059641).

[0357] Indeed, exosomes have been shown to be particularly useful in delivery siRNA, a system with some parallels to the CRISPR system. For instance, El-Andaloussi S, et al. ("Exosome-mediated delivery of siRNA in vitro and in vivo." Nat Protoc. 2012 December; 7(12):2112-26. doi: 10.1038/nprot.2012.131. Epub 2012 Nov. 15.) describe how exosomes are promising tools for drug delivery across different biological barriers and can be harnessed for delivery of siRNA in vitro and in vivo. Their approach is to generate targeted exosomes through transfection of an expression vector, comprising an exosomal protein fused with a peptide ligand. The exosomes are then purify and characterized from transfected cell supernatant, then siRNA is loaded into the exosomes. Delivery or administration according to the invention can be performed with exosomes, in particular but not limited to the brain.

[0358] Vitamin E (α-tocopherol) may be conjugated with CRISPR Cas and delivered to the brain along with high density lipoprotein (HDL), for example in a similar manner as was done by Uno et al. (HUMAN GENE THERAPY 22:711-719 (June 2011)) for delivering short-interfering RNA (siRNA) to the brain. Mice were infused via Osmotic minipumps (model 1007D; Alzet, Cupertino, CA) filled with phosphate-buffered saline (PBS) or free TocsiBACE or Toc-siBACE/HDL and connected with Brain Infusion Kit 3 (Alzet). A brain-infusion cannula was placed about 0.5 mm posterior to the bregma at midline for infusion into the dorsal third ventricle. Uno et al. found that as little as 3 nmol of Toc-siRNA with HDL could induce a target reduction in comparable degree by the same ICV infusion method. A similar dosage of CRISPR Cas conjugated to α-tocopherol and co-administered with HDL targeted to the brain may be contemplated for humans in the present invention, for example, about 3 nmol to about 3 μmol of CRISPR Cas targeted to the brain may becontemplated.

[0359] Zou et al. ((HUMAN GENE THERAPY 22:465-475 (April 2011)) describes a method of lentiviral-mediated delivery of short-hairpin RNAs targeting PKCγ for in vivo gene silencing in the spinal cord of rats. Zou et al. administered about 10 ul of a recombinant lentivirus having a titer of 1×10^9 transducing units (TU)/ml by an intrathecal catheter. A similar dosage of CRISPR Cas expressed in a lentiviral vector targeted to the brain may be contemplated for humans in the present invention, for example, about 10-50 ml of CRISPR Cas targeted to the brain in a lentivirus having a titer of 1×10^9 transducing units (TU)/ml may becontemplated.

Targeted Deletion, Therapeutic Applications

[0360] 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 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. 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.

[0361] 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.

[0362] 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-1 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.

[0363] 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.

[0364] 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.

[0365] 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.

[0366] 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 B×B1 attB site into the AAVS1 locus of the human ES cell. Then one may use the B×B1 integrase to facilitate the integration of individual guide RNA genes into the B×B1 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 B×B1 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.

[0367] 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).

[0368] 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, Appli-

cants used a multiple approach, with several different sequences targeted at the same time, but independently.

[0369] A suitable example of a protocol for producing AAV, a preferred vector of the invention is provided in Example 34.

[0370] Trinucleotide repeat disorders are preferred conditions to be treated. These are also exemplified herein.

[0371] 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.

[0372] 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).

[0373] 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).

[0374] Non-limiting examples of proteins associated with trinucleotide repeat expansion disorders include AR (androgen receptor), FMRI (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 1 (*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 (pro-

grammed 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.

[0375] 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.

[0376] 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.

[0377] 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

Genome Editing

[0378] 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.

[0379] 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.

Blood

[0380] The present invention also contemplates delivering the CRISPR-Cas system to the blood.

[0381] The plasma exosomes of Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130) were previously described and may be utilized to deliver the CRISPR Cas system to the blood.

[0382] The CRISPR Cas system of the present invention is also contemplated to treat hemoglobinopathies, such as thalassemias and sickle cell disease. See, e.g., International Patent Publication No. WO 2013/126794 for potential targets that may be targeted by the CRISPR Cas system of the present invention.

[0383] US Patent Publication Nos. 20110225664, 20110091441, 20100229252, 20090271881 and 20090222937 assigned to Cellectis, relates to CREI variants, wherein at least one of the two I-CreI monomers has at least two substitutions, one in each of the two functional subdomains of the LAGLIDADG core domain (SEQ ID NO: 60) situated respectively from positions 26 to 40 and 44 to 77 of I-CreI, said variant being able to cleave a DNA target sequence from the human interleukin-2 receptor gamma chain (IL2RG) gene also named common cytokine receptor gamma chain gene or gamma C gene. The target sequences identified in US Patent Publication Nos. 20110225664, 20110091441, 20100229252, 20090271881 and 20090222937 may be utilized for the CRISPR Cas system of the present invention.

[0384] Severe Combined Immune Deficiency (SCID) results from a defect in lymphocytes T maturation, always associated with a functional defect in lymphocytes B (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). Overall incidence is estimated to 1 in 75 000 births. Patients with untreated SCID are subject to multiple opportunist micro-organism infections, and do generally not live beyond one year. SCID can be treated by allogenic hematopoietic

stem cell transfer, from a familial donor. Histocompatibility with the donor can vary widely. In the case of Adenosine Deaminase (ADA) deficiency, one of the SCID forms, patients can be treated by injection of recombinant Adenosine Deaminase enzyme.

[0385] Since the ADA gene has been shown to be mutated in SCID patients (Giblett et al., Lancet, 1972, 2, 1067-1069), several other genes involved in SCID have been identified (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). There are four major causes for SCID: (i) the most frequent form of SCID, SCID-X1 (X-linked SCID or X-SCID), is caused by mutation in the IL2RG gene, resulting in the absence of mature T lymphocytes and NK cells. IL2RG encodes the gamma C protein (Noguchi, et al., Cell, 1993, 73, 147-157), a common component of at least five interleukin receptor complexes. These receptors activate several targets through the JAK3 kinase (Macchi et al., Nature, 1995, 377, 65-68), which inactivation results in the same syndrome as gamma C inactivation; (ii) mutation in the ADA gene results in a defect in purine metabolism that is lethal for lymphocyte precursors, which in turn results in the quasi absence of B, T and NK cells; (iii) V(D)J recombination is an essential step in the maturation of immunoglobulins and T lymphocytes receptors (TCRs). Mutations in Recombination Activating Gene 1 and 2 (RAG1 and RAG2) and Artemis, three genes involved in this process, result in the absence of mature T and B lymphocytes; and (iv) Mutations in other genes such as CD45, involved in T cell specific signaling have also been reported, although they represent a minority of cases (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109).

[0386] Since when their genetic bases have been identified, the different SCID forms have become a paradigm for gene therapy approaches (Fischer et al., Immunol. Rev., 2005, 203, 98-109) for two major reasons. First, as in all blood diseases, an ex vivo treatment can be envisioned. Hematopoietic Stem Cells (HSCs) can be recovered from bone marrow, and keep their pluripotent properties for a few cell divisions. Therefore, they can be treated in vitro, and then reinjected into the patient, where they repopulate the bone marrow. Second, since the maturation of lymphocytes is impaired in SCID patients, corrected cells have a selective advantage. Therefore, a small number of corrected cells can restore a functional immune system. This hypothesis was validated several times by (i) the partial restoration of immune functions associated with the reversion of mutations in SCID patients (Hirschhorn et al., Nat. Genet., 1996, 13, 290-295; Stephan et al., N. Engl. J. Med., 1996, 335, 1563-1567; Bousso et al., Proc. Natl., Acad. Sci. USA, 2000, 97, 274-278; Wada et al., Proc. Natl. Acad. Sci. USA, 2001, 98, 8697-8702; Nishikomori et al., Blood, 2004, 103, 4565-4572), (ii) the correction of SCID-X1 deficiencies in vitro in hematopoietic cells (Candotti et al., Blood, 1996, 87, 3097-3102; Cavazzana-Calvo et al., Blood, 1996, Blood, 88, 3901-3909; Taylor et al., Blood, 1996, 87, 3103-3107; Hacein-Bey et al., Blood, 1998, 92, 4090-4097), (iii) the correction of SCID-X1 (Soudais et al., Blood, 2000, 95, 3071-3077; Tsai et al., Blood, 2002, 100, 72-79), JAK-3 (Bunting et al., Nat. Med., 1998, 4, 58-64; Bunting et al., Hum. Gene Ther., 2000, 11, 2353-2364) and RAG2 (Yates et al., Blood, 2002, 100, 3942-3949) deficiencies in vivo in animal models and (iv) by the result of gene therapy clinical trials (Cavazzana-Calvo et al., Science, 2000, 288, 669-672; Aiuti et al., Nat. Med., 2002; 8, 423-425; Gaspar et al., Lancet, 2004, 364, 2181-2187).

[0387] US Patent Publication No. 20110182867 assigned to the Children's Medical Center Corporation and the President and Fellows of Harvard College relates to methods and uses of modulating fetal hemoglobin expression (HbF) in a hematopoietic progenitor cells via inhibitors of BCL11A expression or activity, such as RNAi and antibodies. The targets disclosed in US Patent Publication No. 20110182867, such as BCL11A, may be targeted by the CRISPR Cas system of the present invention for modulating fetal hemoglobin expression. See also Bauer et al. (Science 11 Oct. 2013: Vol. 342 no. 6155 pp. 253-257) and Xu et al. (Science 18 Nov. 2011: Vol. 334 no. 6058 pp. 993-996) for additional BCL11A targets.

Ears

[0388] The present invention also contemplates delivering the CRISPR-Cas system to one or both ears.

[0389] Researchers are looking into whether gene therapy could be used to aid current deafness treatments—namely, cochlear implants. Deafness is often caused by lost or damaged hair cells that cannot relay signals to auditory neurons. In such cases, cochlear implants may be used to respond to sound and transmit electrical signals to the nerve cells. But these neurons often degenerate and retract from the cochlea as fewer growth factors are released by impaired hair cells.

[0390] U.S. patent application No. 20120328580 describes injection of a pharmaceutical composition into the ear (e.g., auricular administration), such as into the luminae of the cochlea (e.g., the Scala media, Sc vestibulae, and Sc tympani), e.g., using a syringe, e.g., a single-dose syringe. For example, one or more of the compounds described herein can be administered by intratympanic injection (e.g., into the middle ear), and/or injections into the outer, middle, and/or inner ear. Such methods are routinely used in the art, for example, for the administration of steroids and antibiotics into human ears. Injection can be, for example, through the round window of the ear or through the cochlear capsule. Other inner ear administration methods are known in the art (see, e.g., Salt and Plontke, Drug Discovery Today, 10:1299-1306, 2005).

[0391] In another mode of administration, the pharmaceutical composition can be administered in situ, via a catheter or pump. A catheter or pump can, for example, direct a pharmaceutical composition into the cochlear luminae or the round window of the ear and/or the lumen of the colon. Exemplary drug delivery apparatus and methods suitable for administering one or more of the compounds described herein into an ear, e.g., a human ear, are described by McKenna et al., (U.S. Publication No. 2006/0030837) and Jacobsen et al., (U.S. Pat. No. 7,206,639). In some embodiments, a catheter or pump can be positioned, e.g., in the ear (e.g., the outer, middle, and/or inner ear) of a patient during a surgical procedure. In some embodiments, a catheter or pump can be positioned, e.g., in the ear (e.g., the outer, middle, and/or inner ear) of a patient without the need for a surgical procedure.

[0392] Alternatively or in addition, one or more of the compounds described herein can be administered in combination with a mechanical device such as a cochlear implant or a hearing aid, which is worn in the outer ear. An

exemplary cochlear implant that is suitable for use with the present invention is described by Edge et al., (U.S. Publication No. 2007/0093878).

[0393] In some embodiments, the modes of administration described above may be combined in any order and can be simultaneous or interspersed.

[0394] Alternatively or in addition, the present invention may be administered according to any of the Food and Drug Administration approved methods, for example, as described in CDER Data Standards Manual, version number 004 (which is available at fda.give/cder/dsm/DRG/drg00301.htm).

[0395] In general, the cell therapy methods described in U.S. patent application No. 20120328580 can be used to promote complete or partial differentiation of a cell to or towards a mature cell type of the inner ear (e.g., a hair cell) in vitro. Cells resulting from such methods can then be transplanted or implanted into a patient in need of such treatment. The cell culture methods required to practice these methods, including methods for identifying and selecting suitable cell types, methods for promoting complete or partial differentiation of selected cells, methods for identifying complete or partially differentiated cell types, and methods for implanting complete or partially differentiated cells are described below.

[0396] Cells suitable for use in the present invention include, but are not limited to, cells that are capable of differentiating completely or partially into a mature cell of the inner ear, e.g., a hair cell (e.g., an inner and/or outer hair cell), when contacted, e.g., in vitro, with one or more of the compounds described herein. Exemplary cells that are capable of differentiating into a hair cell include, but are not limited to stem cells (e.g., inner ear stem cells, adult stem cells, bone marrow derived stem cells, embryonic stem cells, mesenchymal stem cells, skin stem cells, iPS cells, and fat derived stem cells), progenitor cells (e.g., inner ear progenitor cells), support cells (e.g., Deiters' cells, pillar cells, inner phalangeal cells, tectal cells and Hensen's cells), and/or germ cells. The use of stem cells for the replacement of inner ear sensory cells is described in Li et al., (U.S. Publication No. 2005/0287127) and Li et al., (U.S. patent Ser. No. 11/953,797). The use of bone marrow derived stem cells for the replacement of inner ear sensory cells is described in Edge et al., PCT/US2007/084654. iPS cells are described, e.g., at Takahashi et al., Cell, Volume 131, Issue 5, Pages 861-872 (2007); Takahashi and Yamanaka, Cell 126, 663-76 (2006); Okita et al., Nature 448, 260-262 (2007); Yu, J. et al., Science 318 (5858): 1917-1920 (2007); Nakagawa et al., Nat. Biotechnol. 26:101-106 (2008); and Zaehres and Scholer, Cell 131 (5): 834-835 (2007).

[0397] Such suitable cells can be identified by analyzing (e.g., qualitatively or quantitatively) the presence of one or more tissue specific genes. For example, gene expression can be detected by detecting the protein product of one or more tissue-specific genes. Protein detection techniques involve staining proteins (e.g., using cell extracts or whole cells) using antibodies against the appropriate antigen. In this case, the appropriate antigen is the protein product of the tissue-specific gene expression. Although, in principle, a first antibody (i.e., the antibody that binds the antigen) can be labeled, it is more common (and improves the visualization) to use a second antibody directed against the first (e.g., an anti-IgG). This second antibody is conjugated either with fluorochromes, or appropriate enzymes for colorimetric reactions, or gold beads (for electron microscopy), or with the biotin-avidin system, so that the location of the primary antibody, and thus the antigen, can be recognized.

[0398] The CRISPR Cas molecules of the present invention may be delivered to the ear by direct application of pharmaceutical composition to the outer ear, with compositions modified from US Published application, 20110142917. In some embodiments the pharmaceutical composition is applied to the ear canal. Delivery to the ear may also be referred to as aural or otic delivery.

[0399] In some embodiments the RNA molecules of the invention are delivered in liposome or lipofectin formulations and the like and can be prepared by methods well known to those skilled in the art. Such methods are described, for example, in U.S. Pat. Nos. 5,593,972, 5,589,466, and 5,580,859, which are herein incorporated by reference.

[0400] Delivery systems aimed specifically at the enhanced and improved delivery of siRNA into mammalian cells have been developed, (see, for example, Shen et al FEBS Let. 2003, 539:111-114; Xia et al., Nat. Biotech. 2002, 20:1006-1010; Reich et al., Mol. Vision. 2003, 9: 210-216; Sorensen et al., J. Mol. Biol. 2003, 327: 761-766; Lewis et al., Nat. Gen. 2002, 32:107-108 and Simeoni et al., NAR 2003, 31, 11: 2717-2724) and may be applied to the present invention. siRNA has recently been successfully used for inhibition of gene expression in primates (see for example. Tolentino et al., Retina 24(4):660 which may also be applied to the present invention.

[0401] Qi et al. discloses methods for efficient siRNA transfection to the inner ear through the intact round window by a novel proteidic delivery technology which may be applied to the CRISPR Cas system of the present invention (see, e.g., Qi et al., Gene Therapy (2013), 1-9). In particular, a TAT double stranded RNA-binding domains (TAT-DRBDs), which can transfect Cy3-labeled siRNA into cells of the inner ear, including the inner and outer hair cells, *crista ampullaris, Macula* utriculi and macula sacculi, through intact round-window permeation was successful for delivering double stranded siRNAs in vivo for treating various inner ear ailments and preservation of hearing function. About 40 μl of 10 mM RNA may be contemplated as the dosage for administration to the ear.

[0402] According to Rejali et al. (Hear Res. 2007 June; 228(1-2):180-7), cochlear implant function can be improved by good preservation of the spiral ganglion neurons, which are the target of electrical stimulation by the implant and brain derived neurotrophic factor (BDNF) has previously been shown to enhance spiral ganglion survival in experimentally deafened ears. Rejali et al. tested a modified design of the cochlear implant electrode that includes a coating of fibroblast cells transduced by a viral vector with a BDNF gene insert. To accomplish this type of ex vivo gene transfer, Rejali et al. transduced guinea pig fibroblasts with an adenovirus with a BDNF gene cassette insert, and determined that these cells secreted BDNF and then attached BDNF-secreting cells to the cochlear implant electrode via an agarose gel, and implanted the electrode in the scala tympani. Rejali et al. determined that the BDNF expressing electrodes were able to preserve significantly more spiral ganglion neurons in the basal turns of the cochlea after 48 days of implantation when compared to control electrodes and demonstrated the feasibility of combining cochlear implant therapy with ex vivo gene transfer for enhancing

spiral ganglion neuron survival. Such a system may be applied to the CRISPR Cas system of the present invention for delivery to the ear.

[0403] Mukherjea et al. (Antioxidants & Redox Signaling, Volume 13, Number 5, 2010) document that knockdown of NOX3 using short interfering (si) RNA abrogated cisplatin ototoxicity, as evidenced by protection of OHCs from damage and reduced threshold shifts in auditory brainstem responses (ABRs). Different doses of siNOX3 (0.3, 0.6, and 0.9 μg) were administered to rats and NOX3 expression was evaluated by real time RT-PCR. The lowest dose of NOX3 siRNA used (0.3 μg) did not show any inhibition of NOX3 mRNA when compared to transtympanic administration of scrambled siRNA or untreated *cochleae*. However, administration of the higher doses of NOX3 siRNA (0.6 and 0.9 μg) reduced NOX3 expression compared to control scrambled siRNA. Such a system may be applied to the CRISPR Cas system of the present invention for transtympanic administration with a dosage of about 2 mg to about 4 mg of CRISPR Cas for administration to a human.

[0404] Jung et al. (Molecular Therapy, vol. 21 no. 4, 834-841 April 2013) demonstrate that Hes5 levels in the utricle decreased after the application of siRNA and that the number of hair cells in these utricles was significantly larger than following control treatment. The data suggest that siRNA technology may be useful for inducing repair and regeneration in the inner ear and that the Notch signaling pathway is a potentially useful target for specific gene expression inhibition. Jung et al. injected 8 μg of Hes5 siRNA in 2 μl volume, prepared by adding sterile normal saline to the lyophilized siRNA to a vestibular epithelium of the ear. Such a system may be applied to the CRISPR Cas system of the present invention for administration to the vestibular epithelium of the ear with a dosage of about 1 to about 30 mg of CRISPR Cas for administration to a human.

Eyes

[0405] The present invention also contemplates delivering the CRISPR-Cas system to one or both eyes.

[0406] In yet another aspect of the invention, the CRISPR-Cas system may be used to correct ocular defects that arise from several genetic mutations further described in Genetic Diseases of the Eye, Second Edition, edited by Elias I. Traboulsi, Oxford University Press, 2012.

[0407] For administration to the eye, lentiviral vectors, in particular equine infectious anemia viruses (EIAV) are particularly preferred.

[0408] 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). The vectors are contemplated to have cytomegalovirus (CMV) promoter driving expression of the target gene. Intracameral, subretinal, intraocular and intravitreal injections are all contemplated (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). Intraocular injections may be performed with the aid of an operating microscope. For subretinal and intravitreal injections, eyes may be prolapsed by gentle digital pressure and fundi visualised using a contact lens system consisting of a drop of a coupling medium solution on the cornea covered with a glass microscope slide coverslip. For subretinal injections, the tip of a 10-mm 34-gauge needle, mounted on a 5-μl Hamilton syringe may be advanced under direct visualisation through the superior equatorial sclera tangentially towards the posterior pole until the aperture of the needle was visible in the subretinal space. Then, 2 μl of vector suspension may be injected to produce a superior bullous retinal detachment, thus confirming subretinal vector administration. This approach creates a self-sealing sclerotomy allowing the vector suspension to be retained in the subretinal space until it is absorbed by the RPE, usually within 48 h of the procedure. This procedure may be repeated in the inferior hemisphere to produce an inferior retinal detachment. This technique results in the exposure of approximately 70% of neurosensory retina and RPE to the vector suspension. For intravitreal injections, the needle tip may be advanced through the sclera 1 mm posterior to the corneoscleral limbus and 2 μl of vector suspension injected into the vitreous cavity. For intracameral injections, the needle tip may be advanced through a corneoscleral limbal paracentesis, directed towards the central cornea, and 2 μl of vector suspension may be injected. For intracameral injections, the needle tip may be advanced through a corneoscleral limbal paracentesis, directed towards the central cornea, and 2 μl of vector suspension may be injected. These vectors may be injected at titres of either $1.0\text{-}1.4\times10^{10}$ or $1.0\text{-}1.4\times10^{9}$ transducing units (TU)/ml.

[0409] 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)). Such a vector may be modified for the CRISPR-Cas system of the present invention. Each eye may be treated with either RetinoStat® at a dose of 1.1×10^{5} transducing units per eye (TU/eye) in a total volume of 100 μl.

[0410] In another embodiment, an E1-, partial E3-, E4-deleted adenoviral vector may be contemplated for delivery to the eye. Twenty-eight patients with advanced neovascular age-related macular degeneration (AMD) were given a single intravitreous injection of an E1-, partial E3-, E4-deleted adenoviral vector expressing human pigment epithelium-derived factor (AdPEDF.ll) (see, e.g., Campochiaro et al., Human Gene Therapy 17:167-176 (February 2006)). Doses ranging from 10^{6} to $10^{9.5}$ particle units (PU) were investigated and there were no serious adverse events related to AdPEDF.ll and no dose-limiting toxicities (see, e.g., Campochiaro et al., Human Gene Therapy 17:167-176 (February 2006)). Adenoviral vector-mediated ocular gene transfer appears to be a viable approach for the treatment of ocular disorders and could be applied to the CRISPR Cas system.

[0411] In another embodiment, the sd-rxRNA® system of RXi Pharmaceuticals may be used/and or adapted for delivering CRISPR Cas to the eye. In this system, a single intravitreal administration of 3 μg of sd-rxRNA results in sequence-specific reduction of PPIB mRNA levels for 14 days. The the sd-rxRNA® system may be applied to the CRISPR Cas system of the present invention, contemplating a dose of about 3 to 20 mg of CRISPR administered to a human.

[0412] Millington-Ward et al. (Molecular Therapy, vol. 19 no. 4, 642-649 April 2011) describes adeno-associated virus (AAV) vectors to deliver an RNA interference (RNAi)-based rhodopsin suppressor and a codon-modified rhodopsin replacement gene resistant to suppression due to nucleotide alterations at degenerate positions over the RNAi target site. An injection of either 6.0×10^{8} vp or 1.8×10^{10} vp AAV were subretinally injected into the eyes by Millington-Ward et al. The AAV vectors of Millington-Ward et al. may be applied to the CRISPR Cas system of the present invention, contemplating a dose of about 2×10^{11} to about 6×10^{13} vp administered to a human.

[0413] Dalkara et al. (Sci Transl Med 5, 189ra76 (2013)) also relates to in vivo directed evolution to fashion an AAV vector that delivers wild-type versions of defective genes throughout the retina after noninjurious injection into the eyes' vitreous humor. Dalkara describes a a 7 mer peptide display library and an AAV library constructed by DNA shuffling of cap genes from AAV1, 2, 4, 5, 6, 8, and 9. The rcAAV libraries and rAAV vectors expressing GFP under a CAG or Rho promoter were packaged and and deoxyribonuclease-resistant genomic titers were obtained through quantitative PCR. The libraries were pooled, and two rounds of evolution were performed, each consisting of initial library diversification followed by three in vivo selection steps. In each such step, P30 rho-GFP mice were intravitreally injected with 2 ml of iodixanol-purified, phosphate-buffered saline (PBS)-dialyzed library with a genomic titer of about 1×10^{12} vg/ml. The AAV vectors of Dalkara et al. may be applied to the CRISPR Cas system of the present invention, contemplating a dose of about 1×10^{15} to about 1×10^{16} vg/ml administered to a human.

[0414] In another embodiment, the rhodopsin gene may be targeted for the treatment of retinitis pigmentosa (RP), wherein the system of US Patent Publication No. 20120204282 assigned to Sangamo BioSciences, Inc. may be modified in accordance of the CRISPR Cas system of the present invention.

[0415] In another embodiment, the methods of US Patent Publication No. 20130183282 assigned to Cellectis, which is directed to methods of cleaving a target sequence from the human rhodopsin gene, may also be modified to the CRISPR Cas system of the present invention.

[0416] US Patent Publication No. 20130202678 assigned to Academia *Sinica* relates to methods for treating retinopathies and sight-threatening ophthalmologic disorders relating to delivering of the Puf-A gene (which is expressed in retinal ganglion and pigmented cells of eye tissues and displays a unique anti-apoptotic activity) to the sub-retinal or intravitreal space in the eye. In particular, desirable targets are zgc:193933, prdm1a, spata2, tex10, rbb4, ddx3, zp2.2, Blimp-1 and HtrA2, all of which may be targeted by the CRISPR Cas system of the present invention.

[0417] Wu (Cell Stem Cell, 13:659-62, 2013) designed a guide RNA that led Cas9 to a single base pair mutation that causes cataracts in mice, where it induced DNA cleavage. Then using either the other wild-type allele or oligos given to the zygotes repair mechanisms corrected the sequence of the broken allele and corrected the cataract-causing genetic defect in mutant mouse.

[0418] US Patent Publication No. 20120159653, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with macular degeration (MD). Macular degeneration (MD) is the primary cause of visual impairment in the elderly, but is also a hallmark symptom of childhood diseases such as Stargardt disease, Sorsby fundus, and fatal childhood neurodegenerative diseases, with an age of onset as young as infancy. Macular degeneration results in a loss of vision in the center of the visual field (the macula) because of damage to the retina. Currently existing animal models do not recapitulate major hallmarks of the disease as it is observed in humans. The available animal models comprising mutant genes encoding proteins associated with MD also produce highly variable phenotypes, making translations to human disease and therapy development problematic.

[0419] One aspect of US Patent Publication No. 20120159653 relates to editing of any chromosomal sequences that encode proteins associated with MD which may be applied to the CRISPR Cas system of the present invention. The proteins associated with MD are typically selected based on an experimental association of the protein associated with MD to an MD disorder. For example, the production rate or circulating concentration of a protein associated with MD may be elevated or depressed in a population having an MD disorder relative to a population lacking the MD 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 MD 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).

[0420] By way of non-limiting example, proteins associated with MD include but are not limited to the following proteins: (ABCA4) ATP-binding cassette, sub-family A (ABC1), member 4 ACHM1 achromatopsia (rod monochromacy) 1 ApoE Apolipoprotein E (ApoE) C1QTNF5 (CTRP5) C1q and tumor necrosis factor related protein 5 (C1QTNF5) C2 Complement component 2 (C2) C3 Complement components (C3) CCL2 Chemokine (C-C motif) Ligand 2 (CCL2) CCR2 Chemokine (C-C motif) receptor 2 (CCR2) CD36 Cluster of Differentiation 36 CFB Complement factor B CFH Complement factor CFH H CFHR1 complement factor H-related 1 CFHR3 complement factor H-related 3 CNGB3 cyclic nucleotide gated channel beta 3 CP ceruloplasmin (CP) CRP C reactive protein (CRP) CST3 cystatin C or cystatin 3 (CST3) CTSD Cathepsin D (CTSD) CX3CR1 chemokine (C-X3-C motif) receptor 1 ELOVL4 Elongation of very long chain fatty acids 4 ERCC6 excision repair cross-complementing rodent repair deficiency, complementation group 6 FBLN5 Fibulin-5 FBLN5 Fibulin 5 FBLN6 Fibulin 6 FSCN2 fascin (FSCN2) HMCN1 Hemicentrin 1 HMCN1 hemicentin 1 HTRA1 HtrA serine peptidase 1 (HTRA1) HTRA1 HtrA serine peptidase 1 IL-6 Interleukin 6 IL-8 Interleukin 8 LOC387715 Hypothetical protein PLEKHA1 Pleckstrin homology domain-containing family A member 1 (PLEKHA1) PROM1 Prominin 1 (PROM1 or CD133) PRPH2 Peripherin-2 RPGR retinitis pigmentosa GTPase regulator SERPING1 serpin peptidase inhibitor, clade G, member 1 (C1-inhibitor) TCOF1 Treacle TIMP3 Metalloproteinase inhibitor 3 (TIMP3) TLR3 Toll-like receptor 3

[0421] The identity of the protein associated with MD whose chromosomal sequence is edited can and will vary. In

preferred embodiments, the proteins associated with MD whose chromosomal sequence is edited may be 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, the chemokine (C-C motif) Ligand 2 protein (CCL2) encoded by the CCL2 gene, the chemokine (C-C motif) receptor 2 protein (CCR2) encoded by the CCR2 gene, the ceruloplasmin protein (CP) encoded by the CP gene, the cathepsin D protein (CTSD) encoded by the CTSD gene, or the metalloproteinase inhibitor 3 protein (TIMP3) encoded by the TIMP3 gene. In an exemplary embodiment, the genetically modified animal is a rat, and the edited chromosomal sequence encoding the protein associated with MD may be: (ABCA4) ATP-binding cassette, NM_000350 sub-family A (ABC1), member 4 APOE Apolipoprotein E NM_138828 (APOE) CCL2 Chemokine (C-C NM_031530 motif) Ligand 2 (CCL2) CCR2 Chemokine (C-C NM_021866 motif) receptor 2 (CCR2) CP ceruloplasmin (CP) NM_012532 CTSD Cathepsin D (CTSD) NM_134334 TIMP3 Metalloproteinase NM_012886 inhibitor 3 (TIMP3) The animal or cell may comprise 1, 2, 3, 4, 5, 6, 7 or more disrupted chromosomal sequences encoding a protein associated with MD and zero, 1, 2, 3, 4, 5, 6, 7 or more chromosomally integrated sequences encoding the disrupted protein associated with MD.

[0422] The edited or integrated chromosomal sequence may be modified to encode an altered protein associated with MD. Several mutations in MD-related chromosomal sequences have been associated with MD. Non-limiting examples of mutations in chromosomal sequences associated with MD include those that may cause MD including in the ABCR protein, E471K (i.e. glutamate at position 471 is changed to lysine), R1129L (i.e. arginine at position 1129 is changed to leucine), T1428M (i.e. threonine at position 1428 is changed to methionine), R1517S (i.e. arginine at position 1517 is changed to serine), I1562T (i.e. isoleucine at position 1562 is changed to threonine), and G1578R (i.e. glycine at position 1578 is changed to arginine); in the CCR2 protein, V64I (i.e. valine at position 192 is changed to isoleucine); in CP protein, G969B (i.e. glycine at position 969 is changed to asparagine or aspartate); in TIMP3 protein, S156C (i.e. serine at position 156 is changed to cysteine), G166C (i.e. glycine at position 166 is changed to cysteine), G167C (i.e. glycine at position 167 is changed to cysteine), Y168C (i.e. tyrosine at position 168 is changed to cysteine), S170C (i.e. serine at position 170 is changed to cysteine), Y172C (i.e. tyrosine at position 172 is changed to cysteine) and S181C (i.e. serine at position 181 is changed to cysteine). Other associations of genetic variants in MD-associated genes and disease are known in the art.

Heart

[0423] The present invention also contemplates delivering the CRISPR-Cas system to the heart. For the heart, a myocardium tropic adena-associated virus (AAVM) is preferred, in particular AAVM41 which showed preferential gene transfer in the heart (see, e.g., Lin-Yanga et al., PNAS, Mar. 10, 2009, vol. 106, no. 10). Administration may be systemic or local. A dosage of about $1\text{-}10\times10^{14}$ vector genomes are contemplated for systemic administration. See also, e.g., Eulalio et al. (2012) Nature 492:376 and Somasuntharam et al. (2013) Biomaterials 34: 7790.

[0424] For example, US Patent Publication No. 20110023139, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with cardiovascular disease. Cardiovascular diseases generally include high blood pressure, heart attacks, heart failure, and stroke and TIA. Any chromosomal sequence involved in cardiovascular disease or the protein encoded by any chromosomal sequence involved in cardiovascular disease may be utilized in the methods described in this disclosure. The cardiovascular-related proteins are typically selected based on an experimental association of the cardiovascular-related protein to the development of cardiovascular disease. For example, the production rate or circulating concentration of a cardiovascular-related protein may be elevated or depressed in a population having a cardiovascular disorder relative to a population lacking the cardiovascular 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 cardiovascular-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).

[0425] By way of example, the chromosomal sequence may comprise, but is not limited to, 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), CTSK (cathepsin K), PTGIR (prostaglandin 12 (prostacyclin) receptor (IP)), KCNJ11 (potassium inwardly-rectifying channel, subfamily J, member 11), INS (insulin), CRP (C-reactive protein, pentraxin-related), PDGFRB (platelet-derived growth factor receptor, beta polypeptide), CCNA2 (cyclin A2), PDGFB (platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog)), KCNJ5 (potassium inwardly-rectifying channel, subfamily J, member 5), KCNN3 (potassium intermediate/small conductance calcium-activated channel, subfamily N, member 3), CAPN10 (calpain 10), PTGES (prostaglandin E synthase), ADRA2B (adrenergic, alpha-2B-, receptor), ABCG5 (ATP-binding cassette, sub-family G (WHITE), member 5), PRDX2 (peroxiredoxin 2), CAPN5 (calpain 5), PARP14 (poly (ADP-ribose) polymerase family, member 14), MEX3C (mex-3 homolog C (*C. elegans*)), ACE angiotensin I converting enzyme (peptidyl-dipeptidase A) 1), TNF (tumor necrosis factor (TNF superfamily, member 2)), IL6 (interleukin 6 (interferon, beta 2)), STN (statin), SERPINE1 (serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1), ALB (albumin), ADIPOQ (adiponectin, C1Q and collagen domain containing), APOB (apolipoprotein B (including Ag (x) antigen)), APOE (apolipoprotein E), LEP (leptin), MTHFR (5,10-methylenetetrahydrofolate reductase (NADPH)), APOA1 (apolipoprotein A-I), EDN1 (endothelin 1), NPPB (natriuretic peptide precursor B), NOS3 (nitric oxide synthase 3 (endothelial cell)), PPARG (peroxisome proliferator-activated receptor gamma), PLAT (plasminogen activator, tissue), PTGS2 (prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)), CETP (cholesteryl ester transfer protein, plasma), AGTR1 (angio-

tensin II receptor, type 1), HMGCR (3-hydroxy-3-methylglutaryl-Coenzyme A reductase), IGF1 (insulin-like growth factor 1 (somatomedin C)), SELE (selectin E), REN (renin), PPARA (peroxisome proliferator-activated receptor alpha), PON1 (paraoxonase 1), KNG1 (kininogen 1), CCL2 (chemokine (C-C motif) ligand 2), LPL (lipoprotein lipase), VWF (von Willebrand factor), F2 (coagulation factor II (thrombin)), ICAM1 (intercellular adhesion molecule 1), TGFB1 (transforming growth factor, beta 1), NPPA (natriuretic peptide precursor A), IL10 (interleukin 10), EPO (erythropoietin), SOD1 (superoxide dismutase 1, soluble), VCAM1 (vascular cell adhesion molecule 1), IFNG (interferon, gamma), LPA (lipoprotein, Lp(a)), MPO (myeloperoxidase), ESR1 (estrogen receptor 1), MAPK1 (mitogen-activated protein kinase 1), HP (haptoglobin), F3 (coagulation factor III (thromboplastin, tissue factor)), CST3 (cystatin C), COG2 (component of oligomeric golgi complex 2), MMP9 (matrix metallopeptidase 9 (gelatinase B, 92 kDa gelatinase, 92 kDa type IV collagenase)), SERPINC1 (serpin peptidase inhibitor, clade C (antithrombin), member 1), F8 (coagulation factor VIII, procoagulant component), HMOX1 (heme oxygenase (decycling) 1), APOC3 (apolipoprotein C-III), IL8 (interleukin 8), PROK1 (prokineticin 1), CBS (cystathionine-beta-synthase), NOS2 (nitric oxide synthase 2, inducible), TLR4 (toll-like receptor 4), SELP (selectin P (granule membrane protein 140 kDa, antigen CD62)), ABCA1 (ATP-binding cassette, sub-family A (ABC1), member 1), AGT (angiotensinogen (serpin peptidase inhibitor, clade A, member 8)), LDLR (low density lipoprotein receptor), GPT (glutamic-pyruvate transaminase (alanine aminotransferase)), VEGFA (vascular endothelial growth factor A), NR3C2 (nuclear receptor subfamily 3, group C, member 2), IL18 (interleukin 18 (interferon-gamma-inducing factor)), NOS1 (nitric oxide synthase 1 (neuronal)), NR3C1 (nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)), FGB (fibrinogen beta chain), HGF (hepatocyte growth factor (hepapoietin A; scatter factor)), ILIA (interleukin 1, alpha), RETN (resistin), AKTI (v-akt murine thymoma viral oncogene homolog 1), LIPC (lipase, hepatic), HSPD1 (heat shock 60 kDa protein 1 (chaperonin)), MAPK14 (mitogen-activated protein kinase 14), SPP1 (secreted phosphoprotein 1), ITGB3 (integrin, beta 3 (platelet glycoprotein 111a, antigen CD61)), CAT (catalase), UTS2 (urotensin 2), THBD (thrombomodulin), F10 (coagulation factor X), CP (ceruloplasmin (ferroxidase)), TNFRSF11B (tumor necrosis factor receptor superfamily, member 11b), EDNRA (endothelin receptor type A), EGFR (epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)), MMP2 (matrix metallopeptidase 2 (gelatinase A, 72 kDa gelatinase, 72 kDa type IV collagenase)), PLG (plasminogen), NPY (neuropeptide Y), RHOD (ras homolog gene family, member D), MAPK8 (mitogen-activated protein kinase 8), MYC (v-myc myelocytomatosis viral oncogene homolog (avian)), FN1 (fibronectin 1), CMA1 (chymase 1, mast cell), PLAU (plasminogen activator, urokinase), GNB3 (guanine nucleotide binding protein (G protein), beta polypeptide 3), ADRB2 (adrenergic, beta-2-, receptor, surface), APOA5 (apolipoprotein A-V), SOD2 (superoxide dismutase 2, mitochondrial), F5 (coagulation factor V (proaccelerin, labile factor)), VDR (vitamin D (1,25-dihydroxyvitamin D3) receptor), ALOX5 (arachidonate 5-lipoxygenase), HLA-DRB1 (major histocompatibility complex, class II, DR beta 1), PARP1 (poly(ADP-ribose) polymerase 1), CD40LG (CD40 ligand), PON2 (paraoxonase 2), AGER (advanced glycosylation end product-specific receptor), IRS1 (insulin receptor substrate 1), PTGS1 (prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)), ECE1 (endothelin converting enzyme 1), F7 (coagulation factor VII (serum prothrombin conversion accelerator)), URN (interleukin 1 receptor antagonist), EPHX2 (epoxide hydrolase 2, cytoplasmic), IGFBP1 (insulin-like growth factor binding protein 1), MAPK10 (mitogen-activated protein kinase 10), FAS (Fas (TNF receptor superfamily, member 6)), ABCB1 (ATP-binding cassette, subfamily B (MDR/TAP), member 1), JUN (jun oncogene), IGFBP3 (insulin-like growth factor binding protein 3), CD14 (CD14 molecule), PDE5A (phosphodiesterase 5A, cGMP-specific), AGTR2 (angiotensin II receptor, type 2), CD40 (CD40 molecule, TNF receptor superfamily member 5), LCAT (lecithin-cholesterol acyltransferase), CCR5 (chemokine (C-C motif) receptor 5), MMP1 (matrix metallopeptidase 1 (interstitial collagenase)), TIMP1 (TIMP metallopeptidase inhibitor 1), ADM (adrenomedullin), DYT10 (dystonia 10), STAT3 (signal transducer and activator of transcription 3 (acute-phase response factor)), MMP3 (matrix metallopeptidase 3 (stromelysin 1, progelatinase)), ELN (elastin), USF1 (upstream transcription factor 1), CFH (complement factor H), HSPA4 (heat shock 70 kDa protein 4), MMP12 (matrix metallopeptidase 12 (macrophage elastase)), MME (membrane metallo-endopeptidase), F2R (coagulation factor II (thrombin) receptor), SELL (selectin L), CTSB (cathepsin B), ANXA5 (annexin A5), ADRB1 (adrenergic, beta-1-, receptor), CYBA (cytochrome b-245, alpha polypeptide), FGA (fibrinogen alpha chain), GGT1 (gamma-glutamyltransferase 1), LIPG (lipase, endothelial), HIF1A (hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)), CXCR4 (chemokine (C—X—C motif) receptor 4), PROC (protein C (inactivator of coagulation factors Va and VIIIa)), SCARB1 (scavenger receptor class B, member 1), CD79A (CD79a molecule, immunoglobulin-associated alpha), PLTP (phospholipid transfer protein), ADD1 (adducin 1 (alpha)), FGG (fibrinogen gamma chain), SAA1 (serum amyloid A1), KCNH2 (potassium voltage-gated channel, subfamily H (eag-related), member 2), DPP4 (dipeptidyl-peptidase 4), G6PD (glucose-6-phosphate dehydrogenase), NPR1 (natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A)), VTN (vitronectin), KIAA0101 (KIAA0101), FOS (FBJ murine osteosarcoma viral oncogene homolog), TLR2 (toll-like receptor 2), PPIG (peptidylprolyl isomerase G (cyclophilin G)), ILIR1 (interleukin 1 receptor, type I), AR (androgen receptor), CYP1A1 (cytochrome P450, family 1, subfamily A, polypeptide 1), SERPINA1 (serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1), MTR (5-methyltetrahydrofolate-homocysteine methyltransferase), RBP4 (retinol binding protein 4, plasma), APOA4 (apolipoprotein A-IV), CDKN2A (cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)), FGF2 (fibroblast growth factor 2 (basic)), EDNRB (endothelin receptor type B), ITGA2 (integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)), CABIN1 (calcineurin binding protein 1), SHBG (sex hormone-binding globulin), HMGB1 (high-mobility group box 1), HSP90B2P (heat shock protein 90 kDa beta (Grp94), member 2 (pseudogene)), CYP3A4 (cytochrome P450, family 3, subfamily A, polypeptide 4), GJA1 (gap junction protein, alpha 1, 43 kDa), CAV1 (caveolin 1,

caveolae protein, 22 kDa), ESR2 (estrogen receptor 2 (ER beta)), LTA (lymphotoxin alpha (TNF superfamily, member 1)), GDF15 (growth differentiation factor 15), BDNF (brain-derived neurotrophic factor), CYP2D6 (cytochrome P450, family 2, subfamily D, polypeptide 6), NGF (nerve growth factor (beta polypeptide)), SP1 (Sp1 transcription factor), TGIF1 (TGFB-induced factor homeobox 1), SRC (v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)), EGF (epidermal growth factor (beta-urogastrone)), PIK3CG (phosphoinositide-3-kinase, catalytic, gamma polypeptide), HLA-A (major histocompatibility complex, class I, A), KCNQ1 (potassium voltage-gated channel, KQT-like subfamily, member 1), CNR1 (cannabinoid receptor 1 (brain)), FBN1 (fibrillin 1), CHKA (choline kinase alpha), BEST1 (bestrophin 1), APP (amyloid beta (A4) precursor protein), CTNNB1 (catenin (cadherin-associated protein), beta 1, 88 kDa), IL2 (interleukin 2), CD36 (CD36 molecule (thrombospondin receptor)), PRKAB1 (protein kinase, AMP-activated, beta 1 non-catalytic subunit), TPO (thyroid peroxidase), ALDH7A1 (aldehyde dehydrogenase 7 family, member A1), CX3CR1 (chemokine (C-X3-C motif) receptor 1), TH (tyrosine hydroxylase), F9 (coagulation factor IX), GH1 (growth hormone 1), TF (transferrin), HFE (hemochromatosis), IL17A (interleukin 17A), PTEN (phosphatase and tensin homolog), GSTMI (glutathione S-transferase mu 1), DMD (dystrophin), GATA4 (GATA binding protein 4), F13A1 (coagulation factor XIII, A1 polypeptide), TTR (transthyretin), FABP4 (fatty acid binding protein 4, adipocyte), PON3 (paraoxonase 3), APOC1 (apolipoprotein C-I), INSR (insulin receptor), TNFRSF1B (tumor necrosis factor receptor superfamily, member 1B), HTR2A (5-hydroxytryptamine (serotonin) receptor 2A), CSF3 (colony stimulating factor 3 (granulocyte)), CYP2C9 (cytochrome P450, family 2, subfamily C, polypeptide 9), TXN (thioredoxin), CYP11B2 (cytochrome P450, family 11, subfamily B, polypeptide 2), PTH (parathyroid hormone), CSF2 (colony stimulating factor 2 (granulocyte-macrophage)), KDR (kinase insert domain receptor (a type III receptor tyrosine kinase)), PLA2G2A (phospholipase A2, group IIA (platelets, synovial fluid)), B2M (beta-2-microglobulin), THBS1 (thrombospondin 1), GCG (glucagon), RHOA (ras homolog gene family, member A), ALDH2 (aldehyde dehydrogenase 2 family (mitochondrial)), TCF7L2 (transcription factor 7-like 2 (T-cell specific, HMG-box)), BDKRB2 (bradykinin receptor B2), NFE2L2 (nuclear factor (erythroid-derived 2)-like 2), NOTCH1 (Notch homolog 1, translocation-associated (*Drosophila*)), UGT1A1 (UDP glucuronosyltransferase 1 family, polypeptide A1), IFNA1 (interferon, alpha 1), PPARD (peroxisome proliferator-activated receptor delta), SIRT1 (sirtuin (silent mating type information regulation 2 homolog) 1 (*S. cerevisiae*)), GNRH1 (gonadotropin-releasing hormone 1 (luteinizing-releasing hormone)), PAPPA (pregnancy-associated plasma protein A, pappalysin 1), ARR3 (arrestin 3, retinal (X-arrestin)), NPPC (natriuretic peptide precursor C), AHSP (alpha hemoglobin stabilizing protein), PTK2 (PTK2 protein tyrosine kinase 2), IL13 (interleukin 13), MTOR (mechanistic target of rapamycin (serine/threonine kinase)), ITGB2 (integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)), GSTT1 (glutathione S-transferase theta 1), IL6ST (interleukin 6 signal transducer (gp130, oncostatin M receptor)), CPB2 (carboxypeptidase B2 (plasma)), CYP1A2 (cytochrome P450, family 1, subfamily A, polypeptide 2), HNF4A (hepatocyte nuclear factor 4, alpha), SLC6A4 (solute carrier family 6 (neurotransmitter transporter, serotonin), member 4), PLA2G6 (phospholipase A2, group VI (cytosolic, calcium-independent)), TNFSF11 (tumor necrosis factor (ligand) superfamily, member 11), SLC8A1 (solute carrier family 8 (sodium/calcium exchanger), member 1), F2RL1 (coagulation factor II (thrombin) receptor-like 1), AKR1A1 (aldo-keto reductase family 1, member A1 (aldehyde reductase)), ALDH9A1 (aldehyde dehydrogenase 9 family, member A1), BGLAP (bone gamma-carboxyglutamate (gla) protein), MTTP (microsomal triglyceride transfer protein), MTRR (5-methyltetrahydrofolate-homocysteine methyltransferase reductase), SULT1A3 (sulfotransferase family, cytosolic, 1A, phenol-preferring, member 3), RAGE (renal tumor antigen), C4B (complement component 4B (Chido blood group), P2RY12 (purinergic receptor P2Y, G-protein coupled, 12), RNLS (renalase, FAD-dependent amine oxidase), CREB1 (CAMP responsive element binding protein 1), POMC (proopiomelanocortin), RAC1 (ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)), LMNA (lamin NC), CD59 (CD59 molecule, complement regulatory protein), SCN5A (sodium channel, voltage-gated, type V, alpha subunit), CYP1B1 (cytochrome P450, family 1, subfamily B, polypeptide 1), MIF (macrophage migration inhibitory factor (glycosylation-inhibiting factor)), MMP13 (matrix metallopeptidase 13 (collagenase 3)), TIMP2 (TIMP metallopeptidase inhibitor 2), CYP19A1 (cytochrome P450, family 19, subfamily A, polypeptide 1), CYP21A2 (cytochrome P450, family 21, subfamily A, polypeptide 2), PTPN22 (protein tyrosine phosphatase, non-receptor type 22 (lymphoid)), MYH14 (myosin, heavy chain 14, non-muscle), MBL2 (mannose-binding lectin (protein C) 2, soluble (opsonic defect)), SELPLG (selectin P ligand), AOC3 (amine oxidase, copper containing 3 (vascular adhesion protein 1)), CTSL1 (cathepsin L1), PCNA (proliferating cell nuclear antigen), IGF2 (insulin-like growth factor 2 (somatomedin A)), ITGB1 (integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)), CAST (calpastatin), CXCL12 (chemokine (C—X—C motif) ligand 12 (stromal cell-derived factor 1)), IGHE (immunoglobulin heavy constant epsilon), KCNE1 (potassium voltage-gated channel, Isk-related family, member 1), TFRC (transferrin receptor (p90, CD71)), COL1A1 (collagen, type I, alpha 1), COL1A2 (collagen, type I, alpha 2), IL2RB (interleukin 2 receptor, beta), PLA2G10 (phospholipase A2, group X), ANGPT2 (angiopoietin 2), PROCR (protein C receptor, endothelial (EPCR)), NOX4 (NADPH oxidase 4), HAMP (hepcidin antimicrobial peptide), PTPN11 (protein tyrosine phosphatase, non-receptor type 11), SLC2A1 (solute carrier family 2 (facilitated glucose transporter), member 1), IL2RA (interleukin 2 receptor, alpha), CCL5 (chemokine (C-C motif) ligand 5), IRF1 (interferon regulatory factor 1), CFLAR (CASP8 and FADD-like apoptosis regulator), CALCA (calcitonin-related polypeptide alpha), EIF4E (eukaryotic translation initiation factor 4E), GSTP1 (glutathione S-transferase pi 1), JAK2 (Janus kinase 2), CYP3A5 (cytochrome P450, family 3, subfamily A, polypeptide 5), HSPG2 (heparan sulfate proteoglycan 2), CCL3 (chemokine (C-C motif) ligand 3), MYD88 (myeloid differentiation primary response gene (88)), VIP (vasoactive intestinal peptide), SOAT1 (sterol O-acyltransferase 1), ADRBK1 (adrenergic, beta, receptor kinase 1), NR4A2 (nuclear receptor subfamily 4, group A, member 2), MMP8 (matrix metallopeptidase 8 (neutrophil collagenase)), NPR2 (natriuretic peptide receptor B/gua-

nylate cyclase B (atrionatriuretic peptide receptor B)), GCH1 (GTP cyclohydrolase 1), EPRS (glutamyl-prolyl-tRNA synthetase), PPARGC1A (peroxisome proliferator-activated receptor gamma, coactivator 1 alpha), F12 (coagulation factor XII (Hageman factor)), PECAM1 (platelet/endothelial cell adhesion molecule), CCL4 (chemokine (C-C motif) ligand 4), SERPINA3 (serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3), CASR (calcium-sensing receptor), GJA5 (gap junction protein, alpha 5, 40 kDa), FABP2 (fatty acid binding protein 2, intestinal), TTF2 (transcription termination factor, RNA polymerase II), PROS1 (protein S (alpha)), CTF1 (cardiotrophin 1), SGCB (sarcoglycan, beta (43 kDa dystrophin-associated glycoprotein)), YME1L1 (YME1-like 1 (*S. cerevisiae*)), CAMP (cathelicidin antimicrobial peptide), ZC3H12A (zinc finger CCCH-type containing 12A), AKR1B1 (aldo-keto reductase family 1, member B1 (aldose reductase)), DES (desmin), MMP7 (matrix metallopeptidase 7 (matrilysin, uterine)), AHR (aryl hydrocarbon receptor), CSF1 (colony stimulating factor 1 (macrophage)), HDAC9 (histone deacetylase 9), CTGF (connective tissue growth factor), KCNMA1 (potassium large conductance calcium-activated channel, subfamily M, alpha member 1), UGT1A (UDP glucuronosyltransferase 1 family, polypeptide A complex locus), PRKCA (protein kinase C, alpha), COMT (catechol-. beta.-methyltransferase), S100B (S100 calcium binding protein B), EGR1 (early growth response 1), PRL (prolactin), IL15 (interleukin 15), DRD4 (dopamine receptor D4), CAMK2G (calcium/calmodulin-dependent protein kinase II gamma), SLC22A2 (solute carrier family 22 (organic cation transporter), member 2), CCL11 (chemokine (C-C motif) ligand 11), PGF (B321 placental growth factor), THPO (thrombopoietin), GP6 (glycoprotein VI (platelet)), TACR1 (tachykinin receptor 1), NTS (neurotensin), HNF1A (HNF1 homeobox A), SST (somatostatin), KCND1 (potassium voltage-gated channel, Shal-related subfamily, member 1), LOC646627 (phospholipase inhibitor), TBXAS1 (thromboxane A synthase 1 (platelet)), CYP2J2 (cytochrome P450, family 2, subfamily J, polypeptide 2), TBXA2R (thromboxane A2 receptor), ADH1C (alcohol dehydrogenase 1C (class I), gamma polypeptide), ALOX12 (arachidonate 12-lipoxygenase), AHSG (alpha-2-HS-glycoprotein), BHMT (betaine-homocysteine methyltransferase), GJA4 (gap junction protein, alpha 4, 37 kDa), SLC25A4 (solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4), ACLY (ATP citrate lyase), ALOX5AP (arachidonate 5-lipoxygenase-activating protein), NUMA1 (nuclear mitotic apparatus protein 1), CYP27B1 (cytochrome P450, family 27, subfamily B, polypeptide 1), CYSLTR2 (cysteinyl leukotriene receptor 2), SOD3 (superoxide dismutase 3, extracellular), LTC4S (leukotriene C4 synthase), UCN (urocortin), GHRL (ghrelin/obestatin prepropeptide), APOC2 (apolipoprotein C-II), CLEC4A (C-type lectin domain family 4, member A), KBTBD10 (kelch repeat and BTB (POZ) domain containing 10), TNC (tenascin C), TYMS (thymidylate synthetase), SHCI (SHC (Src homology 2 domain containing) transforming protein 1), LRP1 (low density lipoprotein receptor-related protein 1), SOCS3 (suppressor of cytokine signaling 3), ADH1B (alcohol dehydrogenase 1B (class I), beta polypeptide), KLK3 (kallikrein-related peptidase 3), HSD11B1 (hydroxysteroid (11-beta) dehydrogenase 1), VKORC1 (vitamin K epoxide reductase complex, subunit 1), SERPINB2 (serpin peptidase inhibitor, clade B (ovalbumin), member 2), TNS1 (tensin 1), RNF19A (ring finger protein 19A), EPOR (erythropoietin receptor), ITGAM (integrin, alpha M (complement component 3 receptor 3 subunit)), PITX2 (paired-like homeodomain 2), MAPK7 (mitogen-activated protein kinase 7), FCGR3A (Fc fragment of IgG, low affinity 111a, receptor (CD16a)), LEPR (leptin receptor), ENG (endoglin), GPX1 (glutathione peroxidase 1), GOT2 (glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)), HRH1 (histamine receptor H1), NR112 (nuclear receptor subfamily 1, group I, member 2), CRH (corticotropin releasing hormone), HTR1A (5-hydroxytryptamine (serotonin) receptor 1A), VDAC1 (voltage-dependent anion channel 1), HPSE (heparanase), SFTPD (surfactant protein D), TAP2 (transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)), RNF123 (ring finger protein 123), PTK2B (PTK2B protein tyrosine kinase 2 beta), NTRK2 (neurotrophic tyrosine kinase, receptor, type 2), IL6R (interleukin 6 receptor), ACHE (acetylcholinesterase (Yt blood group)), GLP1R (glucagon-like peptide 1 receptor), GHR (growth hormone receptor), GSR (glutathione reductase), NQO1 (NAD(P)H dehydrogenase, quinone 1), NR5A1 (nuclear receptor subfamily 5, group A, member 1), GJB2 (gap junction protein, beta 2, 26 kDa), SLC9A1 (solute carrier family 9 (sodium/hydrogen exchanger), member 1), MAOA (monoamine oxidase A), PCSK9 (proprotein convertase subtilisin/kexin type 9), FCGR2A (Fc fragment of IgG, low affinity IIa, receptor (CD32)), SERPINF1 (serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1), EDN3 (endothelin 3), DHFR (dihydrofolate reductase), GAS6 (growth arrest-specific 6), SMPD1 (sphingomyelin phosphodiesterase 1, acid lysosomal), UCP2 (uncoupling protein 2 (mitochondrial, proton carrier)), TFAP2A (transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)), C4BPA (complement component 4 binding protein, alpha), SERPINF2 (serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 2), TYMP (thymidine phosphorylase), ALPP (alkaline phosphatase, placental (Regan isozyme)), CXCR2 (chemokine (C—X—C motif) receptor 2), SLC39A3 (solute carrier family 39 (zinc transporter), member 3), ABCG2 (ATP-binding cassette, sub-family G (WHITE), member 2), ADA (adenosine deaminase), JAK3 (Janus kinase 3), HSPA1A (heat shock 70 kDa protein 1A), FASN (fatty acid synthase), FGF1 (fibroblast growth factor 1 (acidic)), F11 (coagulation factor XI), ATP7A (ATPase, Cu++ transporting, alpha polypeptide), CR1 (complement component (3b/4b) receptor 1 (Knops blood group)), GFAP (glial fibrillary acidic protein), ROCK1 (Rho-associated, coiled-coil containing protein kinase 1), MECP2 (methyl CpG binding protein 2 (Rett syndrome)), MYLK (myosin light chain kinase), BCHE (butyrylcholinesterase), LIPE (lipase, hormone-sensitive), PRDX5 (peroxiredoxin 5), ADORA1 (adenosine A1 receptor), WRN (Werner syndrome, RecQ helicase-like), CXCR3 (chemokine (C—X—C motif) receptor 3), CD81 (CD81 molecule), SMAD7 (SMAD family member 7), LAMC2 (laminin, gamma 2), MAP3K5 (mitogen-activated protein kinase kinase kinase 5), CHGA (chromogranin A (parathyroid secretory protein 1)), IAPP (islet amyloid polypeptide), RHO (rhodopsin), ENPP1 (ectonucleotide pyrophosphatase/phosphodiesterase 1), PTHLH (parathyroid hormone-like hormone), NRG1 (neuregulin 1), VEGFC (vascular endothelial growth factor C), ENPEP (glutamyl aminopep-

tidase (aminopeptidase A)), CEBPB (CCAAT/enhancer binding protein (C/EBP), beta), NAGLU (N-acetylglucosaminidase, alpha-), F2RL3 (coagulation factor II (thrombin) receptor-like 3), CX3CL1 (chemokine (C-X3-C motif) ligand 1), BDKRB1 (bradykinin receptor B1), ADAMTS13 (ADAM metallopeptidase with thrombospondin type 1 motif, 13), ELANE (elastase, neutrophil expressed), ENPP2 (ectonucleotide pyrophosphatase/phosphodiesterase 2), CISH (cytokine inducible SH2-containing protein), GAST (gastrin), MYOC (myocilin, trabecular meshwork inducible glucocorticoid response), ATP1A2 (ATPase, Na+/K+ transporting, alpha 2 polypeptide), NF1 (neurofibromin 1), GJB1 (gap junction protein, beta 1, 32 kDa), MEF2A (myocyte enhancer factor 2A), VCL (vinculin), BMPR2 (bone morphogenetic protein receptor, type II (serine/threonine kinase)), TUBB (tubulin, beta), CDC42 (cell division cycle 42 (GTP binding protein, 25 kDa)), KRT18 (keratin 18), HSF1 (heat shock transcription factor 1), MYB (v-myb myeloblastosis viral oncogene homolog (avian)), PRKAA2 (protein kinase, AMP-activated, alpha 2 catalytic subunit), ROCK2 (Rho-associated, coiled-coil containing protein kinase 2), TFPI (tissue factor pathway inhibitor (lipoprotein-associated coagulation inhibitor)), PRKG1 (protein kinase, cGMP-dependent, type I), BMP2 (bone morphogenetic protein 2), CTNND1 (catenin (cadherin-associated protein), delta 1), CTH (cystathionase (cystathionine gamma-lyase)), CTSS (cathepsin S), VAV2 (vav 2 guanine nucleotide exchange factor), NPY2R (neuropeptide Y receptor Y2), IGFBP2 (insulin-like growth factor binding protein 2, 36 kDa), CD28 (CD28 molecule), GSTA1 (glutathione S-transferase alpha 1), PPIA (peptidylprolyl isomerase A (cyclophilin A)), APOH (apolipoprotein H (beta-2-glycoprotein I)), S100A8 (S100 calcium binding protein A8), IL11 (interleukin 11), ALOX15 (arachidonate 15-lipoxygenase), FBLN1 (fibulin 1), NR1H3 (nuclear receptor subfamily 1, group H, member 3), SCD (stearoyl-CoA desaturase (delta-9-desaturase)), GIP (gastric inhibitory polypeptide), CHGB (chromogranin B (secretogranin 1)), PRKCB (protein kinase C, beta), SRD5A1 (steroid-5-alpha-reductase, alpha polypeptide 1 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 1)), HSD11B2 (hydroxysteroid (11-beta) dehydrogenase 2), CALCRL (calcitonin receptor-like), GALNT2 (UDP-N-acetyl-alpha-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-T2)), ANGPTL4 (angiopoietin-like 4), KCNN4 (potassium intermediate/small conductance calcium-activated channel, subfamily N, member 4), PIK3C2A (phosphoinositide-3-kinase, class 2, alpha polypeptide), HBEGF (heparin-binding EGF-like growth factor), CYP7A1 (cytochrome P450, family 7, subfamily A, polypeptide 1), HLA-DRB5 (major histocompatibility complex, class II, DR beta 5), BNIP3 (BCL2/adenovirus E1B 19 kDa interacting protein 3), GCKR (glucokinase (hexokinase 4) regulator), S100A12 (S100 calcium binding protein A12), PADI4 (peptidyl arginine deiminase, type IV), HSPA14 (heat shock 70 kDa protein 14), CXCR1 (chemokine (C—X—C motif) receptor 1), H19 (H19, imprinted maternally expressed transcript (non-protein coding)), KRTAP19-3 (keratin associated protein 19-3), IDDM2 (insulin-dependent diabetes mellitus 2), RAC2 (ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)), RYR1 (ryanodine receptor 1 (skeletal)), CLOCK (clock homolog (mouse)), NGFR (nerve growth factor receptor (TNFR superfamily, member 16)), DBH (dopamine beta-hydroxylase (dopamine beta-monooxygenase)), CHRNA4 (cholinergic receptor, nicotinic, alpha 4), CACNA1C (calcium channel, voltage-dependent, L type, alpha 1C subunit), PRKAG2 (protein kinase, AMP-activated, gamma 2 non-catalytic subunit), CHAT (choline acetyltransferase), PTGDS (prostaglandin D2 synthase 21 kDa (brain)), NR1H2 (nuclear receptor subfamily 1, group H, member 2), TEK (TEK tyrosine kinase, endothelial), VEGFB (vascular endothelial growth factor B), MEF2C (myocyte enhancer factor 2C), MAPKAPK2 (mitogen-activated protein kinase-activated protein kinase 2), TNFRSF11A (tumor necrosis factor receptor superfamily, member 11a, NFKB activator), HSPA9 (heat shock 70 kDa protein 9 (mortalin)), CYSLTR1 (cysteinyl leukotriene receptor 1), MAT1A (methionine adenosyltransferase I, alpha), OPRL1 (opiate receptor-like 1), IMPA1 (inositol (myo)-1 (or 4)-monophosphatase 1), CLCN2 (chloride channel 2), DLD (dihydrolipoamide dehydrogenase), PSMA6 (proteasome (prosome, macropain) subunit, alpha type, 6), PSMB8 (proteasome (prosome, macropain) subunit, beta type, 8 (large multifunctional peptidase 7)), CHI3L1 (chitinase 3-like 1 (cartilage glycoprotein-39)), ALDH1B1 (aldehyde dehydrogenase 1 family, member B1), PARP2 (poly(ADP-ribose) polymerase 2), STAR (steroidogenic acute regulatory protein), LBP (lipopolysaccharide binding protein), ABCC6 (ATP-binding cassette, sub-family C(CFTR/MRP), member 6), RGS2 (regulator of G-protein signaling 2, 24 kDa), EFNB2 (ephrin-B2), GJB6 (gap junction protein, beta 6, 30 kDa), APOA2 (apolipoprotein A-II), AMPD1 (adenosine monophosphate deaminase 1), DYSF (dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)), FDFT1 (farnesyl-diphosphate farnesyltransferase 1), EDN2 (endothelin 2), CCR6 (chemokine (C-C motif) receptor 6), GJB3 (gap junction protein, beta 3, 31 kDa), IL1RL1 (interleukin 1 receptor-like 1), ENTPD1 (ectonucleoside triphosphate diphosphohydrolase 1), BBS4 (Bardet-Biedl syndrome 4), CELSR2 (cadherin, EGF LAG seven-pass G-type receptor 2 (flamingo homolog, *Drosophila*)), F11R (F11 receptor), RAPGEF3 (Rap guanine nucleotide exchange factor (GEF) 3), HYAL1 (hyaluronoglucosaminidase 1), ZNF259 (zinc finger protein 259), ATOX1 (ATX1 antioxidant protein 1 homolog (yeast)), ATF6 (activating transcription factor 6), KHK (ketohexokinase (fructokinase)), SAT1 (spermidine/spermine N1-acetyltransferase 1), GGH (gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase)), TIMP4 (TIMP metallopeptidase inhibitor 4), SLC4A4 (solute carrier family 4, sodium bicarbonate cotransporter, member 4), PDE2A (phosphodiesterase 2A, cGMP-stimulated), PDE3B (phosphodiesterase 3B, cGMP-inhibited), FADS1 (fatty acid desaturase 1), FADS2 (fatty acid desaturase 2), TMSB4X (thymosin beta 4, X-linked), TXNIP (thioredoxin interacting protein), LIMS1 (LIM and senescent cell antigen-like domains 1), RHOB (ras homolog gene family, member B), LY96 (lymphocyte antigen 96), FOXO1 (forkhead box 01), PNPLA2 (patatin-like phospholipase domain containing 2), TRH (thyrotropin-releasing hormone), GJC1 (gap junction protein, gamma 1, 45 kDa), SLC17A5 (solute carrier family 17 (anion/sugar transporter), member 5), FTO (fat mass and obesity associated), GJD2 (gap junction protein, delta 2, 36 kDa), PSRC1 (proline/serine-rich coiled-coil 1), CASP12 (caspase 12 (gene/pseudogene)), GPBAR1 (G protein-coupled bile acid receptor 1), PXK (PX domain containing serine/threonine kinase), IL33 (interleukin 33), TRIB1 (tribbles homolog 1

(*Drosophila*)), PBX4 (pre-B-cell leukemia homeobox 4), NUPR1 (nuclear protein, transcriptional regulator, 1), 15-September (15 kDa selenoprotein), CILP2 (cartilage intermediate layer protein 2), TERC (telomerase RNA component), GGT2 (gamma-glutamyltransferase 2), MT-CO1 (mitochondrially encoded cytochrome c oxidase I), and UOX (urate oxidase, pseudogene).

[0426] In an additional embodiment, the chromosomal sequence may further be selected from Pon1 (paraoxonase 1), LDLR (LDL receptor), ApoE (Apolipoprotein E), Apo B-100 (Apolipoprotein B-100), ApoA (Apolipoprotein (a)), ApoA1 (Apolipoprotein A1), CBS (Cystathione B-synthase), Glycoprotein IIb/IIb, MTHRF (5,10-methylenetetrahydrofolate reductase (NADPH), and combinations thereof. In one iteration, the chromosomal sequences and proteins encoded by chromosomal sequences involved in cardiovascular disease may be chosen from Cacna1C, Sod1, Pten, Ppar (alpha), Apo E, Leptin, and combinations thereof.

Kidneys

[0427] The present invention also contemplates delivering the CRISPR-Cas system to the kidney. Delivery strategies to induce cellular uptake of the therapeutic nucleic acid include physical force or vector systems such as viral-, lipid- or complex-based delivery, or nanocarriers. From the initial applications with less possible clinical relevance, when nucleic acids were addressed to renal cells with hydrodynamic high pressure injection systemically, a wide range of gene therapeutic viral and non-viral carriers have been applied already to target posttranscriptional events in different animal kidney disease models in vivo (Csaba Révész and Peter Hamar (2011). Delivery Methods to Target RNAs in the Kidney, Gene Therapy Applications, Prof. Chunsheng Kang (Ed.), ISBN: 978-953-307-541-9, InTech, Available from: http://www.intechopen.com/books/gene-therapy-applications/delivery-methods-to-target-rnas-in-the-kidney). Delivery methods to the kidney are summarized as follows:

Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
Hydrodynamic/ Lipid	TransIT In Vivo Gene Delivery System, DOTAP	p85α	Acute renal injury	Ischemia-reperfusion	Uptake, bio-distribution	Larson et al., Surgery, (August 2007), Vol. 142, No. 2, pp. (262-269)
Hydrodynamic/ Lipid	Lipofectamine 2000	Fas	Acute renal injury	Ischemia-reperfusion	Blood urea nitrogen, Fas Immuno-histochemistry apoptosis, histological scoring	Hamar et al., Proc Natl Acad Sci, (October 2004), Vol. 101, No. 41, pp. (14883-14888)
Hydrodynamic	n.a.	Apoptosis cascade elements	Acute renal injury	Ischemia-reperfusion	n.a.	Zheng et al., Am J Pathol, (October 2008), Vol. 173, No. 4, pp. (973-980)
Hydrodynamic	n.a.	Nuclear factor kappa-b (NFkB)	Acute renal injury	Ischemia-reperfusion	n.a.	Feng et al., Transplantation, (May 2009), Vol. 87, No. 9, pp. (1283-1289)
Hydrodynamic/ Viral	Lipofectamine 2000	Apoptosis, antagonizing transcription factor (AATF)	Acute renal injury	Ischemia-reperfusion	Apoptosis oxidative stress, caspase activation, membrane lipid peroxidation	Xie & Guo, Am Soc Nephrol, (December 2006), Vol. 17, No. 12, pp. (3336-3346)
Hydrodynamic	pBAsi mU6 Neo/ TransIt-EE Hydrodynamic Delivery System	Gremlin	Diabetic nephropathy	Streptozotozin-induced diabetes	Proteinuria, serum creatinine, glomerular and tubular diameter, collagen type IV/BMP7 expression	Q. Zhang et al., PloS ONE, (July 2010) Vol. 5, No. 7, e11709, pp. (1-13)
Viral/Lipid	pSUPER vector/ Lipofectamine	TGF-β type II receptor	Interstitial renal fibrosis	Unilateral urethral obstruction	α-SMA expression, collagen content,	Kushibikia et al., J Controlled Release, (July 2005), Vol. 105 No. 3, pp. (318-331)
Viral	Adeno-associated virus-2	Mineral corticoid receptor	Hyper-tension caused renal damage	Cold-induced hypertension	blood pressure, serum albumin, serum urea nitrogen, serum creatinine, kidney weight, urinary sodium	Wang et al., Gene Therapy, (July 2006), Vol. 13, No. 14, pp. (1097-1103)

-continued

Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
Hydrodynamic/ Viral	pU6 vector	Luciferace	n.a.	n.a.	uptake	Kobayashi et al., Journal of Pharmacology and Experimental Therapeutics, (February 2004) Vol. 308, No. 2. pp. (688-693)
Lipid	Lipoproteins, albumin	apoB1, apoM	n.a.	n.a.	Uptake, binding affinity to lipoproteins and albumin	Wolfrum et al., Nature Biotechnology, (September 2007), Vol. 25, No. 10, pp. (1149-1157)
Lipid	Lipofectamine2000	p53	Acute renal injury	Ischemic and cisplatin-induced acute injury	Histological scoring, apoptosis	Molitoris et al., J Am Soc Neprol, (August 2009), Vol. 20, No. 8 pp. (1754-1764)
Lipid	DOTAP/DOPE, DOTAP/DOPE/ DOPE-PEG2000	COX-2	Breast adeno-carcinoma	MDA-MB-231 breast cancer xenograft-bearing mouse	Cell viability, uptake	Mikhaylova et al., Cancer Gene Therapy (March 2011), Vol. 16, No. 3, pp. (217-226)
Lipid	Cholesterol	12/15-lipoxygenase	Diabetic nephro-pathy	Strep-tozotozin-induced diabetes	Albuminuria, urinary creatinine, histology, type I and IV collagen, TGF-β, fibronectin, plasminogen activator inhibitor 1	Yuan et al., Am J Physiol Renal Physiol, (June 2008), Vol. 295, pp. (F605-F617)
Lipid	Lipofectamine 2000	Mitochondrial membrane 44 (TIM44)	Diabetic neuro-pathy	Strep-tozotozin-induced diabetes	Cell proliferation and apoptosis, histology, ROS, mitochondrial import of Mn-SOD and glutathione peroxidase, cellular membrane polarization	Y. Zhang et al., J Am Soc Neprol, (April 2006), Vol. 17, No. 4, pp. (1090-1101)
Hydrodynamic/ Lipid	Proteolipo-some	RLIP76	Renal carcinoma	Caki-2 kidney cancer xenograft-bearing mouse	uptake	Singhal et al. Cancer Res. (May 2009), Vol. 69, No. 10 pp. (4244-4251)
Polymer	PEGylated PEI	Luciferase pGL3	n.a.	n.a.	Uptake, biodistribution, erthrocyte aggregation	Malek et al., Toxicology and Applied Pharmacology (April 2009), Vol. 236, No. 1, pp. (97-108)
Polymer	PEGylated poly-L-lysine	MAPK1	Lupus glomerulo-nephritis	Glomerulo-nephritis	Proteinuria, glomerulosclerosis, TGF-β, fibronectin, plasminogen activator inhibitor 1	Shimizu et al., J Am Soc Nephrology, (April 2010), Vol. 21, No. 4, pp. (622-633)
Polymer/ Nano particle	Hyaluronic acid/ Quantum dot/ PEI	VEGF	Kidney cancer/ melanoma	B16F1 melanoma tumor-bearing mouse	Biodistribution, citotoxicity, tumor volume, endocytosis	Jiang et al., Molecular Pharmaceutics, (May-June 2009) Vol. 6, No. 3, pp. (727-737)

-continued

Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
Polymer/ Nano particle	PEGylated polycapro-lactone nanofiber	GAPDH	n.a.	n.a.	cell viability, uptake	Cao et al, J Controlled Release, (June 2010), Vol. 144, No. 2, pp. (203-212)
Aptamer	Spiegelmer mNOX-E36	CC chemokine ligand 2	Glomerulo-sclerosis	Uninephrecto-mized mouse	urinary albumin, urinary creatinine, histopathology, glomerular filtration rate, macrophage count, serum Ccl2, Mac-2+, Ki-67+	Ninichuk et al., Am J Pathol, (March 2008), Vol. 172, No. 3, pp, (628-637)
Aptamer	Aptamer NOX-F37	vasopressin (AVP)	Congestive heart failure	n.a.	Binding affinity to D-AVP, Inhibition of AVP Signaling, Urine osmolality, and sodium concentration,	Purschke et al., Proc Natl Acad Sci, (March 2006), Vol. 103, No. 13, pp. (5173-5178)

[0428] Yuan et al. (Am J Physiol Renal Physiol 295: F605-F617, 2008) investigated whether in vivo delivery of small interfering RNAs (siRNAs) targeting the 12/15-lipoxygenase (12/15-LO) pathway of arachidonate acid metabolism can ameliorate renal injury and diabetic nephropathy (DN) in a streptozotocininjected mouse model of type 1 diabetes. To achieve greater in vivo access and siRNA expression in the kidney, Yuan et al. used double-stranded 12/15-LO siRNA oligonucleotides conjugated with cholesterol. About 400 μg of siRNA was injected subcutaneously into mice. The method of Yuang et al. may be applied to the CRISPR Cas system of the present invention contemplating a 1-2 g subcutaneous injection of CRISPR Cas conjugated with cholesterol to a human for delivery to the kidneys.

[0429] Molitoris et al. (J Am Soc Nephrol 20:1754-1764, 2009) exploited proximal tubule cells (PTCs), as the site of oligonucleotide reabsorption within the kidney to test the efficacy of siRNA targeted to p53, a pivotal protein in the apoptotic pathway, to prevent kidney injury. Naked synthetic siRNA to p53 injected intravenously 4 h after ischemic injury maximally protected both PTCs and kidney function. Molitoris et al.'s data indicates that rapid delivery of siRNA to proximal tubule cells follows intravenous administration. For dose-response analysis, rats were injected with doses of siP53, 0.33; 1, 3, or 5 mg/kg, given at the same four time points, resulting in cumulative doses of 1.32; 4, 12, and 20 mg/kg, respectively. All siRNA doses tested produced a SCr reducing effect on day one with higher doses being effective over approximately five days compared with PBS-treated ischemic control rats. The 12 and 20 mg/kg cumulative doses provided the best protective effect. The method of Molitoris et al. may be applied to the CRISPR Cas system of the present invention contemplating 12 and 20 mg/kg cumulative doses to a human for delivery to the kidneys.

[0430] Thompson et al. (Nucleic Acid Therapeutics, Volume 22, Number 4, 2012) reports the toxicological and pharmacokinetic properties of the synthetic, small interfering RNA 15NP following intravenous administration in rodents and nonhuman primates. 15NP is designed to act via the RNA interference (RNAi) pathway to temporarily inhibit expression of the pro-apoptotic protein p53 and is being developed to protect cells from acute ischemia/reperfusion injuries such as acute kidney injury that can occur during major cardiac surgery and delayed graft function that can occur following renal transplantation. Doses of 800 mg/kg 15NP in rodents, and 1,000 mg/kg I5NP in nonhuman primates, were required to elicit adverse effects, which in the monkey were isolated to direct effects on the blood that included a sub-clinical activation of complement and slightly increased clotting times. In the rat, no additional adverse effects were observed with a rat analogue of 15NP, indicating that the effects likely represent class effects of synthetic RNA duplexes rather than toxicity related to the intended pharmacologic activity of 15NP. Taken together, these data support clinical testing of intravenous administration of 15NP for the preservation of renal function following acute ischemia/reperfusion injury. The no observed adverse effect level (NOAEL) in the monkey was 500 mg/kg. No effects on cardiovascular, respiratory, and neurologic parameters were observed in monkeys following i.v. administration at dose levels up to 25 mg/kg. Therefore, a similar dosage may be contemplated for intravenous administration of CRISPR Cas to the kidneys of a human.

[0431] Shimizu et al. (J Am Soc Nephrol 21: 622-633, 2010) developed a system to target delivery of siRNAs to glomeruli via poly (ethylene glycol)-poly (L-lysine)-based vehicles. The siRNA/nanocarrier complex was approximately 10 to 20 nm in diameter, a size that would allow it to move across the fenestrated endothelium to access to the mesangium. After intraperitoneal injection of fluorescence-labeled siRNA/nanocarrier complexes, Shimizu et al. detected siRNAs in the blood circulation for a prolonged time. Repeated intraperitoneal administration of a mitogen-activated protein kinase 1 (MAPK1) siRNA/nanocarrier complex suppressed glomerular MAPK1 mRNA and protein expression in a mouse model of glomerulonephritis. For the investigation of siRNA accumulation, Cy5-labeled siRNAs complexed with PIC nanocarriers (0.5 ml, 5 nmol of siRNA content), naked Cy5-labeled siRNAs (0.5 ml, 5 nmol), or Cy5-labeled siRNAs encapsulated in HVJ-E (0.5 ml, 5 nmol

of siRNA content) were administrated to BALB-c mice. The method of Shimizu et al. may be applied to the CRISPR Cas system of the present invention contemplating a dose of about of 10-20 μmol CRISPR Cas complexed with nanocarriers in about 1-2 liters to a human for intraperitoneal administration and delivery to the kidneys.

Lungs

[0432] The present invention also contemplates delivering the CRISPR-Cas system to one or both lungs.

[0433] Although AAV-2-based vectors were originally proposed for CFTR delivery to CF airways, other serotypes such as AAV-1, AAV-5, AAV-6, and AAV-9 exhibit improved gene transfer efficiency in a variety of models of the lung epithelium (see, e.g., Li et al., Molecular Therapy, vol. 17 no. 12, 2067-277 December 2009). AAV-1 was demonstrated to be ~100-fold more efficient than AAV-2 and AAV-5 at transducing human airway epithelial cells in vitro, 5 although AAV-1 transduced murine tracheal airway epithelia in vivo with an efficiency equal to that of AAV-5. Other studies have shown that AAV-5 is 50-fold more efficient than AAV-2 at gene delivery to human airway epithelium (HAE) in vitro and significantly more efficient in the mouse lung airway epithelium in vivo. AAV-6 has also been shown to be more efficient than AAV-2 in human airway epithelial cells in vitro and murine airways in vivo.8 The more recent isolate, AAV-9, was shown to display greater gene transfer efficiency than AAV-5 in murine nasal and alveolar epithelia in vivo with gene expression detected for over 9 months suggesting AAV may enable long-term gene expression in vivo, a desirable property for a CFTR gene delivery vector. Furthermore, it was demonstrated that AAV-9 could be readministered to the murine lung with no loss of CFTR expression and minimal immune consequences. CF and non-CF HAE cultures may be inoculated on the apical surface with 100 ul of AAV vectors for hours (see, e.g., Li et al., Molecular Therapy, vol. 17 no. 12, 2067-277 December 2009). The MOI may vary from 1×10^3 to 4×10^5 vector genomes/cell, depending on virus concentration and purposes of the experiments. The above cited vectors are contemplated for the delivery and/or administration of the invention.

[0434] Zamora et al. (Am J Respir Crit Care Med Vol 183. pp 531-538, 2011) reported an example of the application of an RNA interference therapeutic to the treatment of human infectious disease and also a randomized trial of an antiviral drug in respiratory syncytial virus (RSV)-infected lung transplant recipients. Zamora et al. performed a randomized, double-blind, placebocontrolled trial in LTX recipients with RSV respiratory tract infection. Patients were permitted to receive standard of care for RSV. Aerosolized ALN-RSV01 (0.6 mg/kg) or placebo was administered daily for 3 days. This study demonstrates that an RNAi therapeutic targeting RSV can be safely administered to LTX recipients with RSV infection. Three daily doses of ALN-RSV01 did not result in any exacerbation of respiratory tract symptoms or impairment of lung function and did not exhibit any systemic proinflammatory effects, such as induction of cytokines or CRP. Pharmacokinetics showed only low, transient systemic exposure after inhalation, consistent with preclinical animal data showing that ALN-RSV01, administered intravenously or by inhalation, is rapidly cleared from the circulation through exonucleasemediated digestion and renal excretion. The method of Zamora et al. may be applied to the CRISPR Cas system of the present invention and an aerosolized CRISPR Cas, for example with a dosage of 0.6 mg/kg, may be contemplated for the present invention.

[0435] For an example of CFTRdelta508 chimeric guide RNA, see Example 22 which 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 (CF) related symptoms, using adeno-associated virus (AAV) particles. In particular, they exemplify a repair strategy for Cystic Fibrosis delta F508 mutation. This type of strategy should apply across all organisms. With particular reference to CF, suitable patients may include: Human, non-primate human, canine, feline, bovine, equine and other domestic animals. In this instance, Applicants utilized a CRISPR-Cas system comprising a Cas9 enzyme to target deltaF508 or other CFTR-inducing mutations.

[0436] The treated subjects in this instance receive pharmaceutically effective amount of aerosolized AAV vector system per lung endobronchially delivered while spontaneously breathing. As such, aerosolized delivery is preferred for AAV delivery in general. An adenovirus or an AAV particle may be used for delivery. Suitable gene constructs, each operably linked to one or more regulatory sequences, may be cloned into the delivery vector. In this instance, the following constructs are provided as examples: Cbh or EF1a promoter for Cas9, U6 or H1 promoter for chimeric guide RNA): A preferred arrangement is to use a CFTRdelta508 targeting chimeric guide, a repair template for deltaF508 mutation and a codon optimized Cas9 enzyme (preferred Cas9s are those with nuclease or nickase activity) with optionally one or more nuclear localization signal or sequence(s) (NLS(s)), e.g., two (2) NLSs. Constructs without NLS are also envisaged.

[0437] In order to identify the Cas9 target site, Applicants analyzed the human CFTR genomic locus and identified the Cas9 target site. Preferably, in general and in this CF case, the PAM may contain a NGG or a NNAGAAW motif.

[0438] Accordingly, in the case of CF, the present method comprises manipulation of a target sequence in a genomic locus of interest comprising delivering a non-naturally occurring or engineered composition comprising a viral vector system comprising one or more viral vectors operably encoding a composition for expression thereof, wherein the composition comprises:

- **[0439]** a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising
- **[0440]** I. a first regulatory element operably linked to a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises
- **[0441]** (a) a guide sequence capable of hybridizing to the CF target sequence in a suitable mammalian cell,
- **[0442]** (b) a tracr mate sequence, and
- **[0443]** (c) a tracr sequence, and
- **[0444]** 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, 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. In respect of CF, preferred target DNA sequences comprise the CFTRdelta508 mutation. A preferred PAM is described above. A preferred CRISPR enzyme is any Cas (described herein, but particularly that described in Example 22).

[0445] Alternatives to CF include any genetic disorder and examples of these are well known. Another preferred method or use of the invention is for correcting defects in the EMP2A and EMP2B genes that have been identified to be associated with Lafora disease.

[0446] In some embodiments, a "guide sequence" may be distinct from "guide RNA". A guide sequence may refer to an approx. 20 bp sequence, within the guide RNA, that specifies the target site.

[0447] In some embodiments, the Cas9 is (or is derived from) SpCas9. In such embodiments, preferred mutations are at any or all or positions 10, 762, 840, 854, 863 and/or 986 of SpCas9 or corresponding positions in other Cas9s (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. These are advantageous as they provide nickase activity. Such mutations may be applied to all aspects of the present invention, not only treatment of CF.

[0448] Schwank et al. (Cell Stem Cell, 13:653-58, 2013) used CRISPR/Cas9 to correct a defect associated with cystic fibrosis in human stem cells. The team's target was the gene for an ion channel, cystic fibrosis transmembrane conductor receptor (CFTR). A deletion in CFTR causes the protein to misfold in cystic fibrosis patients. Using cultured intestinal stem cells developed from cell samples from two children with cystic fibrosis, Schwank et al. were able to correct the defect using CRISPR along with a donor plasmid containing the reparative sequence to be inserted. The researchers then grew the cells into intestinal "organoids," or miniature guts, and showed that they functioned normally. In this case, about half of clonal organoids underwent the proper genetic correction.

Muscles

[0449] The present invention also contemplates delivering the CRISPR-Cas system to muscle(s).

[0450] Bortolanza et al. (Molecular Therapy vol. 19 no. 11, 2055-264 November 2011) shows that systemic delivery of RNA interference expression cassettes in the FRG1 mouse, after the onset of facioscapulohumeral muscular dystrophy (FSHD), led to a dose-dependent long-term FRG1 knockdown without signs of toxicity. Bortolanza et al. found that a single intravenous injection of 5×10^{12} vg of rAAV6-sh1FRG1 rescues muscle histopathology and muscle function of FRG1 mice. In detail, 200 ul containing 2×10^{12} or 5×10^{12} vg of vector in physiological solution were injected into the tail vein using a 25-gauge Terumo syringe. The method of Bortolanza et al. may be applied to an AAV expressing CRISPR Cas and injected into humans at a dosage of about 2×10^{15} or 2×10^{16} vg of vector.

[0451] Dumonceaux et al. (Molecular Therapy vol. 18 no. 5, 881-887 May 2010) inhibit the myostatin pathway using the technique of RNA interference directed against the myostatin receptor AcvRIIb mRNA (sh-AcvRIIb). The restoration of a quasi-dystrophin was mediated by the vectorized U7 exon-skipping technique (U7-DYS). Adeno-associated vectors carrying either the sh-AcvrIIb construct alone, the U7-DYS construct alone, or a combination of both constructs were injected in the tibialis anterior (TA) muscle of dystrophic mdx mice. The injections were performed with 10^{11} AAV viral genomes. The method of Dumonceaux et al. may be applied to an AAV expressing CRISPR Cas and injected into humans, for example, at a dosage of about 10^{14} to about 10^{15} vg of vector.

[0452] Kinouchi et al. (Gene Therapy (2008) 15, 1126-1130) report the effectiveness of in vivo siRNA delivery into skeletal muscles of normal or diseased mice through nanoparticle formation of chemically unmodified siRNAs with atelocollagen (ATCOL). ATCOL-mediated local application of siRNA targeting myostatin, a negative regulator of skeletal muscle growth, in mouse skeletal muscles or intravenously, caused a marked increase in the muscle mass within a few weeks after application. These results imply that ATCOL-mediated application of siRNAs is a powerful tool for future therapeutic use for diseases including muscular atrophy. Mst-siRNAs (final concentration, 10 mM) were mixed with ATCOL (final concentration for local administration, 0.5%) (AteloGene, Kohken, Tokyo, Japan) according to the manufacturer's instructions. After anesthesia of mice (20-week-old male C57BL/6) by Nembutal (25 mg/kg, i.p.), the Mst-siRNA/ATCOL complex was injected into the masseter and biceps femoris muscles. The method of Kinouchi et al. may be applied to CRISPR Cas and injected into a human, for example, at a dosage of about 500 to 1000 ml of a 40 μM solution into the muscle.

[0453] Hagstrom et al. (Molecular Therapy Vol. 10, No. 2, August 2004) describe an intravascular, nonviral methodology that enables efficient and repeatable delivery of nucleic acids to muscle cells (myofibers) throughout the limb muscles of mammals. The procedure involves the injection of naked plasmid DNA or siRNA into a distal vein of a limb that is transiently isolated by a tourniquet or blood pressure cuff. Nucleic acid delivery to myofibers is facilitated by its rapid injection in sufficient volume to enable extravasation of the nucleic acid solution into muscle tissue. High levels of transgene expression in skeletal muscle were achieved in both small and large animals with minimal toxicity. Evidence of siRNA delivery to limb muscle was also obtained. For plasmid DNA intravenous injection into a rhesus monkey, a threeway stopcock was connected to two syringe pumps (Model PHD 2000; Harvard Instruments), each loaded with a single syringe. Five minutes after a papaverine injection, pDNA (15.5 to 25.7 mg in 40-100 ml saline) was injected at a rate of 1.7 or 2.0 ml/s. This could be scaled up for plasmid DNA expressing CRISPR Cas of the present invention with an injection of about 300 to 500 mg in 800 to 2000 ml saline for a human. For adenoviral vector injections into a rat, 2×10^{9} infectious particles were injected in 3 ml of normal saline solution (NSS). This could be scaled

up for an adenoviral vector expressing CRISPR Cas of the present invention with an injection of about 1×10^{13} infectious particles were injected in 10 liters of NSS for a human. For siRNA, a rat was injected into the great saphenous vein with 12.5 ug of a siRNA and a primate was injected injected into the great saphenous vein with 750 μg of a siRNA. This could be scaled up for a CRISPR Cas of the present invention, for example, with an injection of about 15 to about 50 mg into the great saphenous vein of a human.

Skin

[0454] The present invention also contemplates delivering the CRISPR-Cas system to the skin.

[0455] Hickerson et al. (Molecular Therapy—Nucleic Acids (2013) 2, e129) relates to a motorized microneedle array skin delivery device for delivering self-delivery (sd)-siRNA to human and murine skin. The primary challenge to translating siRNA-based skin therapeutics to the clinic is the development of effective delivery systems. Substantial effort has been invested in a variety of skin delivery technologies with limited success. In a clinical study in which skin was treated with siRNA, the exquisite pain associated with the hypodermic needle injection precluded enrollment of additional patients in the trial, highlighting the need for improved, more "patient-friendly" (i.e., little or no pain) delivery approaches. Microneedles represent an efficient way to deliver large charged cargos including siRNAs across the primary barrier, the stratum corneum, and are generally regarded as less painful than conventional hypodermic needles. Motorized "stamp type" microneedle devices, including the motorized microneedle array (MMNA) device used by Hickerson et al., have been shown to be safe in hairless mice studies and cause little or no pain as evidenced by (i) widespread use in the cosmetic industry and (ii) limited testing in which nearly all volunteers found use of the device to be much less painful than a flushot, suggesting siRNA delivery using this device will result in much less pain than was experienced in the previous clinical trial using hypodermic needle injections. The MMNA device (marketed as Triple-M or Tri-M by Bomtech Electronic Co, Seoul, South Korea) was adapted for delivery of siRNA to mouse and human skin. sd-siRNA solution (up to 300 μl of 0.1 mg/ml RNA) was introduced into the chamber of the disposable Tri-M needle cartridge (Bomtech), which was set to a depth of 0.1 mm. For treating human skin, deidentified skin (obtained immediately following surgical procedures) was manually stretched and pinned to a cork platform before treatment. All intradermal injections were performed using an insulin syringe with a 28-gauge 0.5-inch needle. The MMNA device and method of Hickerson et al. could be used and/or adapted to deliver the CRISPR Cas of the present invention, for example, at a dosage of up to 300 μl of 0.1 mg/ml CRISPR Cas to the skin.

[0456] Leachman et al. (Molecular Therapy, vol. 18 no. 2, 442-446 February 2010) relates to a phase Ib clinical trial for treatment of a rare skin disorder pachyonychia congenita (PC), an autosomal dominant syndrome that includes a disabling plantar keratoderma, utilizing the first short-interfering RNA (siRNA)-based therapeutic for skin. This siRNA, called TD101, specifically and potently targets the keratin 6a (K6a) N171K mutant mRNA without affecting wild-type K6a mRNA. The dose-escalation schedule is presented below:

Week	Dose no.	Days	Volume (ml)	Concentration of TD101 (mg/ml)	Total dose TD101 (mg)
1	1-2	1-7	0.1	1.0	0.10
2	3-4	8-14	0.25	1.0	0.25
3	5-6	15-21	0.50	1.0	0.50
4	7-8	22-28	1.0	1.0	1.0
5	9-10	29-35	1.5	1.0	1.5
6	11-12	36-42	2.0	1.0	2.0
7	13-14	43-49	2.0	1.5	3.0
8	15-16	50-56	2.0	2.0	4.0
9	17-18	57-63	2.0	2.5	5.0
10	19-20	64-70	2.0	3.0	6.0
11	21-22	71-77	2.0	3.5	7.0
12	23-24	78-84	2.0	4.0	8.0
13	25-26	85-91	2.0	4.5	9.0
14	27-28	92-98	2.0	5.0	10.0
15	29-30	99-105	2.0	6.0	12.0
16	31-32	106-112	2.0	7.0	14.0
17	33	113-119	2.0	8.5	17.0

[0457] Initially, 0.1 ml of a 1.0 mg/ml solution of TD101 or vehicle alone (Dulbecco's phosphate-buffered saline without calcium or magnesium) was administered to symmetric calluses. Six rising dose-volumes were completed without an adverse reaction to the increases: 0.1, 0.25, 0.5, 1.0, 1.5, and 2.0 ml of a 1.0 mg/ml solution of TD101 solution per injection. As the highest planned volume (2.0 ml) was well tolerated, the concentration of TD101 was then increased each week from 1 mg/ml up to a final concentration of 8.5 mg/ml. Similar dosages are contemplated for the administration of a CRISPR Cas that specifically and potently targets the keratin 6a (K6a) N171K mutant mRNA.

[0458] Zheng et al. (PNAS, Jul. 24, 2012, vol. 109, no. 30, 11975-11980) show that spherical nucleic acid nanoparticle conjugates (SNA-NCs), gold cores surrounded by a dense shell of highly oriented, covalently immobilized siRNA, freely penetrate almost 100% of keratinocytes in vitro, mouse skin, and human epidermis within hours after application. Zheng et al. demonstrated that a single application of 25 nM epidermal growth factor receptor (EGFR) SNA-NCs for 60 h demonstrate effective gene knockdown in human skin. A similar dosage may be contemplated for CRISPR Cas immobilized in SNA-NCs for administration to the skin.

Hepatitis Viruses

[0459] The present invention may also be applied to treat hepatitis B virus (HBV). However, the CRISPR Cas system must be adapted to avoid the shortcomings of RNAi, such as the risk of oversatring endogenous small RNA pathways, by for example, optimizing dose and sequence (see, e.g., Grimm et al., Nature vol. 441, 26 May 2006). For example, low doses, such as about $1\text{-}10\times10^{14}$ particles per humane are contemplated.

[0460] In another embodiment, the CRISPR Cas system directed against HBV may be administered in liposomes, such as a stable nucleic-acid-lipid particle (SNALP) (see, e.g., Morrissey et al., Nature Biotechnology, Vol. 23, No. 8, August 2005). Daily intravenous injections of about 1, 3 or 5 mg/kg/day of CRISPR Cas targeted to HBV RNA in a SNALP are contemplated. The daily treatment may be over about three days and then weekly for about five weeks.

[0461] In another embodiment, the system of Chen et al. (Gene Therapy (2007) 14, 11-19) may be used/and or adapted for the CRISPR Cas system of the present invention. Chen et al. use a double-stranded adenoassociated virus

8-pseudotyped vector (dsAAV2/8) to deliver shRNA. A single administration of dsAAV2/8 vector (1×10^{12} vector genomes per mouse), carrying HBV-specific shRNA, effectively suppressed the steady level of HBV protein, mRNA and replicative DNA in liver of HBV transgenic mice, leading to up to 2-3 $\log_{10}$ decrease in HBV load in the circulation. Significant HBV suppression sustained for at least 120 days after vector administration. The therapeutic effect of shRNA was target sequence dependent and did not involve activation of interferon. For the present invention, a CRISPR Cas system directed to HBV may be cloned into an AAV vector, such as a dsAAV2/8 vector and administered to a human, for example, at a dosage of about 1×10^{15} vector genomes to about 1×10^{16} vector genomes per human.

[0462] In another embodiment, the method of Wooddell et al. (Molecular Therapy vol. 21 no. 5, 973-985 May 2013) may be used/and or adapted to the CRISPR Cas system of the present invention. Woodell et al. show that simple coinjection of a hepatocyte-targeted, N-acetylgalactosamine-conjugated melittin-like peptide (NAG-MLP) with a liver-tropic cholesterol-conjugated siRNA (chol-siRNA) targeting coagulation factor VII (F7) results in efficient F7 knockdown in mice and nonhuman primates without changes in clinical chemistry or induction of cytokines. Using transient and transgenic mouse models of HBV infection, Wooddell et al. show that a single coinjection of NAG-MLP with potent chol-siRNAs targeting conserved HBV sequences resulted in multilog repression of viral RNA, proteins, and viral DNA with long duration of effect. Intraveinous coinjections, for example, of about 6 mg/kg of NAG-MLP and 6 mg/kg of HBV specific CRISPR Cas may be envisioned for the present invention. In the alternative, about 3 mg/kg of NAG-MLP and 3 mg/kg of HBV specific CRISPR Cas may be delivered on day one, followed by administration of about about 2-3 mg/kg of NAG-MLP and 2-3 mg/kg of HBV specific CRISPR Cas two weeks later.

[0463] The present invention may also be applied to treat hepatitis C virus (HCV). The methods of Roelvinki et al. (Molecular Therapy vol. 20 no. 9, 1737-1749 September 2012) may be applied to the CRISPR Cas system. For example, an AAV vector such as AAV8 may be a contemplated vector and for example a dosage of about 1.25×10^{11} to 1.25×10^{13} vector genomes per kilogram body weight (vg/kg) may be contemplated.

[0464] It will be readily apparent that a host of other diseases can be treated in a similar fashion. Some examples of genetic diseases caused by mutations are provided herein, but many more are known. The above strategy can be applied to these diseases.

Huntington's Disease (HD)

[0465] RNA interference (RNAi) offers therapeutic potential for this disorder by reducing the expression of HTT, the disease-causing gene of Huntington's disease (see, e.g., McBride et al., Molecular Therapy vol. 19 no. 12 Dec. 2011, pp. 2152-2162), therefore Applicant postulates that it may be used/and or adapted to the CRISPR-Cas system. The CRISPR-Cas system may be generated using an algorithm to reduce the off-targeting potential of antisense sequences. The CRISPR-Cas sequences may target either a sequence in exon 52 of mouse, rhesus or human huntingtin and expressed in a viral vector, such as AAV. Animals, including humans, may be injected with about three microinjections per hemisphere (six injections total): the first 1 mm rostral to the anterior commissure (12 μl) and the two remaining injections (12 μl and 10 μl, respectively) spaced 3 and 6 mm caudal to the first injection with 1e12 vg/ml of AAV at a rate of about 1 μl/minute, and the needle was left in place for an additional 5 minutes to allow the injectate to diffuse from the needle tip.

[0466] DiFiglia et al. (PNAS, Oct. 23, 2007, vol. 104, no. 43, 17204-17209) observed that single administration into the adult striatum of an siRNA targeting Htt can silence mutant Htt, attenuate neuronal pathology, and delay the abnormal behavioral phenotype observed in a rapid-onset, viral transgenic mouse model of HD. DiFiglia injected mice intrastriatally with 2 μl of Cy3-labeled cc-siRNA-Htt or unconjugated siRNA-Htt at 10 μM. A similar dosage of CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 5-10 ml of 10 μM CRISPR Cas targeted to Htt may be injected intrastriatally.

[0467] In another example, Boudreau et al. (Molecular Therapy vol. 17 no. 6 Jun. 2009) injects 5 μl of recombinant AAV serotype 2/1 vectors expressing htt-specific RNAi virus (at 4×10^{12} viral genomes/ml) into the straiatum. A similar dosage of CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 10-20 ml of 4×10^{12} viral genomes/ml) CRISPR Cas targeted to Htt may be injected intrastriatally.

[0468] In another example, a CRISPR Cas targeted to HTT may be administered continuously (see, e.g., Yu et al., Cell 150, 895-908, Aug. 31, 2012). Yu et al. utilizes osmotic pumps delivering 0.25 ml/hr (Model 2004) to deliver 300 mg/day of ss-siRNA or phosphate-buffered saline (PBS) (Sigma Aldrich) for 28 days, and pumps designed to deliver 0.5 μl/hr (Model 2002) were used to deliver 75 mg/day of the positive control MOE ASO for 14 days. Pumps (Durect Corporation) were filled with ss-siRNA or MOE diluted in sterile PBS and then incubated at 37 C for 24 or 48 (Model 2004) hours prior to implantation. Mice were anesthetized with 2.5% isofluorane, and a midline incision was made at the base of the skull. Using stereotaxic guides, a cannula was implanted into the right lateral ventricle and secured with Loctite adhesive. A catheter attached to an Alzet osmotic mini pump was attached to the cannula, and the pump was placed subcutaneously in the midscapular area. The incision was closed with 5.0 nylon sutures. A similar dosage of CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 500 to 1000 g/day CRISPR Cas targeted to Htt may be administered.

[0469] In another example of continuous infusion, Stiles et al. (Experimental Neurology 233 (2012) 463-471) implanted an intraparenchymal catheter with a titanium needle tip into the right putamen. The catheter was connected to a SynchroMed® II Pump (Medtronic Neurological, Minneapolis, MN) subcutaneously implanted in the abdomen. After a 7 day infusion of phosphate buffered saline at 6 uL/day, pumps were re-filled with test article and programmed for continuous delivery for 7 days. About 2.3 to 11.52 mg/d of siRNA were infused at varying infusion rates of about 0.1 to 0.5 uL/min. A similar dosage of CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 20 to 200 mg/day CRISPR Cas targeted to Htt may be administered.

[0470] In another example, the methods of US Patent Publication No. 20130253040 assigned to Sangamo may

also be also be adapted from TALES to the CRISPR Cas system of the present invention for treating Huntington's Disease.

Nucleic Acids, Amino Acids and Proteins

[0471] 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.

[0472] 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.

[0473] 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.

[0474] 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.

[0475] 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.

[0476] "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.

[0477] 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.

[0478] "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.

[0479] 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.

[0480] 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.

[0481] 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.

[0482] 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.

[0483] 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.

[0484] 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.

[0485] 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.

[0486] 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.

[0487] 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.

[0488] 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 p387). Examples of other software than may perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel et al., 1999 *Short Protocols in Molecular Biology,* 4th 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; *FEMS 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).

[0489] 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.

[0490] 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.

[0491] 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	
Hydro-phobic	F W Y H K 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
		Positive-ly charged	H K R
		Negative-ly charged	E D
Small	V C A G S P T N D	Tiny	A G S

[0492] 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.

[0493] 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.

[0494] 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 PROTOCOLS 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)).

Vectors

[0495] In one aspect, the invention provides for vectors that are used in the engineering and optimization of CRISPR-Cas systems.

[0496] 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.

[0497] 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.

[0498] 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 (SEQ ID NO: 61). 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.

[0499] 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.

[0500] 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 expres-

sion 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.

[0501] 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.

[0502] 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).

[0503] 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 (In Vitrogen Corp, San Diego, Calif.).

[0504] 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).

[0505] 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.

[0506] 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 a-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.

Regulatory Elements

[0507] 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, Campylobacter, Wolinella, Acinetobacter, Erwinia, Escherichia, Legionella, Methylococcus, Pasteurella, Photobacterium, Salmonella, Xanthomonas, Yersinia, Treponema*, and *Thermotoga.*

[0508] 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.

[0509] In some embodiments, direct repeats may be identified in silico by searching for repetitive motifs that fulfill any or all of the following criteria:

[0510] 1. found in a 2Kb window of genomic sequence flanking the type II CRISPR locus;

[0511] 2. span from 20 to 50 bp; and

[0512] 3. interspaced by 20 to 50 bp.

[0513] 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.

[0514] In some embodiments, candidate tracrRNA may be subsequently predicted by sequences that fulfill any or all of the following criteria:

[0515] 1. sequence homology to direct repeats (motif search in Geneious with up to 18-bp mismatches);

[0516] 2. presence of a predicted Rho-independent transcriptional terminator in direction of transcription; and

[0517] 3. stable hairpin secondary structure between tracrRNA and direct repeat.

[0518] 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.

[0519] In some embodiments, chimeric synthetic guide RNAs (sgRNAs) designs may incorporate at least 12 bp of duplex structure between the direct repeat and tracrRNA.

[0520] 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 defence in bacteria and archaea, Mole Cell 2010 Jan. 15; 37(1): 7.

[0521] 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). 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.

[0522] 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.

[0523] 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.

[0524] 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, Cas1B, 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.

[0525] 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.

[0526] Preferred orthologs are described herein. A Cas enzyme may be identified 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.

[0527] 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, StlCas9 and so forth.

Codon Optimization

[0528] 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.

[0529] 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.

[0530] 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.

Nuclear Localization Sequences (NLSs)

[0531] 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 (SEQ ID NO: 62); the NLS from nucleoplasmin (e.g. the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK (SEQ ID NO: 63)); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID NO: 64) or RQRRNELKRSP (SEQ ID NO: 65); the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAK-PRNQGGY (SEQ ID NO: 66); the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKD-EQILKRRNV (SEQ ID NO: 67) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID NO: 68) and PPKKARED (SEQ ID NO: 69) of the myoma T protein; the sequence PQPKKKPL (SEQ ID NO: 70) of

human p53; the sequence SALIKKKKKMAP (SEQ ID NO: 71) of mouse c-abl IV; the sequences DRLRR (SEQ ID NO: 72) and PKQKKRK (SEQ ID NO: 73) of the influenza virus NS1; the sequence RKLKKKIKKL (SEQ ID NO: 74) of the Hepatitis virus delta antigen; the sequence REKKKFLKRR (SEQ ID NO: 75) of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 76) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK (SEQ ID NO: 77) of the steroid hormone receptors (human) glucocorticoid.

[0532] In general, the one or more NLSs 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.

Guide Sequence

[0533] 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.novocraft.com), ELAND (Illumina, San Diego, CA), 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.

[0534] 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 MMMMMMMMNNNNNNNNNNNNXGG where NNNNNNNNNNNNXGG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. For the *S. thermophilus* CRISPRI Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXXAGAAW (SEQ ID NO: 78) where NNNNNNNNNNNNXXAGAAW (SEQ ID NO: 79) (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* CRISPR1 Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXXAGAAW (SEQ ID NO: 80) where NNNNNNNNNNNNXXAGAAW (SEQ ID NO: 81) (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 t 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 MMMMMMMMNNNNNNNNNNNNXGGXG where NNNNNNNNNNNNXGGXG (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.

[0535] In some embodiments, a guide sequence is selected to reduce the degree 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).

Tracr Mate Sequence

[0536] 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) NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagataa ggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (SEQ ID NO: 82); (2) NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccg aaatcaacaccctgtcatttttatggcagggtgttttcgttatttaaTTTTTT (SEQ ID NO: 83); (3) NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccg aaatcaacaccctgtcatttttatggcagggtgtTTTTTT (SEQ ID NO: 84); (4) NNNNNNNNNNNNNNNNNNNNgtttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcTTTTTT (SEQ ID NO: 85); (5) NNNNNNNNNNNNNNNNNNNNgtttttagagctaGAAATAGcaagttaaaataaggctagtccgttatcaacttgaaaaagtgTTTTTTT (SEQ ID NO: 86); and (6) NNNNNNNNNNNNNNNNNNNNgtttttagagctagAAATAGcaagttaaaataaggctagtccgttatcaTTTTTTT (SEQ ID NO: 87). 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.

Recombination Template

[0537] 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.

Fusion Protein

[0538] 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, GAL4 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.

Inducible System

[0539] 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.

Delivery

[0540] 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).

[0541] 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).

[0542] 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).

[0543] 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.

[0544] 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).

[0545] 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 epsilon-retroviral 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.

[0546] 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.

[0547] 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).

[0548] 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.

[0549] 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.

[0550] 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.

[0551] 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.

[0552] 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

[0553] 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).

[0554] 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.

[0555] 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, Panc1, PC-3, TF1, CTLL-2, CIR, 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) (Manassas, 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.

[0556] 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.

[0557] 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.

Modifying a Target

[0558] 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.

[0559] 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.

[0560] 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.

[0561] 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.

Kits

[0562] 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.

[0563] 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.

CRISPR Complex

[0564] 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.

[0565] 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.

[0566] 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 modify 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.

[0567] 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.

[0568] 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.

[0569] 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.

[0570] 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.

[0571] 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).

[0572] 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.

[0573] 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.

[0574] 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.

[0575] 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.

[0576] 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.

Disease Models

[0577] 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.

[0578] 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.

[0579] 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.

[0580] 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.

[0581] 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.

[0582] 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.

[0583] 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.

[0584] 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.

[0585] 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.

[0586] 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.

[0587] 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.

[0588] 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.

[0589] 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.

[0590] 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.

[0591] 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.

[0592] 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.

[0593] 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.

[0594] 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.

[0595] 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.

[0596] 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.

[0597] 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 con-

ventional 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.

[0598] 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.

[0599] 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).

[0600] 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.

[0601] 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.

[0602] 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).

[0603] 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.

[0604] 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.

[0605] 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.

[0606] 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 tis-

sues 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.

[0607] 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.

[0608] 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 application 61/736,527 filed Dec. 12, 2012. 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; Uchl3; 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); II-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);

TABLE B-continued

	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, 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 (FSHMD1A, FSHD1A); Muscular dystrophy (FKRP, MDC1C, LGMD2I, 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, Neurexin1, 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,

TABLE B-continued

	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), DMPK (myotonic dystrophy), Atrophin-1 and Atn1 (DRPLA Dx), CBP (Creb-BP-global instability), VLDLR (Alzheimer's), Atxn7, Atxn10).
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;

TABLE C-continued

CELLULAR FUNCTION	GENES
	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; 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; ATF4; 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; RELB; 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

TABLE C-continued

CELLULAR FUNCTION	GENES
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
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; PDGFRB; TNF; INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1; PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10; GSK3B; AKT3; TNFAIP3; IL1R1

TABLE C-continued

CELLULAR FUNCTION	GENES
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; FYN; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; 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
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; KRAS; 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

TABLE C-continued

CELLULAR FUNCTION	GENES
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: G1/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; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN; MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; 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; 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

TABLE C-continued

CELLULAR FUNCTION	GENES
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; UBE2I; 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; SMURF1; MAPK1; SMAD3; SMAD1; FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2; SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2; MAP2K1; TGFBR1; 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
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; AKT1; 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
LXR/RXR 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

TABLE C-continued

CELLULAR FUNCTION	GENES
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	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Interferon Signaling	IRF1; 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
Linoleic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Methionine Metabolism	DNMT1; DNMT3B; AHCY; DNMT3A
Pyruvate Metabolism	GLO1; ALDH1A1; 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, Coumarine and Lignin Biosynthesis	PRDX6; PRDX1; TYR
Antigen Presentation Pathway	CALR; B2M
Biosynthesis of Steroids	NQO1; DHCR7
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	ALDH1A1; LDHA
Selenoamino Acid Metabolism	PRMT5; AHCY

TABLE C-continued

CELLULAR FUNCTION	GENES
Sphingolipid Metabolism	SPHK1; SPHK2
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; Cnr1; 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 (Pou4f1 or Brn3a); Numb; Reln

[0609] 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. McIvor 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.

[0610] 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).

[0611] 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.

[0612] 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.

[0613] 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.

[0614] 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%.

[0615] 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).

[0616] 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-4 X-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 2 OR4M2 Olfactory receptor (AFF2) 4M2 FOXP2 Forkhead box protein P2 OR4N4 Olfactory receptor (FOXP2) 4N4 FXR1 Fragile 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 (GABRA5) 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 TAS2R1 Taste 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)

[0617] 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 (MGLUR5)

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 FXR1 Fragile X mental NM_001012179 retardation, autosomal homolog 1 (FXR1) FXR2 Fragile 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) MGLUR5 Metabotropic glutamate NM_017012 (GRM5) receptor 5 (MGLUR5) NRXN1 Neurexin-1 NM_021767 SEMA5A Semaphorin-5A (SEMA5A) NM_001107659

[0618] 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 1425V 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.

[0619] 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.

[0620] Examples of addiction-related proteins may include ABAT for example.

[0621] 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.

[0622] Examples of cardiovascular diseases associated proteins may include IL1B (interleukin 1, beta), XDH (xanthine dehydrogenase), TP53 (tumor protein p53), PTGIS (prostaglandin I2 (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.

[0623] 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.

[0624] 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).

[0625] 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.

[0626] 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 I-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) HTR7 5-hydroxytryptamine (serotonin) receptor 7 (adenylate cyclase-coupled) IDE Insulin degrading enzyme IF127 IF127 IFI6 Interferon, alpha-inducible protein 6 (IFI6) IFIT2 Interferon-induced protein with tetratricopeptide repeats 2 (IFIT2) IL1RN interleukin-1 receptor antagonist (IL-IRA) 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) ORMI 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)

[0627] 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 (UBEIC) 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) SORL1 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)

[0628] 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.

[0629] 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.

[0630] 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.

[0631] 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.

[0632] Examples of proteins associated Schizophrenia may include NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, DISC1, GSK3B, and combinations thereof.

[0633] 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.

[0634] 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), APHIB (anterior pharynx defective 1 homolog B (*C. elegans*)), PSEN2 (presenilin 2 (Alzheimer disease 4)), or BACE1 (beta-site APP-cleaving enzyme 1), for example.

[0635] 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).

[0636] 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).

[0637] 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), APHIB (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), SP1 (Sp1 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)), ILIA (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*)), ILIR1 (interleukin 1 receptor, type I), PROK1 (prokineticin 1), MAPK3 (mitogen-activated protein kinase 3), NTRK1 (neurotrophic tyrosine kinase, receptor, type 1), IL13 (interleukin 13), MME (membrane metallo-endopep-

tidase), 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), NR112 (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.

[0638] 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.

[0639] 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.

[0640] 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.

[0641] 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).

[0642] 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 (RaIGDS/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 33 carrier family 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, phosphoribosylaminoimi dazole 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 AKTI 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 SLC1A3 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 fms-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) OGGI 8-oxoguanine DNA IL6 interleukin 6 (interferon, glycosylase beta 2).

[0643] 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.

[0644] 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.

[0645] 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-ID adrenoreceptor), for example.

[0646] 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.

[0647] 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.

[0648] 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.

[0649] 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.

[0650] 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 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.

[0651] 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.

[0652] 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.

[0653] 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 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

[0654] 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

[0655] 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.

Cell Culture and Transfection

[0656] 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), 100U/mL penicillin, and 100 μg/mL streptomycin at 37° C. with 5% CO_2.

[0657] 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.

Surveyor Assay and Sequencing Analysis for Genome Modification

[0658] 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.

[0659] 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.

Restriction Fragment Length Polymorphism Assay for Detection of Homologous Recombination.

[0660] 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).

RNA Secondary Structure Prediction and Analysis

[0661] 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).

RNA Purification

[0662] 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.

Northern Blot Analysis of crRNA and tracrRNA Expression in Mammalian Cells

[0663] RNAs were mixed with equal volumes of 2X 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-32P] 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).

Bacterial CRISPR System Construction and Evaluation

[0664] 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

[0665] 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.

[0666] 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 "tracrmate 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.

[0667] 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).

[0668] 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.

[0669] 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.

[0670] 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 (GTTTTAGAGCTA (SEQ ID NO: 88)) and a partial tracrRNA sequence (TAGCAAGTTAAAATAAGGCTAGTCCGTTTTT (SEQ ID NO: 89)). 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**).

[0671] 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 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**).

[0672] 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).

[0673] 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.

[0674] 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 HI promoter.

[0675] 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.

[0676] 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).

[0677] 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.

[0678] 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

[0679] 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

[0680] 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' (SEQ ID NO: 90) 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.

[0681] 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**.

[0682] 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

[0683] 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 (SEQ ID NO: 61) 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	SEQ ID NO:	strand
1	EMX1	GGACATCGATGTCACCTCCAATGACTAGGG	TGG	91	+
2	EMX1	CATTGGAGGTGACATCGATGTCCTCCCCAT	TGG	92	−
3	EMX1	GGAAGGGCCTGAGTCCGAGCAGAAGAAGAA	GGG	93	+
4	PVALB	GGTGGCGAGAGGGGCCGAGATTGGGTGTTC	AGG	94	+
5	PVALB	ATGCAGGAGGGTGGCGAGAGGGGCCGAGAT	TGG	95	+

[0684] 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.

[0685] 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*).

[0686] 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 3' end of the tracrRNA may play a role in enhancing the rate of CRISPR complex formation.

Example 5: Cas9 Diversity

[0687] 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.

[0688] 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).

[0689] 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

[0690] 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.

[0691] 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:

[0692] 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.

[0693] 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

[0694] 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.

[0695] Example primer pair that allows the generation a PCR product consisting of U6-promoter::guideRNA targeting human Emx1 locus:

```
Forward Primer:
                                        (SEQ ID NO: 96)
AAACTCTAGAgagggcctatttcccatgattc

Reverse Primer (carrying the guide RNA, which is
underlined):
                                        (SEQ ID NO: 97)
acctctagAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAA

CGGACTAGCCTTATTTTAACTTGCTATGCTGTTTTGTTTCCAAAACAGC

ATAGCTCTAAAACCCCTAGTCATTGGAGGTGACGGTGTTTCGTCCTTTC

CACaag
```

Example 8: Methodological Improvement to Improve Activity

[0696] 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.

[0697] One example of this system may involve introduction of three pieces of DNA:

[0698] 1. expression vector for Cas9

[0699] 2. expression vector for T7 polymerase

[0700] 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

[0701] 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:

```
Forward Primer (to add on T7 promoter for in
vitro transcription):
                                        (SEQ ID NO: 98)
TAATACGACTCACTATAGGAAGTGCGCCACCATGGCCCCAAAGAAGAAG

CGG

Reverse Primer (to add on polyA tail):
                                        (SEQ ID NO: 99)
GGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTttcttaCTTTTTCTTTT

TTGCCTGGCCG
```

[0702] 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

[0703] 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

[0704] 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.

[0705] 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)

(SEQ ID NO: 100)

MARILAFDIGISSIGWAFSENDELKGCGVRIFTKVENPKTGESLALPRR

LARSARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSL

ISPYELRFRALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILK

AIKQNEEKLANYQSVGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIA

QSFLKDELKLIFKKQREFGFSFSKKFEEEVLSVAFYKRALKDFSHLVGN

CSFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKNTEGILYTKDDLNAL

LNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKALG

EHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKD

HLNISFKALKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLPAFNET

YYKDEVTNPVVLRAIKEYRKVLNALLKKYGKVHKINIELAREVGHNHSQ

RAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAY

SGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQ

TPFEAFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNL

NDTRYIARLVLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGML

TSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESN

SAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPS

-continued

GALHEETFRKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVKNGDMFRV

DIFKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILMDENY

EFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSSTVSLIVSKHDNKFETL

SKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQRE

DFKK.

The Putative tracrRNA Element for this CjCas9 is:

(SEQ ID NO: 101)

TATAATCTCATAAGAAATTTAAAAAGGGACTAAAATAAAGAGTTTGCGG

GACTCTGCGGGGTTACAATCCCCTAAAACCGCTTTTAAAATT

The Direct Repeat Sequence is:

(SEQ ID NO: 102)

ATTTTACCATAAAGAAATTTAAAAAGGGACTAAAAC

An Example of a Chimeric guideRNA for CjCas9 is:

(SEQ ID NO: 103)

NNNNNNNNNNNNNNNNNNNNGUUUUAGUCCCGAAAGGGACUAAAAUAAA

GAGUUUGCGGGACUCUGCGGGGUUACAAUCCCCUAAAACCGCUUUU

Example 12: Cas9 Optimization

[0706] 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:

[0707] 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

(SEQ ID NO: 104)

MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNR

QGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDEL

SNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKT

PGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQ

QEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDN

IFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQ

KNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTF

EAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGS

FSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTIL

TRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEY

GDFDNIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN

TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSID

NKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA

-continued

ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV

KVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK

LESEFVYGDYKVYDVRKMIAKSEQEIGKATAFYFFYSNIMNFFKTEITLA

NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG

GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK

SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSL

FELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNE

QKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR

EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIT

GLYETRIDLSQLGGD

>Sp(N)St1(C)Cas9

(SEQ ID NO: 105)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKEFKVLGNTDRHSIKKNLI

GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF

FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST

DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL

FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL

SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA

KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQ

LPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV

KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKI

EKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ

SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA

FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF

NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL

KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKS

DGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIK

KGILQTVKVVDELVKVMGRHKPENIVIEMARE**TNEDDEKKAIQKIQKAN**

KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT

GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQR

TPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRK

KFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHW

GIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETG

ELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISD

ATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFL

MYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRK

YSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWARADVY

FNKTTGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEF

-continued

KFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEG

GEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKP

KLDF

[0708] The benefit of making chimeric Cas9 include:

[0709] reduce toxicity

[0710] improve expression in eukaryotic cells

[0711] enhance specificity

[0712] reduce molecular weight of protein, make protein smaller by combining the smallest domains from different Cas9 homologs.

Altering the PAM Sequence Requirement

Example 13: Utilization of Cas9 as a Generic DNA Binding Protein

[0713] 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.

[0714] 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.

[0715] 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.

[0716] Applicants designed a second generation of constructs. (Table below).

---

pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-hSpCsn(D10A, H840A)-NLS ("6xHis" disclosed as SEQ ID NO: 370)

pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-hSpCsn1(D10A, H840A) ("6xHis" disclosed as SEQ ID NO: 370)

pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-NLS-hSpCsn1 (D10A, H840A) ("6xHis" disclosed as SEQ ID NO: 370)

pLenti-EF1a-GFP-2A-6xHis-NLS-hSpCsn1(D10A, H840A)-NLA ("6xHis" disclosed as SEQ ID NO: 370)

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

-continued

("6xHis" disclosed as SEQ ID NO: 370)
pLenti-EF1a-GFP-2A-6xHis-NLS-NLS-hSpCsn1(D10A, H840A) ("6xHis" disclosed as SEQ ID NO: 370)

[0717] 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

[0718] 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.

[0719] 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.

[0720] 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.

[0721] 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.

[0722] The two constructs were electroporated into RI 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.

[0723] 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

[0724] 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 cr-RNA (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.

[0725] 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).

[0726] Applicants have also optimized Cas9 guide RNA using in vitro methods.

Example 16: Cas9 Mutations

[0727] In this example, Applicants show that the following mutations can convert SpCas9 into a nicking enzyme: D10A, E762A, H840A, N854A, N863A, D986A.

[0728] 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.

[0729] 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.

[0730] The catalytic residue in 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

Cas9 Transcriptional Activation

[0731] 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-His antibody, nuclease activity using a Surveyor nuclease assay, and DNA binding affinity using a gel shift assay.

Cas Repressor

[0732] 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.

[0733] 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).

[0734] Further work investigates: repeating the above experiment, targeting different genes, utilizing other gRNAs to determine the optimal targeting position, and multiplexed repression.

TABLE 1

(Table 1 discloses "GGGGS3" as SEQ ID NO: 106, "EAAAK3" as SEQ ID NO: 107 and "GGGGGS3" as SEQ ID NO: 108)

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

TABLE 1-continued

(Table 1 discloses "GGGGS3" as SEQ ID NO: 106, "EAAAK3" as SEQ ID NO: 107 and "GGGGGS3" as SEQ ID NO: 108)

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-EAANK$_3$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-KRAB-NTS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE

pXRP052-pLenti2-EFTa-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-Linker-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-ptenti2-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-LTR-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

TABLE 1-continued

(Table 1 discloses "GGGGS3" as SEQ ID NO: 106, "EAAAK3" as SEQ ID NO: 107 and "GGGGGS3" as SEQ ID NO: 108)

pXRP035-ptenti2-LTR-VP64-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP036-ptenti2-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-EAANK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP048-pLenti2-EF1a-SID4x-NLS-FLAG-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-GFP-WPRE
pXRP053-pLenti2-EF1a-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-ptenti2-EF1a-Cas9-GGGGGS$_3$-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP059-pLenti2-EF1a-Cas9-GGGGGS$_3$-NLS-KRAB-OAK-2A-GFP-WPRE
pXRP060-pLenti2-EF1a-Cas9-EAAAK$_3$-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP061-pLenti2-EF1a-Cas9-EAAAK$_3$-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP062-pLenti2-EF1a-Cas9-EAAAK$_3$-NLS-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

[0735] 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.

[0736] Study Design: 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.

[0737] 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.

Identification and Design of Gene Targets

[0738] 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.

[0739] 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.

Cloning of Guide Sequences into a Delivery System

[0740] 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:

Virus-Based Delivery Methods

AAV-Based Vectors (PX260, 330, 334, 335) have been Described Elsewhere

[0741] 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.

Virus Production is Described Elsewhere.

Nanoparticle-Based RNA Delivery Methods

[0742] 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.

[0743] 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.

Hydrodynamic Tail Vein Delivery Methods (for Mouse)

[0744] Guide sequences are cloned into AAV plasmids as described above and elsewhere in this application.

In Vitro Validation on Cell Lines

Transfection

[0745] 1. DNA plasmid transfection

[0746] 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, lug of total DNA is transfected into a single well using Fugene HD.

2. RNA Transfection

[0747] 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**.

Assay of Indel Formation In Vitro

[0748] Cells are harvested 72-hours post-transfection and assayed for indel formation as an indication of double-stranded breaks.

[0749] 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.

In Vivo Proof of Principle in Animal

Delivery Mechanisms

[0750] AAV or Lentivirus production is described elsewhere.

Nanoparticle Formulation: RNA Mixed into Nanoparticle Formulation

[0751] Hydrodynamic tail vein injections with DNA plasmids in mice are conducted using a commercial kit

[0752] 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.

[0753] 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,

Assay for Indel Formation

[0754] DNA is extracted from tissue using commercial kits; indel assay will be performed as described for in vitro demonstration.

[0755] 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.

Examples of a Single Guide-RNA to Introduce Targeted Indels at a Gene Locus

Disease	GENE	SPACER	PAM	SEQ ID NO:	Mechanism	References
Hypercholesterol-emia	HMG-CR	GCCAAATTG GACGACCCT CG	CGG	109	Knockout	Fluvastatin: a review of its pharmacology and use in the management of hypercholest-erolaemia. (Plosker GL et al. Drugs 1996, 51(3): 433-459)
Hypercholesterol-emia	SQLE	CGAGGAGAC CCCCGTTTC GG	TGG	110	Knockout	Potential role of nonstatin cholesterol lowering agents (Trapani et al. IUBMB Life, Volume 63, Issue 11, pages 964-971, November 2011)

-continued

Disease	GENE	SPACER	PAM	SEQ ID NO:	Mechanism	References
Hyperlipidemia	DGAT 1	CCCGCCGCC GCCGTGGCT CG	AGG	111	Knockout	DGAT1 inhibitors as anti-obesity and anti-diabetic agents. (Birch AM et al. Current Opinion in Drug Discovery & Development [2010, 13(4): 489-496)
Leukemia	BCR-ABL	TGAGCTCTA CGAGATCCA CA	AGG	112	Knockout	Killing of leukemic cells with a BCR/ABL fusion gene by RNA interference (RNAi). (Fuchs et al. Oncogene 2002, 21(37): 5716-5724)

Examples of a Pair of Guide-RNA to Introduce Chromosomal Microdeletion at a Gene Locus

Disease	GENE	SPACER	PAM	SEQ ID NO:	Mechanism	References
Hyperlipidemia	PLIN2 guide1	CTCAAAATT CATACCGGT TG	TGG	113	Microdeletion	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	114	Microdeletion	
Hyperlipidemia	SREBP guide1	TTCACCCCG CGGCGCTGA AT	ggg	115	Microdeletion	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	116	Microdeletion	

Example 19: Targeted Integration of Repair for Genes Carrying Disease-Causing Mutations; Reconstitution of Enzyme Deficiencies and Other Related Diseases

Study Design

[0756] I Identification and design of gene targets

[0757] Described in Example 22

[0758] II. Cloning of guide sequences and repair templates into a delivery system

[0759] Described above in Example 22

[0760] Applicants clone DNA repair templates to include homology arms with diseased allele as well a wild-type repair template

[0761] III. In vitro validation on cell lines

[0762] a. Transfection is described above in Example 22; Cas9, guide RNAs, and repair template are co-transfected into relevant cell types.

[0763] b. Assay for repair in vitro

[0764] i. Applicants harvest cells 72-hours post-transfection and assay for repair

[0765] ii. Briefly, Applicants amplify genomic region around repair template PCR using high-fidelity polymerase. Applicants sequence products for decreased incidence of mutant allele.

[0766] IV. In vivo proof of principle in animal

[0767] a. Delivery mechanisms are described above Examples 22 and 34.

[0768] b. Assay for repair in vivo

[0769] i. Applicants perform the repair assay as described in the in vitro demonstration.

[0770] V. Therapeutic applications

[0771] 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.

[0772] Example of one single missense mutation with repair template:

Disease	GENE	SPACER	PAM
Familial amyloid polyneuropathy	TTR	AGCCTTTCTGAACACATGCA (SEQ ID NO: 117)	CGG

Mechanism	References
V30M repair	Transthyretin mutations in health and disease (Joao et al. Human Mutation, Volume 5, Issue 3, pages 191-196, 1995)
V30M allele	CCTGCCATCAATGTGGCCATGCATGTGTTCA GAAAGGCT (SEQ ID NO: 118)
WT allele	CCTGCCATCAATGTGGCCGTGCATGTGTTCA GAAAGGCT (SEQ ID NO: 119)

Example 20: Therapeutic Application of the CRISPR-Cas System in Glaucoma, Amyloidosis, and Huntington's Disease

[0773] 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.

[0774] 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.

[0775] 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.

[0776] 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.

[0777] 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).

[0778] 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 I.

Example 21: Therapeutic Application of the CRISPR-Cas System in HIV

[0779] 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.

[0780] 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.

[0781] 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.

[0782] 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.

[0783] 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.

[0784] 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.

[0785] 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.

[0786] 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	(SEQ ID NO: 120)
TCACCAGCAATATTCGCTCGAGG	(SEQ ID NO: 121)
CACCAGCAATATTCCGCTCGAGG	(SEQ ID NO: 122)
TAGCAACAGACATACGCTCGAGG	(SEQ ID NO: 123)
GGGCAGTAGTAATACGCTCGAGG	(SEQ ID NO: 124)
CCAATTCCCATACATTATTGTAC	(SEQ ID NO: 125)

Example 22: Targeted Correction of deltaF508 or Other Mutations in Cystic Fibrosis

[0787] 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.

[0788] 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.

[0789] 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.

[0790] 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.

[0791] Applicants use an adenovirus or an AAV particle.

[0792] 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.

Identification of Cas9 Target Site

[0793] 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).

Gene Repair Strategy

[0794] 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

[0795] This example demonstrates how to generate a library of cells where each cell has a single gene knocked out:

[0796] 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.

[0797] 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 B×B1 attB site into the AAVS1 locus of the human ES cell. Then Applicants use the B×B1 integrase to facilitate the integration of individual guide RNA genes into the B×B1 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 B×B1 will recombine the attB site in the genome with the attP site on the guide RNA containing plasmid.

[0798] 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 who 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

Methods of Delivering Cas9

[0799] 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.

[0800] 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.

[0801] 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.

[0802] For Homologous recombination, Applicants provide an additional homology directed repair template.

[0803] Sequence for a cassette driving the expression of Cas9 under the control of beta-2 tubulin promoter, followed by 3' UTR of Cop1.

(SEQ ID NO: 126)

TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCG

ATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAA

TAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCT

ACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAA

CATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCGACAAGAAGTACAG

CATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATT

CAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAAC

AGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGA

GATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGA

TAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTA

CCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGAT

CAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCT

GGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGC

-continued

CAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAA

CCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCT

GAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGC

CGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAG

CGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCC

TGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGA

AGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGA

GGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCAT

TCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCAT

CCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCAC

CCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAA

GAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAA

AGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCT

GTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGT

GGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAA

GGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGA

GATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATA

CACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTT

CCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACAT

CCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAA

GAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGT

GATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGA

GGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTA

CCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGT

GGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCG

GGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAA

GCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTT

CATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAA

GTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAA

GGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGG

AACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAA

GATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTT

CAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGAT

CGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGAC

CGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAA

GGACTGGGACCCTAAGAAGTACGGCGGCCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGG

AAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGA

-continued

AGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACT

CCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCC

TGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGC

AGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAG

TGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGG

CCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCG

ACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGA

CACGGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGAGGCCAGCTAAGGATCCGGCA

AGACTGGCCCCGCTTGGCAACGCAACAGTGAGCCCCTCCCTAGTGTGTTTGGGGATGTGACTATGTATTCGTGTGTTG

GCCAACGGGTCAACCCGAACAGATTGATACCCGCCTTGGCATTTCCTGTCAGAATGTAACGTCAGTTGATGGTACT

[0804] Sequence for a cassette driving the expression of T7 polymerase under the control of beta-2 tubulin promoter, followed by 3' UTR of Cop1:

(SEQ ID NO: 127)

TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGA

CGGCTTCCCGGCGTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGC

TCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGT

TTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAG

CCATATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGC

TTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCA

CAACCCGCAAACatgcctaagaagaagaggaaggttaacacgattaacat

cgctaagaacgacttctctgacatcgaactggctgctatcccgttcaaca

ctctggctgaccattacggtgagcgtttagctcgcgaacagttggccctt

gagcatgagtcttacgagatgggtgaagcacgcttccgcaagatgtttga

gcgtcaacttaaagctggtgaggttgcggataacgctgccgccaagcctc

tcatcactaccctactccctaagatgattgcacgcatcaacgactggttt

gaggaagtgaaagctaagcgcggcaagcgcccgacagccttccagttcct

gcaagaaatcaagccggaagccgtagcgtacatcaccattaagaccactc

tggcttgcctaaccagtgctgacaatacaaccgttcaggctgtagcaagc

gcaatcggtcgggccattgaggacgaggctcgcttcggtcgtatccgtga

ccttgaagctaagcacttcaagaaaaacgttgaggaacaactcaacaagc

gcgtagggcacgtctacaagaaagcatttatgcaagttgtcgaggctgac

atgctctctaagggtctactcggtggcgaggcgtggtcttcgtggcataa

ggaagactctattcatgtaggagtacgctgcatcgagatgctcattgagt

caaccggaatggttagcttacaccgccaaaatgctggcgtagtaggtcaa

gactctgagactatcgaactcgcacctgaatacgctgaggctatcgcaac

ccgtgcaggtgcgctggctggcatctctccgatgttccaaccttgcgtag

ttcctcctaagccgtggactggcattactggtggtggctattgggctaac

-continued

ggtcgtcgtcctctggcgctggtgcgtactcacagtaagaaagcactgat

gcgctacgaagacgtttacatgcctgaggtgtacaaagcgattaacattg

cgcaaaacaccgcatggaaaatcaacaagaaagtcctagcggtcgccaac

gtaatcaccaagtggaagcattgtccggtcgaggacatccctgcgattga

gcgtgaagaactcccgatgaaaccggaagacatcgacatgaatcctgagg

ctctcaccgcgtggaaacgtgctgccgctgctgtgtaccgcaaggacaag

gctcgcaagtctcgccgtatcagccttgagttcatgcttgagcaagccaa

taagtttgctaaccataaggccatctggttcccttacaacatggactggc

gcggtcgtgtttacgctgtgtcaatgttcaacccgcaaggtaacgatatg

accaaaggactgcttacgctggcgaaaggtaaaccaatcggtaaggaagg

ttactactggctgaaaatccacggtgcaaactgtgcgggtgtcgacaagg

ttccgttccctgagcgcatcaagttcattgaggaaaaccacgagaacatc

atggcttgcgctaagtctccactggagaacacttggtgggctgagcaaga

ttctccgttctgcttccttgcgttctgctttgagtacgctggggtacagc

accacggcctgagctataactgctcccttccgctggcgtttgacgggtct

tgctctggcatccagcacttctccgcgatgctccgagatgaggtaggtgg

tcgcgcggttaacttgcttcctagtgaaaccgttcaggacatctacggga

ttgttgctaagaaagtcaacgagattctacaagcagacgcaatcaatggg

accgataacgaagtagttaccgtgaccgatgagaacactggtgaaatctc

tgagaaagtcaagctgggcactaaggcactggctggtcaatggctggctt

acggtgttactcgcagtgtgactaagcgttcagtcatgacgctggcttac

gggtccaaagagttcggcttccgtcaacaagtgctggaagataccattca

gccagctattgattccggcaagggtctgatgttcactcagccgaatcagg

ctgctggatacatggctaagctgatttgggaatctgtgagcgtgacggtg

gtagctgcggttgaagcaatgaactggcttaagtctgctgctaagctgct

ggctgctgaggtcaaagataagaagactggagagattcttcgcaagcgtt

-continued

gcgctgtgcattgggtaactcctgatggtttccctgtgtggcaggaatac

aagaagcctattcagacgcgcttgaacctgatgttcctcggtcagttccg

cttacagcctaccattaacaccaacaaagatagcgagattgatgcacaca

aacaggagtctggtatcgctcctaactttgtacacagccaagacggtagc

caccttcgtaagactgtagtgtgggcacacgagaagtacggaatcgaatc

ttttgcactgattcacgactccttcggtacgattccggctgacgctgcga

acctgttcaaagcagtgcgcgaaactatggttgacacatatgagtcttgt

gatgtactggctgatttctacgaccagttcgctgaccagttgcacgagtc

tcaattggacaaaatgccagcacttccggctaaaggtaacttgaacctcc

gtgacattcttagagtcggacttcgcgttcgcgtaaGGATCCGGCAAGAC

TGGCCCCGCTTGGCAACGCAACAGTGAGCCCCTCCCTAGTGTGTTTGGGG

ATGTGACTATGTATTCGTGTGTTGGCCAACGGGTCAACCCGAACAGATTG

ATACCCGCCTTGGCATTTCCTGTCAGAATGTAACGTCAGTTGATGGTACT

[0805] Sequence of guide RNA driven by the T7 promoter (T7 promoter, Ns represent targeting sequence):

(SEQ ID NO: 128)

gaaat**TAATACGACTCACTATA**NNNNNNNNNNNNNNNNNNNNgttttag

agctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaa

aagtggcaccgagtcggtgctttttt

Gene Delivery:

[0806] *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.

[0807] 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.

pChlamy1-Cas9:

(SEQ ID NO: 129)

TGCGGTATTTCACACCGCATCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAAT

ACATTCAAATATGTATCCGCTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTT

TAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGC

GATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATC

TGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGG

AAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGT

AAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGG

TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAG

CTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAA

TTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTG

TATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCT

CATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCAC

TCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGC

CGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTA

TCAGGGTTATTGTCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGAT

CAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGT

GGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATAC

TGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAAT

CCTGTTACCAGTGGCTGTTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA

GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATA

CCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGT

CGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCT

-continued

CTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTT

ACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTAT

TACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGT

CGCTGAGGCTTGACATGATTGGTGCGTATGTTTGTATGAAGCTACAGGACTGATTTGGCGGGCTATGAGGGCGGGGGA

AGCTCTGGAAGGGCCGCGATGGGGCGCGCGGCGTCCAGAAGGCGCCATACGGCCCGCTGGCGGCACCCATCCGGTATA

AAAGCCCGCGACCCCGAACGGTGACCTCCACTTTCAGCGACAAACGAGCACTTATACATACGCGACTATTCTGCCGCT

ATACATAACCACTCAGCTAGCTTAAGATCCCATCAAGCTTGCATGCCGGGCGCGCCAGAAGGAGCGCAGCCAAACCAG

GATGATGTTTGATGGGGTATTTGAGCACTTGCAACCCTTATCCGGAAGCCCCCTGGCCCACAAAGGCTAGGCGCCAAT

GCAAGCAGTTCGCATGCAGCCCCTGGAGCGGTGCCCTCCTGATAAACCGGCCAGGGGGCCTATGTTCTTTACTTTTTT

ACAAGAGAAGTCACTCAACATCTTAAAATGGCCAGGTGAGTCGACGAGCAAGCCCGGCGGATCAGGCAGCGTGCTTGC

AGATTTGACTTGCAACGCCCGCATTGTGTCGACGAAGGCTTTTGGCTCCTCTGTCGCTGTCTCAAGCAGCATCTAACC

CTGCGTCGCCGTTTCCATTTGCAGGAGATTCGAGGTACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAG

AAAAAGCGCAAGGTCGAAGCGTCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCC

GTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAG

AACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGA

TACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTC

TTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTG

GACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCC

GACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAAC

CCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATC

AACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCC

CAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAG

AGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTG

GCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATC

CTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGAC

CTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAAC

GGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATG

GACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGC

ATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGAC

AACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGA

TTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCC

GCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTG

CTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTC

CTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAA

GAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTG

GGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAA

GATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTC

-continued

GACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGC

ATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAG

CTCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGC

ACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGC

TCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGG

GACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAAC

ACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGG

ACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACT

CCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGA

AGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGG

CCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAA

AGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAG

TGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACT

ACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCG

AGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTA

CCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGA

AGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGA

AAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAGAGTCTATCCT

GCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCC

CACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCT

GCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGA

AGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGC

CTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCA

CTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGA

CGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGC

CTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGG

AGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGC

CACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCAA

GAAGAAGAGAAAGGTGGAGGCCAGCTAACATATGATTCGAATGTCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTA

GGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCT

GCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGA

GCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCT

AAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACATGACACAAGAATCCCTGTTACTTCTCGACCGTA

TTGATTCGGATGATTCCTACGCGAGCCTGCGGAACGACCAGGAATTCTGGGAGGTGAGTCGACGAGCAAGCCCGGCGG

ATCAGGCAGCGTGCTTGCAGATTTGACTTGCAACGCCCGCATTGTGTCGACGAAGGCTTTTGGCTCCTCTGTCGCTGT

CTCAAGCAGCATCTAACCCTGCGTCGCCGTTTCCATTTGCAGCCGCTGGCCCGCCGAGCCCTGGAGGAGCTCGGGCTG

CCGGTGCCGCCGGTGCTGCGGGTGCCCGGCGAGAGCACCAACCCCGTACTGGTCGGCGAGCCCGGCCCGGTGATCAAG

-continued

CTGTTCGGCGAGCACTGGTGCGGTCCGGAGAGCCTCGCGTCGGAGTCGGAGGCGTACGCGGTCCTGGCGGACGCCCCG

GTGCCGGTGCCCCGCCTCCTCGGCCGCGGCGAGCTGCGGCCCGGCACCGGAGCCTGGCCGTGGCCCTACCTGGTGATG

AGCCGGATGACCGGCACCACCTGGCGGTCCGCGATGGACGGCACGACCGACCGGAACGCGCTGCTCGCCCTGGCCCGC

GAACTCGGCCGGGTGCTCGGCCGGCTGCACAGGGTGCCGCTGACCGGGAACACCGTGCTCACCCCCCATTCCGAGGTC

TTCCCGGAACTGCTGCGGGAACGCCGCGCGGCGACCGTCGAGGACCACCGCGGGTGGGGCTACCTCTCGCCCCGGCTG

CTGGACCGCCTGGAGGACTGGCTGCCGGACGTGGACACGCTGCTGGCCGGCCGCGAACCCCGGTTCGTCCACGGCGAC

CTGCACGGGACCAACATCTTCGTGGACCTGGCCGCGACCGAGGTCACCGGGATCGTCGACTTCACCGACGTCTATGCG

GGAGATCTCCCGCTACAGCCTGGTGCAACTGCATCTCAACGCCTTCCGGGGCGACCGCGAGATCCTGGCCGCGCTGCT

CGACGGGGCGCAGTGGAAGCGGACCGAGGACTTCGCCCGCGAACTGCTCGCCTTCACCTTCCTGCACGACTTCGAGGT

GTTCGAGGAGACCCCGCTGGATCTCTCCGGCTTCACCGATCCGGAGGAACTGGCGCAGTTCCTCTGGGGGCCGCCGGA

CACCGCCCCCGGCGCCTGATAAGGATCCGGCAAGACTGGCCCCGCTTGGCAACGCAACAGTGAGCCCCTCCCTAGTGT

GTTTGGGGATGTGACTATGTATTCGTGTGTTGGCCAACGGGTCAACCCGAACAGATTGATACCCGCCTTGGCATTTCC

TGTCAGAATGTAACGTCAGTTGATGGTACT

[0808] 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

Diseases that Involve Mutations in Protein Coding Sequence:

[0809] 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).

[0810] 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.

[0811] 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).

Diseases that Involve Non-Coding Sequences

[0812] 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

[0813] Applicants also use Cas9 to disrupt functional motifs in 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

[0814] 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.

[0815] 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 (fragments below disclosed as SEQ ID NOS 130-137, respectively, in order of appearance).

```
HTT locus
with frag-
ment ex-     . . . CCGTGCCGGGCGGGAGACCGCCATGG           GGCCCGGCTGTGGCTGAGGAGC. . .
cised by     . . . GGCACGGCCCGCCCTCTGCC             TGGGCCGGGCCGACACCGACTCCTCG. . .
Cas9 nick-
ase and two
pairs of
guide RNAs
                                                +
exogenous
DNA
fragment
with
fewer
number
of CAG
repeats
also           CGACCCTGGAAA. . . . . reduced number of CAG repeats. . . . . CCCCGCCGCCACCC
cleaved  GGTACCGCTGGGACCTTT. . . . .                                . . . . . GGGGCGGCGG
by Cas9
nicakse
and
the two
pairs of
guide RNAs
```

[0816] 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

[0817] 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.

[0818] 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.

[0819] 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

[0820] 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).

[0821] 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.

[0822] 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 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 heterochromatin which turn off the frataxin gene (FIG. **32**A).

[0823] Competitive Advantage over other therapeutic strategies are listed below:

[0824] 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.

[0825] 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 ZEN or TALEN is difficult to achieve).

[0826] 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.

[0827] 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).

[0828] Applicants deliver a combination of guide RNAs targeting the intron 1 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.

[0829] 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.

Alternative Strategies:

[0830] 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

[0831] 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.

[0832] 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.

[0833] The following list of targets in tabular form are for genes involved in the following diseases:

[0834] Lafora's Disease—target GSY1 or PPP1R3C (PTG) to reduce glycogen in neurons.

Hypercholesterolemia—Target PCSK9

[0835] 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	(SEQ ID NO: 138)
	GGCC-R	GAACGCAGTGCTCTTCGAAG	(SEQ ID NO: 139)
	GGNCC-L	CTCACGCCCTGCTCCGTGTA	(SEQ ID NO: 140)
	GGNCC-R	GGCGACAACTACTTCCTGGT	(SEQ ID NO: 141)
	GGNNCC-L	CTCACGCCCTGCTCCGTGTA	(SEQ ID NO: 142)
	GGNNCC-R	GGGCGACAACTACTTCCTGG	(SEQ ID NO: 143)
	GGNNNCC-L	CCTCTTCAGGGCCGGGGTGG	(SEQ ID NO: 144)
	GGNNNCC-R	GAGGACCCAGGTGGAACTGC	(SEQ ID NO: 145)
PCSK9 (human)	GGCC-L	TCAGCTCCAGGCGGTCCTGG	(SEQ ID NO: 146)
	GGCC-R	AGCAGCAGCAGCAGTGGCAG	(SEQ ID NO: 147)
	GGNCC-L	TGGGCACCGTCAGCTCCAGG	(SEQ ID NO: 148)
	GGNCC-R	CAGCAGTGGCAGCGGCCACC	(SEQ ID NO: 149)
	GGNNCC-L	ACCTCTCCCCTGGCCCTCAT	(SEQ ID NO: 150)
	GGNNCC-R	CCAGGACCGCCTGGAGCTGA	(SEQ ID NO: 151)
	GGNNNCC-L	CCGTCAGCTCCAGGCGGTCC	(SEQ ID NO: 152)
	GGNNNCC-R	AGCAGCAGCAGCAGTGGCAG	(SEQ ID NO: 153)
PPP1R3C (PTG) (human)	GGCC-L	ATGTGCCAAGCAAAGCCTCA	(SEQ ID NO: 154)
	GGCC-R	TTCGGTCATGCCCGTGGATG	(SEQ ID NO: 155)
	GGNCC-L	GTCGTTGAAATTCATCGTAC	(SEQ ID NO: 156)
	GGNCC-R	ACCACCTGTGAAGAGTTTCC	(SEQ ID NO: 157)
	GGNNCC-L	CGTCGTTGAAATTCATCGTA	(SEQ ID NO: 158)
	GGNNCC-R	ACCACCTGTGAAGAGTTTCC	(SEQ ID NO: 159)

-continued

Gys1 (mouse)	GGCC-L	GAACGCAGTGCTTTTCGAGG	(SEQ ID NO: 160)
	GGCC-R	ACCCTTGTTGGCCACCTCCC	(SEQ ID NO: 161)
	GGNCC-L	GGTGACAACTACTATCTGGT	(SEQ ID NO: 162)
	GGNCC-R	CTCACACCCTGCTCCGTGTA	(SEQ ID NO: 163)
	GGNNCC-L	GGGTGACAACTACTATCTGG	(SEQ ID NO: 164)
	GGNNCC-R	CTCACACCCTGCTCCGTGTA	(SEQ ID NO: 165)
	GGNNNCC-L	CGAGAACGCAGTGCTTTTCG	(SEQ ID NO: 166)
	GGNNNCC-R	ACCCTTGTTGGCCACCTCCC	(SEQ ID NO: 167)
PPP1R3C (PTG) (mouse)	GGCC-L	ATGAGCCAAGCAAATCCTCA	(SEQ ID NO: 168)
	GGCC-R	TTCCGTCATGCCCGTGGACA	(SEQ ID NO: 169)
	GGNCC-L	CTTCGTTGAAAACCATTGTA	(SEQ ID NO: 170)
	GGNCC-R	CCACCTCTGAAGAGTTTCCT	(SEQ ID NO: 171)
	GGNNCC-L	CTTCGTTGAAAACCATTGTA	(SEQ ID NO: 172)
	GGNNCC-R	ACCACCTCTGAAGAGTTTCC	(SEQ ID NO: 173)
	GGNNNCC-L	CTTCCACTCACTCTGCGATT	(SEQ ID NO: 174)
	GGNNNCC-R	ACCATGTCTCAGTGTCAAGC	(SEQ ID NO: 175)
PCSK9 (mouse)	GGCC-L	GGCGGCAACAGCGGCAACAG	(SEQ ID NO: 176)
	GGCC-R	ACTGCTCTGCGTGGCTGCGG	(SEQ ID NO: 177)
	GGNNCC-L	CCGCAGCCACGCAGAGCAGT	(SEQ ID NO: 178)
	GGNNCC-R	GCACCTCTCCTCGCCCCGAT	(SEQ ID NO: 179)

[0836] Alternative strategies for improving stability of guide RNA and increasing specificity

[0837] 1. Nucleotides in 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.

[0838] 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

[0839] 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.

[0840] 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.

[0841] 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.

[0842] 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.

[0843] 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.

[0844] 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.

[0845] 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. Retargeting 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 TALEN domains.

[0846] 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.

[0847] 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.

[0848] 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.

[0849] 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.

[0850] 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.

[0851] 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 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.

[0852] The detailed information on computationally predicted off-target sites is as follows:

[0853] 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.

[0854] 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 3' end of the guide sequence are less tolerant of mismatches than 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.

[0855] 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.

[0856] 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 5' of the sgRNA where the 20-nt guide sequence does not begin with G.

[0857] 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.

[0858] 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.

[0859] In addition to PCR and plasmid-based delivery methods, both Cas9 and sgRNA can be introduced into cells as RNA.

[0860] 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.

[0861] 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.

Functional Testing

[0862] 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.

[0863] 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.

[0864] 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.

[0865] 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.

Materials and Reagents

sgRNA Preparation:

[0866] UltraPure DNaseRNase-free distilled water (Life Technologies, cat. no. 10977-023)

[0867] Herculase II fusion polymerase (Agilent Technologies, cat. no. 600679)

[0868] 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.

[0869] Herculase II reaction buffer (5×; Agilent Technologies, included with polymerase)

[0870] dNTP solution mix (25 mM each; Enzymatics, cat. no. N205L)

[0871] MgCl2 (25 mM; ThermoScientific, cat. no. R0971)

[0872] QIAquick gel extraction kit (Qiagen, cat. no. 28704)

[0873] QIAprep spin miniprep kit (Qiagen, cat. no. 27106)

[0874] UltraPure TBE buffer (10×; Life Technologies, cat. no. 15581-028)

[0875] SeaKem LE agarose (Lonza, cat. no. 50004)

[0876] SYBR Safe DNA stain (10,000×; Life Technologies, cat. no. S33102)

[0877] 1-kb Plus DNA ladder (Life Technologies, cat. no. 10787-018)

[0878] TrackIt CyanOrange loading buffer (Life Technologies, cat. no. 10482-028)
[0879] FastDigest BbsI (BpiI) (Fermentas/ThermoScientific, cat. no. FD1014)
[0880] Fermentas Tango Buffer (Fermentas/ThermoScientific, cat. no. BY5)
[0881] DL-dithiothreitol (DTT; Fermentas/ThermoScientific, cat. no. R0862)
[0882] T7 DNA ligase (Enzymatics, cat. no. L602L)
[0883] 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.
[0884] T7 2X Rapid Ligation Buffer (included with T7 DNA ligase, Enzymatics, cat. no. L602L)
[0885] T4 Polynucleotide Kinase (New England Biolabs, cat. no M0201S)
[0886] T4 DNA Ligase Reaction Buffer (10X; New England Biolabs, cat. no B0202S)
[0887] Adenosine 5'-triphosphate (10 mM; New England Biolabs, cat. no. P0756S)
[0888] PlasmidSafe ATP-dependent DNase (Epicentre, cat. no. E3101K)
[0889] One Shot Stb13 chemically competent *Escherichia coli* (*E. coli*) (Life Technologies, cat. no. C7373-03)
[0890] SOC medium (New England Biolabs, cat. no. B9020S)
[0891] LB medium (Sigma, cat. no. L3022)
[0892] LB agar medium (Sigma, cat. no. L2897)
[0893] Ampicillin, sterile filtered (100 mg ml-1; Sigma, cat. no. A5354)

Mammalian Cell Culture:

[0894] HEK293FT cells (Life Technologies, cat. no. R700-07)
[0895] Dulbecco's minimum Eagle's medium (DMEM, 1×, high glucose; Life Technologies, cat. no. 10313-039)
[0896] Dulbecco's minimum Eagle's medium (DMEM, 1×, high glucose, no phenol red; Life Technologies, cat. no. 31053-028)
[0897] Dulbecco's phosphate-buffered saline (DPBS, 1×; Life Technologies, cat. no. 14190-250)
[0898] Fetal bovine serum, qualified and heat inactivated (Life Technologies, cat. no. 10438-034)
[0899] Opti-MEM I reduced-serum medium (FBS; Life Technologies, cat. no. 11058-021)
[0900] Penicillin-streptomycin (100x; Life Technologies, cat. no. 15140-163)
[0901] TrypLE™ Express (1X, no Phenol Red; Life Technologies, cat. no. 12604-013)
[0902] Lipofectamine 2000 transfection reagent (Life Technologies, cat. no. 11668027)
[0903] Amaxa SF Cell Line 4D-Nucleofector® X Kit S (32 RCT; Lonza, cat. no V4XC-2032)
[0904] HUES 9 cell line (HARVARD STEM CELL SCIENCE)
[0905] Geltrex LDEV-Free Reduced Growth Factor Basement Membrane Matrix (Life Technologies, cat. no. A1413201)
[0906] mTeSR1 medium (Stemcell Technologies, cat. no. 05850)
[0907] Accutase cell detachment solution (Stemcell Technologies, cat. no. 07920)
[0908] ROCK Inhibitor (Y-27632; Millipore, cat. no. SCM075)
[0909] Amaxa P3 Primary Cell 4D-Nucleofector® X Kit S (32 RCT; Lonza cat. no. V4XP-3032)

Genotyping Analysis:

[0910] QuickExtract DNA extraction solution (Epicentre, cat. no. QE09050)
[0911] PCR primers for SURVEYOR, RFLP analysis, or sequencing (see Primer table)
[0912] Herculase II fusion polymerase (Agilent Technologies, cat. no. 600679)
[0913] 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.
[0914] Herculase II reaction buffer (5x; Agilent Technologies, included with polymerase)
[0915] dNTP solution mix (25 mM each; Enzymatics, cat. no. N205L)
[0916] QIAquick gel extraction kit (Qiagen, cat. no. 28704)
[0917] Taq Buffer (10x; Genscript, cat. no. B0005)
[0918] SURVEYOR mutation detection kit for standard gel electrophoresis (Transgenomic, cat. no. 706025)
[0919] UltraPure TBE buffer (10x; Life Technologies, cat. no. 15581-028)
[0920] SeaKem LE agarose (Lonza, cat. no. 50004)
[0921] 4-20% TBE Gels 1.0 mm, 15 Well (Life Technologies, cat. no. EC62255BOX)
[0922] Novex® Hi-Density TBE Sample Buffer (5X; Life Technologies, cat. no. LC6678)
[0923] SYBR Gold Nucleic Acid Gel Stain (10,000X; Life Technologies, cat. no. S-11494)
[0924] 1-kb Plus DNA ladder (Life Technologies, cat. no. 10787-018)
[0925] TrackIt CyanOrange loading buffer (Life Technologies, cat. no. 10482-028)
[0926] FastDigest HindIII (Fermentas/ThermoScientific, cat. no. FD0504)

Equipment

[0927] Filtered sterile pipette tips (Corning)
[0928] Standard 1.5 ml microcentrifuge tubes (Eppendorf, cat. no. 0030 125.150)
[0929] Axygen 96-well PCR plates (VWR, cat. no. PCR-96M2-HSC)
[0930] Axygen 8-Strip PCR tubes (Fischer Scientific, cat. no. 14-222-250)
[0931] Falcon tubes, polypropylene, 15 ml (BD Falcon, cat. no. 352097)
[0932] Falcon tubes, polypropylene, 50 ml (BD Falcon, cat. no. 352070)
[0933] Round-bottom Tube with cell strainer cap, 5 ml (BD Falcon, cat. no. 352235)
[0934] Petri dishes (60 mm×15 mm; BD Biosciences, cat. no. 351007)
[0935] Tissue culture plate (24 well; BD Falcon, cat. no. 353047)
[0936] Tissue culture plate (96 well, flat bottom; BD Falcon, cat no. 353075)

[0937] Tissue culture dish (100 mm; BD Falcon, 353003)

[0938] 96-well thermocycler with programmable temperature stepping functionality (Applied Biosystems Veriti, cat. no. 4375786).

[0939] Desktop microcentrifuges 5424, 5804 (Eppendorf)

[0940] 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)

[0941] Novex XCell SureLock Mini-Cell (Life Technologies, cat. no. EI0001)

[0942] Digital gel imaging system (GelDoc E Z, Bio-Rad, cat. no. 170-8270, and blue sample tray, Bio-Rad, cat. no. 170-8273)

[0943] Blue light transilluminator and orange filter goggles (SafeImager 2.0; Invitrogen, cat. no. G6600)

[0944] 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)

Reagent Setup

[0945] Tris-borate EDTA (TBE) electrophoresis solution Dilute TBE buffer in distilled water to 1X 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.

[0946] 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.

[0947] 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.

[0948] D10 culture medium For culture of HEK293FT cells, prepare D10 culture medium by supplementing DMEM with 1X 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.

[0949] mTeSR1 culture medium For culture of human embryonic stem cells, prepare mTeSR1 medium by supplementing the 5X supplement (included with mTeSR1 basal medium), and 100 μg/ml Normocin.

Procedure

Design of Targeting Components and Use of the Online Tool·Timing 1 d

[0950] 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.

[0951] 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

[0952] 3| To generate the sgRNA expression construct, either the PCR- or plasmid-based protocol can be used.

(A) Via PCR Amplification·Timing 2 h

[0953] (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.

[0954] (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).

[0955] (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
Herculase II Fusion polymerase	0.5	
Distilled water	36	
Total	50	

[0956] (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

[0957] (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.

[0958] (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.

(B) Cloning sgRNA into Cas9-Containing Bicistronic Expression Vector·Timing 3 d

[0959] (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 μM. 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

[0960] (ii) Anneal in a thermocycler using the following parameters:

[0961] 37° C. for 30 m

[0962] 95° C. for 5 m

[0963] Ramp down to 25° C. at 5° C. per m

[0964] (iii) Applicants diluted phosphorylated and annealed oligos 1:200 by add 1 ul of oligo to 199 μl room temperature ddH_2O.

[0965] (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

[0966] (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

[0967] (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

[0968] (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.

[0969] (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 μl of the product from step (vii) into 20 μl 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 μl of SOC medium is added, and this is plated onto an LB plate containing 100 μg/ml ampicillin with incubation overnight at 37° C.

[0970] (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.

[0971] (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 μg/ml ampicillin. Incubate and shake at 37° C. overnight.

[0972] (xi) Day 3: Applicants isolated plasmid DNA from overnight cultures using a QiAprep Spin miniprep kit according to the manufacturer's instructions.

[0973] (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 (SEQ ID NO: 180)	Amplify U6-sgRNA
U6-Rev	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACNNNNNNNNNNNNNNNNNNNNCCGGTGTTTCGTCCTTTCCACAAG (SEQ ID NO: 181)	Amplify U6-sgRNA; N is reverse complement of target
sgRNA-top	CACCGNNNNNNNNNNNNNNNNNNNN (SEQ ID NO: 182)	Clone sgRNA into PX330
sgRNA-bottom	AAACNNNNNNNNNNNNNNNNNNNC (SEQ ID NO: 183)	Clone sgRNA into PX330
U6-EMX1-Rev	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCCTAGTCATTGGAGGTGACCGGTGTTTCGTCCTTTCCACAAG (SEQ ID NO: 184)	Amplify U6-EMX1 sgRNA
EMX1-top	CACCGTCACCTCCAATGACTAGGG (SEQ ID NO: 185)	Clone EMX1 sgRNA into PX330

-continued

Primer	Sequence (5' to 3')	Purpose
EMX1-bottom	AAACCCCTAGTCATTGGAGGTGAC (SEQ ID NO: 186)	Clone EMX1 sgRNA into PX330
ssODN-sense	CAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAG GCCAATGGGGAGGACATCGATGTCACCTCCAATGACAAGCTTGCTAGCGGTGGG CAACCACAAACCCACGAGGGCAGAGTGCTGCTTGCTGCTGGCCAGGCCCCTGCG TGGGCCCAAGCTGGACTCTGGCCACTCCCT (SEQ ID NO: 187)	EMX1 HDR (sense; insertion underlined)
ssODN-antisense	AGGGAGTGGCCAGAGTCCAGCTTGGGCCCACGCAGGGGCCTGGCCAGCAGCAAG CAGCACTCTGCCCTCGTGGGTTTGTGGTTGCCCACCGCTAGCAAGCTTGTCATT GGAGGTGACATCGATGTCCTCCCCATTGGCCTGCTTCGTGGCAATGCGCCACCG GTTGATGTGATGGGAGCCCTTCTTCTTCTG (SEQ ID NO: 188)	EMX1 HDR (antisense; insertion underlined)
EMX1-SURV-F	CCATCCCCTTCTGTGAATGT (SEQ ID NO: 189)	EMX1 SURVEYOR assay PCR, sequencing
EMX1-SURV-R	GGAGATTGGAGACACGGAGA (SEQ ID NO: 190)	EMX1 SURVEYOR assay PCR, sequencing
EMX1-HDR-F	GGCTCCCTGGGTTCAAAGTA (SEQ ID NO: 191)	EMX1 RFLP analysis PCR, sequencing
EMX1-HDR-R	AGAGGGGTCTGGATGTCGTAA (SEQ ID NO: 192)	EMX1 RFLP analysis PCR, sequencing
pUC19-F	CGCCAGGGTTTTCCCAGTCACGAC (SEQ ID NO: 193)	pUC19 multiple cloning site F primer, for Sanger sequencing

[0974] 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.

(Optional) Design of ssODN Template·Timing 3 d Planning Ahead

[0975] 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.

[0976] 4| Applicants resuspended and diluted ssODN ultramers to a final concentration of 10 μM. Do not combine or anneal the sense and antisense ssODNs. Store at −20° C.

[0977] 5| Note for HDR applications, Applicants recommend cloning sgRNA into the PX330 plasmid.

Functional Validation of sgRNAs: Cell Culture and Transfections. Timing 3-4 d

[0978] 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.

[0979] 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.

[0980] 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.

[0981] 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 μl. 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.

[0982] 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.

[0983] (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.

[0984] (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

[0985] 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.

[0986] 11| Applicants added Lipofectamine complex to cells gently as HEK293FT cells may detach easily from plate easily and result in lower transfection efficiency.

[0987] 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.

[0988] 13| Applicants supplemented the culture medium with an additional 500 μl 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.

[0989] 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.

(Optional) Co-Transfection of CRISPR Plasmids and ssODNs or Targeting Plasmids for HR·Timing 3-4 d

[0990] 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.

[0991] 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.

[0992] 17| Applicants purified linearized plasmid with the QIAQuick PCR Purification kit.

[0993] 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.

[0994] 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.

[0995] 20| Nucleofection. Applicants transfected HEK293FT cells according to the Amaxa SF Cell Line Nucleofector 4D Kit manufacturer's instructions, adapted in the steps below.

[0996] 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

[0997] 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

[0998] For transfection controls, see previous section. In addition, Applicants recommend transfecting ssODN or targeting plasmid alone as a negative control.

[0999] 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 μl aliquot from the suspension and diluted into 90 μl 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.

[1000] 23| Applicants spun down the new tube at 200×g for 5 m.

[1001] Applicants prepared the transfection solution by mixing the SF solution and S1 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.

[1002] 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.

[1003] 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.

[1004] Electroporate cells using the Nucleofector 4D program recommended by Amaxa, CM-130.

[1005] 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.

Human Embryonic Stem Cell (HUES 9) Culture and Transfection. Timing 3-4 d

[1006] 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 μM 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.

[1007] 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 μM Rock Inhibitor. Change to normal mTeSR1 medium 24 h after plating.

[1008] 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.

[1009] 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

[1010] Clonal isolation may be performed 24 h post-transfection by FACS or by serial dilution.

[1011] 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 1X penicillin/streptinomycin and filter through a 0.22 μm Steriflip filter.

[1012] 55| Prepare 96 well plates. Applicants added 100 μl of D10 media supplemented with 1× penicillin/streptinomycin per well and prepared the number of plates as needed for the desired number of clones.

[1013] 56| Prepare cells for FACS. Applicants dissociated cells by aspirating the medium completely and adding 100 μl TrypLE per well of a 24-well plate. Incubate for 5 m and add 400 μl warm D10 media.

[1014] 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.

[1015] 58| Spin down cells at 200×g for 5 minutes.

[1016] 59| Applicants aspirated the media, and resuspended the cells in 200 μl of FACS media.

[1017] 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.

[1018] 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.

[1019] NOTE. The remainder of the cells may be kept and used for genotyping at the population level to gauge overall modification efficiency.

[1020] 62| Applicants returned cells into the incubator and allowed them to expand for 2-3 weeks. 100 μl of warm D10 medium is added 5 d post sorting. Change 100 μl of medium every 3-5 d as necessary.

[1021] 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.

[1022] 64| When cells are more than 60% confluent, Applicants prepared a set of replica plates for passaging. 100 μl 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.

[1023] 65| Use the remainder 80% of cells for DNA isolation and genotyping.

Optional: Isolation of Clonal Cell Lines by Dilution. Timing. 2-3 h Hands-on; 2-3 Weeks Expansion

[1024] 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.

[1025] 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 μl. 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.

[1026] 68| Multichannel pipette was used to pipette 100 μl of diluted cells to each well of a 96 well plate.

[1027] NOTE. The remainder of the cells may be kept and used for genotyping at the population level to gauge overall modification efficiency.

[1028] 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.

[1029] 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.

SURVEYOR Assay for CRISPR Cleavage Efficiency. Timing. 5-6 h

[1030] 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.

[1031] 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.

[1032] 72| Aspirate the supernatant completely.

[1033] 73| Applicants used QuickExtract DNA extraction solution according to the manufacturer's instructions. Applicants typically used 50 μl of the solution for each well of a 24 well plate and 10 μl for a 96 well plate.

[1034] 74| Applicants normalized extracted DNA to a final concentration of 100-200 ng/ul with ddH2O. Pause point: Extracted DNA may be stored at −20° C. for several months.

[1035] 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)	

[1036] 76| Applicants added 100-200 ng of normalized genomic DNA template from step 74 for each reaction.

[1037] 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

[1038] 78| Applicants ran 2-5 μl 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, MgCl2 concentration, and/or the annealing temperature.

[1039] 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.

[1040] 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

[1041] 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

[1042] 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, 0.15M	2.5	15 mM
ddH_2O	0.5	
SURVEYOR nuclease S	1	1X
SURVEYOR enhancer S	1	1X
Total	5	

[1043] 83| Vortex well and spin down. Incubate the reaction at 42° C. for 1 h.

[1044] 84| Optional: 2 μl 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.

[1045] 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 μl 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.

[1046] 86| Applicants stained the gel with 1X SYBR Gold dye diluted in TBE. The gel was gently rocked for 15 m.

[1047] 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.

[1048] 88| Estimate the cleavage intensity. Applicants quantified the integrated intensity of each band using ImageJ or other gel quantification software.

[1049] 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. 90| Cleavage efficiency may be estimated using the following formula, based on the binomial probability distribution of duplex formation:

$$91|\ \text{indel}(\%) = 100 \times \left(1 - \sqrt{(1 - f_{cut})}\right)$$

Sanger Sequencing for Assessing CRISPR Cleavage Efficiency. Timing·3 d

[1050] 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.

[1051] 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

[1052] 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

[1053] 94| Applicants purified the digestion reactions using the QIAQuick PCR purification kit. Pause point: Purified PCR product may be stored at −20° C.

[1054] 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

[1055] 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, 5ul 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.

[1056] 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.

[1057] 98| Day 3: Applicants isolated plasmid DNA from overnight cultures using a QIAprep Spin miniprep kit according to the manufacturer's instructions.

[1058] 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.

Genotyping for Microdeletion. Timing. 2-3 d Hands on; 2-3 Weeks Expansion

[1059] 100| Cells were transfected as described above with a pair of sgRNAs targeting the region to be deleted.

[1060] 101| 24 h post-transfection, clonal lines are isolated by FACS or serial dilution as described above.

[1061] 102| Cells are expanded for 2-3 weeks.

[1062] 103| Applicants harvested DNA from clonal lines as described above using 10 μl QuickExtract solution and normalized genomic DNA with ddH2O to a final concentration of 50-100 ng/ul.

[1063] 104| POR 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)	

[1064] 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.

[1065] 105| To screen for inversions, a PCR reaction is set up as follows:

Component:	Amount (ul)	Final concentration
Herculase II PGR 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)	

[1066] Note: primers are paired either as Out-Fwd+In Fwd, or Out-Rev+In-Rev.

[1067] 106| Applicants added 100-200 ng of normalized genomic DNA template from step 103 for each reaction.

[1068] 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

[1069] 108| Applicants run 2-5 μl 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, MgCl2 concentration, and/or the annealing temperature.

Genotyping for Targeted Modifications Via HDR. Timing. 2-3 d, 2-3 h Hands on

[1070] 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.

[1071] 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)	

[1072] 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

[1073] 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, MgCl2 concentration, and/or the annealing temperature.

[1074] 113| Applicants purified the PCR reactions using the QIAQuick PCR purification kit.

[1075] 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

[1076] i. The DNA is digested for 10 m at 37° C.:

[1077] ii. Applicants ran 10 μl 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.

[1078] iii. Applicants stained the gel with 1X SYBR Gold dye while rocking for 15 m.

[1079] 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.

[1080] 115| Alternatively, purified PCR amplicons from step 113 may be cloned and genotyped using Sanger sequencing or NGS.

Deep Sequencing and Off-Target Analysis·Timing 1-2 d

[1081] 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.

[1082] 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).

[1083] 117| Harvest genomic DNA from Cas9-targeted cells. Normalize QuickExtract genomic DNA to 100-200 ng/ul with ddH2O.

[1084] 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 III Fusion polymerase	1	
MgCl2 (25 mM)	2	1 mM
ddH2O	33	
Total	49 (for each reaction)	

[1085] 119| Add 100-200 ng of normalized genomic DNA template for each reaction.

[1086] 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	72° C., 15 s
22			72° C., 3 min

[1087] 121| Run 2-5 μl 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, MgCl2 concentration, and/or the annealing temperature.

[1088] 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.

[1089] 123| Nextera XT DNA Sample Preparation Kit. Following the manufacturer's protocol, generate Miseq sequencing-ready libraries with unique barcodes for each sample.

[1090] 124| Analyze sequencing data. Off-target analysis may be performed through read alignment programs such as ClustalW, Geneious, or simple sequence analysis scripts.

Timing

[1091] Steps 1-2 Design and synthesis of sgRNA oligos and ssODNs: 1-5 d, variable depending on supplier

[1092] Steps 3-5 Construction of CRISPR plasmid or PCR expression cassette: 2 h to 3 d Steps 6-53 Transfection into cell lines: 3 d (1 h hands-on time)

[1093] Steps 54-70 Optional derivation of clonal lines: 1-3 weeks, variable depending on cell type

[1094] Steps 71-91 Functional validation of NHEJ via SURVEYOR: 5-6 h

[1095] 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

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

-continued

Situation	Solution
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

DISCUSSION

[1096] 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

[1097] 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.

[1098] 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.

[1099] To more robustly understand the influence of NLS elements on Cas9 Applicants made 16 Cas9-GFP fusions by adding the same alpha importin NLS 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.

[1100] 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

Material and Reagents

[1101] Herculase II fusion polymerase (Agilent Technologies, cat. no. 600679)

[1102] 10×NEBuffer 4 (NEB, cat. No. B7004S)

[1103] BsaI HF (NEB, cat. No. R3535S)

[1104] T7 DNA ligase (Enzymatics, cat. no. L602L)

[1105] Fast Digest buffer, 10X (ThermoScientific, cat. No. B64)

[1106] FastDigest NotI (ThermoScientific, cat. No. FD0594)

[1107] FastAP Alkaline Phosphatase (ThermoScientific, cat. No. EF0651)

[1108] Lipofectamine2000 (Life Technologies, cat. No. 11668-019)

[1109] Trypsin (Life Technologies, cat. No. 15400054)

[1110] Forceps #4 (Sigma, cat. No. Z168777-1EA)

[1111] Forceps #5 (Sigma, cat. No. F6521-1EA)

[1112] 10× Hank's Balanced Salt Solution (Sigma, cat. No. H4641-500ML)

[1113] Penicillin/Streptomycin solution (Life Technologies, cat. No. P4333)

[1114] Neurobasal (Life Technologies, cat. No. 21103049)

[1115] B27 Supplement (Life Technologies, cat. No. 17504044)

[1116] L-glutamine (Life Technologies, cat. No. 25030081)

[1117] Glutamate (Sigma, cat. No. RES5063G-A7)

[1118] B-mercaptoethanol (Sigma, cat. No. M6250-100ML)

[1119] HA rabbit antibody (Cell Signaling, cat. No. 3724S)

[1120] LIVE/DEAD® Cell Imaging Kit (Life Technologies, cat. No. R37601)

[1121] 30G World Precision Instrument syringe (World Precision Instruments, cat. No. NANOFIL)

[1122] Stereotaxic apparatus (Kopf Instruments)

[1123] UltraMicroPump3 (World Precision Instruments, cat. No. UMP3-4)

[1124] Sucrose (Sigma, cat. No. S7903)

[1125] Calcium chloride (Sigma, cat. No. C1016)

[1126] Magnesium acetate (Sigma, cat. No. M0631)

[1127] Tris-HCl (Sigma, cat. no T5941)

[1128] EDTA (Sigma, cat. No. E6758)

[1129] NP-40 (Sigma, cat. No. NP40)

[1130] Phenylmethanesulfonyl fluoride (Sigma, cat. No. 78830)

[1131] Magnesium chloride (Sigma, cat. No. M8266)

[1132] Potassium chloride (Sigma, cat. No. P9333)

[1133] β-glycerophosphate (Sigma, cat. No. G9422)

[1134] Glycerol (Sigma, cat. No. G9012)

[1135] Vybrant® DyeCycle™ Ruby Stain (Life technologies, cat. No. S4942)

[1136] FACS Aria Flu-act-cell sorter (Koch Institute of MIT, Cambridge US)

[1137] DNAeasy Blood & Tissue Kit (Qiagen, cat. No. 69504)

Procedure

Constructing gRNA Multiplexes for Using In Vivo in the Brain

[1138] 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).

[1139] All 4 parts of the system were PCR amplified using Herculase II Fusion polymerase using following primers:

```
1st U6 Fw:
                                        (SEQ ID NO: 194)
gagggtctcgtccttgcggccgcgctagcgagggcctatttcccatgat

tc

1st gRNA Rv:
                                        (SEQ ID NO: 195)
ctcggtctcggtAAAAAAgcaccgactcggtgccactttttcaagttga

taacggactagccttattttaacttgctaTTTCtagctctaaaacNNNN

NNNNNNNNNNNNNNNNGGTGTTTCGTCCTTTCCAC

2nd U6 Fw:
                                        (SEQ ID NO: 196)
gagggtctcTTTaccggtgagggcctatttcccatgattcc

2nd gRNA Rv:
                                        (SEQ ID NO: 197)
ctcggtctccctcAAAAAAgcaccgactcggtgccactttttcaagttg

ataacggactagc cttattttaacttgctaTTTCtagctctaaaacNN

NNNNNNNNNNNNNNNNNNGGTGTTTCGTCCTTTCCAC

3rd U6 Fw:
                                        (SEQ ID NO: 198)
gagggtctcTTTgagctcgagggcctatttcccatgattc

3rd gRNA Rv:
                                        (SEQ ID NO: 199)
ctcggtctcgcgtAAAAAAgcaccgactcggtgccactttttcaagttg

ataacggactag ccttattttaacttgctaTTTCtagctctaaaacNN

NNNNNNNNNNNNNNNNNNGGTGTTTCGTCCTTTCCA

hSyn_GFP-kash Fw:
                                        (SEQ ID NO: 200)
gagggtctcTTacgcgtgtgtctagac

hSyn_GFP-kash Rev:
                                        (SEQ ID NO: 201)
ctcggtctcAaggaCAGGGAAGGGAGCAGTGGTTCACGCCTGTAATCCC

AGCAATTTGGGA GGCCAAGGTGGGTAGATCACCTGAGATTAGGAGTTG

C

(NNNNNNNNNNNNNNNNNNNN is a reverse compliment

targeted genomic sequence)
```

[1140] Applicants used Golden Gate strategy to assemble all parts (1:1 molecular ratio) of the system in a single step reaction:

1st U6_gRNA	18 ng
2nd U6_gRNA	18 ng
3rd 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

[1141] Golden Gate reaction product was PCR amplified using Herculase II fusion polymerase and following primers:

```
Fw
                                        (SEQ ID NO: 202)
5' cctgtccttgcggccgcgctagcgagggcc

Rv
                                        (SEQ ID NO: 203)
            5' cacgcggccgcaaggacagggaagggagcag
```

[1142] PCR product was cloned into AAV backbone, between ITR sequences using NotI restriction sites:

PCR Product Digestion:

Fast Digest buffer, 10X	3 μl
FastDigest NotI	1 μl
DNA	1 μg
ddH_2O	up to 30 μl

AAV Backbone Digestion:

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

[1143] 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_2O	up to 10 μl

[1144] After transformation of bacteria with ligation reaction product, applicants confirmed obtained clones with Sanger sequencing.

[1145] Positive DNA clones were tested in N2a cells after co-transfection with Cas9 construct (FIGS. **35** and **36**).

Design of New Cas9 Constructs for AAV Delivery

[1146] 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**).

Testing AAV Multiplex System in Primary Neurons

[1147] 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).

[1148] 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

HBSS

[1149] 435 ml H2O
[1150] 50 ml 10× Hank's Balanced Salt Solution
[1151] 16.5 ml 0.3M HEPES pH 7.3
[1152] 5 ml penicillin-streptomycin solution
[1153] Filter (0.2 μm) and store 4° C.

Neuron Plating Medium (100 ml)

[1154] 97 ml Neurobasal
[1155] 2 ml B27 Supplement
[1156] 1 ml penicillin-streptomycin solution
[1157] 250 μl glutamine
[1158] 125 μl glutamate

[1159] 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.

AAV-Driven Expression of the System

[1160] 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 B-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.

[1161] Expression of Syn-GFP-kash from gRNA multiplex AAV was confirmed with fluorescent microscopy (FIG. **50**).

Toxicity

[1162] 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 flourophores 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.

Virus Production

[1163] Concentrated virus was produced according to the methods described in McClure et al., 2011. Supernatant virus production occurred in HEK293FT cells.

Brain Surgeries

[1164] 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 μl was injected. The injection was monitored by a 'World Precision Instruments UltraMicroPump3' injection pump at a flow rate of 0.5 μl/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.

Sorting Nuclei and In Vivo Results

[1165] 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:

(SEQ ID NO: 204)

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICT

TGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIF

FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHN

VYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNH

YLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKSGLRSREEEEE

TDSRMPHLDSPGSSQPRRSFLSRVIRAALPLQLLLLLLLLLACLLPASED

DYSCTQANNFARSFYPMLRYTNGPPPT

[1166] 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 B-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.

[1167] 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

[1168] 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.

[1169] 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

[1170] 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.

[1171] AAV1/2 (serotype AAV1/2, i.e., hybrid or mosaic AAV1/AAV2 capsid AAV) heparin purified concentrated virus protocol

[1172] Media: D10+HEPES

[1173] 500 ml bottle DMEM high glucose+Glutamax (GIBCO)

[1174] 50 ml Hyclone FBS (heat-inactivated) (Thermo Fischer)

[1175] 5.5 ml HEPES solution (1M, GIBCO)

[1176] 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)

Transfection Reagent: Polyethylenimine (PEI) "Max"

[1177] Dissolve 50 mg PEI "Max" in 50 ml sterile Ultrapure H20

[1178] Adjust pH to 7.1

[1179] Filter with 0.22 um fliptop filter

[1180] Seal tube and wrap with parafilm

[1181] Freeze aliquots at −20° C. (for storage, can also be used immediately)

Cell Culture

[1182] Culture low passage HEK293FT in D10+HEPES

[1183] Passage everyday between 1:2 and 1:2.5

[1184] Advantageously do not allow cells to reach more than 85% confluency

For T75

[1185] Warm 10 ml HBSS (—Mg2+, —Ca2+, GIBCO)+1 ml TrypLE Express (GIBCO) per flask to 37° C. (Waterbath)

Aspirate Media Fully

[1186] Add 10 ml warm HBSS gently (to wash out media completely)

[1187] Add 1 ml TrypLE per Flask

[1188] Place flask in incubator (37° C.) for 1 min

[1189] Rock flask to detach cells

[1190] Add 9 ml D10+HEPES media (37° C.)

[1191] Pipette up and down 5 times to generate single cell suspension

[1192] 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)

[1193] transfer to T225 as soon as enough cells are present (for ease of handling large amounts of cells)

AAV Production (5*15 cm Dish Scale Per Construct):

[1194] Plate 10 million cells in 21.5 ml media into a 15 cm dish

[1195] Incubate for 18-22 hours at 37° C.

[1196] Transfection is ideal at 80% confluence

Per Plate

[1197] Prewarm 22 ml media (D10+HEPES)

Prepare Tube with DNA Mixture (Use Endofree Maxiprep DNA);

[1198] 5.2 ug vector of interest plasmid

[1199] 4.35 ug AAV 1 serotype plasmid

[1200] 4.35 ug AAV 2 serotype plasmid

[1201] 10.4 ug pDF6 plasmid (adenovirus helper genes) Vortex to mix

[1202] Add 434 uL DMEM (no serum!)

[1203] Add 130 μl PEI solution

[1204] Vortex 5-10 seconds

[1205] Add DNA/DMEM/PEI mixture to prewarmed media

[1206] Vortex briefly to mix

[1207] Replace media in 15 cm dish with DNA/DMEM/PEI mixture
[1208] Return to 37° C. incubator
[1209] Incubate 48 h before harvesting (make sure medium isn't turning too acidic)

Virus Harvest:

[1210] 1. aspirate media carefully from 15 cm dish dishes (advantageously do not dislodge cells)
[1211] 2. Add 25 ml RT DPBS (Invitrogen) to each plate and gently remove cells with a cell scraper. Collect suspension in 50 ml tubes.
[1212] 3. Pellet cells at 800×g for 10 minutes.
[1213] 4. Discard supernatant

Pause Point: Freeze Cell Pellet at −80 C if Desired

[1214] 5. resuspend pellet in 150 mM NaCl, 20 mM Tris pH 8.0, use 10 ml per tissue culture plate.
[1215] 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.
[1216] 7. Incubate at 37° C. for 1 hour (Waterbath).
[1217] 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.

Heparin Column Purification of AAV1/2:

[1218] 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.
[1219] 2. Equilibrate the column with 10 ml 150 mM NaCl, 20 mM Tris, pH 8.0 using the peristaltic pump.
[1220] 3. Binding of virus: Apply 50 ml virus solution to column and allow to flow through.
[1221] 4. Wash step 1: column with 20 ml 100 mM NaCl, 20 mM Tris, pH 8.0. (using the peristaltic pump)
[1222] 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.
[1223] Discard the flow-through.
[1224] (prepare the syringes with different buffers during the 50 min flow through of virus solution above)
[1225] 6. Elution Using 5 ml syringes and gentle pressure (flow rate of <1 ml/min) elute the virus from the column by applying:
[1226] 1.5 ml 400 mM NaCl, 20 mM Tris, pH 8.0
[1227] 3.0 ml 450 mM NaCl, 20 mM Tris, pH 8.0
[1228] 1.5 ml 500 mM NaCl, 20 mM Tris, pH 8.0
[1229] Collect these in a 15 ml centrifuge tube.

Concentration of AAV1/2:

[1230] 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.
[1231] 2. buffer exchange: Add 1 ml sterile DPBS to filter unit, centrifuge in 1 min intervals until correct volume (500ul) is reached.
[1232] 3. Concentration step 2: Add 500 μl concentrate to an Amicon Ultra 0.5 ml 100K filter unit. Centrifuge at 6000g for 2 min. Check concentrated volume-it should be approximately 100 μl. If necessary, centrifuge in 1 min intervals until correct volume is reached.
[1233] 4. Recovery: Invert filter insert and insert into fresh collection tube. Centrifuge at 1000g for 2 min.
[1234] Aliquot and freeze at −80° C.
[1235] 1 ul is typically required per injection site, small aliquots (e.g. 5ul) are therefore recommended (avoid freeze-thaw of virus).
[1236] determine DNaseI-resistant GC particle titer using qPCR (see separate protocol)

Materials

[1237] Amicon Ultra, 0.5 ml, 100K; MILLIPORE; UFC510024
[1238] Amicon Ultra, 15 ml, 100K; MILLIPORE; UFC910024
[1239] Benzonase nuclease; Sigma-Aldrich, E1014
[1240] HiTrap Heparin cartridge; Sigma-Aldrich; 54836
[1241] Sodium deoxycholate; Sigma-Aldrich; D5670

AAV1 Supernatant Production Protocol

[1242] Media: D10+HEPES
[1243] 500 ml bottle DMEM high glucose+Glutamax (Invitrogen)
[1244] 50 ml Hyclone FBS (heat-inactivated) (Thermo Fischer)
[1245] 5.5 ml HEPES solution (1M, GIBCO)
[1246] Cells: low passage HEK293FT (passage <10 at time of virus production)
[1247] Thaw new cells of passage 2-4 for virus production, grow up for 2-5 passages
[1248] Transfection reagent: Polyethylenimine (PEI) "Max"
[1249] Dissolve 50 mg PEI "Max" in 50 ml sterile Ultrapure H20
[1250] Adjust pH to 7.1
[1251] Filter with 0.22 um fliptop filter
[1252] Seal tube and wrap with parafilm
[1253] Freeze aliquots at −20° C. (for storage, can also be used immediately)
[1254] Cell Culture
[1255] Culture low passage HEK293FT in D10+HEPES Passage everyday between 1:2 and 1:2.5
[1256] Advantageously do let cells reach more than 85% confluency
[1257] For T75
[1258] Warm 10 ml HBSS (—Mg2+, —Ca2+, GIBCO)+1 ml TrypLE Express (GIBCO) per flask to 37° C. (Waterbath)
[1259] Aspirate media fully
[1260] Add 10 ml warm HBSS gently (to wash out media completely)
[1261] Add 1 ml TrypLE per Flask

[1262] Place flask in incubator (37° C.) for 1 min
[1263] Rock flask to detach cells
[1264] Add 9 ml D10+HEPES media (37° C.)
[1265] Pipette up and down 5 times to generate single cell suspension
[1266] 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)
[1267] transfer to T225 as soon as enough cells are present (for ease of handling large amounts of cells)
[1268] AAV production (single 15 cm dish scale)
[1269] Plate 10 million cells in 21.5 ml media into a 15 cm dish
[1270] Incubate for 18-22 hours at 37° C.
[1271] Transfection is ideal at 80% confluence per plate
[1272] Prewarm 22 ml media (D10+HEPES)
[1273] Prepare tube with DNA mixture (use endofree maxiprep DNA):
[1274] 5.2 μg vector of interest plasmid
[1275] 8.7 ug AAV 1 serotype plasmid
[1276] 10.4 ug DF6 plasmid (adenovirus helper genes)
[1277] Vortex to mix
[1278] Add 434 uL DMEM (no serum!) Add 130 μl PEI solution
[1279] Vortex 5-10 seconds
[1280] Add DNA/DMEM/PEI mixture to prewarmed media
[1281] Vortex briefly to mix
[1282] Replace media in 15 cm dish with DNA/DMEM/PEI mixture
[1283] Return to 37° C. incubator
[1284] Incubate 48 h before harvesting (advantageously monitor to ensure medium is not turning too acidic)

Virus Harvest:

[1285] Remove supernatant from 15 cm dish
[1286] Filter with 0.45 μm filter (low protein binding) Aliquot and freeze at −80° C.
[1287] Transduction (primary neuron cultures in 24-well format, 5 DIV)
[1288] 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)
[1289] Thaw AAV supernatant in 37° C. waterbath
[1290] Let equilibrate in incubator for 30 min
[1291] Add 250 μl AAV supernatant to each well
[1292] Incubate 24 h at 37° C.
[1293] Remove media/supernatant and replace with fresh complete neurobasal
[1294] Expression starts to be visible after 48h, saturates around 6-7 Days Post Infection
[1295] Constructs for pAAV plasmid with GOI should not exceed 4.8 kb including both ITRS.

[1296] Example of a human codon optimized sequence (i.e. being optimized for expression in humans) sequence: SaCas9 is provided below:

(SEQ ID NO: 205)

```
ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTAGCGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATG

AAAAGGAACTACATTCTGGGGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTATTGACTATGAAACAAGGGACG

TGATCGACGCAGGCGTCAGACTGTTCAAGGAGGCCAACGTGGAAAACAATGAGGGACGGAGAAGCAAGAGGGGAGCCAG

GCGCCTGAAACGACGGAGAAGGCACAGAATCCAGAGGGTGAAGAAACTGCTGTTCGATTACAACCTGCTGACCGACCAT

TCTGAGCTGAGTGGAATTAATCCTTATGAAGCCAGGGTGAAAGGCCTGAGTCAGAAGCTGTCAGAGGAAGAGTTTTCCG

CAGCTCTGCTGCACCTGGCTAAGCGCCGAGGAGTGCATAACGTCAATGAGGTGGAAGAGGACACCGGCAACGAGCTGTC

TACAAAGGAACAGATCTCACGCAATAGCAAAGCTCTGGAAGAGAAGTATGTCGCAGAGCTGCAGCTGGAACGGCTGAAG

AAAGATGGCGAGGTGAGAGGGTCAATTAATAGGTTCAAGACAAGCGACTACGTCAAAGAAGCCAAGCAGCTGCTGAAAG

TGCAGAAGGCTTACCACCAGCTGGATCAGAGCTTCATCGATACTTATATCGACCTGCTGGAGACTCGGAGAACCTACTA

TGAGGGACCAGGAGAAGGGAGCCCCTTCGGATGGAAAGACATCAAGGAATGGTACGAGATGCTGATGGGACATTGCACC

TATTTTCCAGAAGAGCTGAGAAGCGTCAAGTACGCTTATAACGCAGATCTGTACAACGCCCTGAATGACCTGAACAACC

TGGTCATCACCAGGGATGAAAACGAGAAACTGGAATACTATGAGAAGTTCCAGATCATCGAAAACGTGTTTAAGCAGAA

GAAAAAGCCTACACTGAAACAGATTGCTAAGGAGATCCTGGTCAACGAAGAGGACATCAAGGGCTACCGGGTGACAAGC

ACTGGAAAACCAGAGTTCACCAATCTGAAAGTGTATCACGATATTAAGGACATCACAGCACGGAAAGAAATCATTGAGA

ACGCCGAACTGCTGGATCAGATTGCTAAGATCCTGACTATCTACCAGAGCTCCGAGGACATCCAGGAAGAGCTGACTAA

CCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGATTAGTAATCTGAAGGGGTACACCGGAACACACAACCTGTCC

CTGAAAGCTATCAATCTGATTCTGGATGAGCTGTGGCATACAAACGACAATCAGATTGCAATCTTTAACCGGCTGAAGC

TGGTCCCAAAAAAGGTGGACCTGAGTCAGCAGAAAGAGATCCCAACCACACTGGTGGACGATTTCATTCTGTCACCCGT

GGTCAAGCGGAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAATGATATCATT

ATCGAGCTGGCTAGGGAGAAGAACAGCAAGGACGCACAGAAGATGATCAATGAGATGCAGAAACGAAACCGGCAGACCA

ATGAACGCATTGAAGAGATTATCCGAACTACCGGGAAAGAGAACGCAAAGTACCTGATTGAAAAAATCAAGCTGCACGA
```

-continued

TATGCAGGAGGGAAAGTGTCTGTATTCTCTGGAGGCCATCCCCCTGGAGGACCTGCTGAACAATCCATTCAACTACGAG

GTCGATCATATTATCCCCAGAAGCGTGTCCTTCGACAATTCCTTTAACAACAAGGTGCTGGTCAAGCAGGAAGAGAACT

CTAAAAAGGGCAATAGGACTCCTTTCCAGTACCTGTCTAGTTCAGATTCCAAGATCTCTTACGAAACCTTTAAAAAGCA

CATTCTGAATCTGGCCAAAGGAAAGGGCCGCATCAGCAAGACCAAAAAGGAGTACCTGCTGGAAGAGCGGGACATCAAC

AGATTCTCCGTCCAGAAGGATTTTATTAACCGGAATCTGGTGGACACAAGATACGCTACTCGCGGCCTGATGAATCTGC

TGCGATCCTATTTCCGGGTGAACAATCTGGATGTGAAAGTCAAGTCCATCAACGGCGGGTTCACATCTTTTCTGAGGCG

CAAATGGAAGTTTAAAAAGGAGCGCAACAAAGGGTACAAGCACCATGCCGAAGATGCTCTGATTATCGCAAATGCCGAC

TTCATCTTTAAGGAGTGGAAAAAGCTGGACAAAGCCAAGAAAGTGATGGAGAACCAGATGTTCGAAGAGAAGCAGGCCG

AATCTATGCCCGAAATCGAGACAGAACAGGAGTACAAGGAGATTTTCATCACTCCTCACCAGATCAAGCATATCAAGGA

TTTCAAGGACTACAAGTACTCTCACCGGGTGGATAAAAAGCCCAACAGAGAGCTGATCAATGACACCCTGTATAGTACA

AGAAAAGACGATAAGGGGAATACCCTGATTGTGAACAATCTGAACGGACTGTACGACAAAGATAATGACAAGCTGAAAA

AGCTGATCAACAAAAGTCCCGAGAAGCTGCTGATGTACCACCATGATCCTCAGACATATCAGAAACTGAAGCTGATTAT

GGAGCAGTACGGCGACGAGAAGAACCCACTGTATAAGTACTATGAAGAGACTGGGAACTACCTGACCAAGTATAGCAAA

AAGGATAATGGCCCCGTGATCAAGAAGATCAAGTACTATGGGAACAAGCTGAATGCCCATCTGGACATCACAGACGATT

ACCCTAACAGTCGCAACAAGGTGGTCAAGCTGTCACTGAAGCCATACAGATTCGATGTCTATCTGGACAACGGCGTGTA

TAAATTTGTGACTGTCAAGAATCTGGATGTCATCAAAAAGGAGAACTACTATGAAGTGAATAGCAAGTGCTACGAAGAG

GCTAAAAAGCTGAAAAAGATTAGCAACCAGGCAGAGTTCATCGCCTCCTTTTACAACAACGACCTGATTAAGATCAATG

GCGAACTGTATAGGGTCATCGGGGTGAACAATGATCTGCTGAACCGCATTGAAGTGAATATGATTGACATCACTTACCG

AGAGTATCTGGAAAACATGAATGATAAGCGCCCCCCTCGAATTATCAAAACAATTGCCTCTAAGACTCAGAGTATCAAA

AAGTACTCAACCGACATTCTGGGAAACCTGTATGAGGTGAAGAGCAAAAAGCACCCTCAGATTATCAAAAAGGGCTAAG

AATTC

Example 35: Minimizing Off-Target Cleavage Using Cas9 Nickase and Two Guide RNAs

[1297] 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.

[1298] 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 ore 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.

[1299] 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.

[1300] 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**).

[1301] 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 5' overhang. Efficient cleavage is reported for 5' overhang of at least 26 base pairs. In a preferred embodiment of the invention, 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**).

Example 36: Behavior Protocols

[1302] Elevated Plus Maze. The Elevated Plus-Maze is used to assess anxiety-like behavior, exploiting the conflict between the innate fear that rodents have of open areas versus their desire to explore novel environments.

[1303] The apparatus used for the elevated plus maze test is made of stainless steel and consists of four arms (two open without walls and two enclosed by 15 cm high walls) 35 cm long and 5 cm wide. Each arm of the maze is attached to sturdy metal legs such that it is elevated 40 cm off of the floor:

[1304] Mice are housed up to three per cage in a room with a 12 hr light/dark cycle with ad libitum access to food and water.

[1305] Behavioral testing is performed between 9:00 A.M. and 6:00 P.M in the brightly lighted room. All the cages containing mice are transferred to the behavior testing room 30 min before the first trial begins.

[1306] A mouse is taken out of its cage and place at the junction of the open and closed arms, facing the open arm opposite to where the experimenter is and let for freely explore the maze for 5 min. During that time animal is recorded and behavior is tracked. The total distance traveled in the open arms, time spent in open arms, the total number of transitions between arms, and the latency to enter the open arm are measured automatically with the Noldus software. At the end of the 5-min test, remove the rodent from the plus maze and place into a transport cage. Place back inside its homecage. The elevated plus maze is cleaned carefully and dry with paper towels before testing with another mouse.

Fear Conditioning

[1307] Animals: 4-weeks old male mice (BL6J/57); 10 per group

[1308] Foot shock: 2 s; 0.5 mA

[1309] Protocol involved three phases:

[1310] 1. Exposure, which allowed the mice to acclimate and become familiar with the training chamber. For unpaired group: 2 min of acclimation, followed by presentation of the 20-sec tone cue, followed by an 80-sec interval, repeated six times. For paired group the acclimation period was 12 min.

[1311] 2. Training; mice were presented with the CS and or US stimuli. Acclimation period-4 min. Six conditioning trials involving a tone (CS) and a shock (US) were run in trial epochs that lasted 100 sec (including the intertrial interval). For the paired group, each trial consisted of a 20-sec "baseline" interval, a 20-sec tone presentation, an 18-sec trace interval, a 2-sec shock, and a 40-sec post-shock interval. Trials were the same for the unpaired groups of mice except that the tone was omitted.

[1312] 3. Testing; mice were observed for freezing in response to each context and to the tone and trace-CS. Each phase occurred at 24-h intervals. Mice were allowed to move freely in the training (familiar) and test (altered; flat floor with a grid, bw patterns on the walls; vanillin scent) contexts for 3 min. All mice received four 100-sec testing trials. For all mice, each trial began with a 20-sec interval and the 20-sec tone presentation was followed by a 60-sec interval.

[1313] Scoring: All scoring was done based on video. The % of freezing was observed in 5-sec increments throughout each context test exposure and each 100-sec trial epoch. Freezing was scored as 1 or 0 in each time bin, if animal froze for at least 2s.

[1314] The data presented on the column charts is the average for the group of 7 animals, error bars represent SEM. The data on the line charts representing training/testing process in time, and each point is the average of 7 animals, from 6 or 4 trials (training and testing protocol, respectively). Error bars represent SEM.

[1315] Open field. The Open Field apparatus is broadly used to assess exploratory behaviour and is validated for use in the measurement of anxiety related behaviours. Animals are placed in the square arena (40×40 cm) for 10 min while their behavior is recorded. The total distance moved, velocity and time spent in the center of arena are scored.

[1316] Novel object recognition. The Novel Object Recognition test (NOR) is a popular test to study learning and memory in rodents. It is based on their tendency to interact more with a novel object than with a familiar object. In general, animals are first placed in an apparatus and allowed to explore an object. After a prescribed interval, the animal is returned to the apparatus, which now contains the familiar object and a novel object. Object recognition is distinguished by more time spent interacting with the novel object. The NOR task is particularly attractive because it requires no external motivation, reward, or punishment but a little training or habituation is required, and it can be completed in a relatively short time.

I Pretraining:

[1317] 1. Following arrival in the colony, the animals are allowed to acclimatize for at least 3-7 d.

[1318] 2. At lest 3 days before starting experiment animals are handled daily for at least 3 min per day and exposed to the transport routine.

[1319] 3. Familiarization with the testing environment—24 h before experiment animals are exposed to testing arena (40×40×35 cm, clear walls) in the behavior room for 10 min.

II Object Recognition Training:

[1320] 1. Two 'identical' to-be-familiarized (sample) objects are placed in the back left and right corners of the apparatus

[1321] 2. Animal is placed at the mid-point of the wall opposite the sample objects, what prevents any unintentional bias in placing the animal such that it is oriented more towards a particular side/object

[1322] 3. The experimenter is recording the animal behavior using camera, staying behind the curtain so as not to serve as a cue for the animal or to introduce unintentional bias into the study.

[1323] 4. After the planned sample-object exposure time (5-10 min), the animal is removed from the apparatus and return to the colony for planned training-to-testing interval. The apparatus is cleaned with 70% ethanol between animals.

III Delay Phase. Commonly Used Training-to-Testing Intervals Vary from 1h (for Robust Object Recognition) to 24 h.

IV Object Recognition Test:

[1324] 1. To test for object recognition after the training-to-testing interval, one of the familiar sample objects is placed in one back corner of the apparatus; the novel object is placed in the other back corner.

[1325] 2. Animal is placed in the apparatus as in step II.2. The experimenter is recording animal behavior, staying behind the curtain.

[1326] 3. After the planned objects exposure time (3-5 min) animals is removed from apparatus and returned to colony. The apparatus is cleaned with 70% ethanol between animals.

[1327] Data analysis: Recorded videos are used for later data analysis: a) directed contact scoring (total duration of animal contact with an object); b) within area scoring (an animal as "interacting" with the object when its nose is in contact with the object or directed at the object within minimal distance of 2 cm). Commonly used measures include time with familiar object versus novel object, a difference score (novel object interaction-familiar object interaction), a discrimination ratio (novel object interaction/total interaction with both objects). Object recognition in these measures is reflected by more time interacting with the novel than familiar object, a positive difference score or a discrimination ratio above 0.5, respectively.

[1328] Barnes maze. Carol Barnes developed a dry-land maze test for spatial learning and memory in 1979 where animals escaped from a brightly lit, exposed circular open platform surface to a small dark recessed chamber located under one of the 18 holes around the perimeter of the platform. Although it was initially invented for rats, the Barnes maze (BM) has become more popular to assess spatial memory in mice, taking advantage of their superior abilities to find and escape through small holes.

Procedure:

[1329] 1. Adaptation. Mouse is placed in a cylindrical black start chamber in the middle of the maze. After 10 s have elapsed chamber is removed and the buzzer is turned on together with strong light. Mouse is gently guided to the escape box. Once the mouse is inside the box, buzzer and light are turned off. The mouse should stay in the escape box for 2 min.

[1330] 2. Spatial acquisition. Mouse is placed in the cylindrical black start chamber in the middle of the maze. After 10 s have the chamber is lifted and the buzzer and light are on. Mouse is allowed to explore the maze for 3 minutes. The trial ends when the mouse enters the goal tunnel or after 3 min have elapsed. Immediately after the mouse enters the tunnel, the buzzer and the light are turned off and the mouse is allowed to stay in the tunnel for 1 min. If the mouse does not reach the goal within 3 minutes the experimenter is guiding it gently to the escape box and leave the mouse inside for 1 min. Mouse stays in the home cage until next trial. Whole trial is repeated 4 times in 15 minutes intervals during next 3-4 days.

[1331] 3. Reference memory trial. 24 h after the last training day, the probe trial is conducted. The target hole must be closed. Animal is placed in the middle of the maze under the cylindrical black start chamber and after 10 s the chamber is removed, the buzzer and the light are turned on. Mouse is removed from the maze after 90 s. The probe trial is done in order to determine if the animal remembers where the target goal was located. Number of pokes (errors) in each hole and latency and path length to reach the virtually target hole are measured.

Example 37: SaCas9 Update

[1332] FIG. **69** shows an AAV-Sa-Cas9 vector, a liver-specific AAV-Sa-Cas9 vector and an alternate AAV-Sa-Cas9 vector.

[1333] FIG. **70** shows data on optimized CMV-SaCas9-NLS-U6-sgRNA vector (submitted vector design last time); new data compares N'-term vs C'-term tagged SaCas9 and shows enhanced cleavage efficiency using C'-term NLS tagging.

[1334] New targets were chosen in exons 4 and 5 of the mouse Pcsk9 gene (sequences and locations below). Exons 4-5 lie downstream of a Pcsk9 N-terminal pro-domain region that is proteolytically cleaved upon protein maturation, and indels in this downstream region are expected to lead to protein degradation. Pcsk9 is involved in cycling and negative regulation of the LDL-receptor, and loss of Pcsk9 should indirectly lead to lowered plasma cholesterol levels.

[1335] SgRNAs were cloned into AAV vector and tested for indel activity in Hepa1-6 mouse hepatocyte cell line in single vector transfection. 500 ng of vector was transfected into 200,000 Hepa1-6 cells by Lipofectamine 2000 and DNA collected for SURVEYOR assay, showing cleavage of Pcsk9.

```
pAAV-CMV-SaCas9-NLS-U6-sgRNA(231:
atctcttagataccagcatc (SEQ ID NO: 206))

pAAV-CMV-SaCas9-NLS-U6-sgRNA(232:
tcaatctcccgatgggcacc (SEQ ID NO: 207))

pAAV-CMV-SaCas9-NLS-U6-sgRNA(233:
gcccatcgggagattgaggg (SEQ ID NO: 208))

pAAV-CMV-SaCas9-NLS-U6-sgRNA(234:
acttcaacagcgtgccggag (SEQ ID NO: 209))

pAAV-CMV-SaCas9-NLS-U6-sgRNA(235:
ccgctgaccacacctgccag (SEQ ID NO: 210))

pAAV-CMV-SaCas9-NLS-U6-sgRNA(236:
tggcaggtgtggtcagcggc (SEQ ID NO: 211))
```

[1336] FIG. **71** shows SURVEYOR image showing indels generated by new Pcsk9 targets.

[1337] FIG. **72** shows SaCas9 specificity: genome-wide off target sites (GWOTs) are predicted based on 2 criteria: they contain 4 or fewer mismatched bases to intended SaCas9 target and bear the least restrictive PAM for SaCas9,

NNGRR. HEK 293FT cells are transfected with either SpCas9 or SaCas9 with their corresponding sgRNAs at a target site (EMX1: TAGGGTTAGGGGCCCCAGGC (SEQ ID NO: 2)) that has CGGGGT as a PAM (sequence including PAM disclosed as SEQ ID NO: 369) so that it can be cut by either SpCas9 (CGG) or SaCas9 (CGGGGT). DNAs from cells are harvested and analyzed for indels by Illumina sequencing at on-target and 41 predicted off-target loci (following protocols from Hsu et al. Nature Biotech 2013 and data analysis pipeline developed by David Scott and Josh Weinstein).

[1338] FIG. **73** shows that that SaCas9 may have a higher level of off-target activity than SpCas9 at certain loci.

[1339] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention.

REFERENCES

[1340] Banker G, Goslin K. Developments in neuronal cell culture. Nature. 1988 Nov. 10; 336(6195):185-6.

[1341] Bedell, V. M. et al. In vivo genome editing using a high-efficiency TALEN system. Nature 491, 114-U133 (2012).

[1342] 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).

[1343] 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).

[1344] Boch, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326, 1509-1512 (2009).

[1345] Bogenhagen, D. F. & Brown, D. D. Nucleotide sequences in *Xenopus* 5S DNA required for transcription termination. Cell 24, 261-270 (1981).

[1346] 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).

[1347] Carlson, D. F. et al. Efficient TALEN-mediated gene knockout in livestock. Proc Natl Acad Sci USA 109, 17382-17387 (2012).

[1348] Chen, F. Q. et al. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat Methods 8, 753-U796 (2011).

[1349] 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).

[1350] Christian, M. et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186, 757-761 (2010).

[1351] Cong, L. et al. Multiplex genome engineering using CRISPR-Cas systems. Science 339, 819-823 (2013).

[1352] Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-607 (2011).

[1353] Deveau, H., Garneau, J. E. & Moineau, S. CRISPR-Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol 64, 475-493 (2010).

[1354] Ding, Q. et al. A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell 12, 238-251 (2013).

[1355] Garneau, J. E. et al. The CRISPR-Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67-71 (2010).

[1356] 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).

[1357] Geurts, A. M. et al. Knockout Rats via Embryo Microinjection of Zinc-Finger Nucleases. Science 325, 433-433 (2009).

[1358] 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.

[1359] Guschin, D. Y. et al. A rapid and general assay for monitoring endogenous gene modification. Methods Mol Biol 649, 247-256 (2010).

[1360] 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).

[1361] Horvath, P. & Barrangou, R. CRISPR-Cas, the immune system of bacteria and archaea. Science 327, 167-170 (2010).

[1362] Hsu, P. D. & Zhang, F. Dissecting neural function using targeted genome engineering technologies. ACS Chem Neurosci 3, 603-610 (2012).

[1363] Hwang, W. Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31, 227-229 (2013).

[1364] 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).

[1365] Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821 (2012).

[1366] Jinek, M. et al. RNA-programmed genome editing in human cells. eLife 2, e00471 (2013).

[1367] 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.

[1368] Levitt N. Briggs D. Gil A. Proudfoot N.J. Definition of an efficient synthetic poly(A) site. Genes Dev. 1989; 3:1019-1025.

[1369] 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.

[1370] 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]

[1371] 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).

[1372] Makarova, K. S. et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9, 467-477 (2011).

[1373] Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013).

[1374] 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.

[1375] Michaelis, L. M., Maud "Die kinetik der invertinwirkung.". Biochem. z(1913).

[1376] Miller, J. C. et al. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25, 778-785 (2007).

[1377] Miller, J. C. et al. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29, 143-148 (2011).

[1378] 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).

[1379] 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).

[1380] 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.

[1381] Oliveira, T. Y. et al. Translocation capture sequencing: a method for high throughput mapping of chromosomal rearrangements. J Immunol Methods 375, 176-181 (2012).

[1382] 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).

[1383] Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173-1183 (2013).

[1384] REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991)

[1385] Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol 30, 460-465 (2012).

[1386] 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).

[1387] Sander, J. D. et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods 8, 67-69 (2011).

[1388] Sanjana, N. E. et al. A transcription activator-like effector toolbox for genome engineering. Nat Protoc 7, 171-192 (2012).

[1389] Sapranauskas, R. et al. The *Streptococcus thermophilus* CRISPR-Cas system provides immunity in *Escherichia coli*. Nucleic Acids Res 39, 9275-9282 (2011).

[1390] Shen, B. et al. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. Cell Res 23, 720-723 (2013).

[1391] 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).

[1392] Soldner, F. et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 146, 318-331 (2011).

[1393] Takasu, Y. et al. Targeted mutagenesis in the silkworm *Bombyx mori* using zinc finger nuclease mRNA injection. Insect Biochem Molec 40, 759-765 (2010).

[1394] Tangri S, et al., Rationally engineered therapeutic proteins with reduced immunogenicity, J Immunol. 2005 Mar. 15; 174(6):3187-96.

[1395] 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).

[1396] Tuschl, T. Expanding small RNA interference. Nat Biotechnol 20, 446-448 (2002).

[1397] 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).

[1398] Valton, J. et al. Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. J Biol Chem 287, 38427-38432 (2012).

[1399] 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).

[1400] Watanabe, T. et al. Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. Nat Commun 3 (2012).

[1401] Wilson, E. B. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 22, 209-212 (1927).

[1402] Wood, A. J. et al. Targeted genome editing across species using ZFNs and TALENs. Science 333, 307 (2011).

[1403] 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).

[1404] Zhang, F. et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 29, 149-153 (2011).

SEQUENCE LISTING

```
Sequence total quantity: 443
SEQ ID NO: 1            moltype = DNA  length = 27
FEATURE                 Location/Qualifiers
misc_feature            1..27
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..27
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 1
gcactgaggg cctatttccc atgattc                                        27

SEQ ID NO: 2            moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 2
tagggttagg ggccccaggc                                                20

SEQ ID NO: 3            moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 3
gagtccgagc agaagaagaa                                                20

SEQ ID NO: 4            moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 4
gagtcctagc aggagaagaa                                                20

SEQ ID NO: 5            moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 5
gagtctaagc agaagaagaa                                                20

SEQ ID NO: 6            moltype =   length =
SEQUENCE: 6
000

SEQ ID NO: 7            moltype =   length =
SEQUENCE: 7
000

SEQ ID NO: 8            moltype =   length =
SEQUENCE: 8
000

SEQ ID NO: 9            moltype =   length =
SEQUENCE: 9
000

SEQ ID NO: 10           moltype =   length =
SEQUENCE: 10
000

SEQ ID NO: 11           moltype =   length =
SEQUENCE: 11
000

SEQ ID NO: 12           moltype =   length =
SEQUENCE: 12
000

SEQ ID NO: 13           moltype =   length =
SEQUENCE: 13
000
```

-continued

SEQ ID NO: 14 moltype = length =
SEQUENCE: 14
000

SEQ ID NO: 15 moltype = length =
SEQUENCE: 15
000

SEQ ID NO: 16 moltype = length =
SEQUENCE: 16
000

SEQ ID NO: 17 moltype = length =
SEQUENCE: 17
000

SEQ ID NO: 18 moltype = length =
SEQUENCE: 18
000

SEQ ID NO: 19 moltype = length =
SEQUENCE: 19
000

SEQ ID NO: 20 moltype = length =
SEQUENCE: 20
000

SEQ ID NO: 21 moltype = length =
SEQUENCE: 21
000

SEQ ID NO: 22 moltype = length =
SEQUENCE: 22
000

SEQ ID NO: 23 moltype = length =
SEQUENCE: 23
000

SEQ ID NO: 24 moltype = length =
SEQUENCE: 24
000

SEQ ID NO: 25 moltype = length =
SEQUENCE: 25
000

SEQ ID NO: 26 moltype = length =
SEQUENCE: 26
000

SEQ ID NO: 27 moltype = length =
SEQUENCE: 27
000

SEQ ID NO: 28 moltype = length =
SEQUENCE: 28
000

SEQ ID NO: 29 moltype = length =
SEQUENCE: 29
000

SEQ ID NO: 30 moltype = length =
SEQUENCE: 30
000

SEQ ID NO: 31 moltype = length =
SEQUENCE: 31
000

SEQ ID NO: 32 moltype = length =
SEQUENCE: 32
000

-continued

SEQ ID NO: 33 moltype = length =
SEQUENCE: 33
000

SEQ ID NO: 34 moltype = length =
SEQUENCE: 34
000

SEQ ID NO: 35 moltype = length =
SEQUENCE: 35
000

SEQ ID NO: 36 moltype = length =
SEQUENCE: 36
000

SEQ ID NO: 37 moltype = length =
SEQUENCE: 37
000

SEQ ID NO: 38 moltype = length =
SEQUENCE: 38
000

SEQ ID NO: 39 moltype = length =
SEQUENCE: 39
000

SEQ ID NO: 40 moltype = length =
SEQUENCE: 40
000

SEQ ID NO: 41 moltype = length =
SEQUENCE: 41
000

SEQ ID NO: 42 moltype = length =
SEQUENCE: 42
000

SEQ ID NO: 43 moltype = length =
SEQUENCE: 43
000

SEQ ID NO: 44 moltype = length =
SEQUENCE: 44
000

SEQ ID NO: 45 moltype = length =
SEQUENCE: 45
000

SEQ ID NO: 46 moltype = length =
SEQUENCE: 46
000

SEQ ID NO: 47 moltype = length =
SEQUENCE: 47
000

SEQ ID NO: 48 moltype = length =
SEQUENCE: 48
000

SEQ ID NO: 49 moltype = length =
SEQUENCE: 49
000

SEQ ID NO: 50 moltype = length =
SEQUENCE: 50
000

SEQ ID NO: 51 moltype = length =
SEQUENCE: 51
000

-continued

```
SEQ ID NO: 52           moltype =   length =
SEQUENCE: 52
000

SEQ ID NO: 53           moltype =   length =
SEQUENCE: 53
000

SEQ ID NO: 54           moltype =   length =
SEQUENCE: 54
000

SEQ ID NO: 55           moltype =   length =
SEQUENCE: 55
000

SEQ ID NO: 56           moltype =   length =
SEQUENCE: 56
000

SEQ ID NO: 57           moltype =   length =
SEQUENCE: 57
000

SEQ ID NO: 58           moltype =   length =
SEQUENCE: 58
000

SEQ ID NO: 59           moltype =   length =
SEQUENCE: 59
000

SEQ ID NO: 60           moltype = AA  length = 9
FEATURE                 Location/Qualifiers
REGION                  1..9
                        note = Description of Unknown: 'LAGLIDADG' family motif
                         peptide
source                  1..9
                        mol_type = protein
                        organism = unidentified
SEQUENCE: 60
LAGLIDADG                                                          9

SEQ ID NO: 61           moltype = RNA  length = 12
FEATURE                 Location/Qualifiers
misc_feature            1..12
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..12
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 61
gttttagagc ta                                                      12

SEQ ID NO: 62           moltype = AA  length = 7
FEATURE                 Location/Qualifiers
source                  1..7
                        mol_type = protein
                        organism = Simian virus 40
SEQUENCE: 62
PKKKRKV                                                            7

SEQ ID NO: 63           moltype = AA  length = 16
FEATURE                 Location/Qualifiers
REGION                  1..16
                        note = Description of Unknown: Nucleoplasmin bipartite NLS
                         sequence
source                  1..16
                        mol_type = protein
                        organism = unidentified
SEQUENCE: 63
KRPAATKKAG QAKKKK                                                  16
```

-continued

```
SEQ ID NO: 64           moltype = AA  length = 9
FEATURE                 Location/Qualifiers
REGION                  1..9
                        note = Description of Unknown: C-myc NLS sequence
source                  1..9
                        mol_type = protein
                        organism = unidentified
SEQUENCE: 64
PAAKRVKLD                                                           9

SEQ ID NO: 65           moltype = AA  length = 11
FEATURE                 Location/Qualifiers
REGION                  1..11
                        note = Description of Unknown: C-myc NLS sequence
source                  1..11
                        mol_type = protein
                        organism = unidentified
SEQUENCE: 65
RQRRNELKRS P                                                        11

SEQ ID NO: 66           moltype = AA  length = 38
FEATURE                 Location/Qualifiers
source                  1..38
                        mol_type = protein
                        organism = Homo sapiens
SEQUENCE: 66
NQSSNFGPMK GGNFGGRSSG PYGGGGQYFA KPRNQGGY                           38

SEQ ID NO: 67           moltype = AA  length = 42
FEATURE                 Location/Qualifiers
REGION                  1..42
                        note = Description of Unknown: IBB domain from
                         importin-alpha sequence
source                  1..42
                        mol_type = protein
                        organism = unidentified
SEQUENCE: 67
RMRIZFKNKG KDTAELRRRR VEVSVELRKA KKDEQILKRR NV                      42

SEQ ID NO: 68           moltype = AA  length = 8
FEATURE                 Location/Qualifiers
REGION                  1..8
                        note = Description of Unknown: Myoma T protein sequence
source                  1..8
                        mol_type = protein
                        organism = unidentified
SEQUENCE: 68
VSRKRPRP                                                            8

SEQ ID NO: 69           moltype = AA  length = 8
FEATURE                 Location/Qualifiers
REGION                  1..8
                        note = Description of Unknown: Myoma T protein sequence
source                  1..8
                        mol_type = protein
                        organism = unidentified
SEQUENCE: 69
PPKKARED                                                            8

SEQ ID NO: 70           moltype = AA  length = 8
FEATURE                 Location/Qualifiers
source                  1..8
                        mol_type = protein
                        organism = Homo sapiens
SEQUENCE: 70
PQPKKKPL                                                            8

SEQ ID NO: 71           moltype = AA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = protein
                        organism = Mus musculus
SEQUENCE: 71
SALIKKKKKM AP                                                       12
```

-continued

```
SEQ ID NO: 72           moltype = AA  length = 5
FEATURE                 Location/Qualifiers
source                  1..5
                        mol_type = protein
                        organism = Influenza virus
SEQUENCE: 72
DRLRR                                                           5

SEQ ID NO: 73           moltype = AA  length = 7
FEATURE                 Location/Qualifiers
source                  1..7
                        mol_type = protein
                        organism = Influenza virus
SEQUENCE: 73
PKQKKRK                                                         7

SEQ ID NO: 74           moltype = AA  length = 10
FEATURE                 Location/Qualifiers
source                  1..10
                        mol_type = protein
                        organism = Hepatitis delta virus
SEQUENCE: 74
RKLKKKIKKL                                                      10

SEQ ID NO: 75           moltype = AA  length = 10
FEATURE                 Location/Qualifiers
source                  1..10
                        mol_type = protein
                        organism = Mus musculus
SEQUENCE: 75
REKKKFLKRR                                                      10

SEQ ID NO: 76           moltype = AA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = protein
                        organism = Homo sapiens
SEQUENCE: 76
KRKGDEVDGV DEVAKKKSKK                                           20

SEQ ID NO: 77           moltype = AA  length = 17
FEATURE                 Location/Qualifiers
source                  1..17
                        mol_type = protein
                        organism = Homo sapiens
SEQUENCE: 77
RKCLQAGMNL EARKTKK                                              17

SEQ ID NO: 78           moltype =   length =
SEQUENCE: 78
000

SEQ ID NO: 79           moltype =   length =
SEQUENCE: 79
000

SEQ ID NO: 80           moltype =   length =
SEQUENCE: 80
000

SEQ ID NO: 81           moltype =   length =
SEQUENCE: 81
000

SEQ ID NO: 82           moltype = DNA  length = 137
FEATURE                 Location/Qualifiers
misc_feature            1..137
                        note = Description of Artificial Sequence: Synthetic
                         polynucleotide
misc_difference         1..20
                        note = a, c, t, g, unknown or other
source                  1..137
                        mol_type = other DNA
                        organism = synthetic construct
```

-continued

```
SEQUENCE: 82
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt tagaaataaa tcttgcagaa  60
gctacaaaga taaggcttca tgccgaaatc aacaccctgt cattttatgg cagggtgttt  120
tcgttattta atttttt                                                 137

SEQ ID NO: 83          moltype = DNA  length = 123
FEATURE                Location/Qualifiers
misc_feature           1..123
                       note = Description of Artificial Sequence: Synthetic
                        polynucleotide
misc_difference        1..20
                       note = a, c, t, g, unknown or other
source                 1..123
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 83
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag  60
gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttcgt tatttaattt  120
ttt                                                                123

SEQ ID NO: 84          moltype = DNA  length = 110
FEATURE                Location/Qualifiers
misc_feature           1..110
                       note = Description of Artificial Sequence: Synthetic
                        polynucleotide
misc_difference        1..20
                       note = a, c, t, g, unknown or other
source                 1..110
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 84
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag  60
gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgtttttt              110

SEQ ID NO: 85          moltype = DNA  length = 102
FEATURE                Location/Qualifiers
misc_feature           1..102
                       note = Description of Artificial Sequence: Synthetic
                        polynucleotide
misc_difference        1..20
                       note = a, c, t, g, unknown or other
source                 1..102
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 85
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc  60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tt                     102

SEQ ID NO: 86          moltype = DNA  length = 88
FEATURE                Location/Qualifiers
misc_feature           1..88
                       note = Description of Artificial Sequence: Synthetic
                        oligonucleotide
misc_difference        1..20
                       note = a, c, t, g, unknown or other
source                 1..88
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 86
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc  60
cgttatcaac ttgaaaaagt gtttttt                                      88

SEQ ID NO: 87          moltype = DNA  length = 76
FEATURE                Location/Qualifiers
misc_feature           1..76
                       note = Description of Artificial Sequence: Synthetic
                        oligonucleotide
misc_difference        1..20
                       note = a, c, t, g, unknown or other
source                 1..76
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 87
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc  60
cgttatcatt tttttt                                                  76
```

-continued

```
SEQ ID NO: 88          moltype = DNA  length = 12
FEATURE                Location/Qualifiers
misc_feature           1..12
                       note = Description of Artificial Sequence: Synthetic
                        oligonucleotide
source                 1..12
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 88
gttttagagc ta                                                  12

SEQ ID NO: 89          moltype = DNA  length = 31
FEATURE                Location/Qualifiers
misc_feature           1..31
                       note = Description of Artificial Sequence: Synthetic
                        oligonucleotide
source                 1..31
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 89
tagcaagtta aaataaggct agtccgtttt t                             31

SEQ ID NO: 90          moltype =   length =
SEQUENCE: 90
000

SEQ ID NO: 91          moltype = DNA  length = 33
FEATURE                Location/Qualifiers
source                 1..33
                       mol_type = unassigned DNA
                       organism = Homo sapiens
SEQUENCE: 91
ggacatcgat gtcacctcca atgactaggg tgg                           33

SEQ ID NO: 92          moltype = DNA  length = 33
FEATURE                Location/Qualifiers
source                 1..33
                       mol_type = unassigned DNA
                       organism = Homo sapiens
SEQUENCE: 92
cattggaggt gacatcgatg tcctccccat tgg                           33

SEQ ID NO: 93          moltype = DNA  length = 33
FEATURE                Location/Qualifiers
source                 1..33
                       mol_type = unassigned DNA
                       organism = Homo sapiens
SEQUENCE: 93
ggaagggcct gagtccgagc agaagaagaa ggg                           33

SEQ ID NO: 94          moltype = DNA  length = 33
FEATURE                Location/Qualifiers
source                 1..33
                       mol_type = unassigned DNA
                       organism = Homo sapiens
SEQUENCE: 94
ggtggcgaga ggggccgaga ttgggtgttc agg                           33

SEQ ID NO: 95          moltype = DNA  length = 33
FEATURE                Location/Qualifiers
source                 1..33
                       mol_type = unassigned DNA
                       organism = Homo sapiens
SEQUENCE: 95
atgcaggagg gtggcgagag gggccgagat tgg                           33

SEQ ID NO: 96          moltype = DNA  length = 32
FEATURE                Location/Qualifiers
misc_feature           1..32
                       note = Description of Artificial Sequence: Synthetic primer
source                 1..32
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 96
aaactctaga gagggcctat ttcccatgat tc                            32
```

-continued

```
SEQ ID NO: 97          moltype = DNA  length = 153
FEATURE                Location/Qualifiers
misc_feature           1..153
                       note = Description of Artificial Sequence: Synthetic primer
source                 1..153
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 97
acctctagaa aaaaagcacc gactcggtgc cactttttca agttgataac ggactagcct  60
tattttaact tgctatgctg ttttgtttcc aaaacagcat agctctaaaa cccctagtca  120
ttggaggtga cggtgtttcg tcctttccac aag                               153

SEQ ID NO: 98          moltype = DNA  length = 52
FEATURE                Location/Qualifiers
misc_feature           1..52
                       note = Description of Artificial Sequence: Synthetic primer
source                 1..52
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 98
taatacgact cactatagga agtgcgccac catggcccca aagaagaagc gg          52

SEQ ID NO: 99          moltype = DNA  length = 60
FEATURE                Location/Qualifiers
misc_feature           1..60
                       note = Description of Artificial Sequence: Synthetic primer
source                 1..60
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 99
ggttttttt tttttttttt tttttttttt ttttcttact ttttcttttt tgcctggccg  60

SEQ ID NO: 100         moltype = AA  length = 984
FEATURE                Location/Qualifiers
source                 1..984
                       mol_type = protein
                       organism = Campylobacter jejuni
SEQUENCE: 100
MARILAFDIG ISSIGWAFSE NDELKDCGVR IFTKVENPKT GESLALPRRL ARSARKRLAR  60
RKARLNHLKH LIANEFKLNY EDYQSFDESL AKAYKGSLIS PYELRFRALN ELLSKQDFAR  120
VILHIAKRRG YDDIKNSDDK EKGAILKAIK QNEEKLANYQ SVGEYLYKEY FQKFKENSKE  180
FTNVRNKKES YERCIAQSFL KDELKLIFKK QREFGFSFSK KFEEEVLSVA FYKRALKDFS  240
HLVGNCSFFT DEKRAPKNSP LAFMFVALTR IINLLNNLKN TEGILYTKDD LNALLNEVLK  300
NGTLTYKQTK KLLGLSDDYE FKGEKGTYFI EFKKYKEFIK ALGEHNLSQD DLNEIAKDIT  360
LIKDEIKLKK ALAKYDLNQN QIDSLSKLEF KDHLNISFKA LKLVTPLMLE GKKYDEACNE  420
LNLKVAINED KKDFLPAFNE TYYKDEVTNP VVLRAIKEYR KVLNALLKKY GKVHKINIEL  480
AREVGKNHSQ RAKIEKEQNE NYKAKKDAEL ECEKLGLKIN SKNILKLRLF KEQKEFCAYS  540
GEKIKISDLQ DEKMLEIDHI YPYSRSFDDS YMNKVLVFTK QNQEKLNQTP FEAFGNDSAK  600
WQKIEVLAKN LPTKKQKRIL DKNYKDKEQK NFKDRNLNDT RYIARLVLNY TKDYLDFLPL  660
SDDENTKLND TQKGSKVHVE AKSGMLTSAL RHTWGFSAKD RNNHLHHAID AVIIAYANNS  720
IVKAFSDFKK EQESNSAELY AKKISELDYK NKRKFFEPFS GFRQKVLDKI DEIFVSKPER  780
KKPSGALHEE TFRKEEEFYQ SYGGKEGVLK ALELGKIRKV NGKIVKNGDM FRVDIFKHKK  840
TNKFYAVPIY TMDFALKVLP NKAVARSKKG EIKDWILMDE NYEFCFSLYK DSLILIQTKD  900
MQEPEFVYYN AFTSSTVSLI VSKHDNKFET LSKNQKILFK NANEKEVIAK SIGIQNLKVF  960
EKYIVSALGE VTKAEFRQRE DFKK                                         984

SEQ ID NO: 101         moltype = DNA  length = 91
FEATURE                Location/Qualifiers
misc_feature           1..91
                       note = Description of Artificial Sequence: Synthetic
                        oligonucleotide
source                 1..91
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 101
tataatctca taagaaattt aaaaagggac taaaataaag agtttgcggg actctgcggg  60
gttacaatcc cctaaaaccg cttttaaaat t                                 91

SEQ ID NO: 102         moltype = DNA  length = 36
FEATURE                Location/Qualifiers
misc_feature           1..36
                       note = Description of Artificial Sequence: Synthetic
                        oligonucleotide
source                 1..36
                       mol_type = other DNA
                       organism = synthetic construct
```

-continued

```
SEQUENCE: 102
attttaccat aaagaaattt aaaaagggac taaaac                          36

SEQ ID NO: 103          moltype = RNA  length = 95
FEATURE                 Location/Qualifiers
misc_feature            1..95
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
misc_difference         1..20
                        note = a, c, u, g, unknown or other
source                  1..95
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 103
nnnnnnnnnn nnnnnnnnnn gttttagtcc cgaaagggac taaaataaag agtttgcggg  60
actctgcggg gttacaatcc cctaaaaccg ctttt                             95

SEQ ID NO: 104          moltype = AA  length = 1115
FEATURE                 Location/Qualifiers
REGION                  1..1115
                        note = Description of Artificial Sequence: Synthetic
                         polypeptide
source                  1..1115
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 104
MSDLVLGLDI GIGSVGVGIL NKVTGEIIHK NSRIFPAAQA ENNLVRRTNR QGRRLARRKK  60
HRRVRLNRLF EESGLITDFT KISINLNPYQ LRVKGLTDEL SNEELFIALK NMVKHRGISY  120
LDDASDDGNS SVGDYAQIVK ENSKQLETKT PGQIQLERYQ TYGQLRGDFT VEKDGKKHRL  180
INVFPTSAYR SEALRILQTQ QEFNPQITDE FINRYLEILT GKRKYYHGPG NEKSRTDYGR  240
YRTSGETLDN IFGILIGKCT FYPDEFRAAK ASYTAQEFNL LNDLNNLTVP TETKKLSKEQ  300
KNQIINYVKN EKAMGPAKLF KYIAKLLSCD VADIKGYRID KSGKAEIHTF EAYRKMKTLE  360
TLDIEQMDRE TLDKLAYVLT LNTEREGIQE ALEHEFADGS FSQKQVDELV QFRKANSSIF  420
GKGWHNFSVK LMMELIPELY ETSEEQMTIL TRLGKQKTTS SSNKTKYIDE KLLTEEIYNP  480
VVAKSVRQAI KIVNAAIKEY GDFDNIVIEM ARENQTTQKG QKNSRERMKR IEEGIKELGS  540
QILKEHPVEN TQLQNEKLYL YYLQNGRDMY VDQELDINRL SDYDVDHIVP QSFLKDDSID  600
NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW RQLLNAKLIT QRKFDNLTKA ERGGLSELDK  660
AGFIKRQLVE TRQITKHVAQ ILDSRMNTKY DENDKLIREV KVITLKSKLV SDFRKDFQFY  720
KVREINNYHH AHDAYLNAVV GTALIKKYPK LESEFVYGDY KVYDVRKMIA KSEQEIGKAT  780
AKYFFYSNIM NFFKTEITLA NGEIRKRPLI ETNGETGEIV WDKGRDFATV RKVLSMPQVN  840
IVKKTEVQTG GFSKESILPK RNSDKLIARK KDWDPKKYGG FDSPTVAYSV LVVAKVEKGK  900
SKKLKSVKEL LGITIMERSS FEKNPIDFLE AKGYKEVKKD LIIKLPKYSL FELENGRKRM  960
LASAGELQKG NELALPSKYV NFLYLASHYE KLKGSPEDNE QKQLFVEQHK HYLDEIIEQI  1020
SEFSKRVILA DANLDKVLSA YNKHRDKPIR EQAENIIHLF TLTNLGAPAA FKYFDTTIDR  1080
KRYTSTKEVL DATLIHQSIT GLYETRIDLS QLGGD                             1115

SEQ ID NO: 105          moltype = AA  length = 1374
FEATURE                 Location/Qualifiers
REGION                  1..1374
                        note = Description of Artificial Sequence: Synthetic
                         polypeptide
source                  1..1374
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 105
MDKKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE  60
ATRLKRTARR RYTRRKNRIC YLQEIFSNEM AKVDDSFFHR LEESFLVEED KKHERHPIFG  120
NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLNPDNSD  180
VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN  240
LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI  300
LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA  360
GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH  420
AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE  480
VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL  540
SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI  600
IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG  660
RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL  720
HEHIANLAGS PAIKKGILQT VKVVDELVKV MGRHKPENIV IEMARETNED DEKKAIQKIQ  780
KANKDEKDAA MLKAANQYNG KAELPHSVFH GHKQLATKIR LWHQQGERCL YTGKTISIHD  840
LINNSNQFEV DHILPLSITF DDSLANKVLV YATANQEKGQ RTPYQALDSM DDAWSFRELK  900
AFVRESKTLS NKKKEYLLTE EDISKFDVRK KFIERNLVDT RYASRVVLNA LQEHFRAHKI  960
DTKVSVVRGQ FTSQLRRHWG IEKTRDTYHH HAVDALIIAA SSQLNLWKKQ KNTLVSYSED  1020
QLLDIETGEL ISDDEYKESV FKAPYQHFVD TLKSKEFEDS ILFSYQVDSK FNRKISDATI  1080
YATRQAKVGK DKADETYVLG KIKDIYTQDG YDAFMKIYKK DKSKFLMYRH DPQTFEKVIE  1140
PILENYPNKQ INEKGKEVPC NPFLKYKEEH GYIRKYSKKG NGPEIKSLKY YDSKLGNHID  1200
ITPKDSNNKV VLQSVSPWRA DVYFNKTTGK YEILGLKYAD LQFEKGTGTY KISQEKYNDI  1260
```

-continued

```
KKKEGVDSDS EFKFTLYKND LLLVKDTETK EQQLFRFLSR TMPKQKHYVE LKPYDKQKFE  1320
GGEALIKVLG NVANSGQCKK GLGKSNISIY KVRTDVLGNQ HIIKNEGDKP KLDF        1374

SEQ ID NO: 106          moltype = AA  length = 15
FEATURE                 Location/Qualifiers
REGION                  1..15
                        note = Description of Artificial Sequence: Synthetic peptide
source                  1..15
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 106
GGGGSGGGGS GGGGS                                                    15

SEQ ID NO: 107          moltype = AA  length = 15
FEATURE                 Location/Qualifiers
REGION                  1..15
                        note = Description of Artificial Sequence: Synthetic peptide
source                  1..15
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 107
EAAAKEAAAK EAAAK                                                    15

SEQ ID NO: 108          moltype = AA  length = 18
FEATURE                 Location/Qualifiers
REGION                  1..18
                        note = Description of Artificial Sequence: Synthetic peptide
source                  1..18
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 108
GGGGGSGGGG GSGGGGGS                                                 18

SEQ ID NO: 109          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 109
gccaaattgg acgaccctcg cgg                                           23

SEQ ID NO: 110          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 110
cgaggagacc cccgtttcgg tgg                                           23

SEQ ID NO: 111          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 111
cccgccgccg ccgtggctcg agg                                           23

SEQ ID NO: 112          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 112
tgagctctac gagatccaca agg                                           23
```

-continued

```
SEQ ID NO: 113          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 113
ctcaaaattc ataccggttg tgg                                          23

SEQ ID NO: 114          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 114
cgttaaacaa caaccggact tgg                                          23

SEQ ID NO: 115          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 115
ttcaccccgc ggcgctgaat ggg                                          23

SEQ ID NO: 116          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 116
accactacca gtccgtccac agg                                          23

SEQ ID NO: 117          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 117
agcctttctg aacacatgca cgg                                          23

SEQ ID NO: 118          moltype = DNA  length = 39
FEATURE                 Location/Qualifiers
source                  1..39
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 118
cctgccatca atgtggccat gcatgtgttc agaaaggct                         39

SEQ ID NO: 119          moltype = DNA  length = 39
FEATURE                 Location/Qualifiers
source                  1..39
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 119
cctgccatca atgtggccgt gcatgtgttc agaaaggct                         39

SEQ ID NO: 120          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
```

-continued

```
                        oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 120
cactgcttaa gcctcgctcg agg                                          23

SEQ ID NO: 121          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 121
tcaccagcaa tattcgctcg agg                                          23

SEQ ID NO: 122          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 122
caccagcaat attccgctcg agg                                          23

SEQ ID NO: 123          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 123
tagcaacaga catacgctcg agg                                          23

SEQ ID NO: 124          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 124
gggcagtagt aatacgctcg agg                                          23

SEQ ID NO: 125          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 125
ccaattccca tacattattg tac                                          23

SEQ ID NO: 126          moltype = DNA  length = 4677
FEATURE                 Location/Qualifiers
misc_feature            1..4677
                        note = Description of Artificial Sequence: Synthetic
                         polynucleotide
source                  1..4677
                        mol_type = other DNA
                        organism = synthetic construct
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
```

-continued

```
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 tcacccccctg  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
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
cagattgata cccgccttgg catttcctgt cagaatgtaa cgtcagttga tggtact     4677
```

SEQ ID NO: 127 moltype = DNA length = 3150
FEATURE Location/Qualifiers

-continued

```
misc_feature            1..3150
                        note = Description of Artificial Sequence: Synthetic
                         polynucleotide
source                  1..3150
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 127
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
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

SEQ ID NO: 128          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic
                         polynucleotide
misc_difference         23..42
                        note = a, c, t, g, unknown or other
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 128
gaaattaata cgactcacta tannnnnnnn nnnnnnnnnn nngttttaga gctagaaata  60
gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga gtcggtgctt  120
ttttt                                                              125
```

-continued

```
SEQ ID NO: 129          moltype = DNA  length = 8452
FEATURE                 Location/Qualifiers
misc_feature            1..8452
                        note = Description of Artificial Sequence: Synthetic
                         polynucleotide
source                  1..8452
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 129
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 ctttttttct  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
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
```

-continued

```
tactacgtgg gccctctggc caggggaaac agcagattcg cctggatgac cagaaagagc  4080
gaggaaacca tcacccccctg 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
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
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 ggggcgcctc 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 gggcgaccgc 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
```

SEQ ID NO: 130 moltype = DNA length = 27

-continued

```
FEATURE                 Location/Qualifiers
source                  1..27
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 130
ccgtgccggg cggggagacc gccatgg                                          27

SEQ ID NO: 131          moltype = DNA  length = 22
FEATURE                 Location/Qualifiers
source                  1..22
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 131
ggcccggctg tggctgagga gc                                               22

SEQ ID NO: 132          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 132
cggtctcccg cccggcacgg                                                  20

SEQ ID NO: 133          moltype = DNA  length = 26
FEATURE                 Location/Qualifiers
source                  1..26
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 133
gctcctcagc cacagccggg ccgggt                                           26

SEQ ID NO: 134          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 134
cgaccctgga aa                                                          12

SEQ ID NO: 135          moltype = DNA  length = 14
FEATURE                 Location/Qualifiers
source                  1..14
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 135
ccccgccgcc accc                                                        14

SEQ ID NO: 136          moltype = DNA  length = 18
FEATURE                 Location/Qualifiers
source                  1..18
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 136
tttccagggt cgccatgg                                                    18

SEQ ID NO: 137          moltype = DNA  length = 10
FEATURE                 Location/Qualifiers
source                  1..10
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 137
ggcggcgggg                                                             10

SEQ ID NO: 138          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 138
acccttgtta gccacctccc                                                  20

SEQ ID NO: 139          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
```

-continued

```
SEQUENCE: 139
gaacgcagtg ctcttcgaag                                                20

SEQ ID NO: 140          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 140
ctcacgccct gctccgtgta                                                20

SEQ ID NO: 141          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 141
ggcgacaact acttcctggt                                                20

SEQ ID NO: 142          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 142
ctcacgccct gctccgtgta                                                20

SEQ ID NO: 143          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 143
gggcgacaac tacttcctgg                                                20

SEQ ID NO: 144          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 144
cctcttcagg gccggggtgg                                                20

SEQ ID NO: 145          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 145
gaggacccag gtggaactgc                                                20

SEQ ID NO: 146          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 146
tcagctccag gcggtcctgg                                                20

SEQ ID NO: 147          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 147
agcagcagca gcagtggcag                                                20

SEQ ID NO: 148          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 148
tgggcaccgt cagctccagg                                                20

SEQ ID NO: 149          moltype = DNA  length = 20
```

-continued

```
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 149
cagcagtggc agcggccacc                                                20

SEQ ID NO: 150          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 150
acctctcccc tggccctcat                                                20

SEQ ID NO: 151          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 151
ccaggaccgc ctggagctga                                                20

SEQ ID NO: 152          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 152
ccgtcagctc caggcggtcc                                                20

SEQ ID NO: 153          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 153
agcagcagca gcagtggcag                                                20

SEQ ID NO: 154          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 154
atgtgccaag caaagcctca                                                20

SEQ ID NO: 155          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 155
ttcggtcatg cccgtggatg                                                20

SEQ ID NO: 156          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 156
gtcgttgaaa ttcatcgtac                                                20

SEQ ID NO: 157          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 157
accacctgtg aagagtttcc                                                20

SEQ ID NO: 158          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
```

-continued

```
SEQUENCE: 158
cgtcgttgaa attcatcgta                                                20

SEQ ID NO: 159          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 159
accacctgtg aagagtttcc                                                20

SEQ ID NO: 160          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 160
gaacgcagtg cttttcgagg                                                20

SEQ ID NO: 161          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 161
acccttgttg gccacctccc                                                20

SEQ ID NO: 162          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 162
ggtgacaact actatctggt                                                20

SEQ ID NO: 163          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 163
ctcacaccct gctccgtgta                                                20

SEQ ID NO: 164          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 164
gggtgacaac tactatctgg                                                20

SEQ ID NO: 165          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 165
ctcacaccct gctccgtgta                                                20

SEQ ID NO: 166          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 166
cgagaacgca gtgcttttcg                                                20

SEQ ID NO: 167          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 167
acccttgttg gccacctccc                                                20

SEQ ID NO: 168          moltype = DNA  length = 20
```

-continued

```
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 168
atgagccaag caaatcctca                                                20

SEQ ID NO: 169          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 169
ttccgtcatg cccgtggaca                                                20

SEQ ID NO: 170          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 170
cttcgttgaa aaccattgta                                                20

SEQ ID NO: 171          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 171
ccacctctga agagtttcct                                                20

SEQ ID NO: 172          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 172
cttcgttgaa aaccattgta                                                20

SEQ ID NO: 173          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 173
accacctctg aagagtttcc                                                20

SEQ ID NO: 174          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 174
cttccactca ctctgcgatt                                                20

SEQ ID NO: 175          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 175
accatgtctc agtgtcaagc                                                20

SEQ ID NO: 176          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 176
ggcggcaaca gcggcaacag                                                20

SEQ ID NO: 177          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
```

-continued

```
SEQUENCE: 177
actgctctgc gtggctgcgg                                                20

SEQ ID NO: 178          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 178
ccgcagccac gcagagcagt                                                20

SEQ ID NO: 179          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 179
gcacctctcc tcgccccgat                                                20

SEQ ID NO: 180          moltype = DNA  length = 23
FEATURE                 Location/Qualifiers
misc_feature            1..23
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..23
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 180
gagggcctat ttcccatgat tcc                                            23

SEQ ID NO: 181          moltype = DNA  length = 126
FEATURE                 Location/Qualifiers
misc_feature            1..126
                        note = Description of Artificial Sequence: Synthetic primer
misc_difference         84..102
                        note = a, c, t, g, unknown or other
source                  1..126
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 181
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacnnnnnnn nnnnnnnnnn nnccggtgtt tcgtcctttc  120
cacaag                                                               126

SEQ ID NO: 182          moltype =   length =
SEQUENCE: 182
000

SEQ ID NO: 183          moltype =   length =
SEQUENCE: 183
000

SEQ ID NO: 184          moltype = DNA  length = 126
FEATURE                 Location/Qualifiers
misc_feature            1..126
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..126
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 184
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacccctagt cattggaggt gaccggtgtt tcgtcctttc  120
cacaag                                                               126

SEQ ID NO: 185          moltype = DNA  length = 24
FEATURE                 Location/Qualifiers
misc_feature            1..24
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..24
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 185
caccgtcacc tccaatgact aggg                                           24

SEQ ID NO: 186          moltype = DNA  length = 24
FEATURE                 Location/Qualifiers
misc_feature            1..24
```

-continued

```
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..24
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 186
aaacccctag tcattggagg tgac                                          24

SEQ ID NO: 187          moltype = DNA  length = 192
FEATURE                 Location/Qualifiers
misc_feature            1..192
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..192
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 187
cagaagaaga agggctccca tcacatcaac cggtggcgca ttgccacgaa gcaggccaat  60
ggggaggaca tcgatgtcac ctccaatgac aagcttgcta gcggtgggca accacaaacc  120
cacgagggca gagtgctgct tgctgctggc caggcccctg cgtgggccca agctggactc  180
tggccactcc ct                                                       192

SEQ ID NO: 188          moltype = DNA  length = 192
FEATURE                 Location/Qualifiers
misc_feature            1..192
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..192
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 188
agggagtggc cagagtccag cttgggccca cgcaggggcc tggccagcag caagcagcac  60
tctgccctcg tgggtttgtg gttgcccacc gctagcaagc ttgtcattgg aggtgacatc  120
gatgtcctcc ccattggcct gcttcgtggc aatgcgccac cggttgatgt gatgggagcc  180
cttcttcttc tg                                                       192

SEQ ID NO: 189          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 189
ccatcccctt ctgtgaatgt                                               20

SEQ ID NO: 190          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 190
ggagattgga gacacggaga                                               20

SEQ ID NO: 191          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 191
ggctccctgg gttcaaagta                                               20

SEQ ID NO: 192          moltype = DNA  length = 21
FEATURE                 Location/Qualifiers
misc_feature            1..21
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..21
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 192
agaggggtct ggatgtcgta a                                             21

SEQ ID NO: 193          moltype = DNA  length = 24
FEATURE                 Location/Qualifiers
misc_feature            1..24
```

-continued

```
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..24
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 193
cgccagggtt ttcccagtca cgac                                          24

SEQ ID NO: 194          moltype = DNA  length = 51
FEATURE                 Location/Qualifiers
misc_feature            1..51
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..51
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 194
gagggtctcg tccttgcggc cgcgctagcg agggcctatt tcccatgatt c            51

SEQ ID NO: 195          moltype = DNA  length = 133
FEATURE                 Location/Qualifiers
misc_feature            1..133
                        note = Description of Artificial Sequence: Synthetic primer
misc_difference         95..114
                        note = a, c, t, g, unknown or other
source                  1..133
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 195
ctcggtctcg gtaaaaaagc accgactcgg tgccactttt tcaagttgat aacggactag  60
ccttatttta acttgctatt tctagctcta aaacnnnnnn nnnnnnnnnn nnnnggtgtt  120
tcgtcctttc cac                                                     133

SEQ ID NO: 196          moltype = DNA  length = 41
FEATURE                 Location/Qualifiers
misc_feature            1..41
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..41
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 196
gagggtctct ttaccggtga gggcctattt cccatgattc c                       41

SEQ ID NO: 197          moltype = DNA  length = 133
FEATURE                 Location/Qualifiers
misc_feature            1..133
                        note = Description of Artificial Sequence: Synthetic primer
misc_difference         95..114
                        note = a, c, t, g, unknown or other
source                  1..133
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 197
ctcggtctcc tcaaaaaagc accgactcgg tgccactttt tcaagttgat aacggactag  60
ccttatttta acttgctatt tctagctcta aaacnnnnnn nnnnnnnnnn nnnnggtgtt  120
tcgtcctttc cac                                                     133

SEQ ID NO: 198          moltype = DNA  length = 40
FEATURE                 Location/Qualifiers
misc_feature            1..40
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..40
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 198
gagggtctct ttgagctcga gggcctattt cccatgattc                         40

SEQ ID NO: 199          moltype = DNA  length = 133
FEATURE                 Location/Qualifiers
misc_feature            1..133
                        note = Description of Artificial Sequence: Synthetic primer
misc_difference         96..115
                        note = a, c, t, g, unknown or other
source                  1..133
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 199
ctcggtctcg cgtaaaaaag caccgactcg gtgccacttt ttcaagttga taacggacta  60
```

-continued

```
gccttatttt aacttgctat ttctagctct aaaacnnnnn nnnnnnnnnn nnnnnggtgt  120
ttcgtccttt cca                                                     133

SEQ ID NO: 200          moltype = DNA  length = 27
FEATURE                 Location/Qualifiers
misc_feature            1..27
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..27
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 200
gagggtctct tacgcgtgtg tctagac                                       27

SEQ ID NO: 201          moltype = DNA  length = 98
FEATURE                 Location/Qualifiers
misc_feature            1..98
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..98
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 201
ctcggtctca aggacaggga agggagcagt ggttcacgcc tgtaatccca gcaatttggg  60
aggccaaggt gggtagatca cctgagatta ggagttgc                          98

SEQ ID NO: 202          moltype = DNA  length = 30
FEATURE                 Location/Qualifiers
misc_feature            1..30
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..30
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 202
cctgtccttg cggccgcgct agcgagggcc                                    30

SEQ ID NO: 203          moltype = DNA  length = 31
FEATURE                 Location/Qualifiers
misc_feature            1..31
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..31
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 203
cacgcggccg caaggacagg gaagggagca g                                  31

SEQ ID NO: 204          moltype = AA  length = 327
FEATURE                 Location/Qualifiers
REGION                  1..327
                        note = Description of Artificial Sequence: Synthetic
                         polypeptide
source                  1..327
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 204
MVSKGEELFT GVVPILVELD GDVNGHKFSV SGEGEGDATY GKLTLKFICT TGKLPVPWPT  60
LVTTLTYGVQ CFSRYPDHMK QHDFFKSAMP EGYVQERTIF FKDDGNYKTR AEVKFEGDTL  120
VNRIELKGID FKEDGNILGH KLEYNYNSHN VYIMADKQKN GIKVNFKIRH NIEDGSVQLA  180
DHYQQNTPIG DGPVLLPDNH YLSTQSALSK DPNEKRDHMV LLEFVTAAGI TLGMDELYKS  240
GLRSREEEEE TDSRMPHLDS PGSSQPRRSF LSRVIRAALP LQLLLLLLLL LACLLPASED  300
DYSCTQANNF ARSFYPMLRY TNGPPPT                                      327

SEQ ID NO: 205          moltype = DNA  length = 3243
FEATURE                 Location/Qualifiers
misc_feature            1..3243
                        note = Description of Artificial Sequence: Synthetic
                         polynucleotide
source                  1..3243
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 205
accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc  60
aaggtcgaag cgtccatgaa aaggaactac 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
```

-continued

```
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
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
atgattgaca tcacttaccg agagtatctg gaaaacatga atgataagcg ccccccctcga  3120
attatcaaaa caattgcctc taagactcag agtatcaaaa agtactcaac cgacattctg  3180
ggaaacctgt atgaggtgaa gagcaaaaag caccctcaga ttatcaaaaa gggctaagaa  3240
ttc                                                                3243

SEQ ID NO: 206          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 206
atctcttaga taccagcatc                                              20

SEQ ID NO: 207          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 207
tcaatctccc gatgggcacc                                              20

SEQ ID NO: 208          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
```

-continued

```
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 208
gcccatcggg agattgaggg                                            20

SEQ ID NO: 209          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 209
acttcaacag cgtgccggag                                            20

SEQ ID NO: 210          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 210
ccgctgacca cacctgccag                                            20

SEQ ID NO: 211          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 211
tggcaggtgt ggtcagcggc                                            20

SEQ ID NO: 212          moltype = DNA  length = 102
FEATURE                 Location/Qualifiers
misc_feature            1..102
                        note = Description of Artificial Sequence: Synthetic
                         polynucleotide
source                  1..102
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 212
gttttagagc tatgctgttt tgaatggtcc caaaacggaa gggcctgagt ccgagcagaa  60
gaagaagttt tagagctatg ctgttttgaa tggtcccaaa ac                     102

SEQ ID NO: 213          moltype = DNA  length = 100
FEATURE                 Location/Qualifiers
source                  1..100
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 213
cggaggacaa agtacaaacg gcagaagctg gaggaggaag ggcctgagtc cgagcagaag  60
aagaagggct cccatcacat caaccggtgg cgcattgcca                        100

SEQ ID NO: 214          moltype = DNA  length = 50
FEATURE                 Location/Qualifiers
source                  1..50
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 214
agctggagga ggaagggcct gagtccgagc agaagaagaa gggctcccac             50

SEQ ID NO: 215          moltype = RNA  length = 30
FEATURE                 Location/Qualifiers
misc_feature            1..30
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..30
                        mol_type = other RNA
                        organism = synthetic construct
```

-continued

```
SEQUENCE: 215
gagtccgagc agaagaagaa gttttagagc                                    30

SEQ ID NO: 216          moltype = DNA  length = 49
FEATURE                 Location/Qualifiers
misc_feature            1..49
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..49
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 216
agctggagga ggaagggcct gagtccgagc agaagagaag ggctcccac               49

SEQ ID NO: 217          moltype = DNA  length = 53
FEATURE                 Location/Qualifiers
source                  1..53
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 217
ctggaggagg aagggcctga gtccgagcag aagagaaggg gctcccatca cat          53

SEQ ID NO: 218          moltype = DNA  length = 52
FEATURE                 Location/Qualifiers
source                  1..52
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 218
ctggaggagg aagggcctga gtccgagcag aagagaaggg ctcccatcac at           52

SEQ ID NO: 219          moltype = DNA  length = 54
FEATURE                 Location/Qualifiers
source                  1..54
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 219
ctggaggagg aagggcctga gtccgagcag aagaaagaag ggctcccatc acat         54

SEQ ID NO: 220          moltype = DNA  length = 50
FEATURE                 Location/Qualifiers
source                  1..50
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 220
ctggaggagg aagggcctga gtccgagcag aagaagggct cccatcacat              50

SEQ ID NO: 221          moltype = DNA  length = 47
FEATURE                 Location/Qualifiers
source                  1..47
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 221
ctggaggagg aagggcctga gcccgagcag aagggctccc atcacat                 47

SEQ ID NO: 222          moltype = DNA  length = 48
FEATURE                 Location/Qualifiers
source                  1..48
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 222
ctggaggagg aagggcctga gtccgagcag aagaagaagg gctcccat                48

SEQ ID NO: 223          moltype = RNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 223
gagtccgagc agaagaagat                                               20

SEQ ID NO: 224          moltype = RNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
```

-continued

```
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 224
gagtccgagc agaagaagta                                              20

SEQ ID NO: 225          moltype = RNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 225
gagtccgagc agaagaacaa                                              20

SEQ ID NO: 226          moltype = RNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 226
gagtccgagc agaagatgaa                                              20

SEQ ID NO: 227          moltype = RNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 227
gagtccgagc agaagtagaa                                              20

SEQ ID NO: 228          moltype = RNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 228
gagtccgagc agatgaagaa                                              20

SEQ ID NO: 229          moltype = RNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 229
gagtccgagc acaagaagaa                                              20

SEQ ID NO: 230          moltype = RNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 230
gagtccgagg agaagaagaa                                              20

SEQ ID NO: 231          moltype = RNA  length = 20
FEATURE                 Location/Qualifiers
```

-continued

```
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 231
gagtccgtgc agaagaagaa                                                20

SEQ ID NO: 232          moltype = RNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 232
gagtcggagc agaagaagaa                                                20

SEQ ID NO: 233          moltype = RNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..20
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 233
gagaccgagc agaagaagaa                                                20

SEQ ID NO: 234          moltype = DNA  length = 24
FEATURE                 Location/Qualifiers
misc_feature            1..24
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..24
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 234
aatgacaagc ttgctagcgg tggg                                           24

SEQ ID NO: 235          moltype = DNA  length = 39
FEATURE                 Location/Qualifiers
misc_feature            1..39
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..39
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 235
aaaacggaag ggcctgagtc cgagcagaag aagaagttt                           39

SEQ ID NO: 236          moltype = DNA  length = 39
FEATURE                 Location/Qualifiers
misc_feature            1..39
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..39
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 236
aaacaggggc cgagattggg tgttcagggc agaggtttt                           39

SEQ ID NO: 237          moltype = DNA  length = 38
FEATURE                 Location/Qualifiers
misc_feature            1..38
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..38
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 237
aaaacggaag ggcctgagtc cgagcagaag aagaagtt                            38
```

-continued

```
SEQ ID NO: 238          moltype = DNA  length = 40
FEATURE                 Location/Qualifiers
misc_feature            1..40
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..40
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 238
aacggaggga ggggcacaga tgagaaactc agggttttag                        40

SEQ ID NO: 239          moltype = DNA  length = 38
FEATURE                 Location/Qualifiers
source                  1..38
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 239
agcccttctt cttctgctcg gactcaggcc cttcctcc                          38

SEQ ID NO: 240          moltype = DNA  length = 40
FEATURE                 Location/Qualifiers
source                  1..40
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 240
cagggaggga ggggcacaga tgagaaactc aggaggcccc                        40

SEQ ID NO: 241          moltype = DNA  length = 80
FEATURE                 Location/Qualifiers
misc_feature            1..80
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..80
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 241
ggcaatgcgc caccggttga tgtgatggga gcccttctag gaggccccca gagcagccac  60
tggggcctca acactcaggc                                              80

SEQ ID NO: 242          moltype = DNA  length = 33
FEATURE                 Location/Qualifiers
source                  1..33
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 242
catcgatgtc ctccccattg gcctgcttcg tgg                               33

SEQ ID NO: 243          moltype = DNA  length = 33
FEATURE                 Location/Qualifiers
source                  1..33
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 243
ttcgtggcaa tgcgccaccg gttgatgtga tgg                               33

SEQ ID NO: 244          moltype = DNA  length = 33
FEATURE                 Location/Qualifiers
source                  1..33
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 244
tcgtggcaat gcgccaccgg ttgatgtgat ggg                               33

SEQ ID NO: 245          moltype = DNA  length = 33
FEATURE                 Location/Qualifiers
source                  1..33
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 245
tccagcttct gccgtttgta ctttgtcctc cgg                               33

SEQ ID NO: 246          moltype = DNA  length = 33
FEATURE                 Location/Qualifiers
source                  1..33
```

-continued

```
                      mol_type = unassigned DNA
                      organism = Homo sapiens
SEQUENCE: 246
ggagggaggg gcacagatga gaaactcagg agg                                33

SEQ ID NO: 247        moltype = DNA  length = 33
FEATURE               Location/Qualifiers
source                1..33
                      mol_type = unassigned DNA
                      organism = Homo sapiens
SEQUENCE: 247
aggggccgag attgggtgtt cagggcagag agg                                33

SEQ ID NO: 248        moltype = DNA  length = 33
FEATURE               Location/Qualifiers
source                1..33
                      mol_type = unassigned DNA
                      organism = Mus musculus
SEQUENCE: 248
caagcactga gtgccattag ctaaatgcat agg                                33

SEQ ID NO: 249        moltype = DNA  length = 33
FEATURE               Location/Qualifiers
source                1..33
                      mol_type = unassigned DNA
                      organism = Mus musculus
SEQUENCE: 249
aatgcatagg gtaccaccca caggtgccag ggg                                33

SEQ ID NO: 250        moltype = DNA  length = 33
FEATURE               Location/Qualifiers
source                1..33
                      mol_type = unassigned DNA
                      organism = Mus musculus
SEQUENCE: 250
acacacatgg gaaagcctct gggccaggaa agg                                33

SEQ ID NO: 251        moltype = DNA  length = 37
FEATURE               Location/Qualifiers
source                1..37
                      mol_type = unassigned DNA
                      organism = Homo sapiens
SEQUENCE: 251
ggaggaggta gtatacagaa acacagagaa gtagaat                            37

SEQ ID NO: 252        moltype = DNA  length = 37
FEATURE               Location/Qualifiers
source                1..37
                      mol_type = unassigned DNA
                      organism = Homo sapiens
SEQUENCE: 252
agaatgtaga ggagtcacag aaactcagca ctagaaa                            37

SEQ ID NO: 253        moltype = DNA  length = 98
FEATURE               Location/Qualifiers
misc_feature          1..98
                      note = Description of Artificial Sequence: Synthetic
                       oligonucleotide
source                1..98
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 253
ggacgaaaca ccggaaccat tcaaaacagc atagcaagtt aaaataaggc tagtccgtta  60
tcaacttgaa aaagtggcac cgagtcggtg ctttttt                            98

SEQ ID NO: 254        moltype = DNA  length = 186
FEATURE               Location/Qualifiers
misc_feature          1..186
                      note = Description of Artificial Sequence: Synthetic
                       polynucleotide
source                1..186
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 254
ggacgaaaca ccggtagtat taagtattgt tttatggctg ataaatttct ttgaatttct  60
ccttgattat ttgttataaa agttataaaa taatcttgtt ggaaccattc aaaacagcat  120
```

-continued

```
agcaagttaa aataaggcta gtccgttatc aacttgaaaa agtggcaccg agtcggtgct  180
tttttt                                                             186

SEQ ID NO: 255         moltype = DNA  length = 95
FEATURE                Location/Qualifiers
misc_feature           1..95
                       note = Description of Artificial Sequence: Synthetic
                        oligonucleotide
source                 1..95
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 255
gggttttaga gctatgctgt tttgaatggt cccaaaacgg gtcttcgaga agacgtttta  60
gagctatgct gttttgaatg gtcccaaaac ttttt                             95

SEQ ID NO: 256         moltype =   length =
SEQUENCE: 256
000

SEQ ID NO: 257         moltype =   length =
SEQUENCE: 257
000

SEQ ID NO: 258         moltype = DNA  length = 84
FEATURE                Location/Qualifiers
misc_feature           1..84
                       note = Description of Artificial Sequence: Synthetic
                        oligonucleotide
source                 1..84
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 258
gtggaaagga cgaaacaccg ggtcttcgag aagacctgtt ttagagctag aaatagcaag  60
ttaaaataag gctagtccgt tttt                                         84

SEQ ID NO: 259         moltype = RNA  length = 46
FEATURE                Location/Qualifiers
misc_feature           1..46
                       note = Description of Artificial Sequence: Synthetic
                        oligonucleotide
misc_difference        1..19
                       note = a, c, u, g, unknown or other
source                 1..46
                       mol_type = other RNA
                       organism = synthetic construct
SEQUENCE: 259
nnnnnnnnnn nnnnnnnnng ttattgtact ctcaagattt attttt                 46

SEQ ID NO: 260         moltype = RNA  length = 91
FEATURE                Location/Qualifiers
misc_feature           1..91
                       note = Description of Artificial Sequence: Synthetic
                        oligonucleotide
source                 1..91
                       mol_type = other RNA
                       organism = synthetic construct
SEQUENCE: 260
gttacttaaa tcttgcagaa gctacaaaga taaggcttca tgccgaaatc aacaccctgt  60
cattttatgg cagggtgttt tcgttattta a                                 91

SEQ ID NO: 261         moltype = DNA  length = 70
FEATURE                Location/Qualifiers
source                 1..70
                       mol_type = unassigned DNA
                       organism = Homo sapiens
SEQUENCE: 261
ttttctagtg ctgagtttct gtgactcctc tacattctac ttctctgtgt ttctgtatac  60
tacctcctcc                                                         70

SEQ ID NO: 262         moltype = DNA  length = 122
FEATURE                Location/Qualifiers
source                 1..122
                       mol_type = unassigned DNA
                       organism = Homo sapiens
SEQUENCE: 262
ggaggaaggg cctgagtccg agcagaagaa gaagggctcc catcacatca accggtggcg  60
```

-continued

```
cattgccacg aagcaggcca atggggagga catcgatgtc acctccaatg actagggtgg  120
gc                                                                 122

SEQ ID NO: 263          moltype = RNA  length = 48
FEATURE                 Location/Qualifiers
misc_feature            1..48
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
misc_difference         3..32
                        note = a, c, u, g, unknown or other
source                  1..48
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 263
acnnnnnnnn nnnnnnnnnn nnnnnnnnnn nngttttaga gctatgct               48

SEQ ID NO: 264          moltype = DNA  length = 67
FEATURE                 Location/Qualifiers
misc_feature            1..67
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..67
                        mol_type = other DNA
                        organism = synthetic construct
misc_feature            1..23
                        note = RNA
misc_feature            24
                        note = DNA
misc_feature            25..67
                        note = RNA
SEQUENCE: 264
agcatagcaa gttaaaataa ggctagtccg ttatcaactt gaaaaagtgg caccgagtcg  60
gtgcttt                                                            67

SEQ ID NO: 265          moltype = RNA  length = 62
FEATURE                 Location/Qualifiers
misc_feature            1..62
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
misc_difference         1..20
                        note = a, c, u, g, unknown or other
source                  1..62
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 265
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc  60
cg                                                                 62

SEQ ID NO: 266          moltype = DNA  length = 73
FEATURE                 Location/Qualifiers
misc_feature            1..73
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..73
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 266
tgaatggtcc caaaacggaa gggcctgagt ccgagcagaa gaagaagttt tagagctatg  60
ctgttttgaa tgg                                                     73

SEQ ID NO: 267          moltype = DNA  length = 69
FEATURE                 Location/Qualifiers
source                  1..69
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 267
ctggtcttcc acctctctgc cctgaacacc caatctcggc ccctctcgcc accctcctgc  60
atttctgtt                                                          69

SEQ ID NO: 268          moltype = DNA  length = 138
FEATURE                 Location/Qualifiers
source                  1..138
                        mol_type = unassigned DNA
                        organism = Mus musculus
SEQUENCE: 268
acccaagcac tgagtgccat tagctaaatg catagggtac cacccacagg tgccaggggc  60
```

-continued

```
ctttcccaaa gttcccagcc ccttctccaa cctttcctgg cccagaggct ttcccatgtg  120
tgtggctgga ccctttga                                                138

SEQ ID NO: 269          moltype = DNA  length = 21
FEATURE                 Location/Qualifiers
misc_feature            1..21
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..21
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 269
aaaaccaccc ttctctctgg c                                             21

SEQ ID NO: 270          moltype = DNA  length = 21
FEATURE                 Location/Qualifiers
misc_feature            1..21
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..21
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 270
ggagattgga gacacggaga g                                             21

SEQ ID NO: 271          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 271
ctggaaagcc aatgcctgac                                               20

SEQ ID NO: 272          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 272
ggcagcaaac tccttgtcct                                               20

SEQ ID NO: 273          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 273
gtgctttgca gaggcctacc                                               20

SEQ ID NO: 274          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 274
cctggagcgc atgcagtagt                                               20

SEQ ID NO: 275          moltype = DNA  length = 22
FEATURE                 Location/Qualifiers
misc_feature            1..22
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..22
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 275
accttctgtg tttccaccat tc                                            22

SEQ ID NO: 276          moltype = DNA  length = 20
FEATURE                 Location/Qualifiers
misc_feature            1..20
```

-continued

```
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..20
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 276
ttggggagtg cacagacttc                                                20

SEQ ID NO: 277          moltype = DNA  length = 30
FEATURE                 Location/Qualifiers
misc_feature            1..30
                        note = Description of Artificial Sequence: Synthetic probe
source                  1..30
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 277
tagctctaaa acttcttctt ctgctcggac                                     30

SEQ ID NO: 278          moltype = DNA  length = 30
FEATURE                 Location/Qualifiers
misc_feature            1..30
                        note = Description of Artificial Sequence: Synthetic probe
source                  1..30
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 278
ctagccttat tttaacttgc tatgctgttt                                     30

SEQ ID NO: 279          moltype = RNA  length = 99
FEATURE                 Location/Qualifiers
misc_feature            1..99
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
misc_difference         1..20
                        note = a, c, u, g, unknown or other
source                  1..99
                        mol_type = other RNA
                        organism = synthetic construct
SEQUENCE: 279
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc  60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgcttt                           99

SEQ ID NO: 280          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 280
tagcgggtaa gc                                                        12

SEQ ID NO: 281          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 281
tcggtgacat gt                                                        12

SEQ ID NO: 282          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 282
actccccgta gg                                                        12

SEQ ID NO: 283          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 283
actgcgtgtt aa                                                        12

SEQ ID NO: 284          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
```

-continued

```
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 284
acgtcgcctg at                                                    12

SEQ ID NO: 285      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 285
taggtcgacc ag                                                    12

SEQ ID NO: 286      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 286
ggcgttaatg at                                                    12

SEQ ID NO: 287      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 287
tgtcgcatgt ta                                                    12

SEQ ID NO: 288      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 288
atggaaacgc at                                                    12

SEQ ID NO: 289      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 289
gccgaattcc tc                                                    12

SEQ ID NO: 290      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 290
gcatggtacg ga                                                    12

SEQ ID NO: 291      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 291
cggtactctt ac                                                    12

SEQ ID NO: 292      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 292
gcctgtgccg ta                                                    12

SEQ ID NO: 293      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 293
tacggtaagt cg                                                    12
```

-continued

```
SEQ ID NO: 294          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 294
cacgaaatta cc                                                      12

SEQ ID NO: 295          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 295
aaccaagata cg                                                      12

SEQ ID NO: 296          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 296
gagtcgatac gc                                                      12

SEQ ID NO: 297          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 297
gtctcacgat cg                                                      12

SEQ ID NO: 298          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 298
tcgtcgggtg ca                                                      12

SEQ ID NO: 299          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 299
actccgtagt ga                                                      12

SEQ ID NO: 300          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 300
caggacgtcc gt                                                      12

SEQ ID NO: 301          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 301
tcgtatccct ac                                                      12

SEQ ID NO: 302          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 302
tttcaaggcc gg                                                      12

SEQ ID NO: 303          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
```

-continued

```
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 303
cgccggtgga at                                                    12

SEQ ID NO: 304      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 304
gaacccgtcc ta                                                    12

SEQ ID NO: 305      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 305
gattcatcag cg                                                    12

SEQ ID NO: 306      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 306
acaccggtct tc                                                    12

SEQ ID NO: 307      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 307
atcgtgccct aa                                                    12

SEQ ID NO: 308      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 308
gcgtcaatgt tc                                                    12

SEQ ID NO: 309      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 309
ctccgtatct cg                                                    12

SEQ ID NO: 310      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 310
ccgattcctt cg                                                    12

SEQ ID NO: 311      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 311
tgcgcctcca gt                                                    12

SEQ ID NO: 312      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 312
taacgtcgga gc                                                    12
```

-continued

```
SEQ ID NO: 313          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 313
aaggtcgccc at                                                      12

SEQ ID NO: 314          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 314
gtcggggact at                                                      12

SEQ ID NO: 315          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 315
ttcgagcgat tt                                                      12

SEQ ID NO: 316          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 316
tgagtcgtcg ag                                                      12

SEQ ID NO: 317          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 317
tttacgcaga gg                                                      12

SEQ ID NO: 318          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 318
aggaagtatc gc                                                      12

SEQ ID NO: 319          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 319
actcgatacc at                                                      12

SEQ ID NO: 320          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 320
cgctacatag ca                                                      12

SEQ ID NO: 321          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 321
ttcataaccg gc                                                      12

SEQ ID NO: 322          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
```

-continued

```
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 322
ccaaacggtt aa                                                    12

SEQ ID NO: 323      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 323
cgattccttc gt                                                    12

SEQ ID NO: 324      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 324
cgtcatgaat aa                                                    12

SEQ ID NO: 325      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 325
agtggcgatg ac                                                    12

SEQ ID NO: 326      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 326
cccctacggc ac                                                    12

SEQ ID NO: 327      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 327
gccaacccgc ac                                                    12

SEQ ID NO: 328      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 328
tgggacaccg gt                                                    12

SEQ ID NO: 329      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 329
ttgactgcgg cg                                                    12

SEQ ID NO: 330      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 330
actatgcgta gg                                                    12

SEQ ID NO: 331      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 331
tcacccaaag cg                                                    12
```

-continued

```
SEQ ID NO: 332          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 332
gcaggacgtc cg                                                      12

SEQ ID NO: 333          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 333
acaccgaaaa cg                                                      12

SEQ ID NO: 334          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 334
cggtgtattg ag                                                      12

SEQ ID NO: 335          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 335
cacgaggtat gc                                                      12

SEQ ID NO: 336          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 336
taaagcgacc cg                                                      12

SEQ ID NO: 337          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 337
cttagtcggc ca                                                      12

SEQ ID NO: 338          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 338
cgaaaacgtg gc                                                      12

SEQ ID NO: 339          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 339
cgtgccctga ac                                                      12

SEQ ID NO: 340          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 340
tttaccatcg aa                                                      12

SEQ ID NO: 341          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
```

-continued

```
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 341
cgtagccatg tt                                                    12

SEQ ID NO: 342      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 342
cccaaacggt ta                                                    12

SEQ ID NO: 343      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 343
gcgttatcag aa                                                    12

SEQ ID NO: 344      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 344
tcgatggtaa ac                                                    12

SEQ ID NO: 345      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 345
cgactttttg ca                                                    12

SEQ ID NO: 346      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 346
tcgacgactc ac                                                    12

SEQ ID NO: 347      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 347
acgcgtcaga ta                                                    12

SEQ ID NO: 348      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 348
cgtacggcac ag                                                    12

SEQ ID NO: 349      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 349
ctatgccgtg ca                                                    12

SEQ ID NO: 350      moltype = DNA  length = 12
FEATURE             Location/Qualifiers
source              1..12
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 350
cgcgtcagat at                                                    12
```

-continued

```
SEQ ID NO: 351          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 351
aagatcggta gc                                                       12

SEQ ID NO: 352          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 352
cttcgcaagg ag                                                       12

SEQ ID NO: 353          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 353
gtcgtggact ac                                                       12

SEQ ID NO: 354          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 354
ggtcgtcatc aa                                                       12

SEQ ID NO: 355          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 355
gttaacagcg tg                                                       12

SEQ ID NO: 356          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 356
tagctaaccg tt                                                       12

SEQ ID NO: 357          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 357
agtaaaggcg ct                                                       12

SEQ ID NO: 358          moltype = DNA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 358
ggtaatttcg tg                                                       12

SEQ ID NO: 359          moltype = DNA  length = 15
FEATURE                 Location/Qualifiers
source                  1..15
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 359
cagaagaaga agggc                                                    15

SEQ ID NO: 360          moltype = DNA  length = 51
FEATURE                 Location/Qualifiers
source                  1..51
```

-continued

```
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 360
ccaatgggga ggacatcgat gtcacctcca atgactaggg tggtgggcaa c          51

SEQ ID NO: 361          moltype = DNA  length = 15
FEATURE                 Location/Qualifiers
source                  1..15
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 361
ctctggccac tccct                                                  15

SEQ ID NO: 362          moltype = DNA  length = 52
FEATURE                 Location/Qualifiers
source                  1..52
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 362
acatcgatgt cacctccaat gacaagcttg ctagcggtgg gcaaccacaa ac         52

SEQ ID NO: 363          moltype =   length =
SEQUENCE: 363
000

SEQ ID NO: 364          moltype =   length =
SEQUENCE: 364
000

SEQ ID NO: 365          moltype = DNA  length = 54
FEATURE                 Location/Qualifiers
misc_feature            1..54
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..54
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 365
aacaccgggt cttcgagaag acctgtttta gagctagaaa tagcaagtta aaat       54

SEQ ID NO: 366          moltype = DNA  length = 54
FEATURE                 Location/Qualifiers
misc_feature            1..54
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..54
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 366
caaaacgggt cttcgagaag acgttttaga gctatgctgt tttgaatggt ccca       54

SEQ ID NO: 367          moltype = DNA  length = 4104
FEATURE                 Location/Qualifiers
source                  1..4104
                        mol_type = unassigned DNA
                        organism = Homo sapiens
CDS                     1..4104
SEQUENCE: 367
atggacaaga agtacagcat cggcctggac atcggcacca actctgtggg ctgggccgtg  60
atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg  120
cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag  180
gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc  240
tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga  300
ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc  360
aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag  420
aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac  480
atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac  540
gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc  600
atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga  660
cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggcaac  720
ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag  780
gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc  840
cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc  900
ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct  960
atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg  1020
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc  1080
```

-continued

```
ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg  1140
gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg  1200
aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac  1260
gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc  1320
gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc  1380
agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa  1440
gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag  1500
aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg  1560
tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg  1620
agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc  1680
gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc  1740
tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt  1800
atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg  1860
ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc  1920
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc  1980
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg  2040
gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac  2100
agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg  2160
cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca  2220
gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg  2280
atcgccatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga  2340
atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc  2400
gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg  2460
gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggacgcc  2520
atcgtgcctc agagctttct gaaggacgac tccatcgacg ccaaggtgct gaccagaagc  2580
gacaaggccc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag  2640
aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg  2700
accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag  2760
ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac  2820
actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc  2880
aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac  2940
taccaccacg cccacgccgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag  3000
taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag  3060
atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc  3120
aacatcatga actttttcaa gaccgagatt accctggcca acggcgagat ccggaagcgg  3180
cctctgatcg agacaaacgg cgaaaccggg gagatcgtgt gggataaggg ccgggatttt  3240
gccaccgtgc ggaaagtgct gagcatgccc caagtgaata tcgtgaaaaa gaccgaggtg  3300
cagacaggcg gcttcagcaa agagtctatc ctgcccaaga ggaacagcga taagctgatc  3360
gccagaaaga aggactggga ccctaagaag tacggcggct tcgacagccc caccgtggcc  3420
tattctgtgc tggtggtggc caaagtggaa aagggcaagt ccaagaaact gaagagtgtg  3480
aaagagctgc tggggatcac catcatggaa agaagcagct tcgagaagaa tcccatcgac  3540
tttctggaag ccaagggcta caaagaagtg aaaaaggacc tgatcatcaa gctgcctaag  3600
tactccctgt tcgagctgga aaacggccgg aagagaatgc tggcctctgc cggcgaactg  3660
cagaagggaa acgaactggc cctgccctcc aaatatgtga acttcctgta cctggccagc  3720
cactatgaga agctgaaggg ctcccccgag gataatgagc agaaacagct gtttgtggaa  3780
cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttctc caagagagtg  3840
atcctggccg acgctaatct ggacaaagtg ctgtccgcct acaacaagca ccgggataag  3900
cccatcagag agcaggccga gaatatcatc cacctgttta ccctgaccaa tctgggagcc  3960
cctgccgcct tcaagtactt tgacaccacc atcgaccgga agaggtacac cagcaccaaa  4020
gaggtgctgg acgccaccct gatccaccag agcatcaccg gcctgtacga gacacggatc  4080
gacctgtctc agctgggagg cgac                                         4104
```

```
SEQ ID NO: 368          moltype = AA  length = 1368
FEATURE                 Location/Qualifiers
source                  1..1368
                        mol_type = protein
                        organism = Homo sapiens
SEQUENCE: 368
MDKKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE  60
ATRLKRTARR RYTRRKNRIC YLQEIFSNEM AKVDDSFFHR LEESFLVEED KKHERHPIFG  120
NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLNPDNSD  180
VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN  240
LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI  300
LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA  360
GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH  420
AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE  480
VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL  540
SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI  600
IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG  660
RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL  720
HEHIANLAGS PAIKKGILQT VKVVDELVKV MGRHKPENIV IAMARENQTT QKGQKNSRER  780
MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVDA  840
IVPQSFLKDD SIDAKVLTRS DKARGKSDNV PSEEVVKKMK NYWRQLLNAK LITQRKFDNL  900
TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS  960
KLVSDFRKDF QFYKVREINN YHHAHAAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK  1020
MIAKSEQEIG KATAKYFFYS NIMNFFKTEI TLANGEIRKR PLIETNGETG EIVWDKGRDF  1080
```

-continued

```
ATVRKVLSMP QVNIVKKTEV QTGGFSKESI LPKRNSDKLI ARKKDWDPKK YGGFDSPTVA  1140
YSVLVVAKVE KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDLIIKLPK  1200
YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE DNEQKQLFVE  1260
QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRDK PIREQAENII HLFTLTNLGA  1320
PAAFKYFDTT IDRKRYTSTK EVLDATLIHQ SITGLYETRI DLSQLGGD               1368

SEQ ID NO: 369          moltype = DNA  length = 26
FEATURE                 Location/Qualifiers
source                  1..26
                        mol_type = unassigned DNA
                        organism = Homo sapiens
SEQUENCE: 369
tagggttagg ggccccaggc cggggt                                       26

SEQ ID NO: 370          moltype = AA  length = 6
FEATURE                 Location/Qualifiers
REGION                  1..6
                        note = Description of Artificial Sequence: Synthetic 6xHis
                         tag
source                  1..6
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 370
HHHHHH                                                             6

SEQ ID NO: 371          moltype = DNA  length = 57
FEATURE                 Location/Qualifiers
source                  1..57
                        mol_type = unassigned DNA
                        organism = Homo sapiens
CDS                     1..57
SEQUENCE: 371
ggcaccatta aagaaaatat cattggtgtt tcctatgatg aatatagata cagaagc     57

SEQ ID NO: 372          moltype = AA  length = 19
FEATURE                 Location/Qualifiers
source                  1..19
                        mol_type = protein
                        organism = Homo sapiens
SEQUENCE: 372
GTIKENIIGV SYDEYRYRS                                               19

SEQ ID NO: 373          moltype = DNA  length = 48
FEATURE                 Location/Qualifiers
misc_feature            1..48
                        note = Description of Artificial Sequence: Synthetic
                         oligonucleotide
source                  1..48
                        mol_type = other DNA
                        organism = synthetic construct
CDS                     1..48
SEQUENCE: 373
attaaagaaa atatcattgg ctttgtttcc tatgatgaat atagatac               48

SEQ ID NO: 374          moltype = AA  length = 16
FEATURE                 Location/Qualifiers
REGION                  1..16
                        note = Description of Artificial Sequence: Synthetic peptide
source                  1..16
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 374
IKENIIGFVS YDEYRY                                                  16

SEQ ID NO: 375          moltype =   length =
SEQUENCE: 375
000

SEQ ID NO: 376          moltype =   length =
SEQUENCE: 376
000

SEQ ID NO: 377          moltype =   length =
SEQUENCE: 377
000
```

-continued

```
SEQ ID NO: 378          moltype =   length =
SEQUENCE: 378
000

SEQ ID NO: 379          moltype =   length =
SEQUENCE: 379
000

SEQ ID NO: 380          moltype =   length =
SEQUENCE: 380
000

SEQ ID NO: 381          moltype =   length =
SEQUENCE: 381
000

SEQ ID NO: 382          moltype =   length =
SEQUENCE: 382
000

SEQ ID NO: 383          moltype =   length =
SEQUENCE: 383
000

SEQ ID NO: 384          moltype =   length =
SEQUENCE: 384
000

SEQ ID NO: 385          moltype =   length =
SEQUENCE: 385
000

SEQ ID NO: 386          moltype =   length =
SEQUENCE: 386
000

SEQ ID NO: 387          moltype =   length =
SEQUENCE: 387
000

SEQ ID NO: 388          moltype =   length =
SEQUENCE: 388
000

SEQ ID NO: 389          moltype =   length =
SEQUENCE: 389
000

SEQ ID NO: 390          moltype =   length =
SEQUENCE: 390
000

SEQ ID NO: 391          moltype =   length =
SEQUENCE: 391
000

SEQ ID NO: 392          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 392
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aactcacatc aaccggtggc gcaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 393          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 393
```

-continued

```
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aactcacatc aaccggtggc gcaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 394          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 394
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgaggaca aagtacaaac ggcggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 395          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 395
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgaggaca aagtacaaac ggcggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 396          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 396
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgtggcgc attgccacga agcggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 397          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 397
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aaccgagggc agagtgctgc ttgggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 398          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 398
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgagtccg agcagaagaa gaaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 399          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 399
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgaggaca aagtacaaac ggcggtgttt cgtcctttcc  120
acaag                                                              125
```

-continued

```
SEQ ID NO: 400          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 400
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacagcagaa gaagaagggc tccggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 401          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 401
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aactcacatc aaccggtggc gcaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 402          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 402
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacccctggc ccaggtgaag gtgggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 403          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 403
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aactccctcc ctggcccagg tgaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 404          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 404
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgaaccgg aggacaaagt acaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 405          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 405
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacaggtgaa ggtgtggttc cagggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 406          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
```

-continued

```
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 406
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacggtgaag gtgtggttcc agaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 407          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 407
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgaaccgg aggacaaagt acaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 408          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 408
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacccctggc ccaggtgaag gtgggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 409          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 409
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacaggtgaa ggtgtggttc cagggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 410          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 410
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgaggaca aagtacaaac ggcggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 411          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 411
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgggaggg aggggcacag atgggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 412          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 412
```

-continued

```
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aaccaccttc acctgggcca gggggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 413          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 413
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacaccctag tcattggagg tgaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 414          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 414
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aaccagagca gccactgggg cctggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 415          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 415
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aaccaccttc acctgggcca gggggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 416          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 416
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacccccatt ggcctgcttc gtgggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 417          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 417
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacattggcc tgcttcgtgg caaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 418          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 418
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aactcctcct ccagcttctg ccgggtgttt cgtcctttcc  120
acaag                                                              125
```

-continued

```
SEQ ID NO: 419          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 419
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aaccctccag cttctgccgt ttgggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 420          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 420
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacattggcc tgcttcgtgg caaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 421          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 421
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgcagcaa gcagcactct gccggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 422          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 422
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacttcttct tctgctcgga ctcggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 423          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 423
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacaccggag gacaaagtac aaaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 424          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 424
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aactcttctt ctgctcggac tcaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 425          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
```

-continued

```
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 425
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgttgatg tgatgggagc cctggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 426          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 426
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgggccag ggagggaggg gcaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 427          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 427
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgggaggg aggggcacag atgggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 428          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 428
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacccggttc tggaaccaca cctggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 429          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 429
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aactcacctg ggccagggag ggaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 430          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 430
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aactcacctg ggccagggag ggaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 431          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 431
```

-continued

```
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgttctgg aaccacacct tcaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 432          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 432
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgggaggg aggggcacag atgggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 433          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 433
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgggccag ggagggaggg gcaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 434          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 434
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgttctgg aaccacacct tcaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 435          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 435
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacaggtgaa ggtgtggttc cagggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 436          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 436
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgaaccgg aggacaaagt acaggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 437          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 437
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aaccaaaccc acgagggcag agtggtgttt cgtcctttcc  120
acaag                                                              125
```

-continued

```
SEQ ID NO: 438          moltype = DNA  length = 125
FEATURE                 Location/Qualifiers
misc_feature            1..125
                        note = Description of Artificial Sequence: Synthetic primer
source                  1..125
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 438
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa  60
cttgctattt ctagctctaa aacgagtttc tcatctgtgc cccggtgttt cgtcctttcc  120
acaag                                                              125

SEQ ID NO: 439          moltype = DNA  length = 684
FEATURE                 Location/Qualifiers
source                  1..684
                        mol_type = unassigned DNA
                        organism = Homo sapiens
CDS                     1..123
CDS                     127..159
CDS                     163..399
CDS                     403..684
SEQUENCE: 439
aaaaccaccc ttctctctgg cccactgtgt cctcttcctg ccctgccatc cccttctgtg  60
aatgttagac ccatgggagc agctggtcag aggggacccc ggcctggggc ccctaaccct  120
atgtagcctc agtcttccca tcaggctctc agctcagcct gagtgttgag gccccagtgg  180
ctgctctggg ggcctcctga gtttctcatc tgtgcccctc cctccctggc ccaggtgaag  240
gtgtggttcc agaaccggag gacaaagtac aaacggcaga agctggagga ggaagggcct  300
gagtccgagc agaagaagaa gggctcccat cacatcaacc ggtggcgcat tgccacgaag  360
caggccaatg gggaggacat cgatgtcacc tccaatgact agggtgggca accacaaacc  420
cacgagggca gagtgctgct tgctgctggc caggcccctg cgtgggccca agctggactc  480
tggccactcc ctggccaggc tttggggagg cctggagtca tggccccaca gggcttgaag  540
cccggggccg ccattgacag agggacaagc aatgggctgg ctgaggcctg ggaccacttg  600
gccttctcct cggagagcct gcctgcctgg gcgggcccgc ccgccaccgc agcctcccag  660
ctgctctccg tgtctccaat ctcc                                         684

SEQ ID NO: 440          moltype = AA  length = 41
FEATURE                 Location/Qualifiers
source                  1..41
                        mol_type = protein
                        organism = Homo sapiens
SEQUENCE: 440
KTTLLSGPLC PLPALPSPSV NVRPMGAAGQ RGPRPGAPNP M                      41

SEQ ID NO: 441          moltype = AA  length = 11
FEATURE                 Location/Qualifiers
source                  1..11
                        mol_type = protein
                        organism = Homo sapiens
SEQUENCE: 441
PQSSHQALSS A                                                       11

SEQ ID NO: 442          moltype = AA  length = 79
FEATURE                 Location/Qualifiers
source                  1..79
                        mol_type = protein
                        organism = Homo sapiens
SEQUENCE: 442
VLRPQWLLWG PPEFLICAPP SLAQVKVWFQ NRRTKYKRQK LEEEGPESEQ KKKGSHHINR  60
WRIATKQANG EDIDVTSND                                               79

SEQ ID NO: 443          moltype = AA  length = 94
FEATURE                 Location/Qualifiers
source                  1..94
                        mol_type = protein
                        organism = Homo sapiens
SEQUENCE: 443
GGQPQTHEGR VLLAAGQAPA WAQAGLWPLP GQALGRPGVM APQGLKPGAA IDRGTSNGLA  60
EAWDHLAFSS ESLPAWAGPP ATAASQLLSV SPIS                              94
```

1. (canceled)

2. A method for editing a mammalian subject in vivo, comprising delivering an engineered CRISPR-Cas system into the mammalian subject, wherein the engineered CRISPR-Cas system comprises:

(a) a Cas9 protein or a polynucleotide encoding the Cas9 protein, wherein the Cas9 protein is *S. pyogenes* Cas9 and is fused with two or more nuclear localization signals (NLSs);

(b) a CRISPR-Cas system chimeric RNA or a polynucleotide encoding the chimeric RNA, wherein the chime-

ric RNA comprises a guide sequence capable of hybridizing to a target sequence adjacent to a protospacer adjacent motif (PAM) in a genomic locus of interests of a mammalian cell, a tracr-mate sequence capable of hybridizing to a tracr sequence, and a tracr sequence comprising 40 or more nucleotides in length;

wherein a CRISPR complex comprising the Cas9 protein and the chimeric RNA is formed in the mammalian cell in vivo, and wherein the guide sequence directs sequence-specific binding of the CRISPR complex to the target sequence adjacent to the PAM in the genomic locus of interests of the mammalian cell.

3. The method of claim **2**, wherein the PAM is NGG.

4. The method of claim **2**, wherein the tracr sequence comprises 50 or more nucleotides in length.

5. The method of claim **2**, wherein the chimeric RNA comprises

NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAUAGCAAGUUAAAAU

AAGGCUAGUCCGUUAUCA.

6. The method of claim **2**, wherein the chimeric RNA further comprises a poly-U sequence.

7. The method of claim **2**, wherein the chimeric RNA is encoded by SEQ ID NO:86.

8. The method of claim **2**, wherein the chimeric RNA comprises one or more modified nucleotides.

9. The method of claim **2**, wherein the chimeric RNA comprises one or more methylated nucleotides or nucleotide analogs.

10. The method of claim **2**, wherein the two or more NLSs are independently selected from the group consisting of PKKKRKV, KRPAATKKAGQAKKKK, PAAKRVKLD, RQRRNELKRSP, NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY, RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV, VSRKRPRP, PPKKARED, PQPKKKPL, SALIKKKKKMAP, DRLRR, PKQKKRK, RKLKKKIKKL, REKKKFLKRR, KRKGDEVDGVDEVAKKKSKK, and RKCLQAGMNLEARKTKK.

11. The method of claim **10**, wherein at least one of the NLSs comprises PKKKRKV.

12. The method of claim **2**, wherein the Cas9 protein comprises a mutation of D10A, H840A, N854A, or N863A.

13. The method of claim **12**, wherein the Cas9 protein is fused with at least one heterologous protein domain.

14. The method of claim **13**, wherein the heterologous protein domain is selected from the group consisting of epitope tags, reporter 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, or nucleic acid binding activity.

15. The method of claim **2**, wherein the polynucleotide encoding the Cas9 protein is codon-optimized for expression in the mammalian cell.

16. The method of claim **2**, wherein the polynucleotide encoding the Cas9 protein comprises a polyadenylation signal.

17. The method of claim **2**, wherein the CRISPR-Cas system is comprised in a liposome for delivery.

18. The method of claim **2**, further comprising delivering into the mammalian cell an exogenous polynucleotide for targeted integration into a DNA break introduced by the CRISPR complex.

19. The method of claim **2**, wherein the CRISPR complex cleaves the genomic locus of interest, and wherein the mammalian cell is modified in vivo with an insertion, deletion, or substitution of one or more nucleotides in the genomic locus of interest.

20. A method for editing a human subject in vivo, comprising delivering an engineered CRISPR-Cas system into the human subject, wherein the engineered CRISPR-Cas system comprises:

(a) a Cas9 protein or a polynucleotide encoding the Cas9 protein, wherein the Cas9 protein is *S. pyogenes* Cas9 and is fused with two or more nuclear localization signals (NLSs) independently selected from the group consisting of PKKKRKV, KRPAATKKAGQAKKKK, PAAKRVKLD, RQRRNELKRSP, NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY, RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV, VSRKRPRP, PPKKARED, PQPKKKPL, SALIKKKKKMAP, DRLRR, PKQKKRK, RKLKKKIKKL, REKKKFLKRR, KRKGDEVDGVDEVAKKKSKK, and RKCLQAGMNLEARKTKK;

(b) a CRISPR-Cas system chimeric RNA or a polynucleotide encoding the chimeric RNA, wherein the chimeric RNA comprises NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCA, wherein NNNNNNNNNNNNNNNNNNNN is a guide sequence capable of hybridizing to a target sequence adjacent to a protospacer adjacent motif (PAM) in a genomic locus of interests of a human cell;

wherein a CRISPR complex comprising the Cas9 protein and the chimeric RNA is formed in the human cell in vivo, and wherein the guide sequence directs sequence-specific binding of the CRISPR complex to the target sequence adjacent to the PAM in the genomic locus of interests of the human cell.

21. A method for editing a human subject in vivo, comprising delivering an engineered CRISPR-Cas system into the human subject, wherein the engineered CRISPR-Cas system comprises:

(a) a polynucleotide encoding a Cas9 protein, wherein the Cas9 protein is *S. pyogenes* Cas9 and is fused with two or more nuclear localization signals (NLSs) independently selected from the group consisting of PKKKRKV, KRPAATKKAGQAKKKK, PAAKRVKLD, RQRRNELKRSP, NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY, RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV, VSRKRPRP, PPKKARED, PQPKKKPL, SALIKKKKKMAP, DRLRR, PKQKKRK, RKLKKKIKKL, REKKKFLKRR, KRKGDEVDGVDEVAKKKSKK, and RKCLQAGMNLEARKTKK, wherein the polynucleotide is codon-optimized for expression in the mammalian cell and comprises a polyadenylation signal;

(b) a CRISPR-Cas system chimeric RNA comprising NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG CUAGU-

CCGUUAUCA, wherein NNNNNNNNNNNNNNNNNNNN is a guide sequence capable of hybridizing to a target sequence adjacent to a protospacer adjacent motif (PAM) in a genomic locus of interests of a human cell;

wherein a CRISPR complex comprising the Cas9 protein and the chimeric RNA is formed in the human cell in vivo, wherein the guide sequence directs sequence-specific binding of the CRISPR complex to the target sequence adjacent to the PAM in the genomic locus of interests, wherein the CRISPR complex cleaves the genomic locus of interest in the human cell.

* * * * *