



US 20250250578A1

(19) United States

(12) Patent Application Publication
Zhang(10) Pub. No.: US 2025/0250578 A1
(43) Pub. Date: Aug. 7, 2025(54) CRISPR-CAS COMPONENT SYSTEMS,
METHODS AND COMPOSITIONS FOR
SEQUENCE MANIPULATION(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.: 19/028,841

(22) Filed: Jan. 17, 2025

Related U.S. Application Data(63) Continuation of application No. 17/034,754, filed on
Sep. 28, 2020, which is a continuation of application
No. 15/967,495, filed on Apr. 30, 2018, which is a
continuation of application No. 14/105,035, 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.
61/768,959, filed on Feb. 25, 2013, provisional ap-
plication No. 61/791,409, filed on Mar. 15, 2013,
provisional application No. 61/835,931, filed on Jun.
17, 2013.**Publication Classification**

(51) Int. Cl.

CI2N 15/74	(2006.01)
CI2N 9/22	(2006.01)
CI2N 15/10	(2006.01)
CI2N 15/113	(2010.01)
CI2N 15/63	(2006.01)

CI2N 15/70 (2006.01)

CI2N 15/79 (2006.01)

CI2N 15/85 (2006.01)

CI2N 15/90 (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	CI2N 15/746 (2013.01); CI2N 9/22 (2013.01); CI2N 15/113 (2013.01); CI2N 15/63 (2013.01); CI2N 15/70 (2013.01); CI2N 15/74 (2013.01); CI2N 15/85 (2013.01); CI2N 15/8509 (2013.01); CI2N 15/907 (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); CI2N 15/102 (2013.01); CI2N 15/1082 (2013.01); CI2N 15/79 (2013.01); CI2N 2310/10 (2013.01); CI2N 2310/20 (2017.05); CI2N 2310/3519 (2013.01); CI2N 2310/531 (2013.01); CI2N 2320/11 (2013.01); CI2N 2320/30 (2013.01); CI2N 2750/14143 (2013.01); CI2N 2800/101 (2013.01); G16B 30/00 (2019.02)
-----------	---

ABSTRACT

The invention provides for systems, methods, and compositions for manipulation of sequences and/or activities of target sequences. Provided are vectors and vector systems, some of which encode one or more components of a CRISPR complex, as well as methods for the design and use of such vectors. Also provided are methods of directing CRISPR complex formation in eukaryotic cells and methods for selecting specific cells by introducing precise mutations utilizing the CRISPR/Cas system.

Specification includes a Sequence Listing.

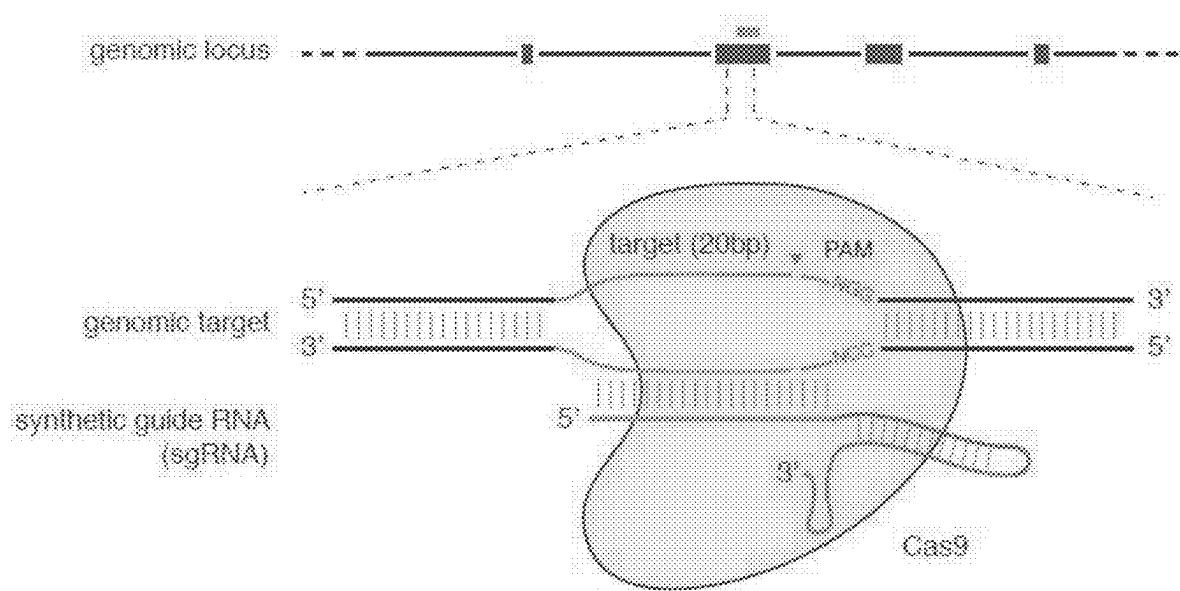


FIG. 1

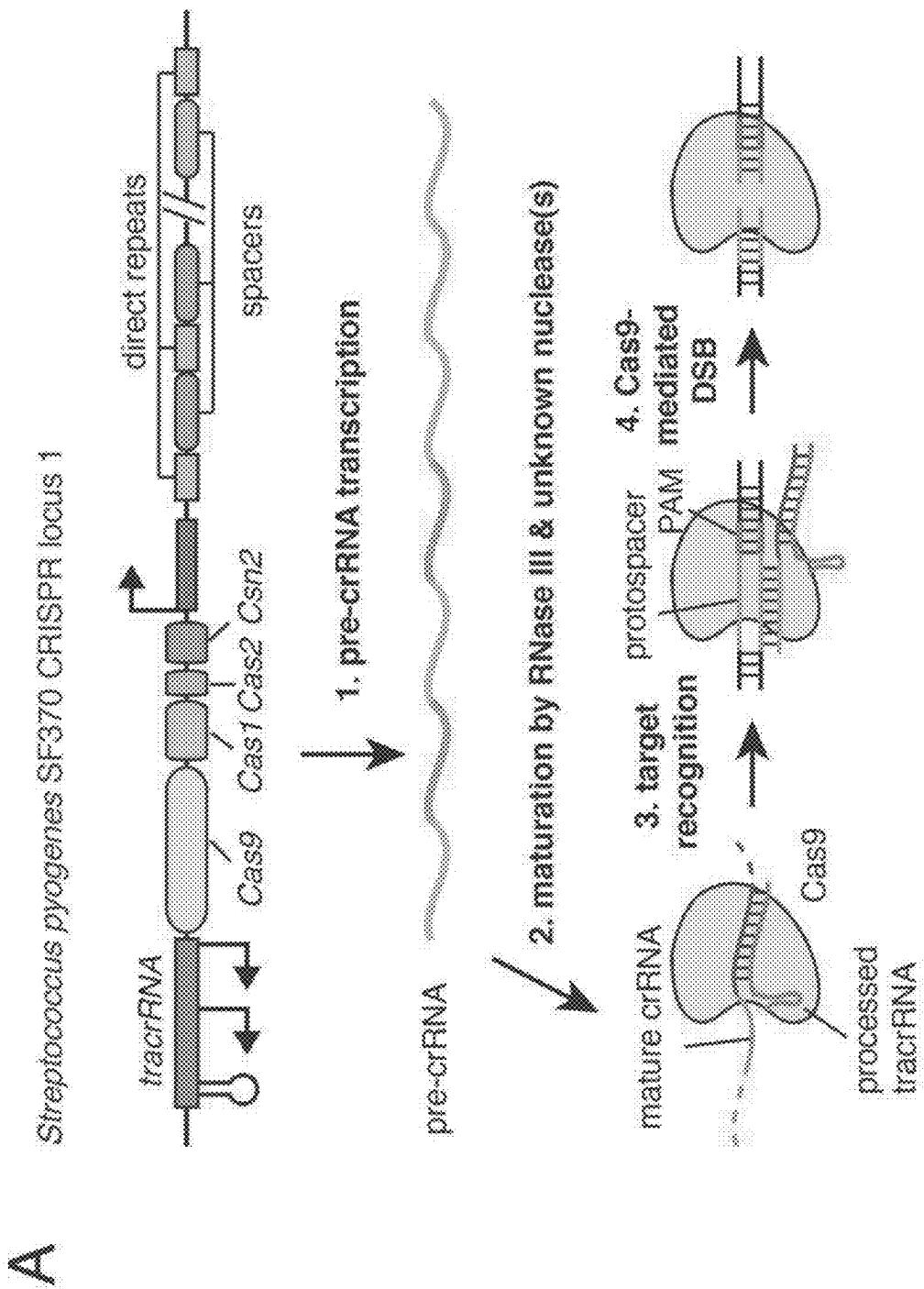


FIG. 2A

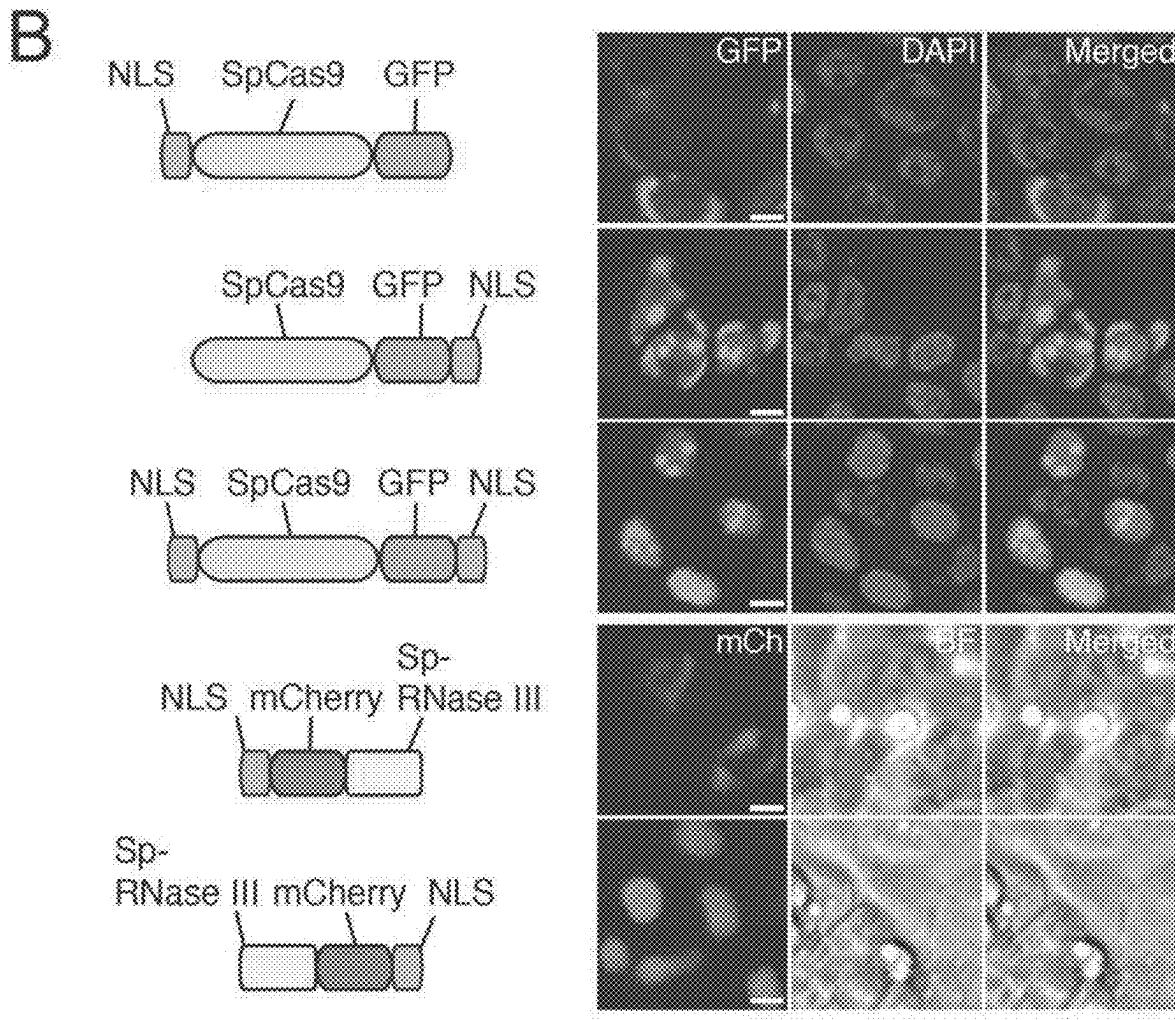


FIG. 2B

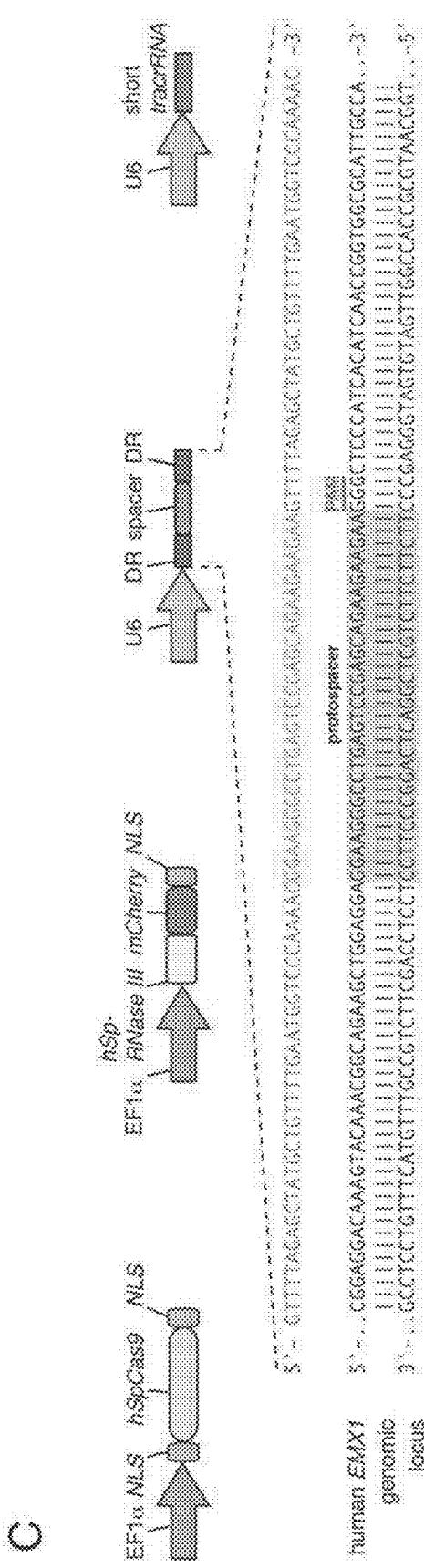


FIG. 2C

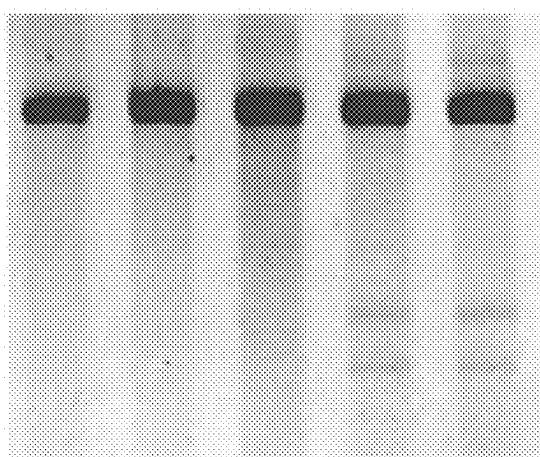
D

2xNLS-SpCas9	+	+	+	+	+
SpRNase III	-	+	+	-	+
short tracrRNA	-	+	-	+	+
DR-EMX1-DR	+	-	+	+	+

684bp ►

367bp ►

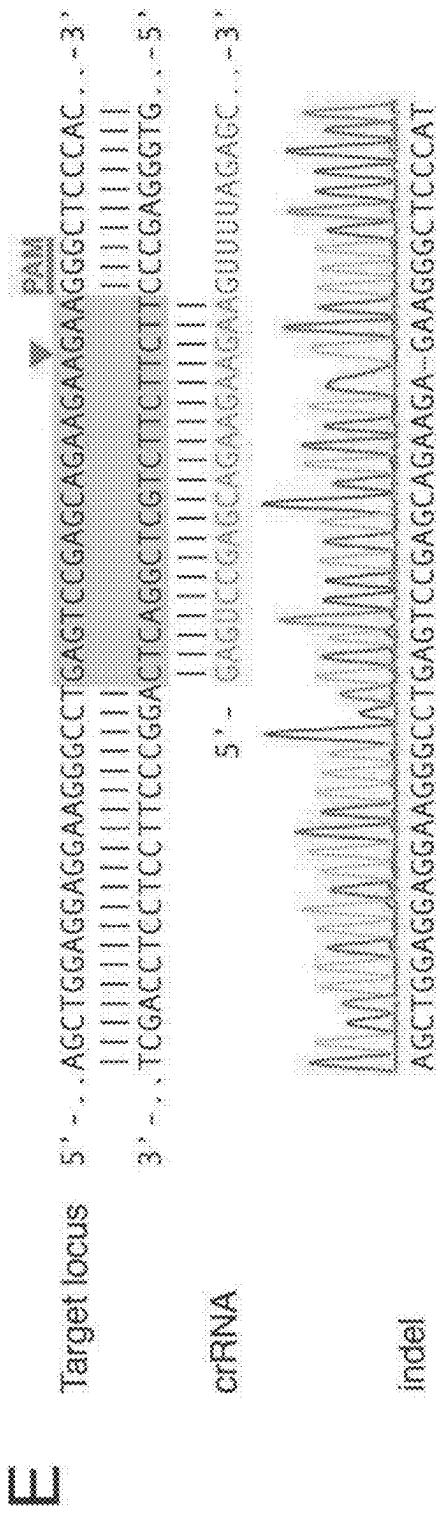
317bp ►



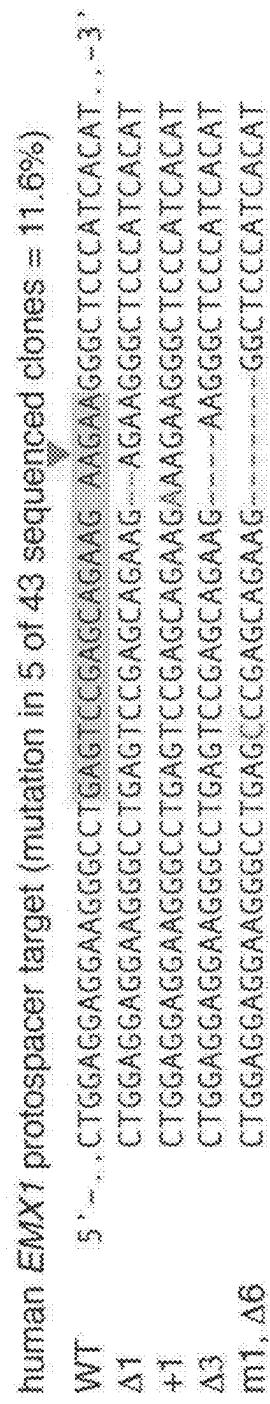
indel (%):

4.7 5.0

FIG. 2D



EIG. 2 E



1

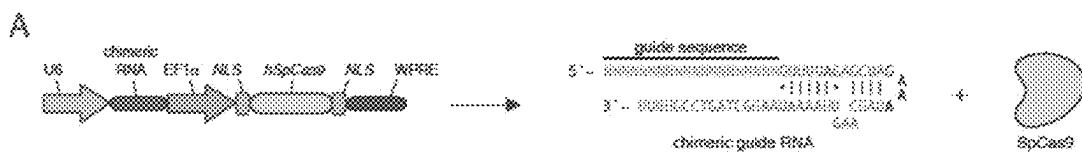


FIG. 3A

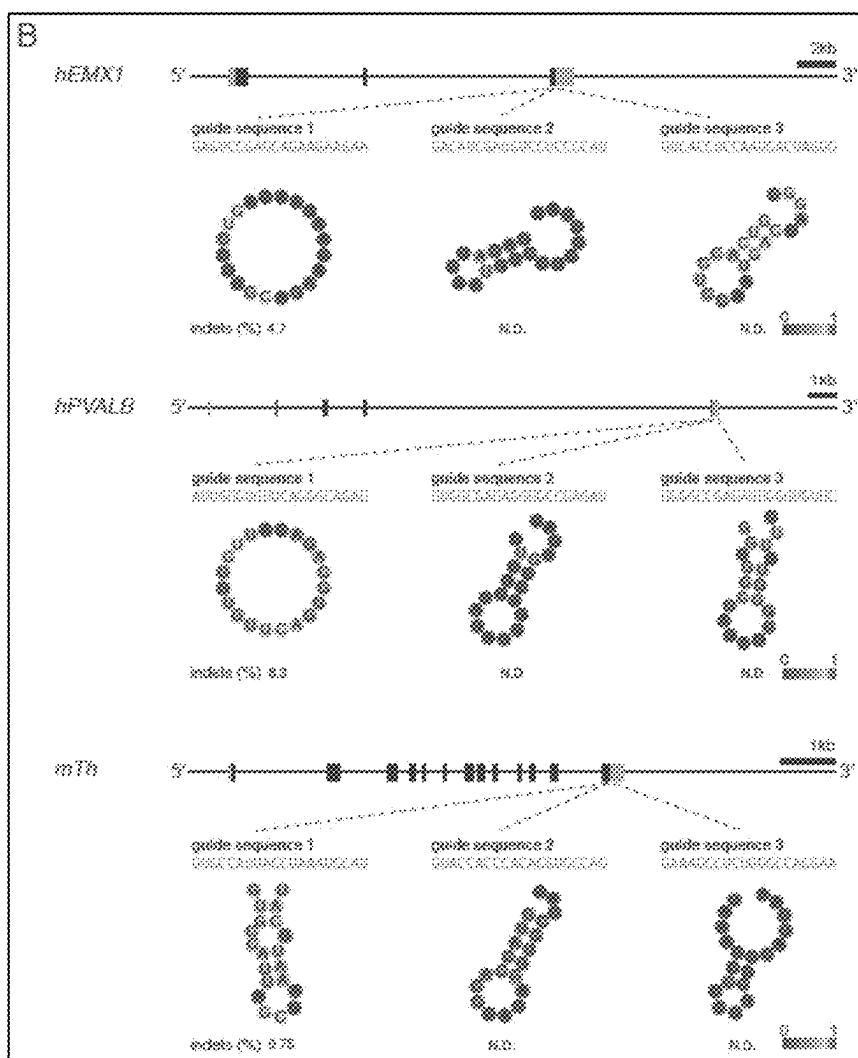


FIG. 3B

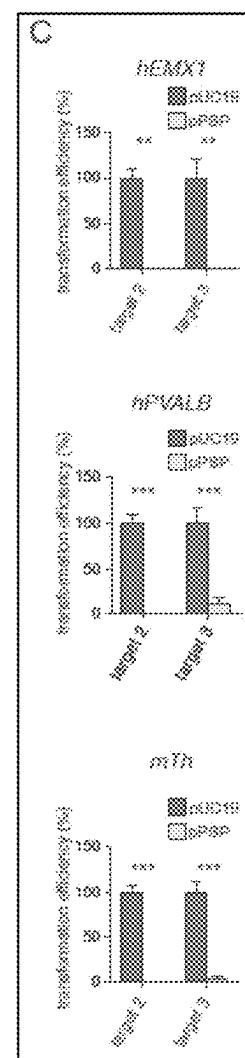


FIG. 3C

A

	protospacer (1) ← PAM
human <i>EMX1</i> locus	5' ~ . CTGGAGGAGGAAGGGCCTGAGTCGGAGCAGAAGAAGAAGGGCTCCAT. . ~ 3' 3' - . GACCTCCTCCCTTCCC GGACTCAGGTTCGTCTTCTTCCTTCCCCGAGGSGTA. . ~ 5'
wt crRNA	5' ~ . GAGUCCGAGCAGAAGAAGAAGAUUUUAGAGC. . ~ 3' m1 GAGUCCGAGCAGAAGAAGAAGAU m2 GAGUCCGAGCAGAAGAAGUA m3 GAGUCCGAGCAGAAGAACAA m4 GAGUCCGAGCAGAAGAAGAA m5 GAGUCCGAGCAGAAGUAGAA m7 GAGUCCGAGCAGAUGAAGAA m9 GAGUCCGAGCACAAGAAGAA m11 GAGUCCGAGGAGAAGAAGAA m13 GAGUCCGUGCAGAAGAAGAA m15 GAGUCCGAGCAGAAGAAGAA m17 GAGACCGAGCAGAAGAAGAA
mismatch-containing guide sequences	

FIG. 4A

B

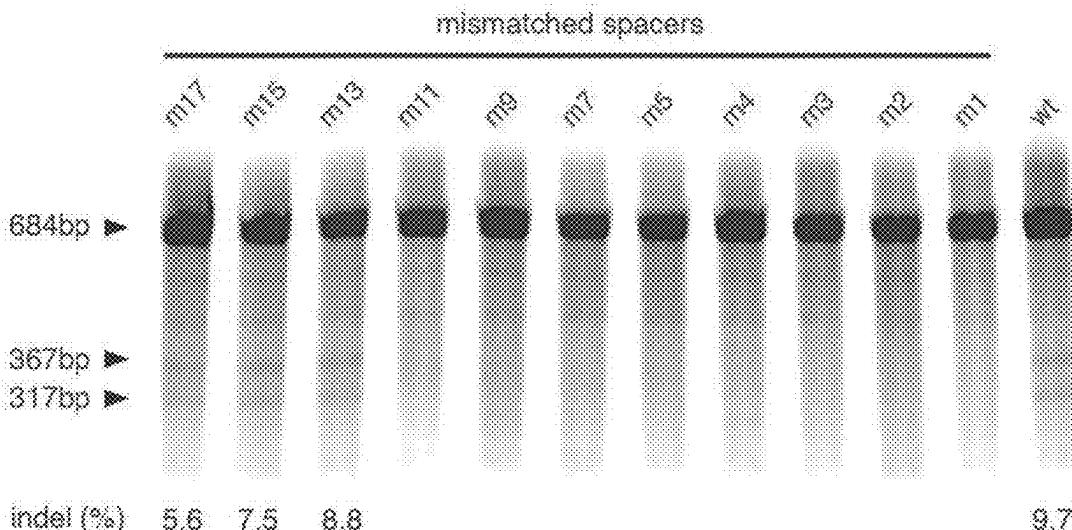


FIG. 4B

6

The diagram illustrates the *EMX1* locus with two TALEN binding sites. The top row shows the **left TALEN binding site** (protospacer 1) and the **PAM** sequence (underlined). The bottom row shows the **right TALEN binding site**. The protospacer sequences are **CTGGAGGAGGAAGGGCTGAGTCCGAGCAGAAGAAGAAGGGCTCCAT** and **GACCTCCCTCTTCCTGGACTCAGGCTCGTCTTCCTTCCTGGAGGGTA**. The PAM sequences are **T** and **T**. The TALEN binding sites are indicated by vertical bars above the DNA sequence.

FIG. 4C

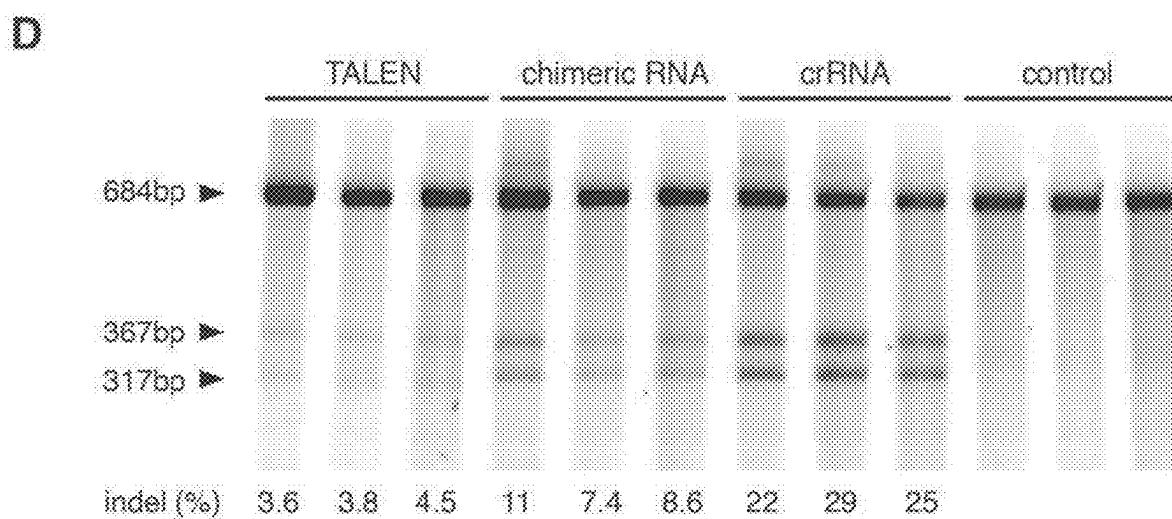
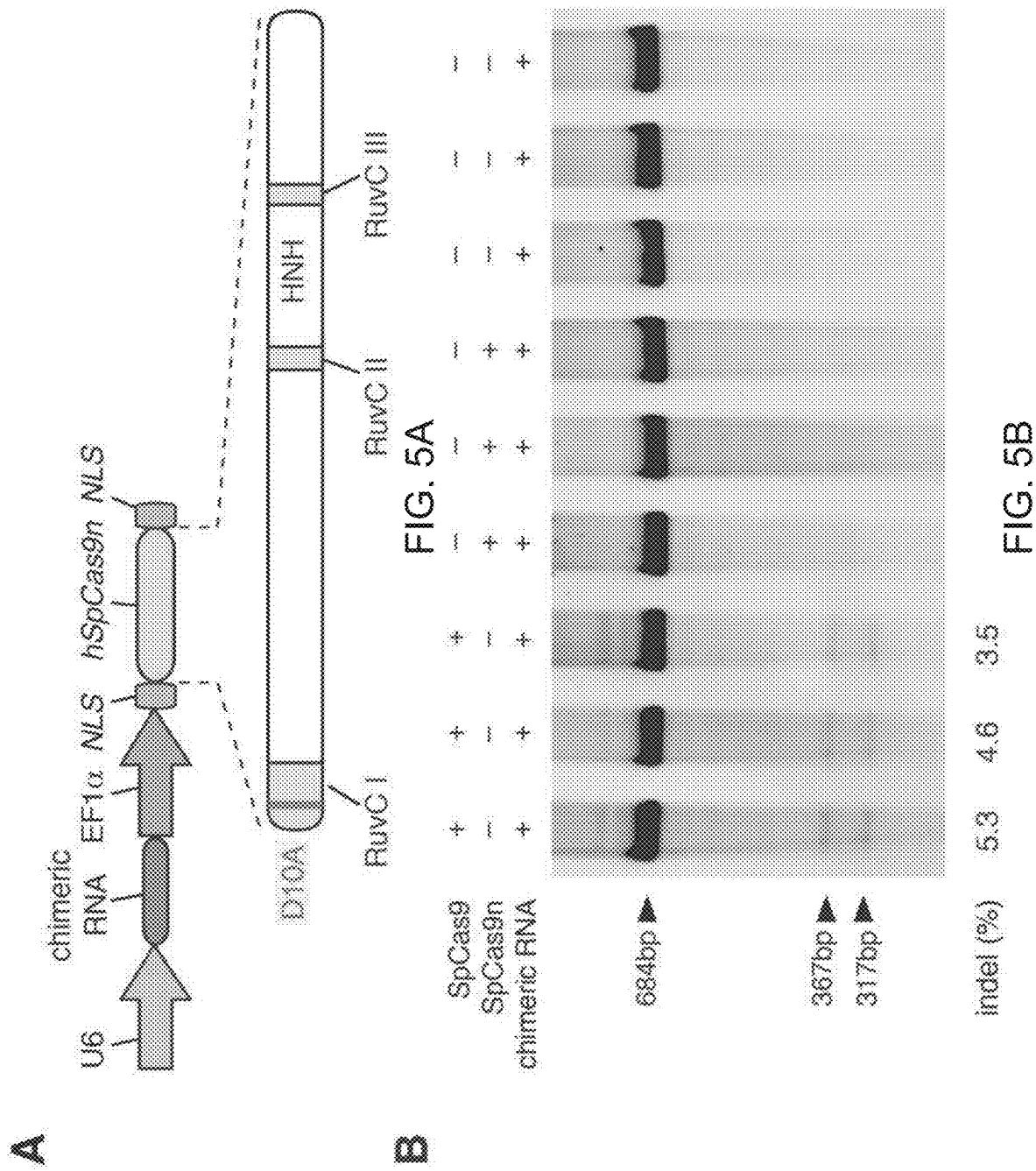


FIG. 4D



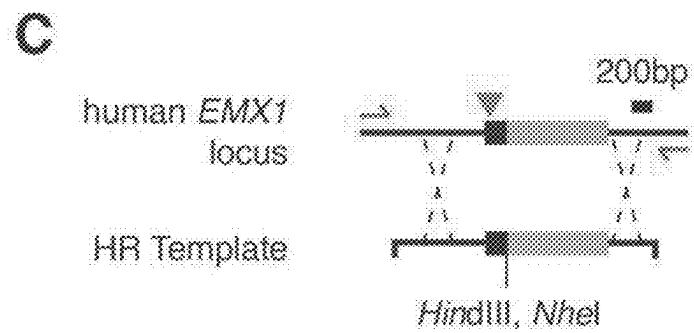


FIG. 5C

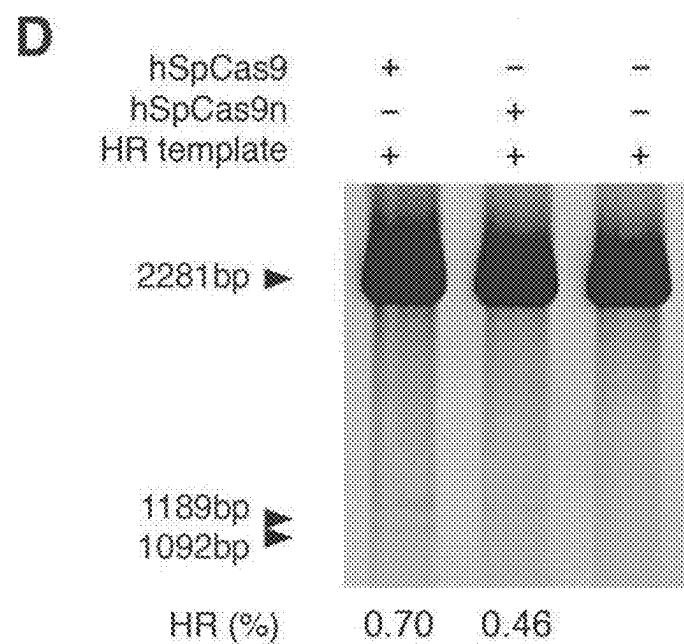


FIG. 5D

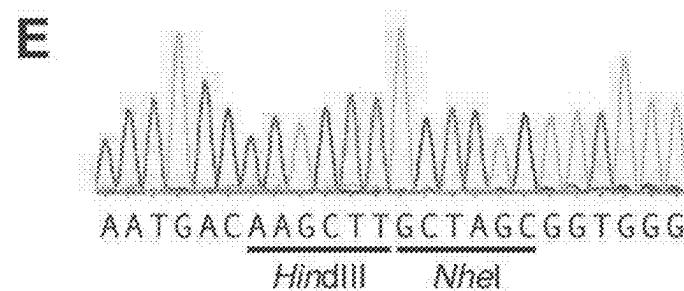
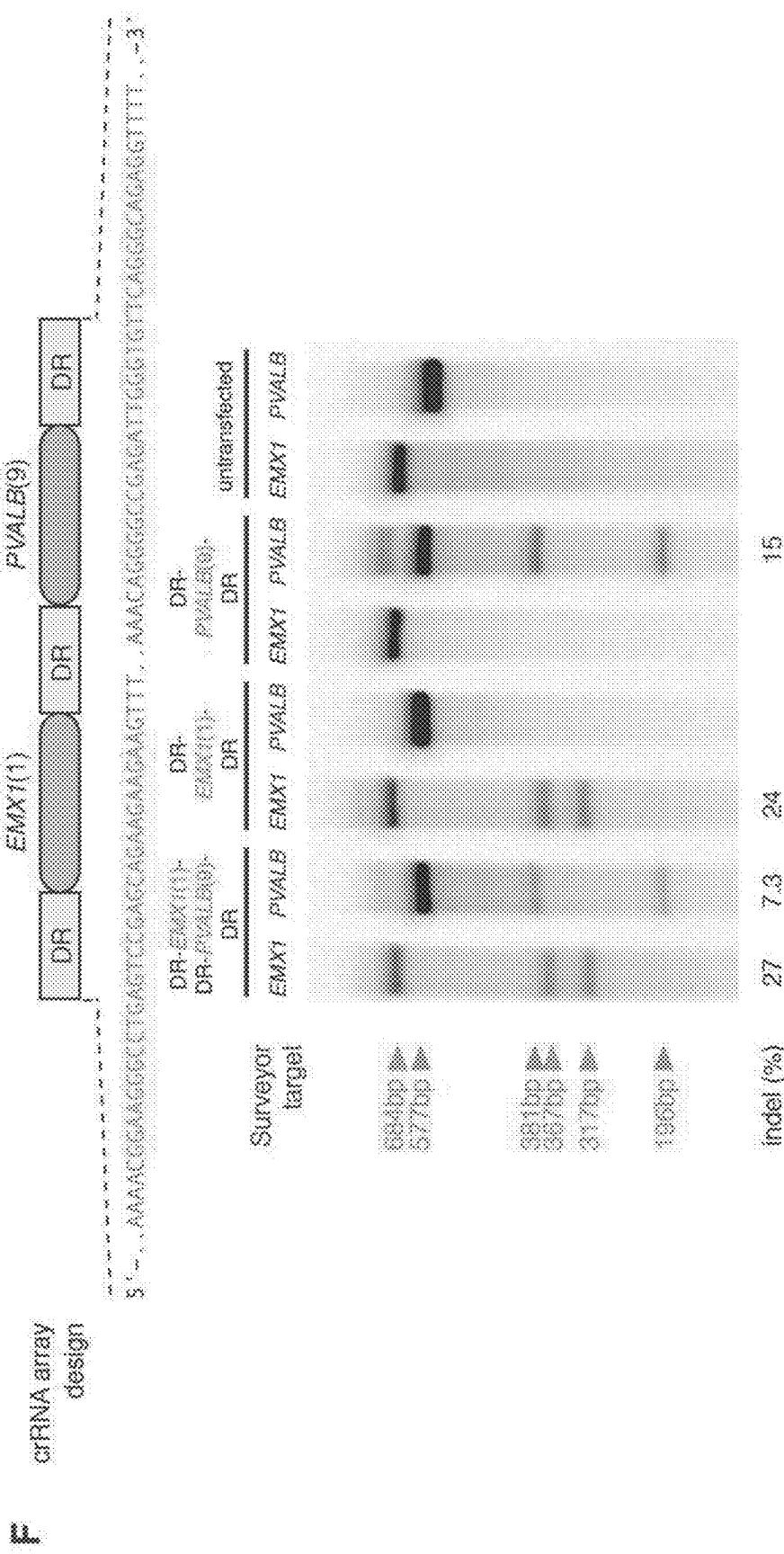
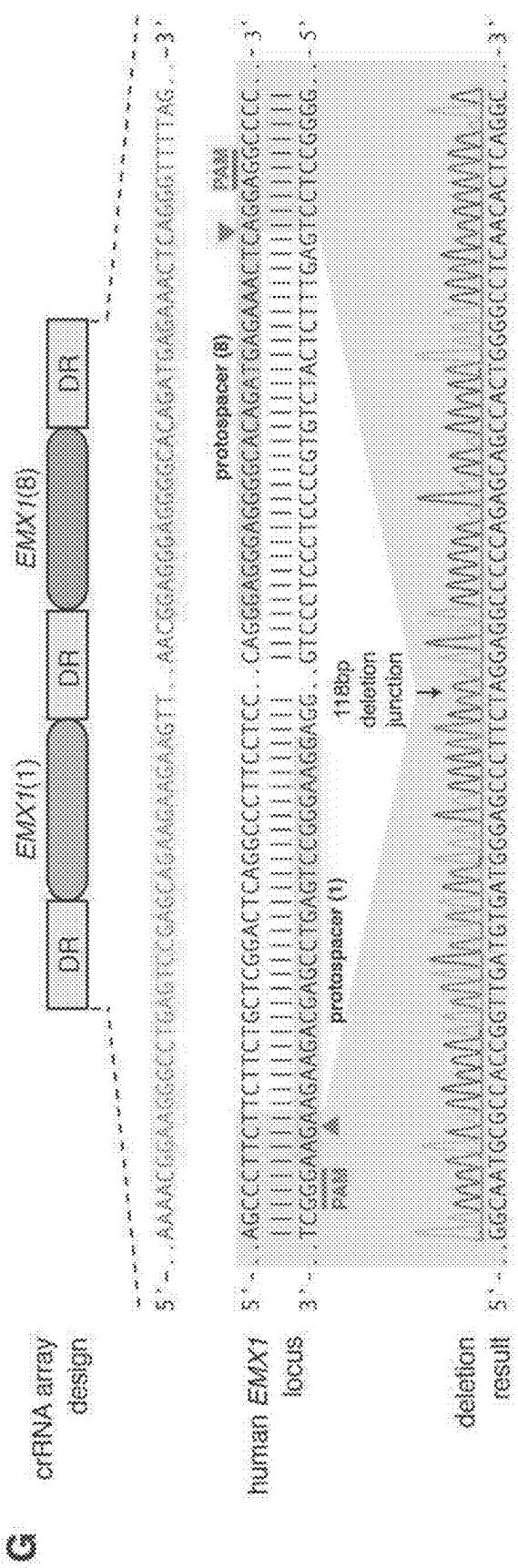


FIG. 5E





Seq #	target species	gene	probeseq	predicted sequence G' to 3'	PAM	strand tested	cell line	% total (pre-mRNA + intr-mRNA) (estimated RNA)	% base
EBV	1	CATGCCCTTCTTCAGGAGAGAGA	GGG	*	238T	20±18	67±02		
Sar	2	CATTGGCTTCAACGATATCCCTCCAT	TGG	*	238T	21±03	ND		
Homo sapiens	3	GGACATGAACTGCTCAAGTGAGGG	TGG	*	238T	14±11	ND		
	4	CATCCATGCTCTTCAATGGCTCTCG	TGG	*	238T	11±17	ND		
	5	TTCGTGTTAATGGGAACTCTTAATA	TGG	*	238T	43±04	21±05		
	6	TCTTCGAAATCCGAGCCCTTCATCAT	GCG	*	238T	40±08	041±025		
	7	TCAACCTCTCCGCTTCTACTTTCCTC	GGG	*	238T	15±01	ND		
	8	GGACGAGGGCCACAUATGAGAACTGG	AGG	*	238T	7.8±0.8	23±12		
Scn type II									
CherryR	9	AACCCCTCAGATTGGTCTTCAGCACAC	AGG	*	238T	21±26	65±020		
Homo sapiens	10	AAGCAGGCTGAGAACGGGCGAG	TGG	*	238T	ND	ND		
	11	GTCGGCGAGGCTGGCAATGGTTTC	AGG	*	238T	ND	ND		
Scn type II									
Mus	12	GAAGCAGTACTGCTTCAGTAAAT	AGG	*	Neuro2A	27±43	4.1±22		
Mus	13	ATTCGATGGCTTCCCAGGCGCCAG	GCG	*	Neuro2A	4.8±1.2	ND		
Mus	14	ACACAGTGGAGGCTCTGGCGACAA	AGG	*	Neuro2A	11.3±1.3	ND		
S. thermophilus									
LacI ChIP&RT	15	GGCGGAGTGTATAGAGGAGGAGA	GCGAA	*	238T	14±0.8	ND		
	16	AGACATAGGAGCTAGAGAGAGAGA	GAGAA	*	238T	7.8±0.7	ND		

FIG. 6

A

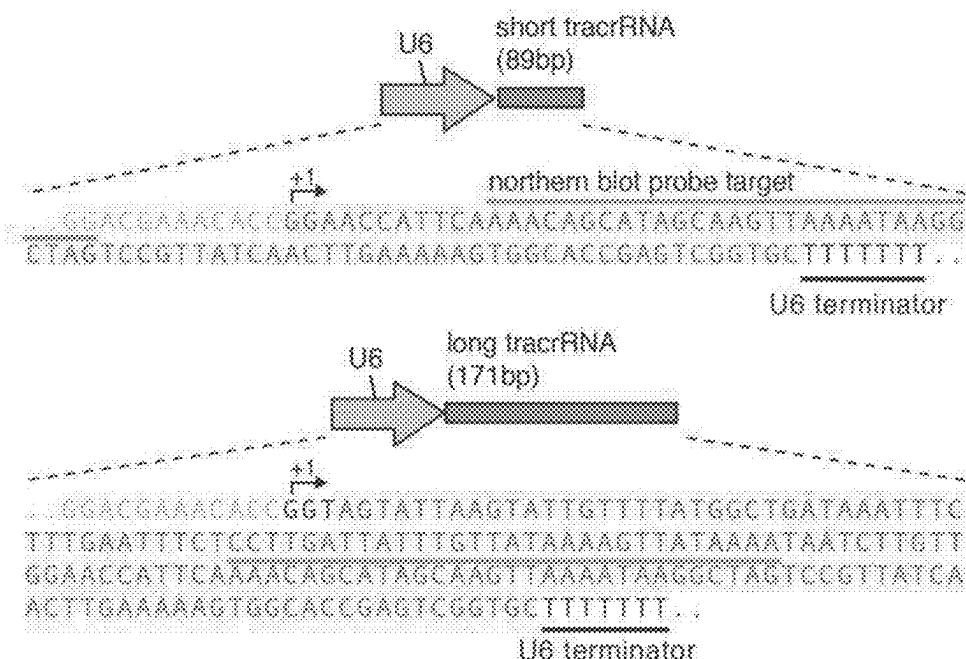


FIG. 7A

B

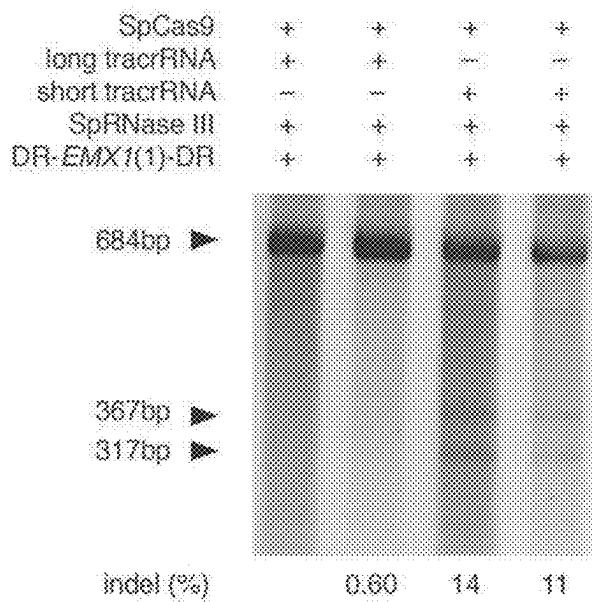


FIG. 7B

C

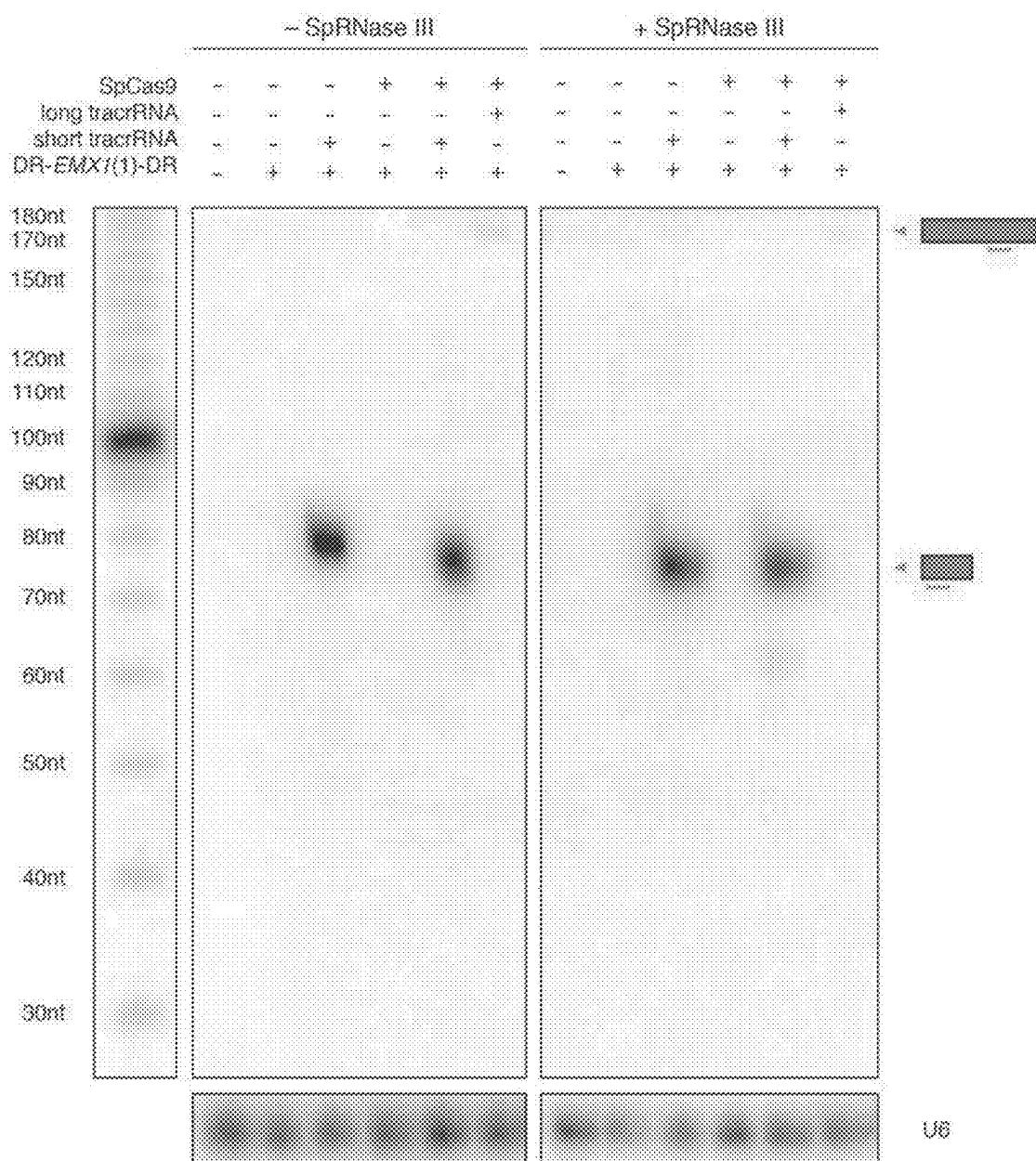
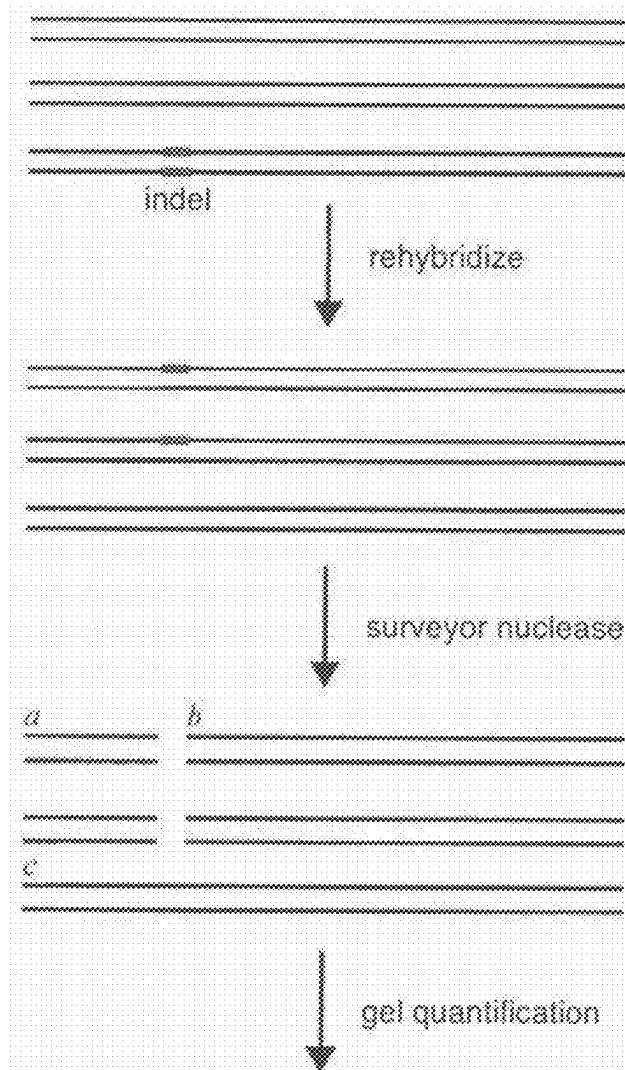
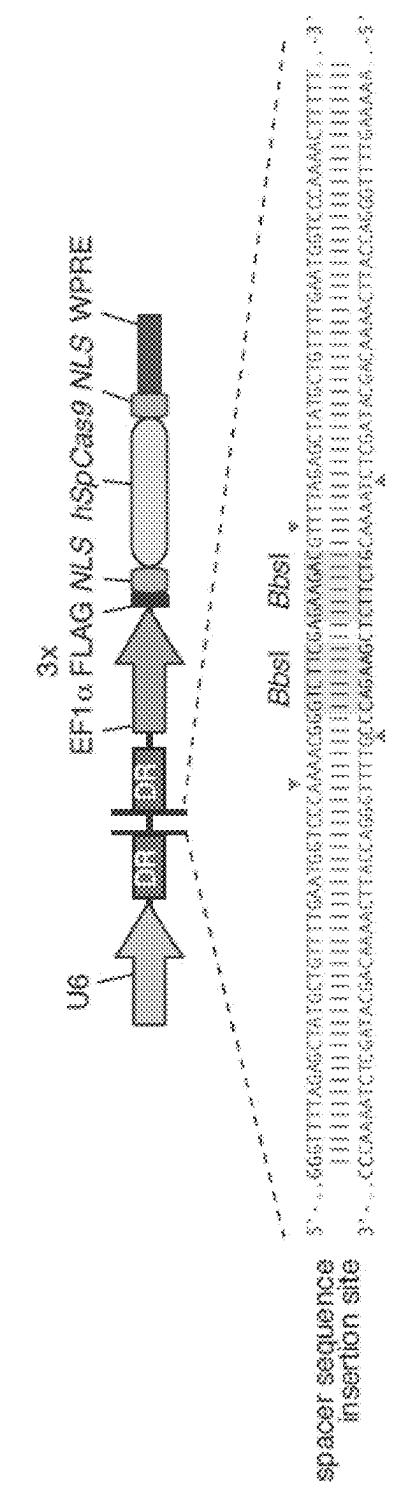


FIG. 7C

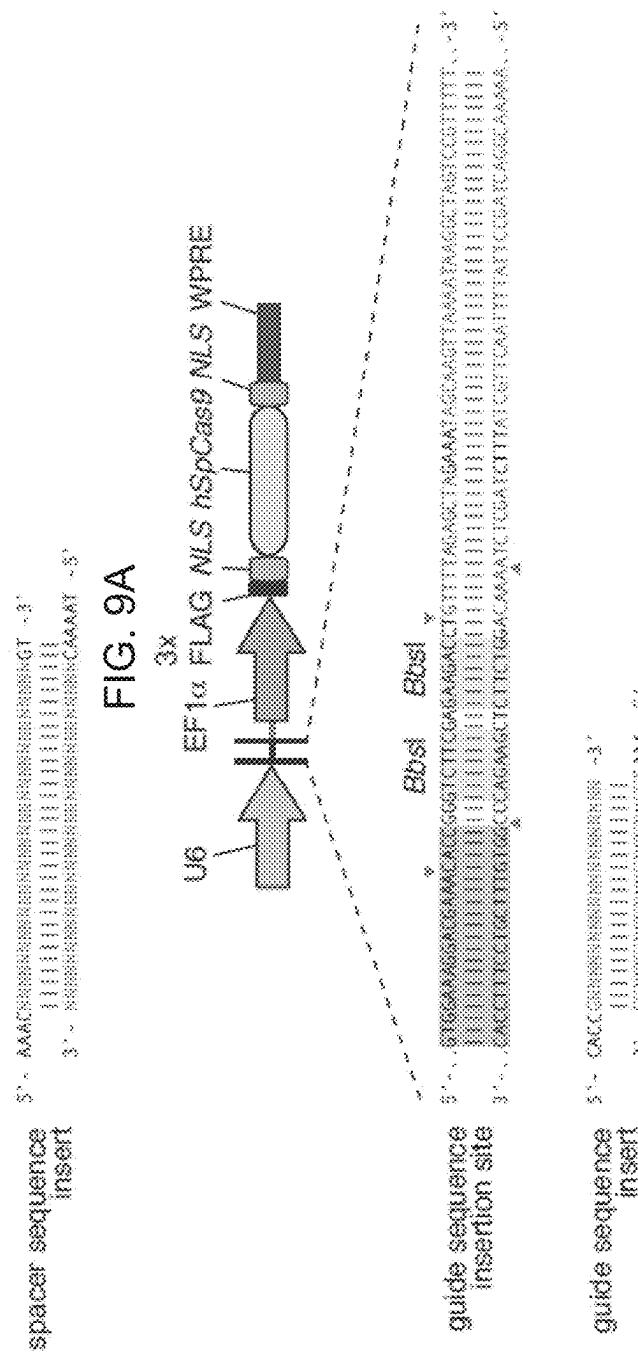


$$\% \text{ indel} = (1 - \sqrt{1 - (a + b)/(a + b + c)}) * 100$$

FIG. 8



94
二
上



三

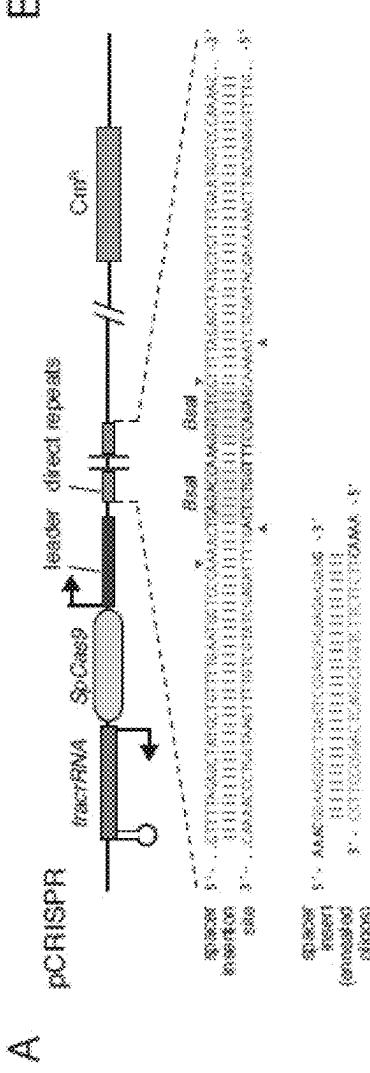


FIG. 10A

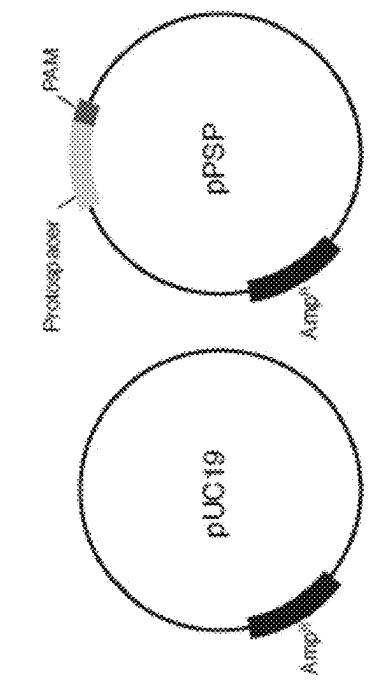
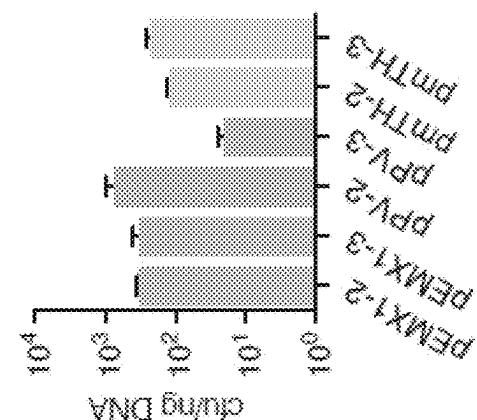
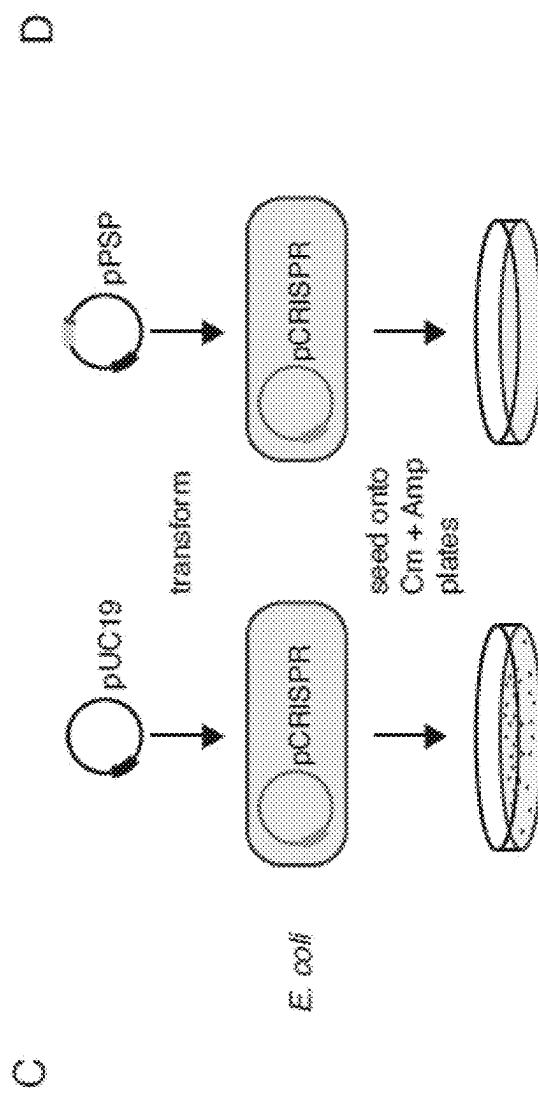


FIG. 10B

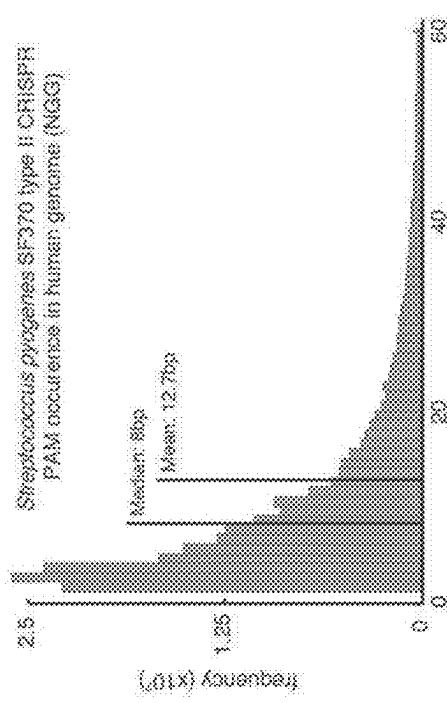


10

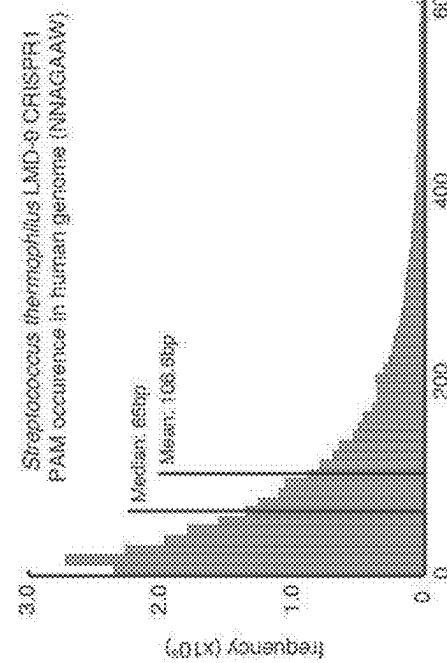


四〇六

A



B



Chr	NNACAAW			NCG		
	median	mean	median	median	mean	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	129.5		
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. 11C

A *Streptococcus thermophilus* LMD-9 CRISPR1

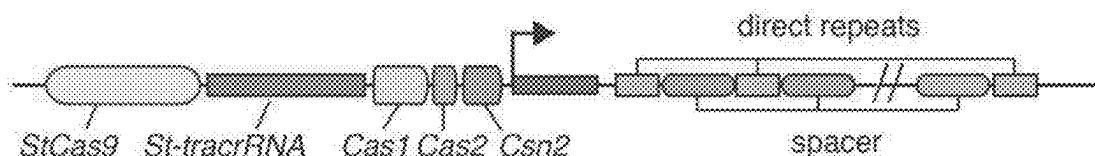


FIG. 12A

B

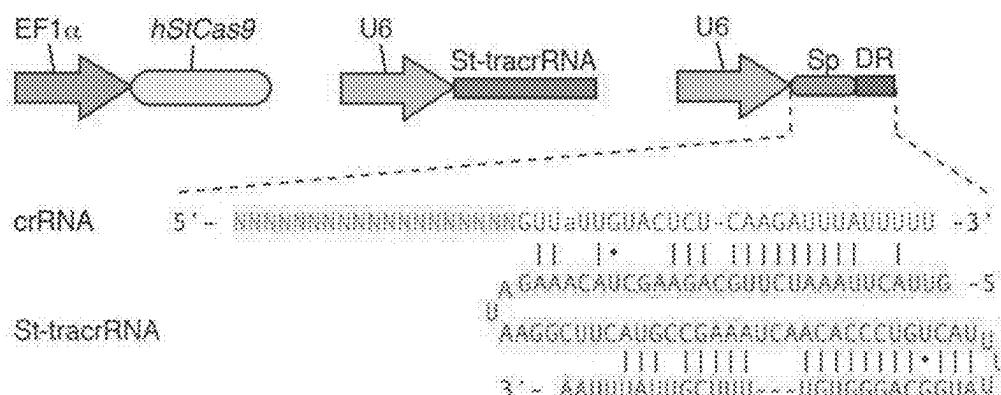


FIG. 12B

C

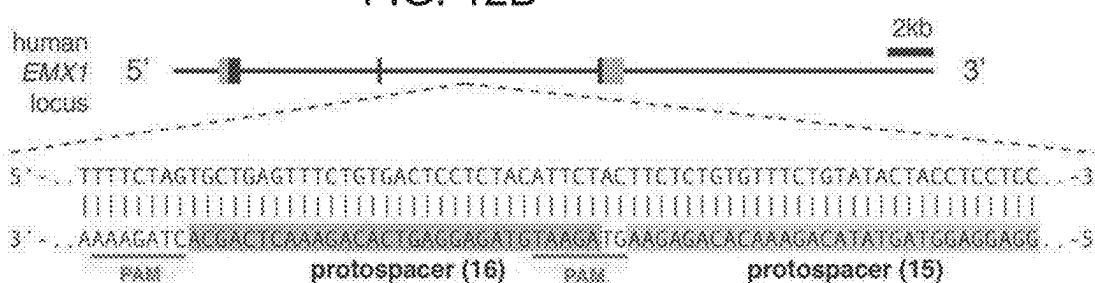
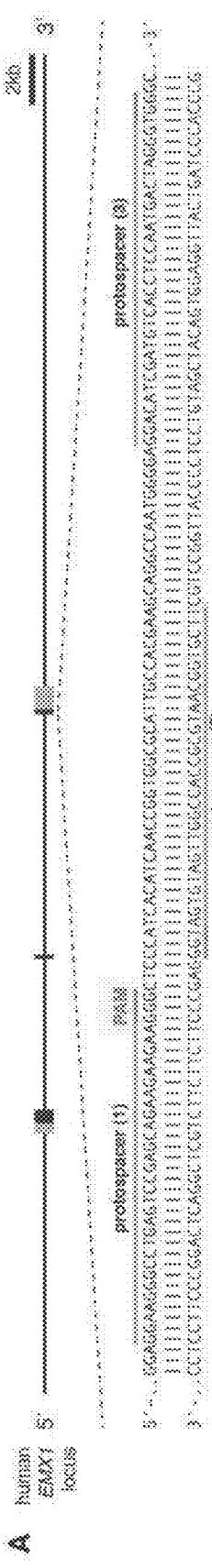
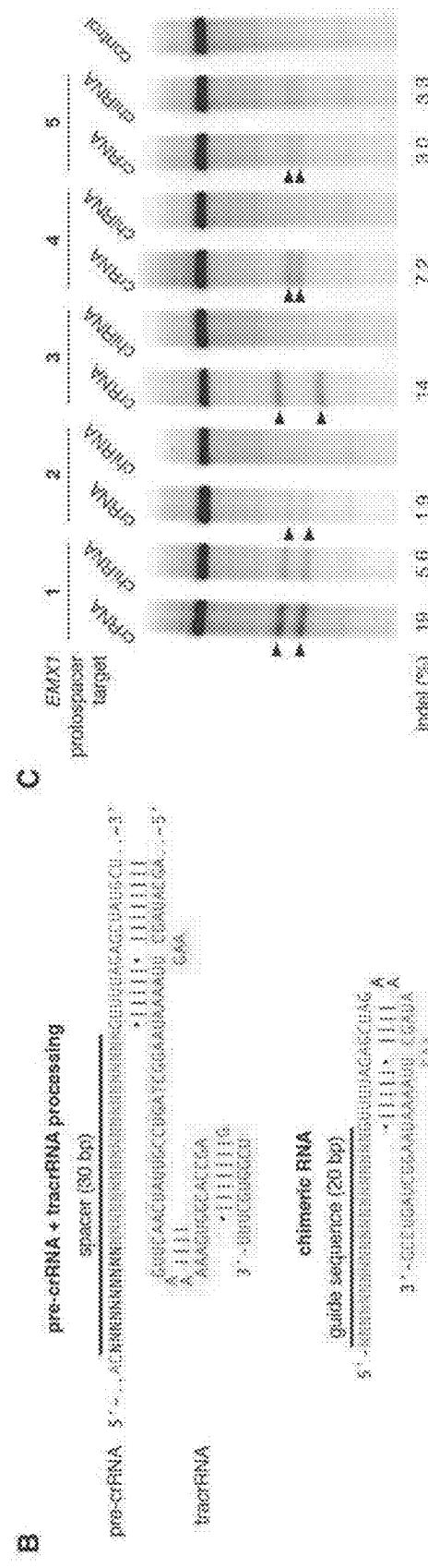


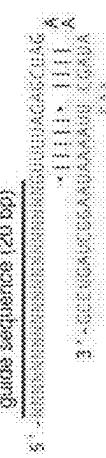
FIG. 12C



EIG. 13A



三



130

A

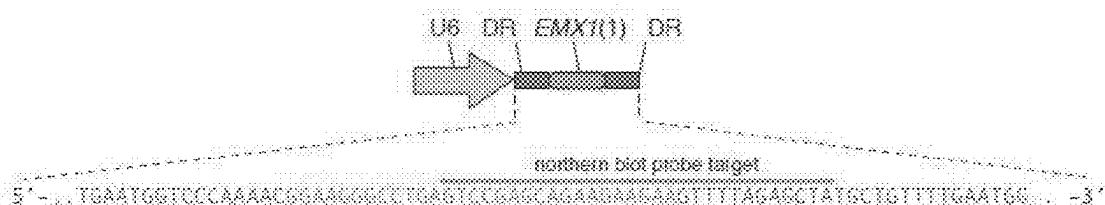


FIG. 14A

B

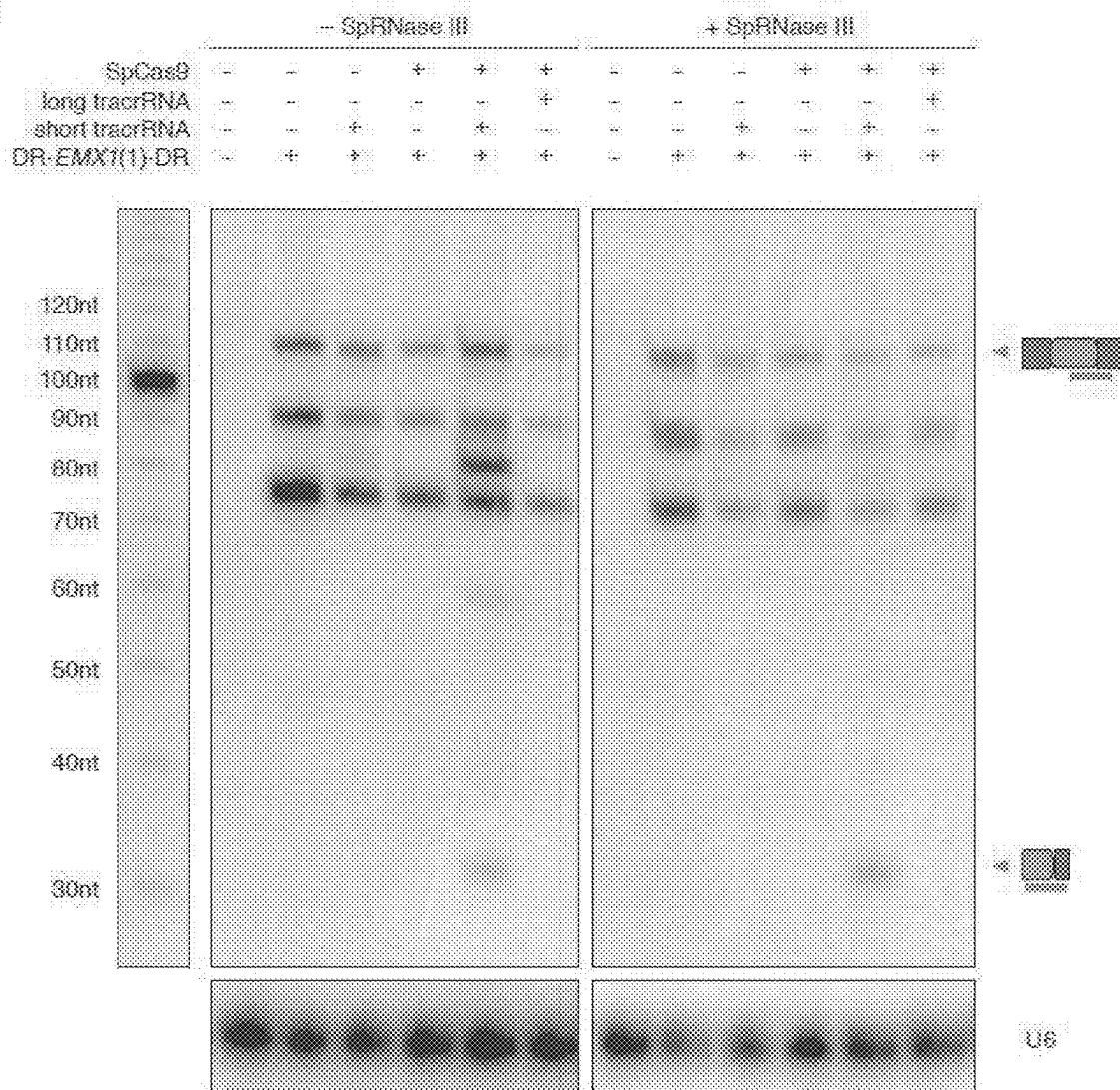


FIG. 14B

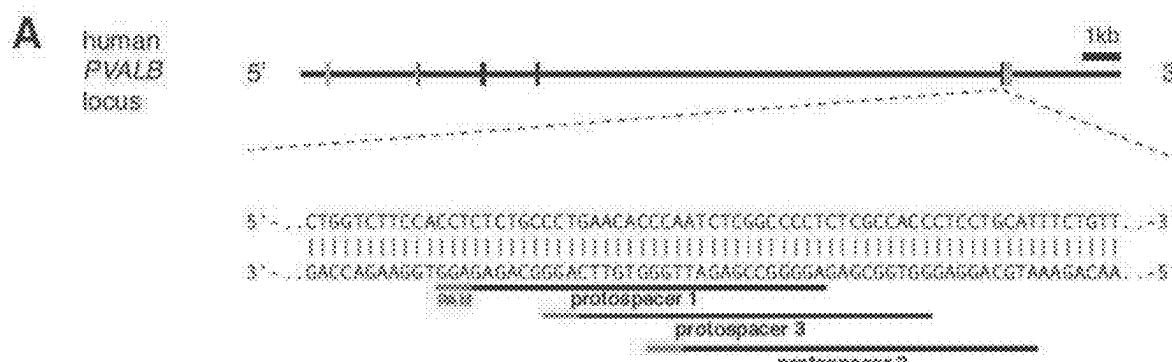


FIG. 15A

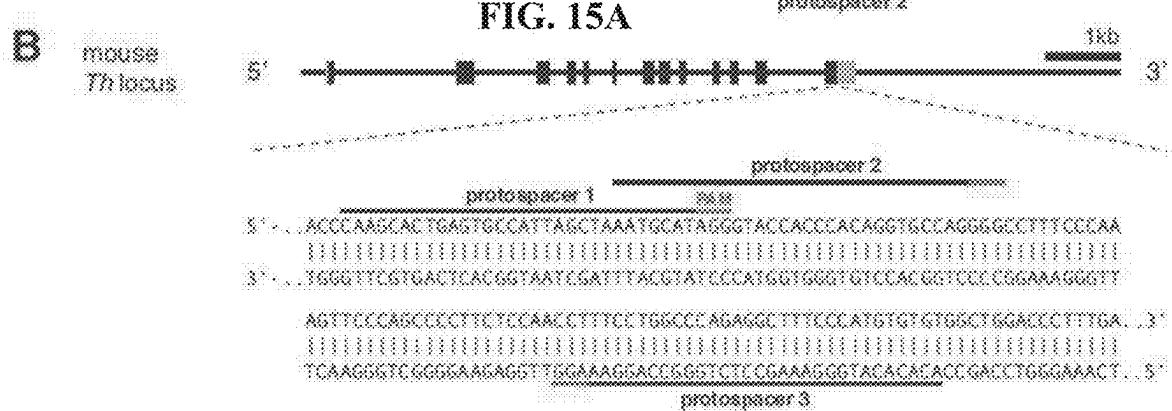


FIG. 15B

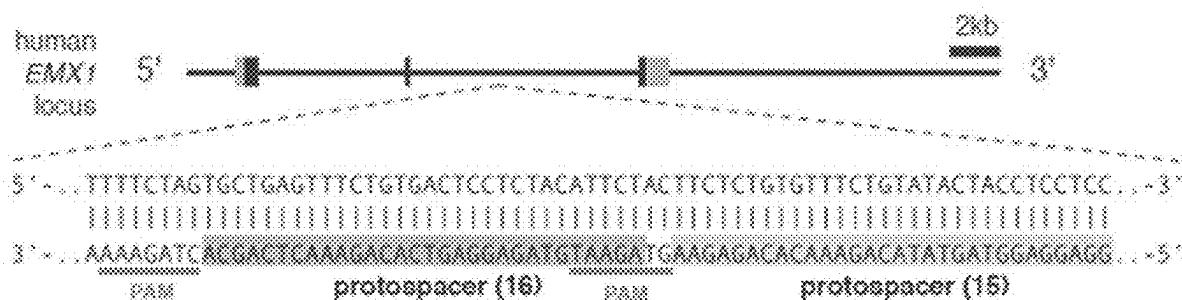


FIG. 16

Primer name	Assay	Genomic Target	Primer sequence
Sp-EMX1-F	SURVEYOR assay, sequencing	<i>EMX1</i>	AAAACCACCCCTCTCTCTGGC
Sp-EMX1-R	SURVEYOR assay, sequencing	<i>EMX1</i>	GGAGATTGGAGACACGGAGAG
Sp-PVALB-F	SURVEYOR assay, sequencing	<i>PVALB</i>	CTGGAAAGCCAATGCCTGAC
Sp-PVALB-R	SURVEYOR assay, sequencing	<i>PVALB</i>	GGCAGCAAACCTCCTGTCT
Sp-Th-F	SURVEYOR assay, sequencing	<i>Th</i>	GTGCTTGAGAGGCCTACC
Sp-Th-R	SURVEYOR assay, sequencing	<i>Th</i>	CCTGGAGCGCATGCAAGTAGT
Sp-EMX1-F	SURVEYOR assay, sequencing	<i>EMX1</i>	ACCTTCTGTGTTCCACCATTC
Sp-EMX1-R	SURVEYOR assay, sequencing	<i>EMX1</i>	TTGGGGAGTGCACAGACTTC
Sp-EMX1-RFLP-F	RFLP, sequencing	<i>EMX1</i>	GGCTCCCTGGGTTCAAAGTA
Sp-EMX1-RFLP-R	RFLP, sequencing	<i>EMX1</i>	AGAGGGGTCTGATGTCGTAA
Pb_EMX1_sp1	Northern Blot Probe	Not applicable	TAGCTCTAAAACTTCTTCTGCTCGGAC
Pb_tracrRNA	Northern Blot Probe	Not applicable	CTAGCCTTATTTAACCTGCTATGCTGTTT

FIG. 17

a

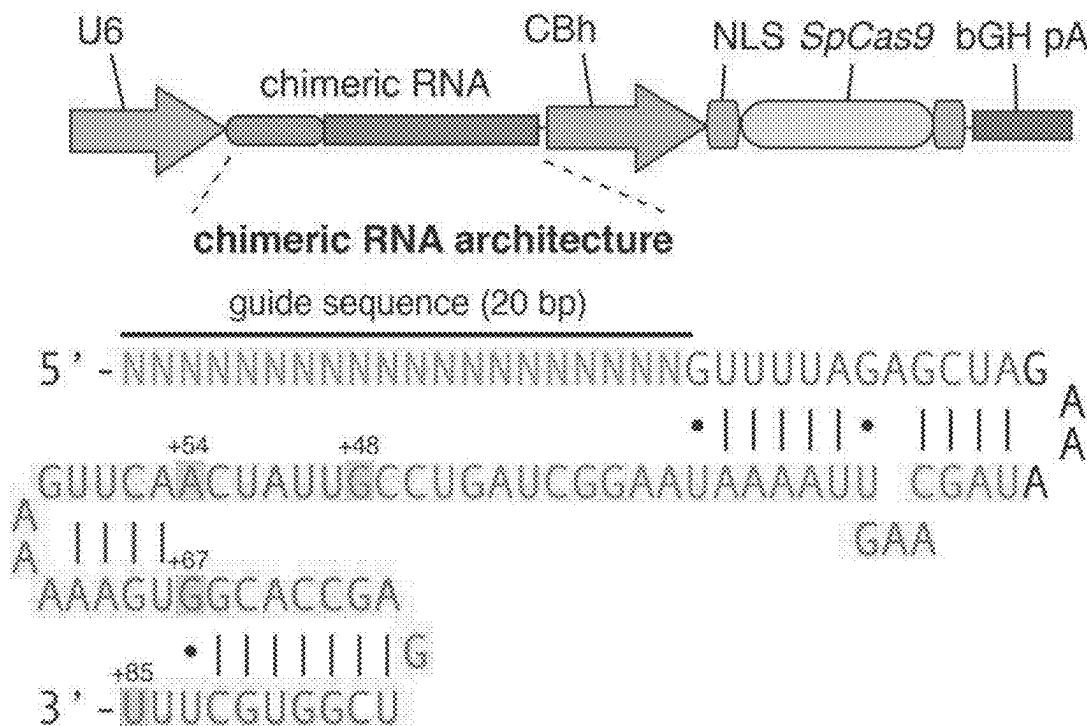


FIG. 18A

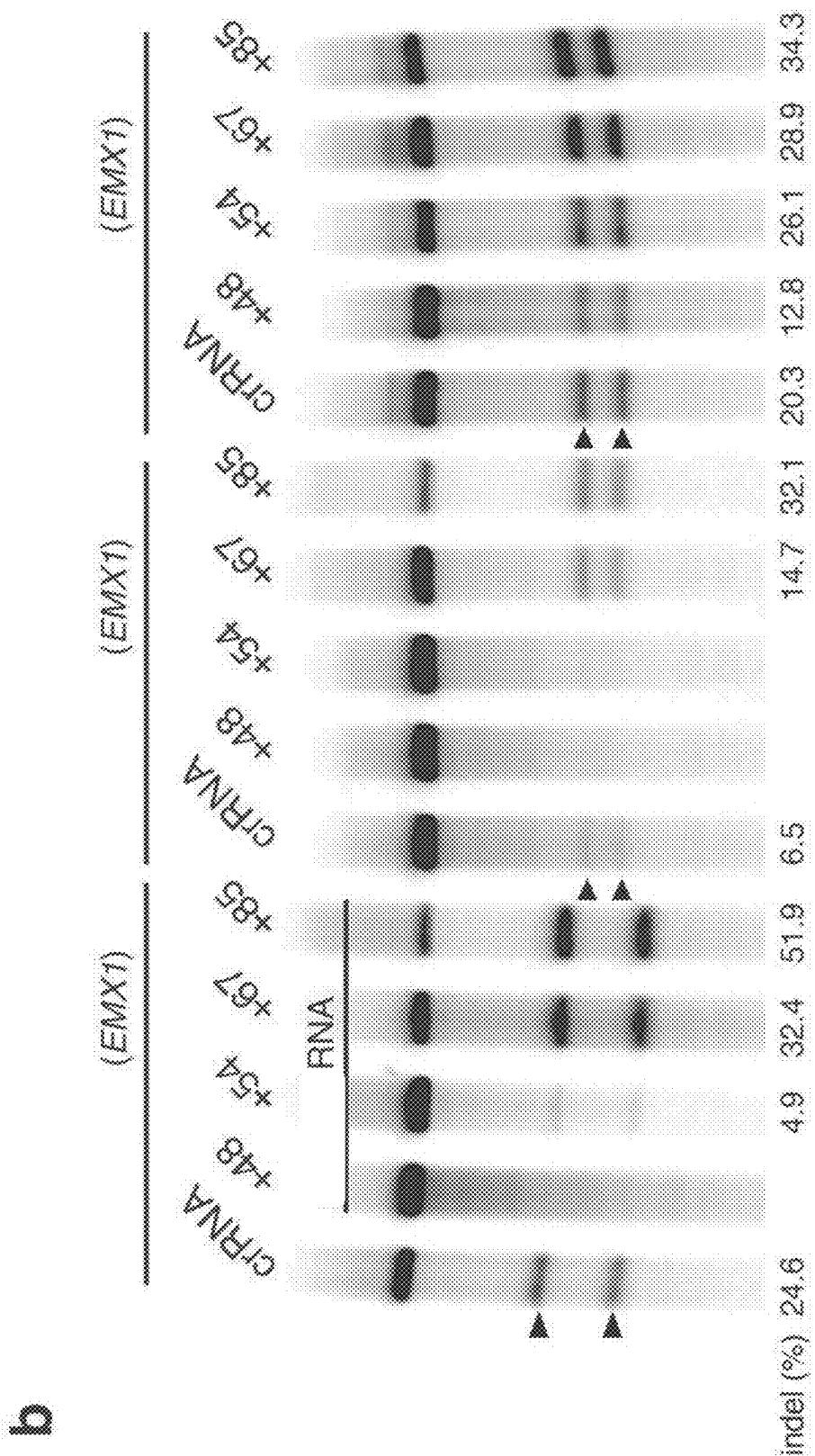


FIG. 18B

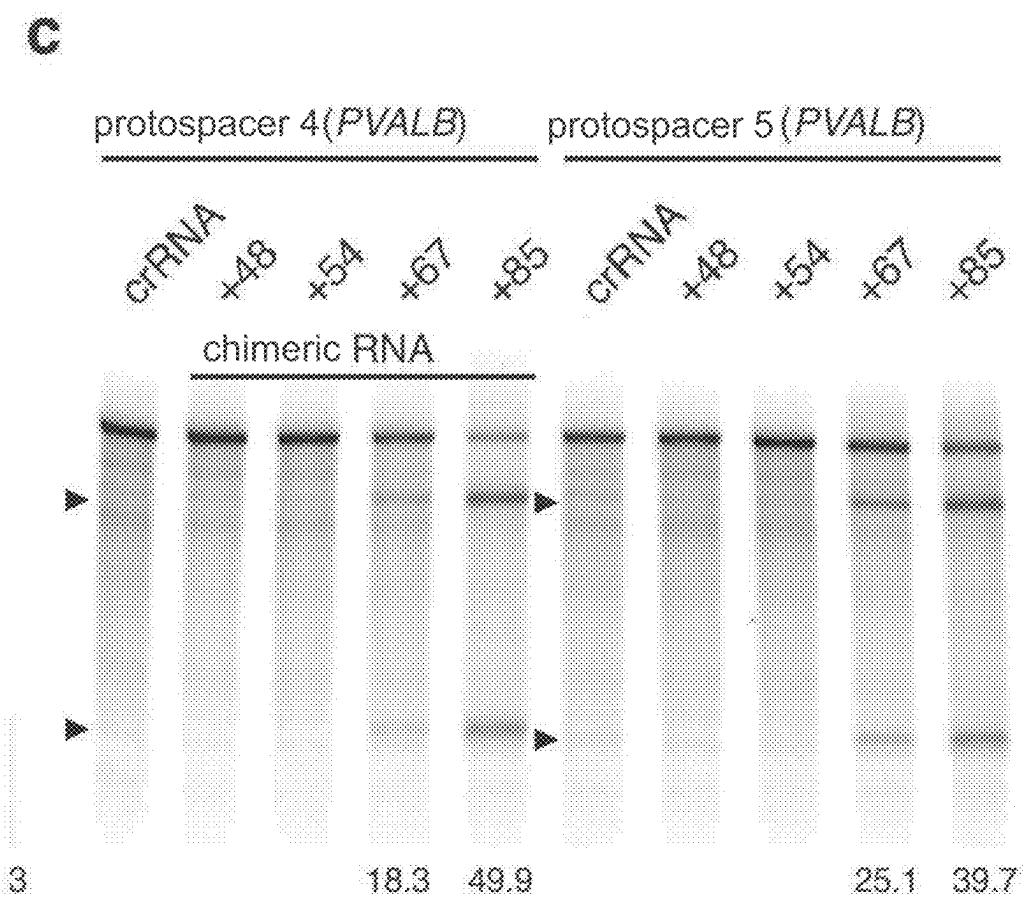


FIG. 18C

b

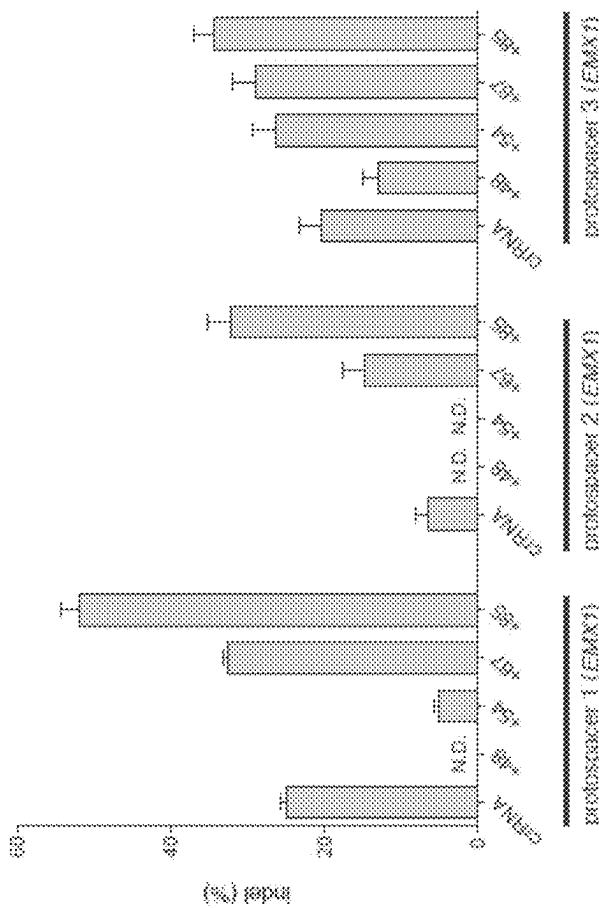


FIG. 19A

c

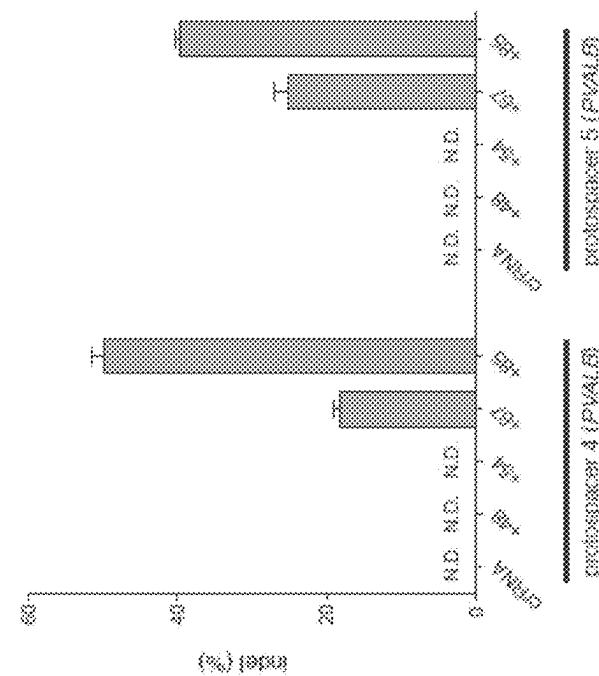
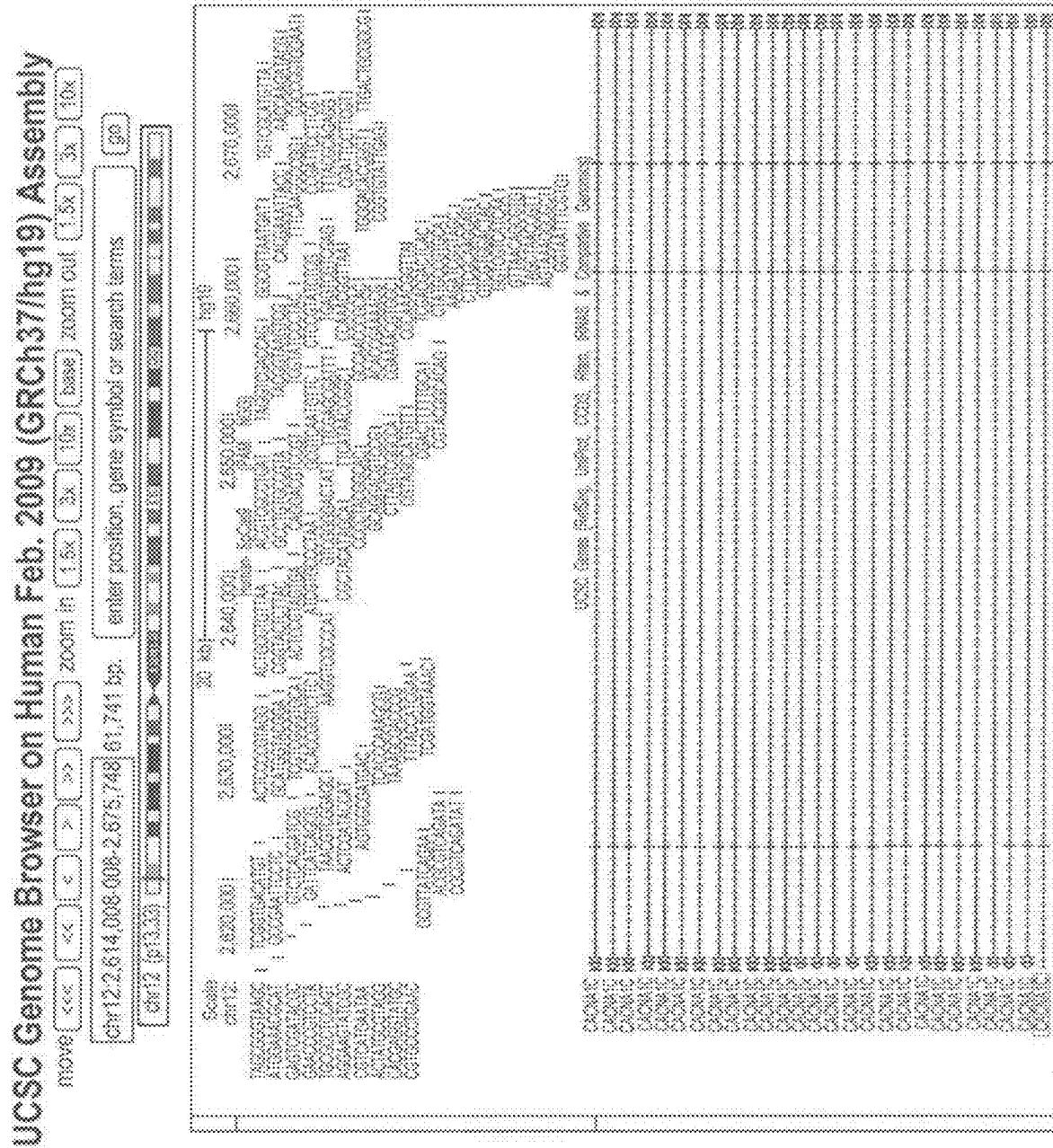


FIG. 19B

FIG. 20



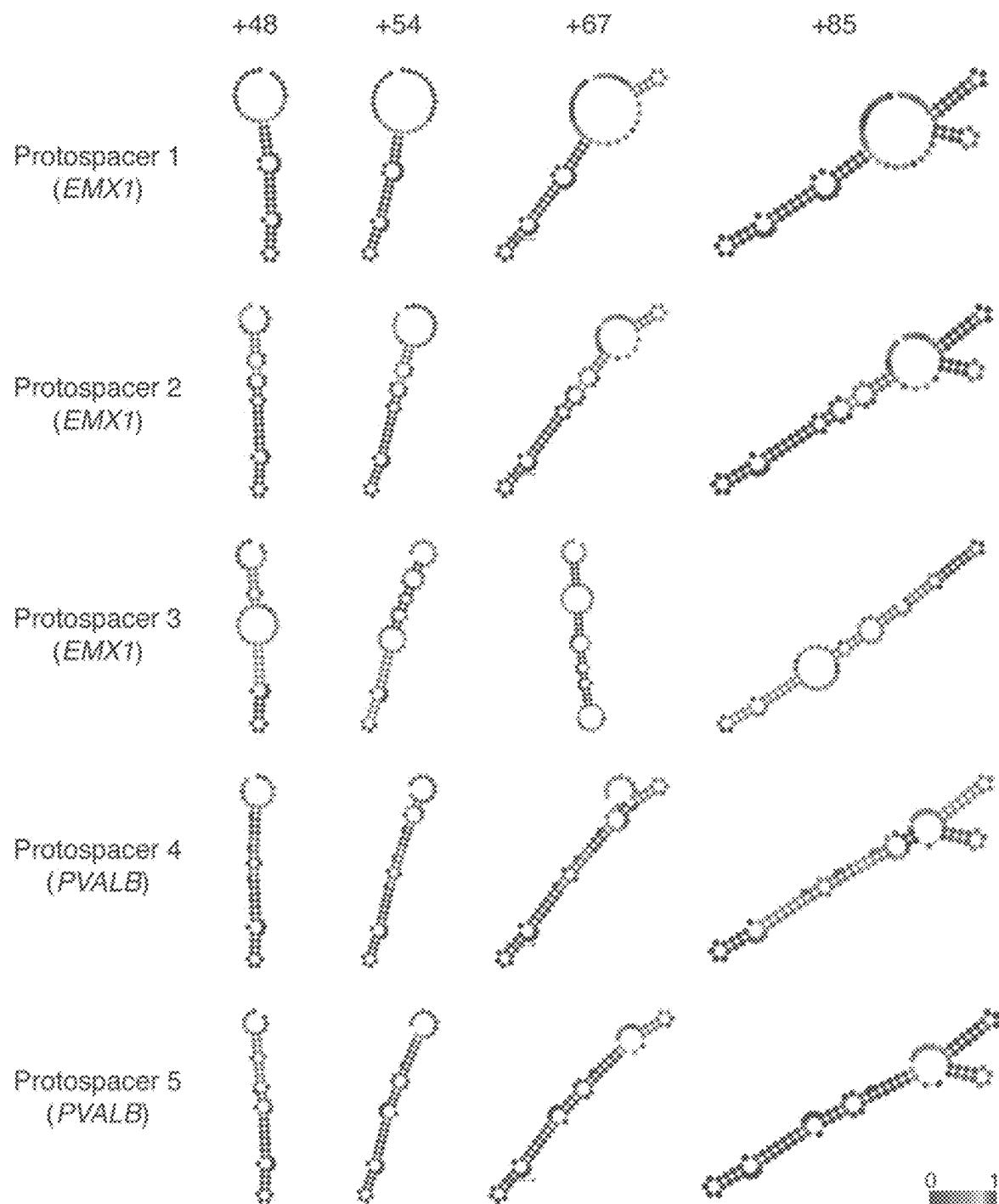


FIG. 21

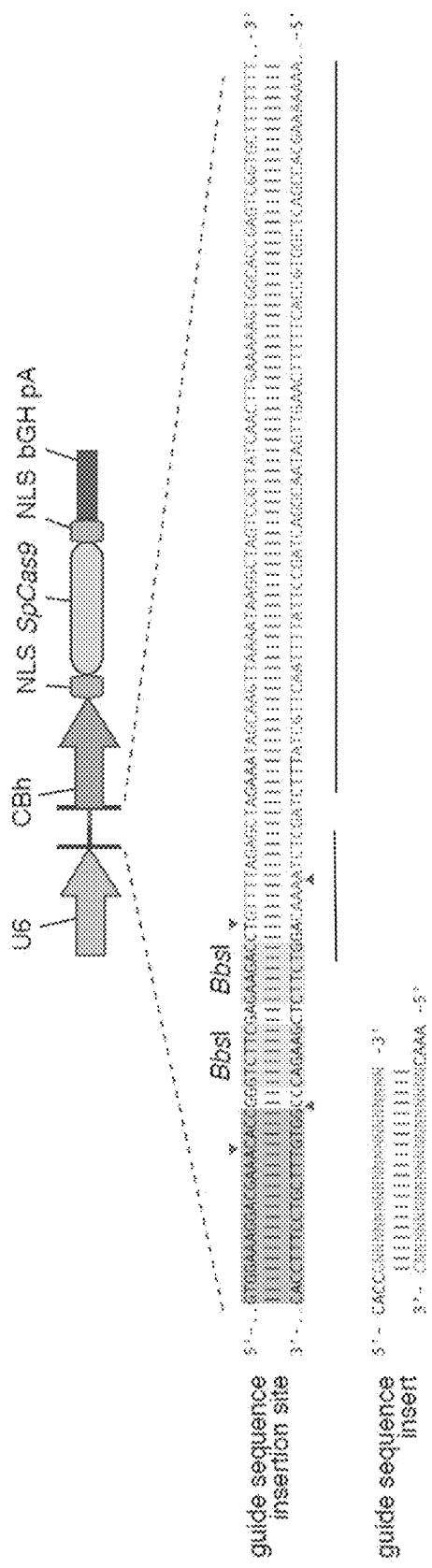


FIG. 22

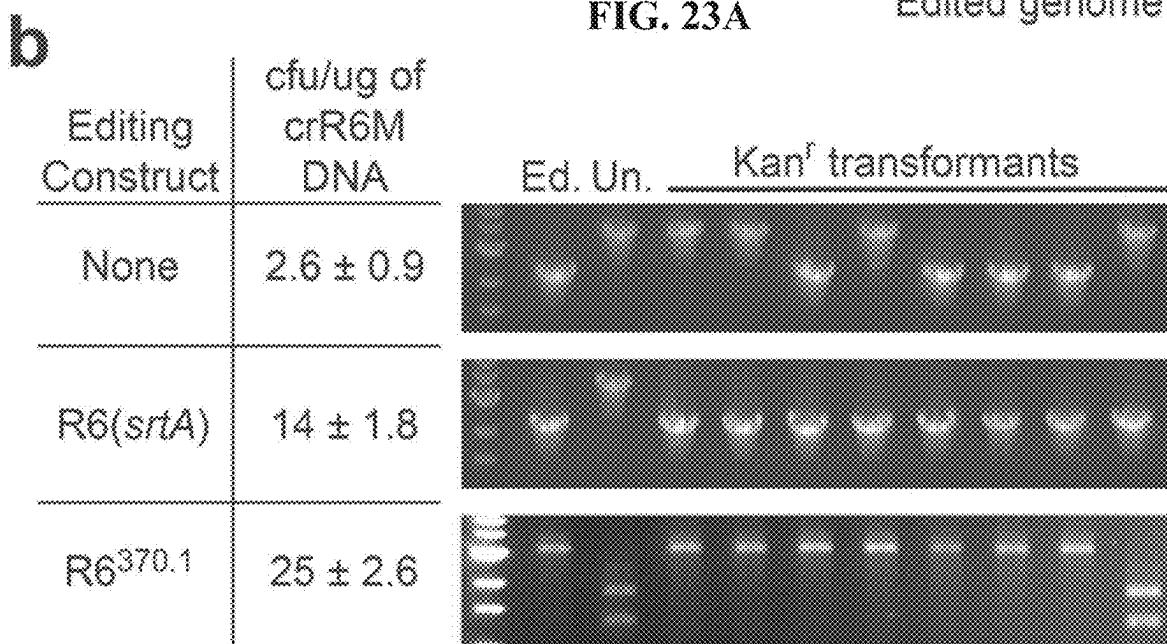
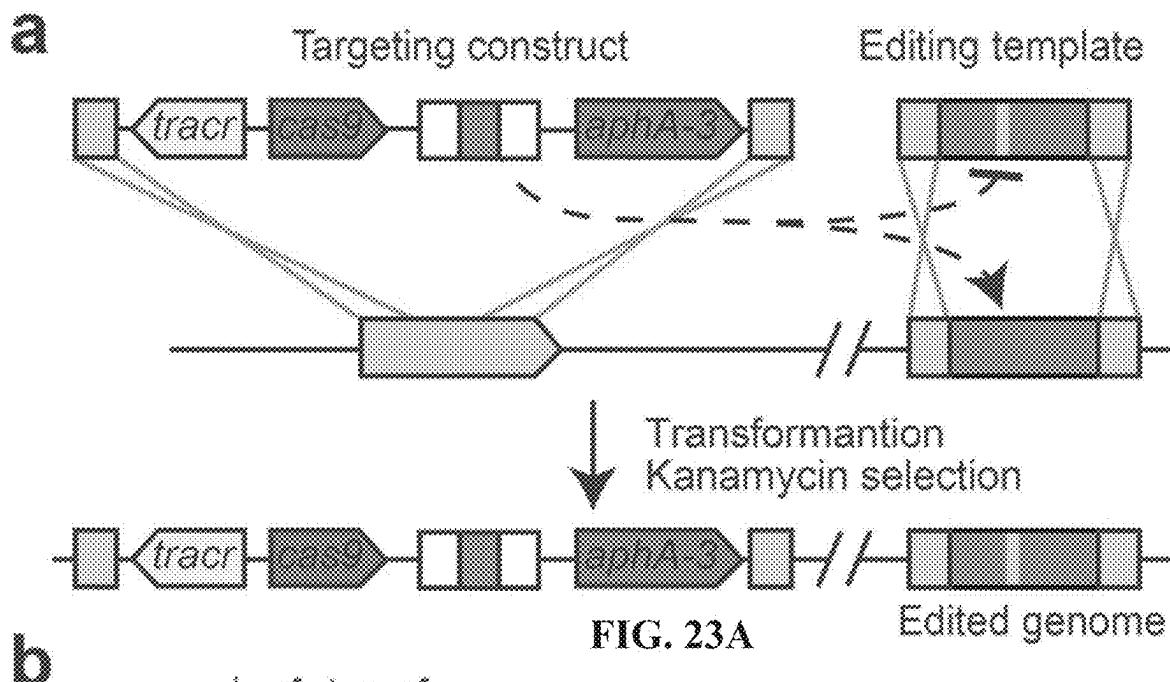


FIG. 23B

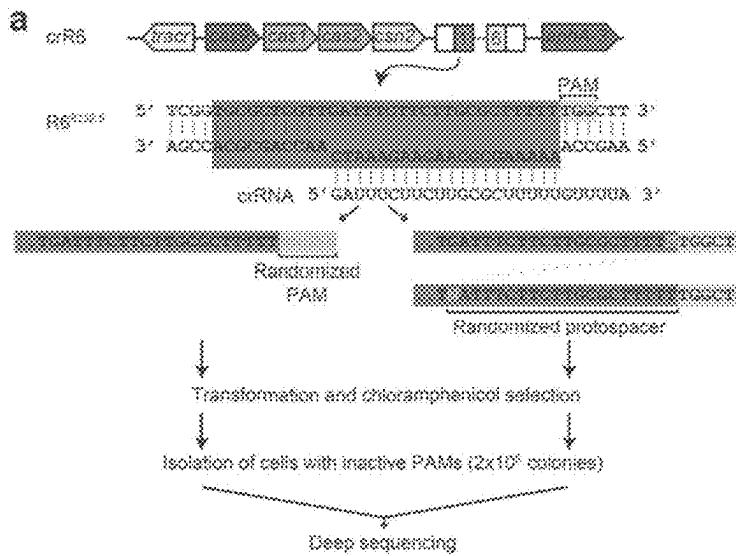


FIG. 24A

		2 nd PAM position								
		A	C	G	T					
5 th PAM position	A	AAX	1.02	A CA	1.06	A GA	0.95	A TX	1.02	A
	C	CAA	1.03	CCA	1.03	CGA	0.97	CTA	1.06	
	G	CAA	1.10	GCA	1.09	GGA	0.93	GTA	1.01	
	T	TAA	1.02	TCA	1.11	TGA	0.81	TTC	1.11	
	A	XAC	1.07	ACC	1.02	AGC	1.00	ATC	1.09	
4 th PAM position	C	CAC	1.06	CCC	1.04	CGC	1.00	CTC	1.09	C
	G	CAC	1.05	GCC	1.02	GGC	1.02	GTC	1.23	
	T	ZAC	1.06	ZCC	1.05	ZGC	1.03	ZTC	1.06	
	A	AAG	0.99	ACG	0.97	AGG	0.95	ATG	0.96	
	C	CAA	0.98	CCC	1.08	CGG	0.95	CTG	1.03	
3 rd PAM position	G	AAA	0.90	GCG	1.09	GGG	0.96	GGG	1.20	G
	T	TTA	0.86	TCG	0.91	TGG	0.97	TTC	1.06	
	A	XAT	1.01	ACT	1.08	AGT	1.04	ATT	1.03	
	C	CAT	1.03	CCT	1.03	CGT	1.07	CTT	1.08	
	G	GAT	1.04	GCT	1.08	G GT	1.08	GTT	1.12	
2 nd PAM position	T	ZAT	1.06	ZCT	1.03	ZGT	1.06	ZTT	1.01	T

FIG. 24B

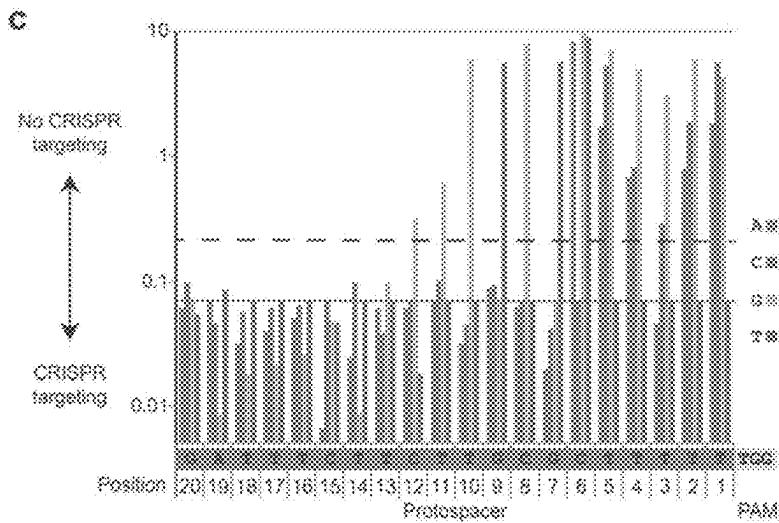


FIG. 24C

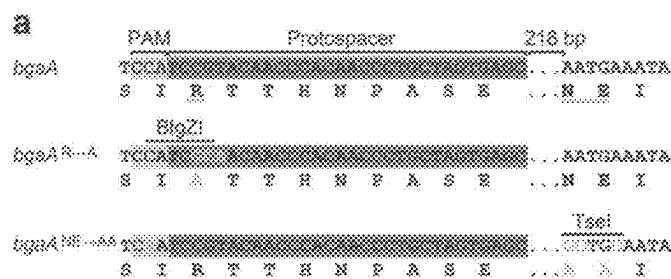


FIG. 25A

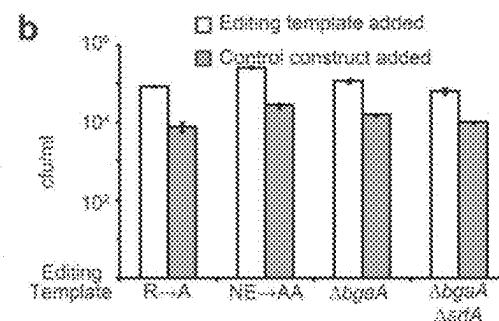


FIG. 25B

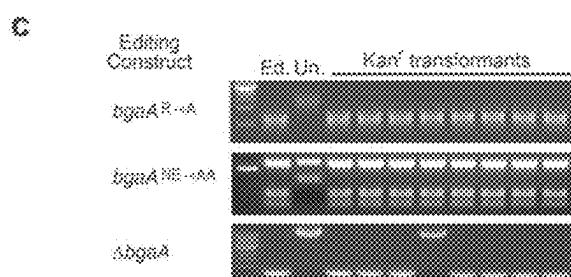


FIG. 25C

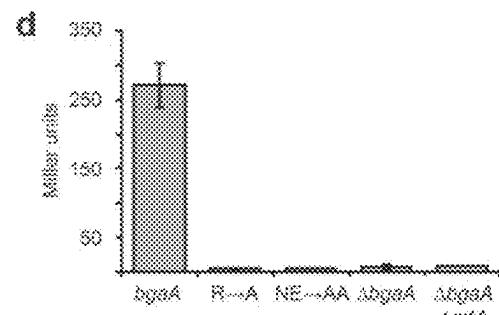


FIG. 25D

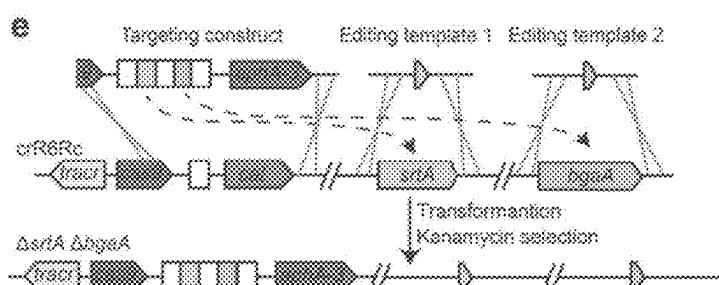


FIG. 25E

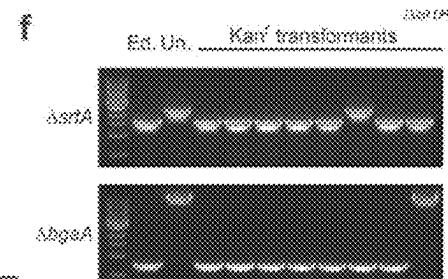


FIG. 25F

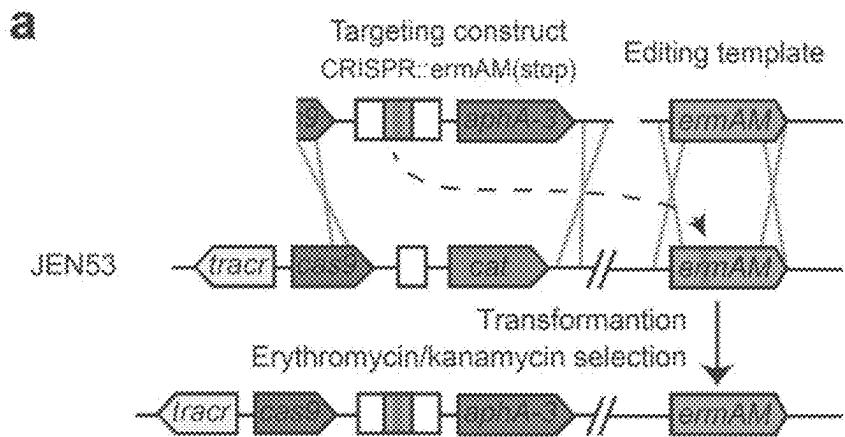


FIG. 26A

b

	Protospacer										PAM			
ermAM (stop)	C	T	A	G	T	C	G	T	A	G	C	TTC	AAAGCA	
	Q	L	N	L	K	E	T	D	T	V	L	*	G	
ermAM	C	T	A	G	T	C	G	T	A	G	C	TCC	AAATTGGA	
	Q	L	N	L	K	E	T	D	T	V	Y	X	I	G

FIG. 26B

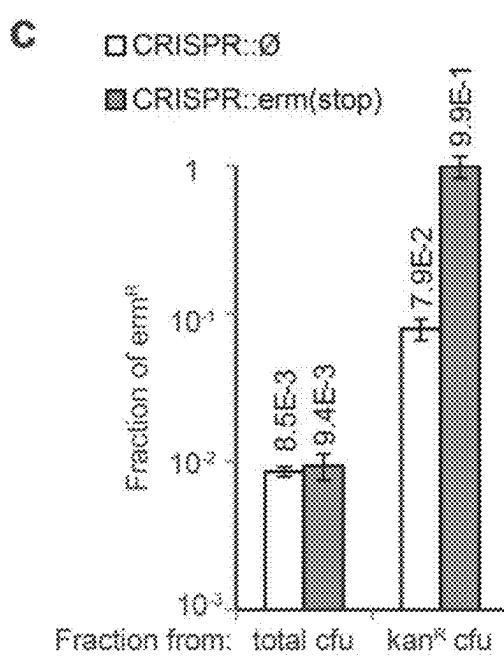


FIG. 26C

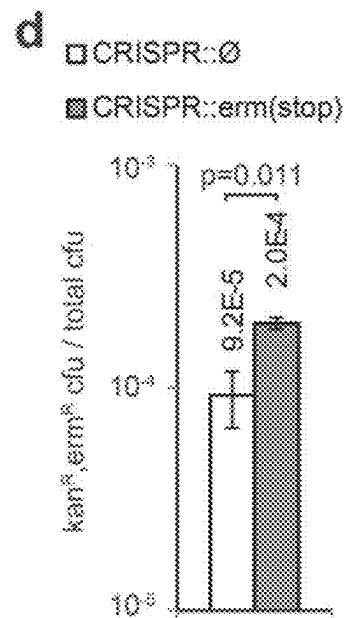


FIG. 26D

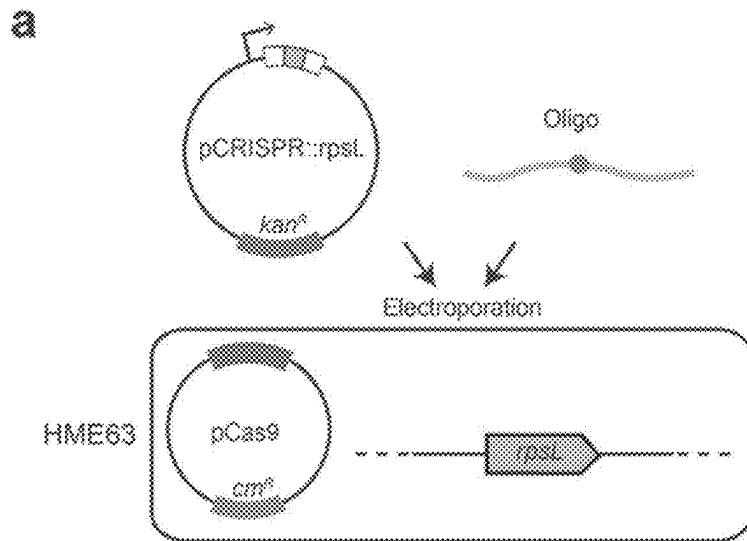


FIG. 27A

	PAM	Protospacer									
<i>rpsL</i>	TCTT	A	G	A	T	C	T	G	A	T	T
		P	K	K	P	N	S	A	L	R	K
<i>rpsL</i> mut	TCTTA	A	G	A	T	C	T	G	A	T	T
		P	T	K	P	N	S	A	L	B	V

FIG. 27B

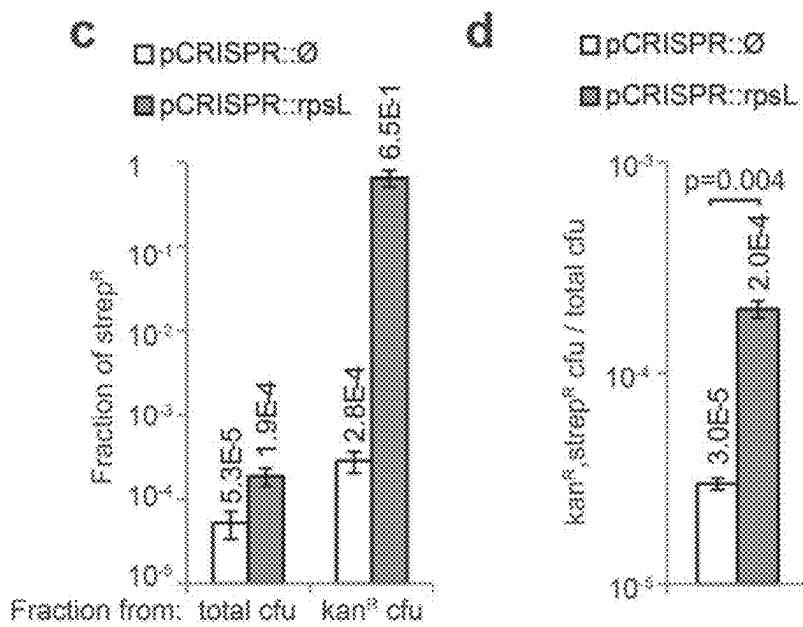


FIG. 27C

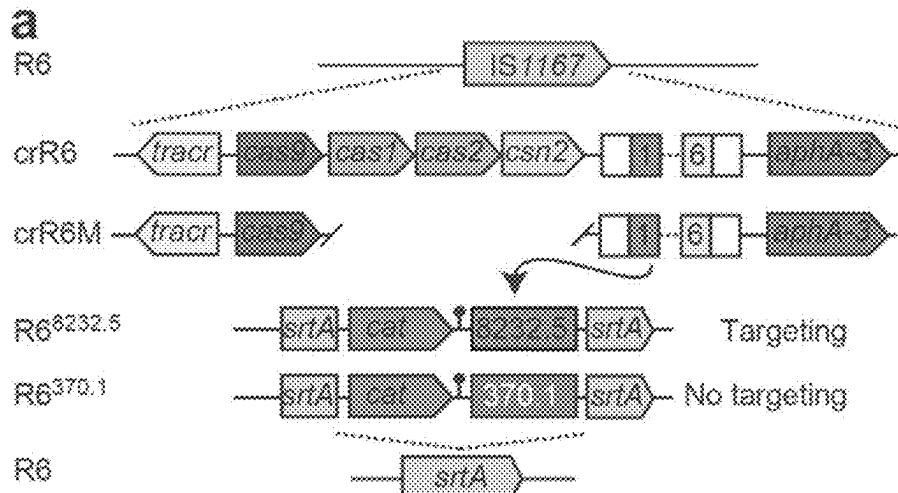


FIG. 28A

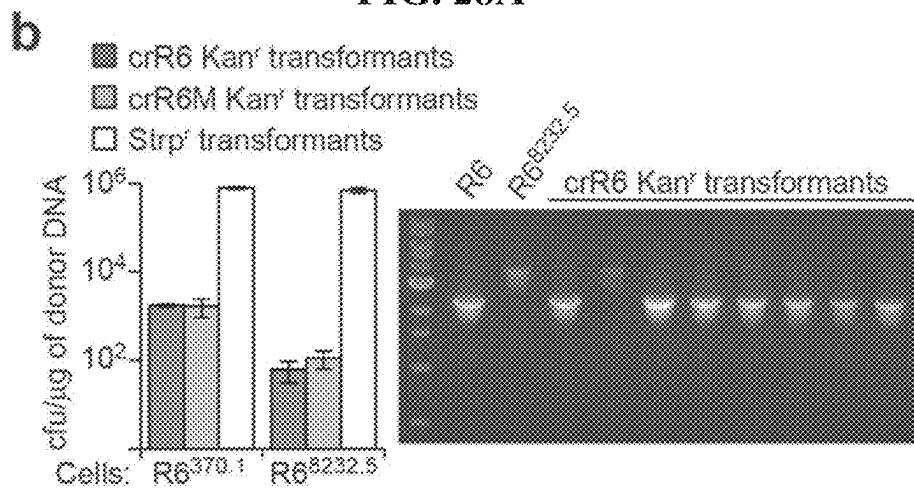


FIG. 28B

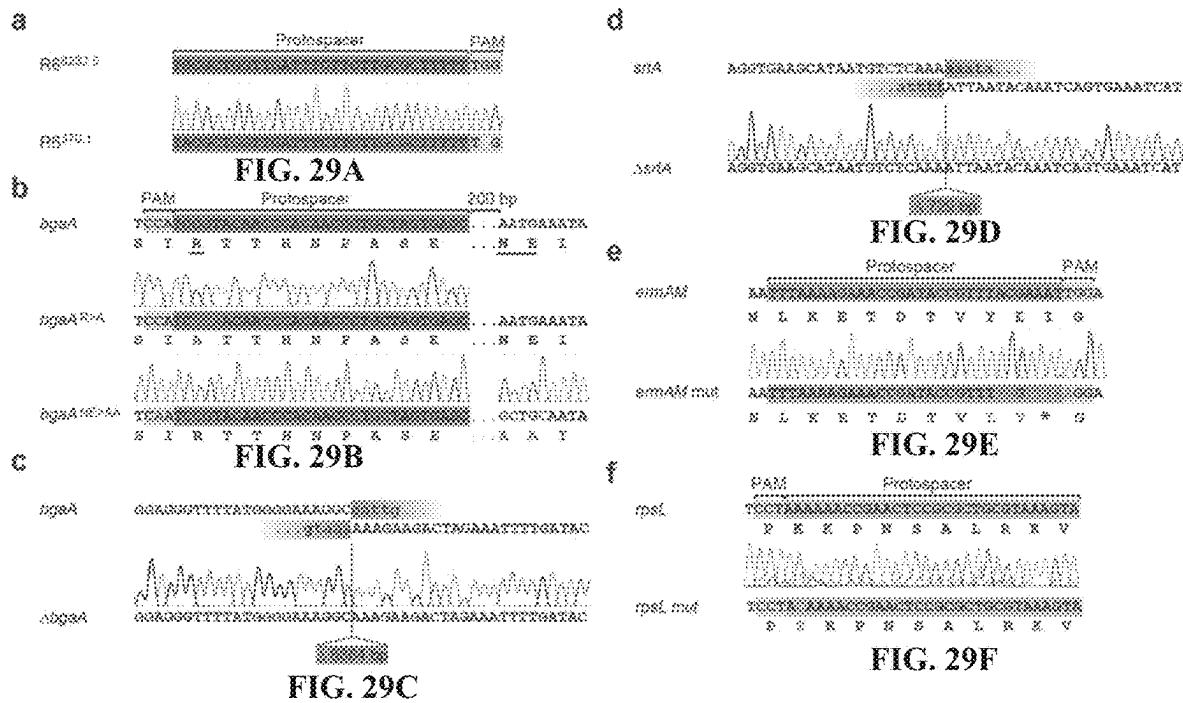


FIG. 29A-F

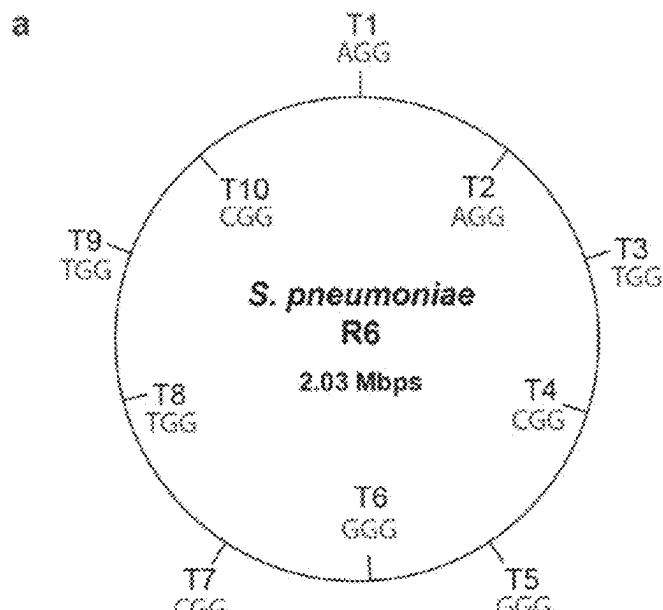


FIG. 30A

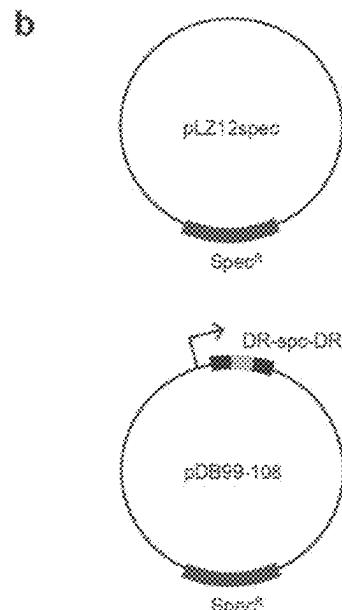


FIG. 30B

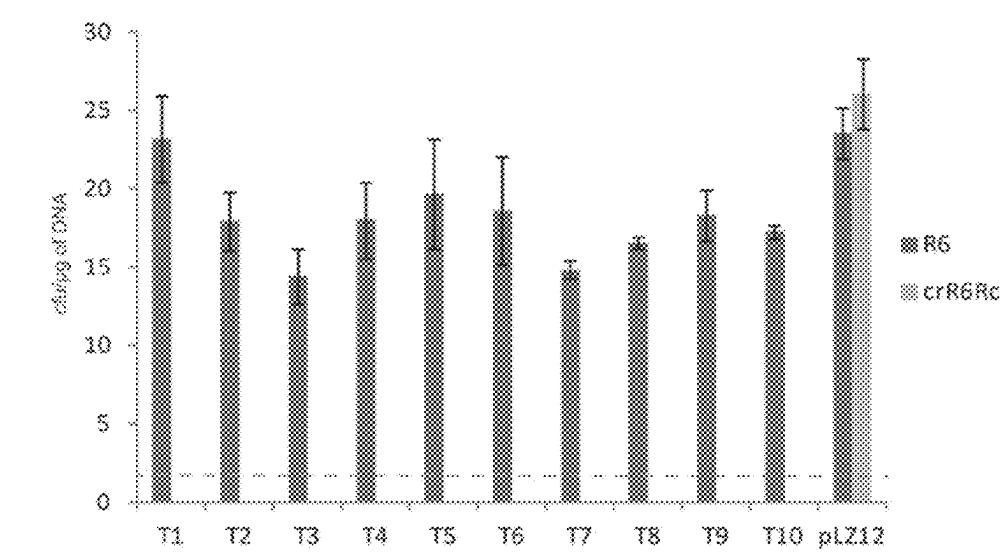


FIG. 30C

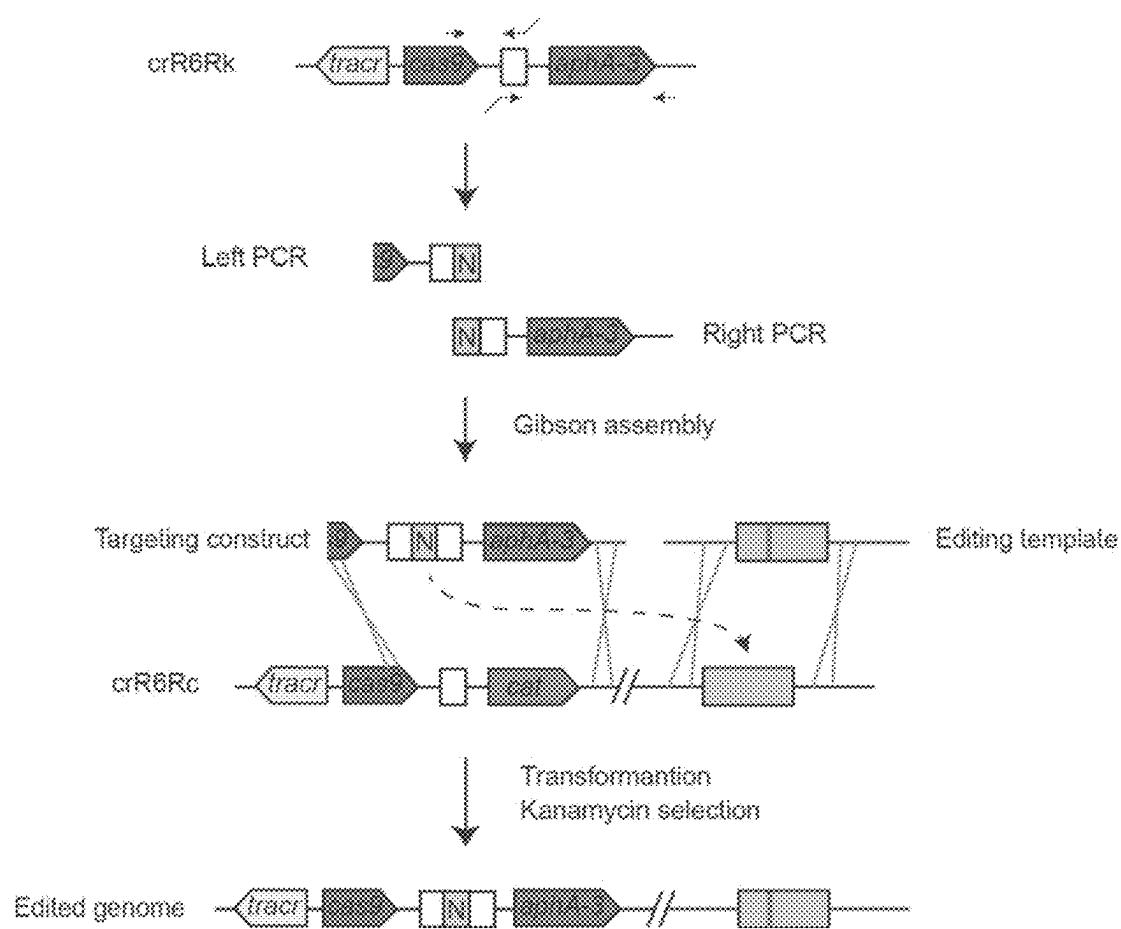


FIG. 31

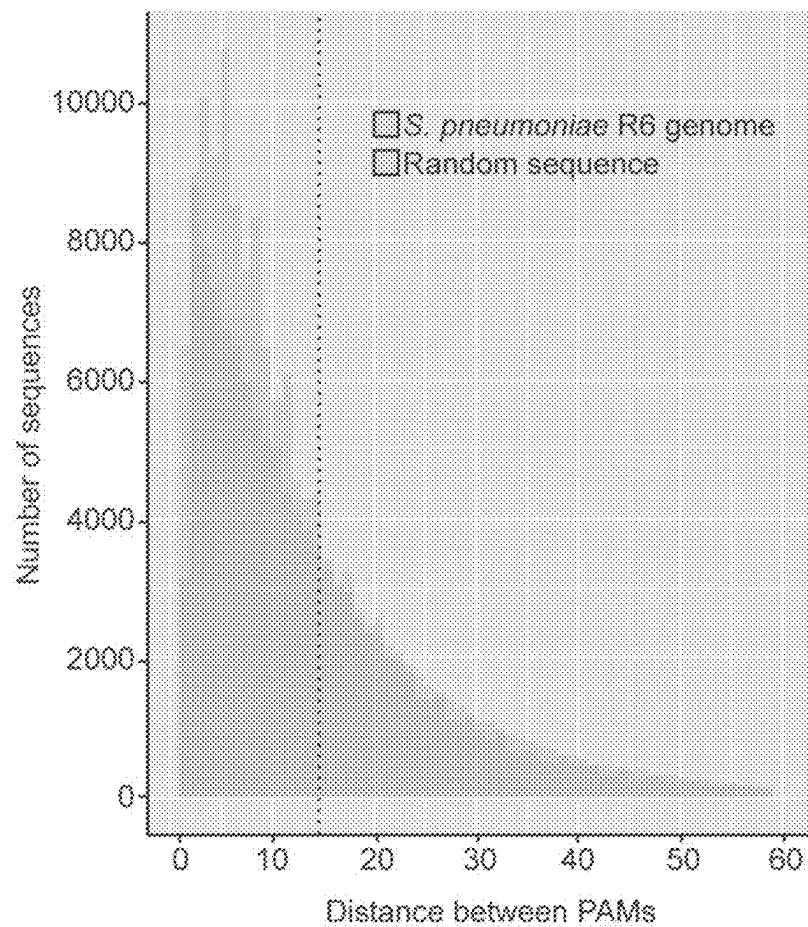


FIG. 32

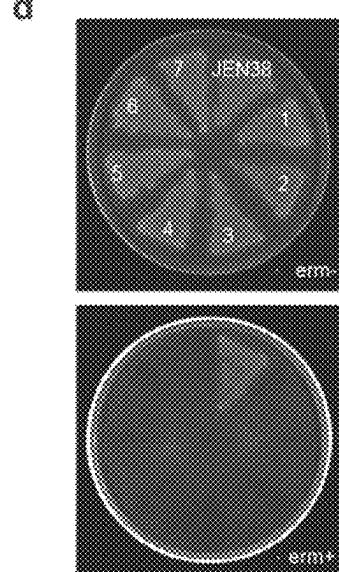
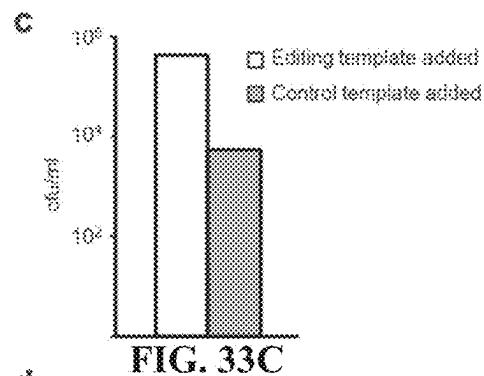
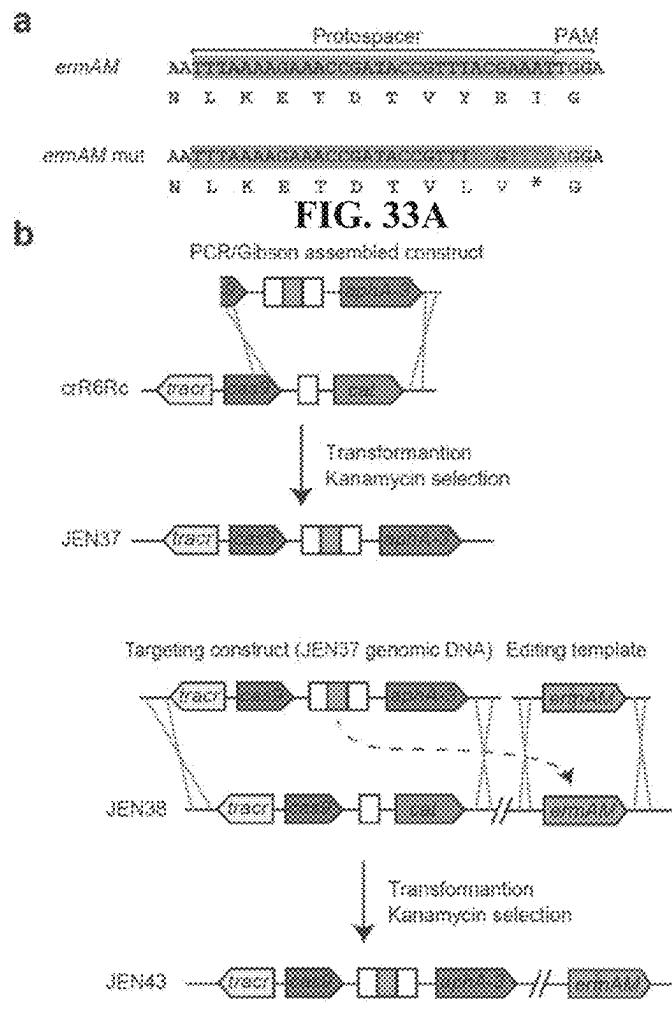


FIG. 33B

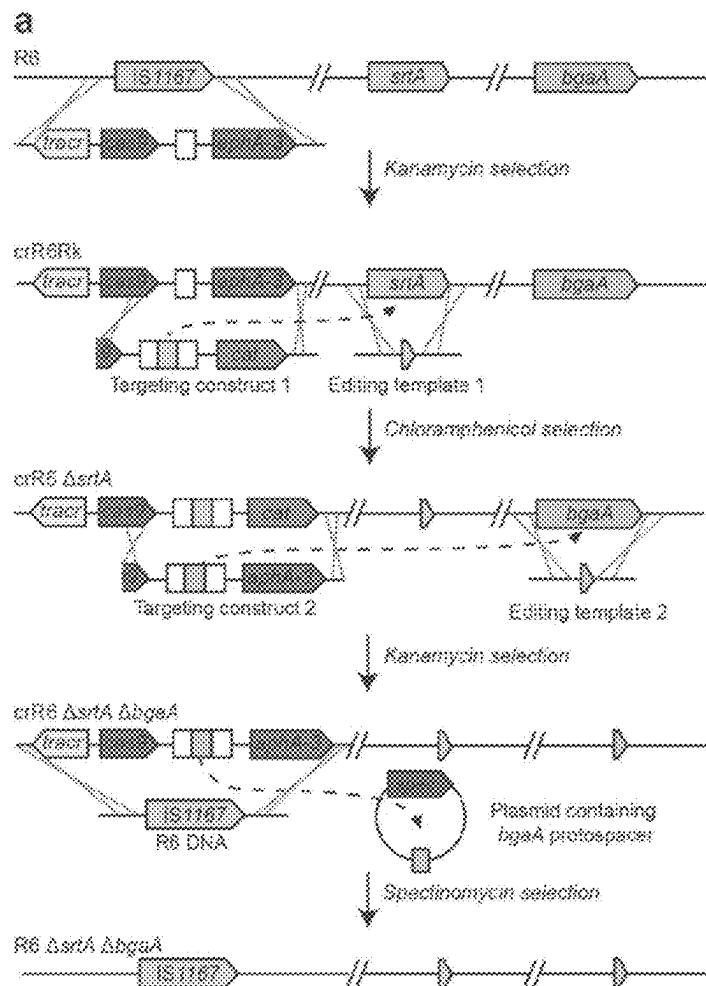


FIG. 34A

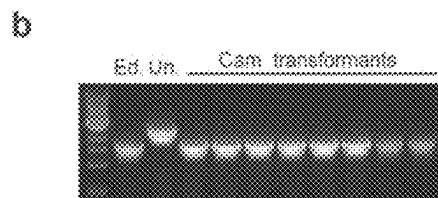


FIG. 34B

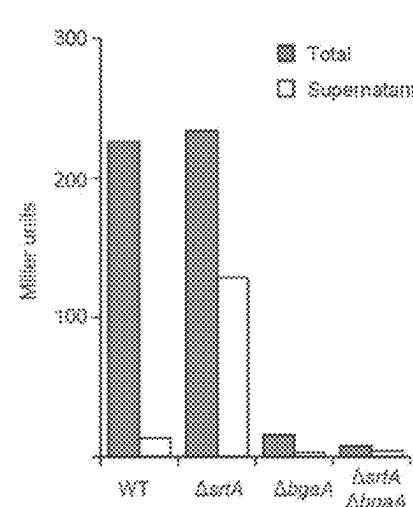


FIG. 34C

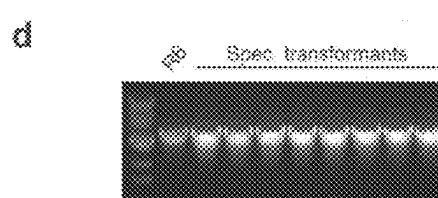


FIG. 34D

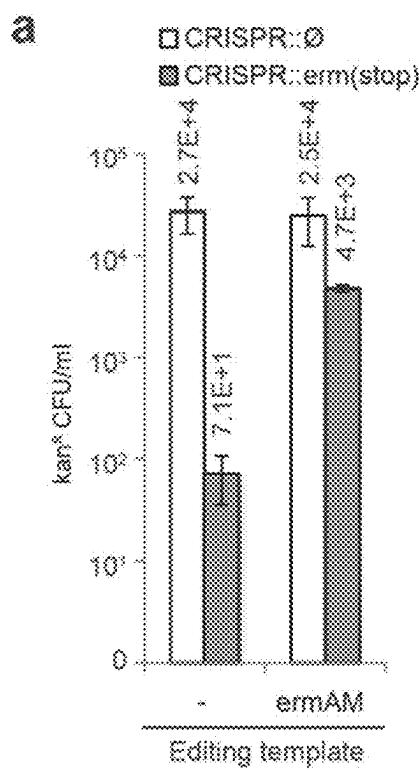


FIG. 35A

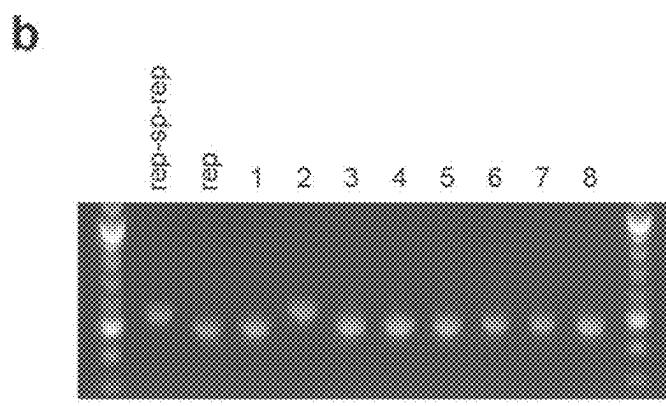


FIG. 35B

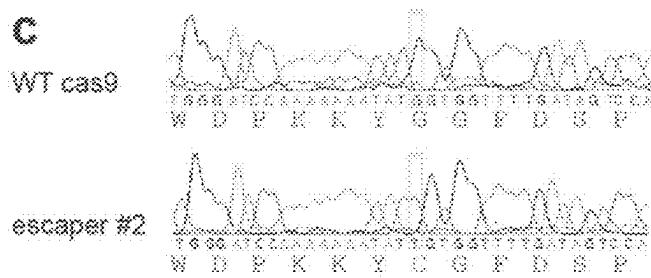


FIG. 35C

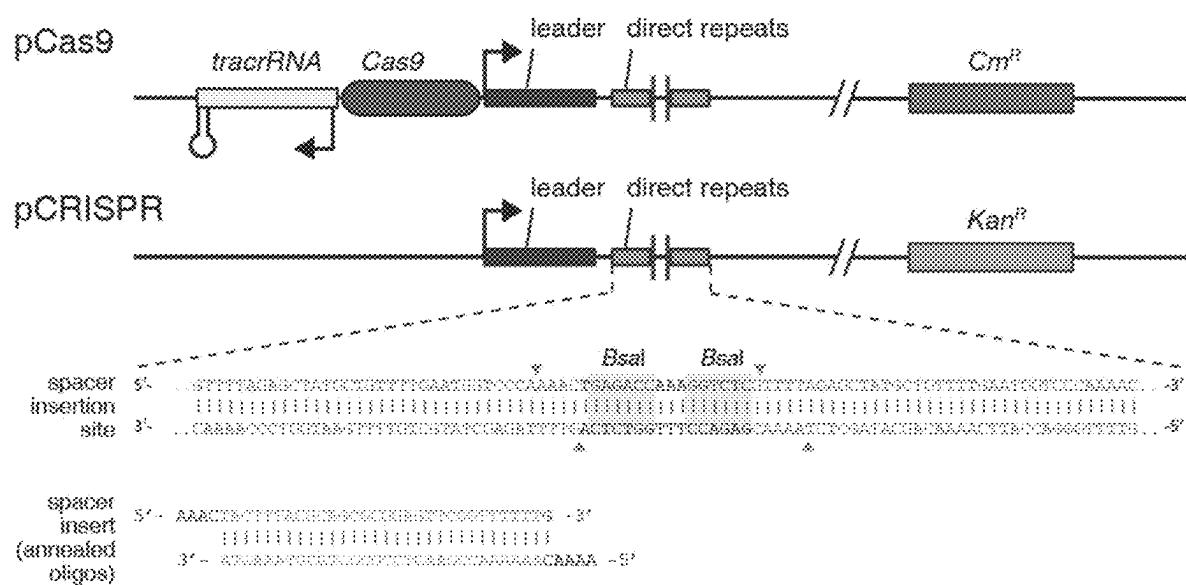


FIG. 36

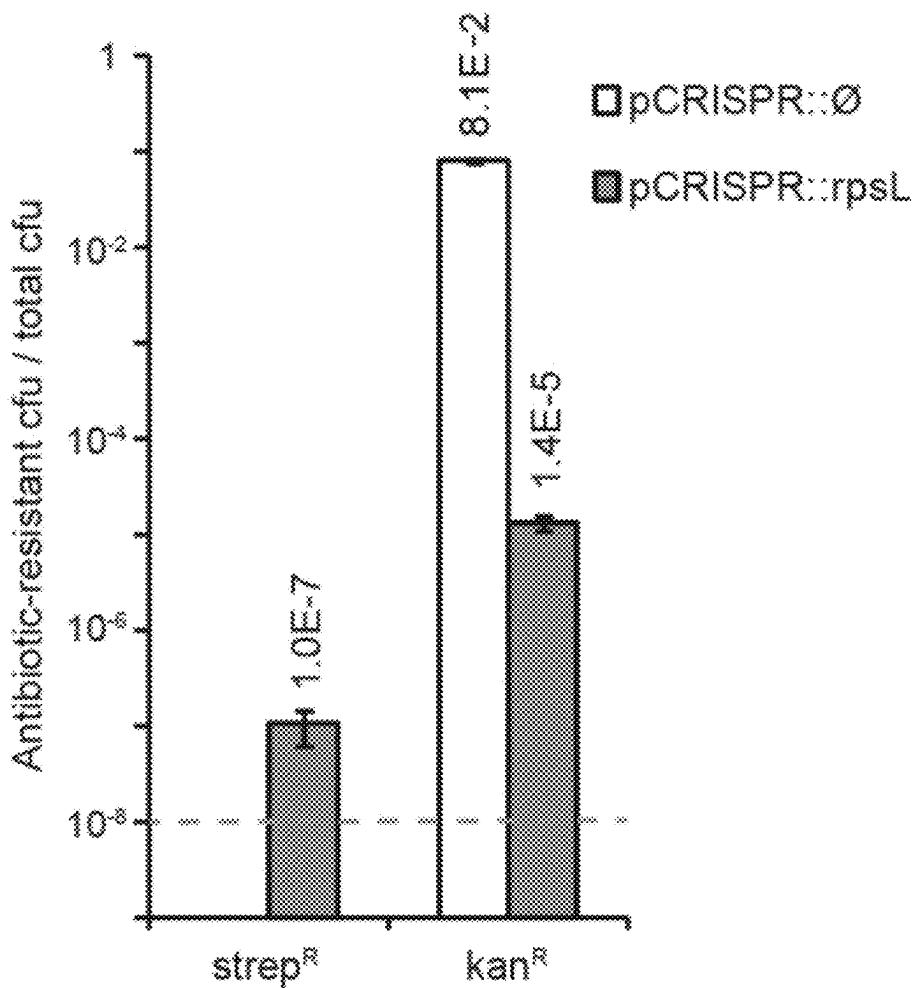


FIG. 37

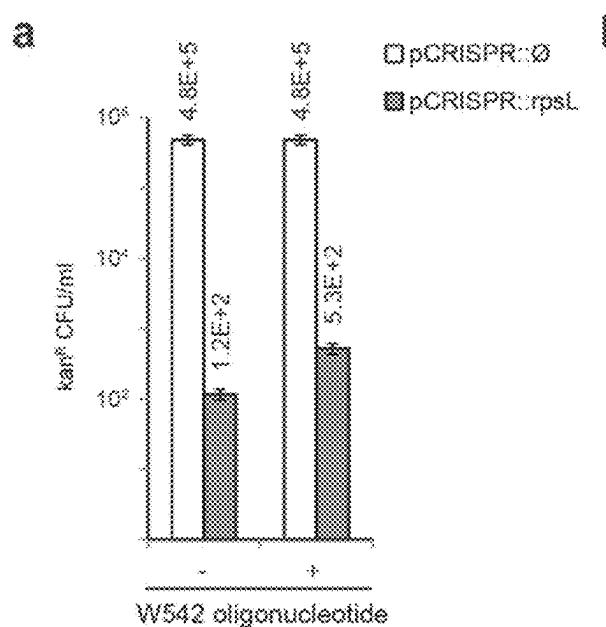


FIG. 38A

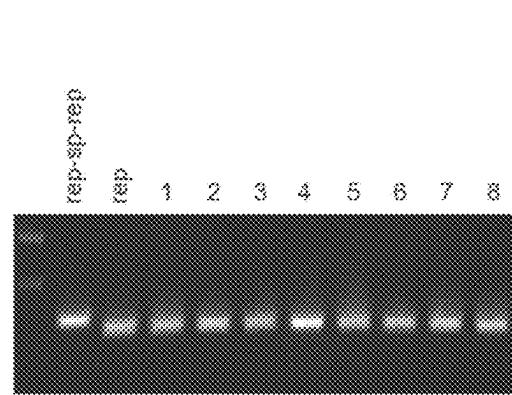
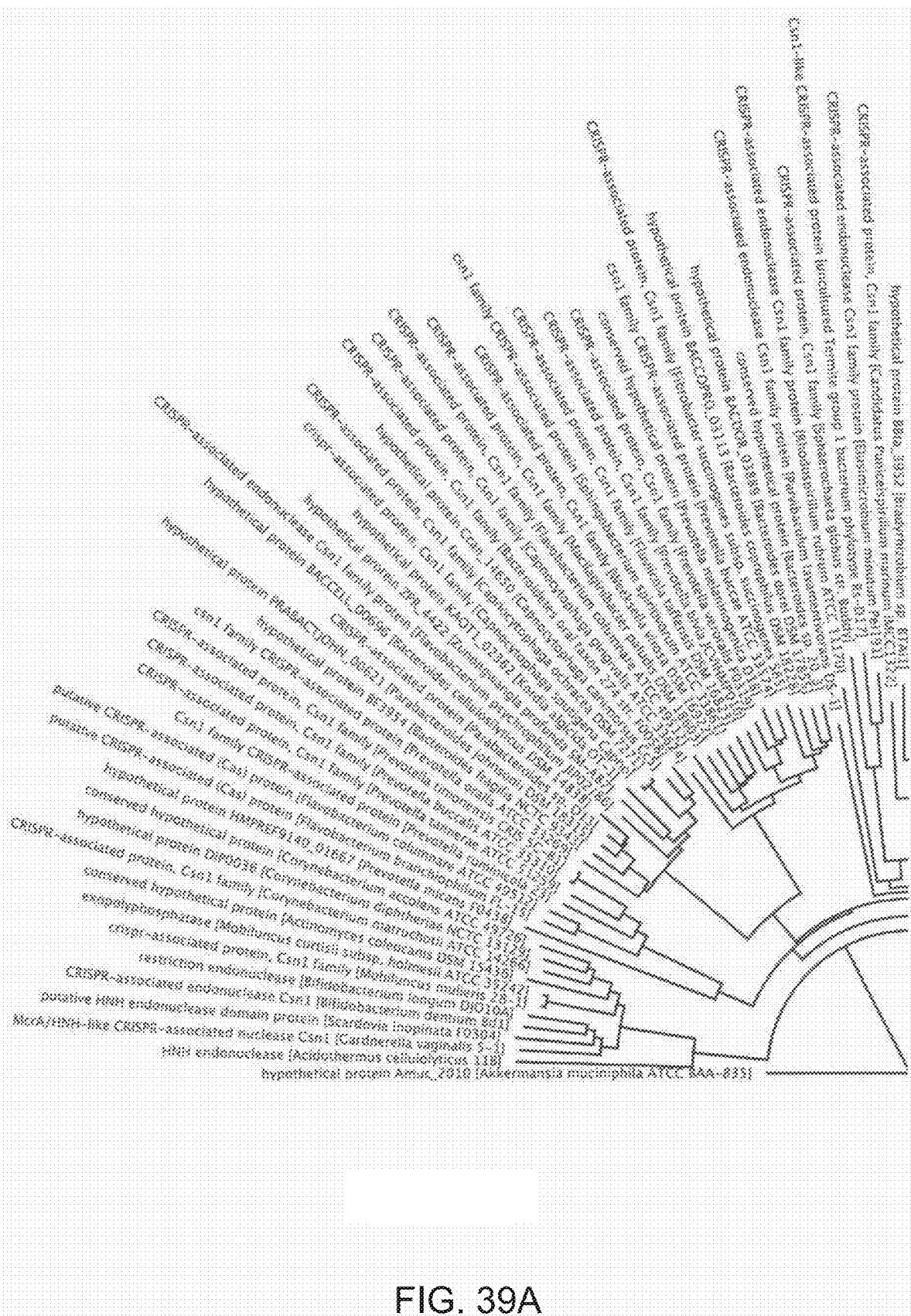
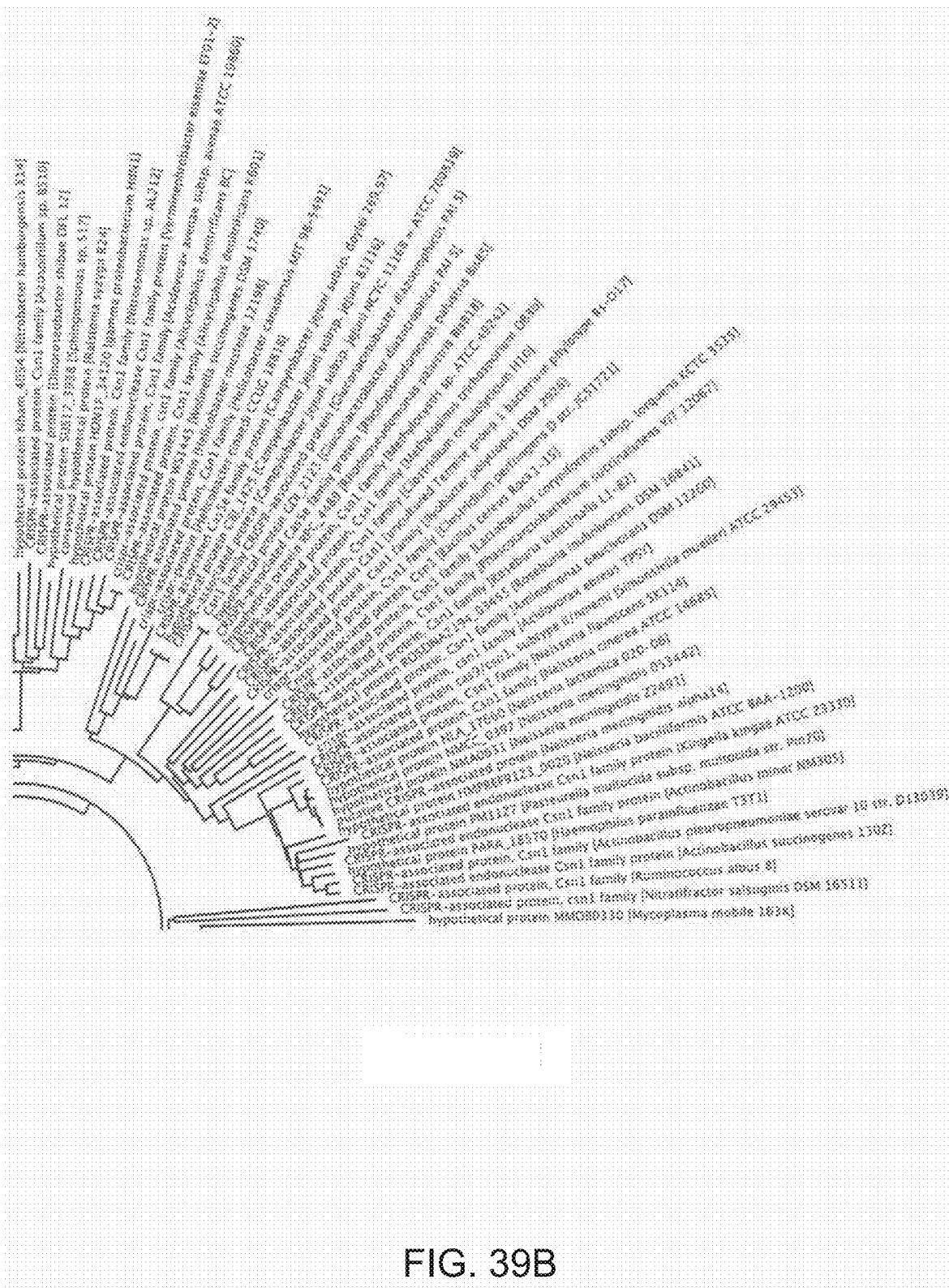
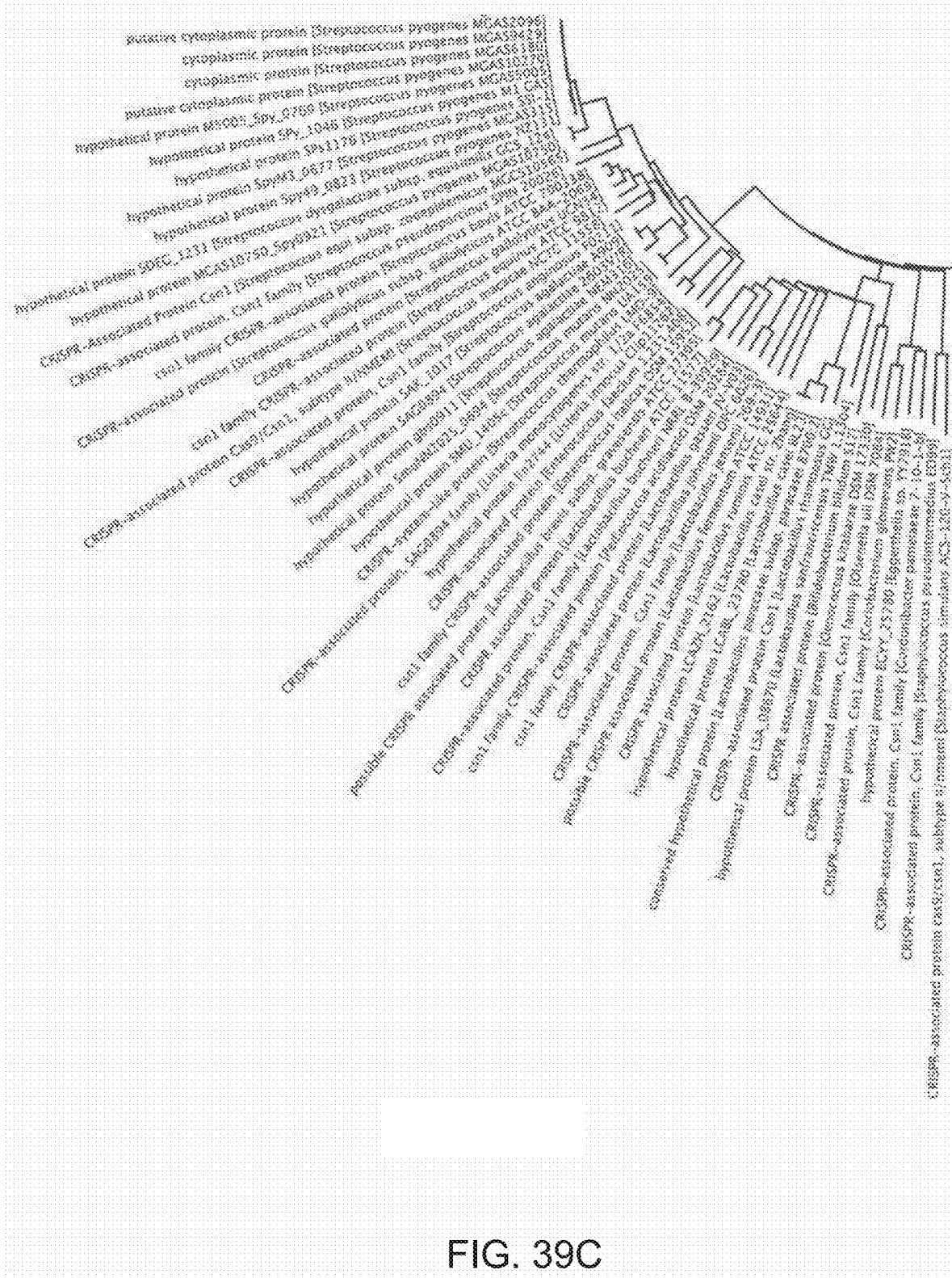


FIG. 38B







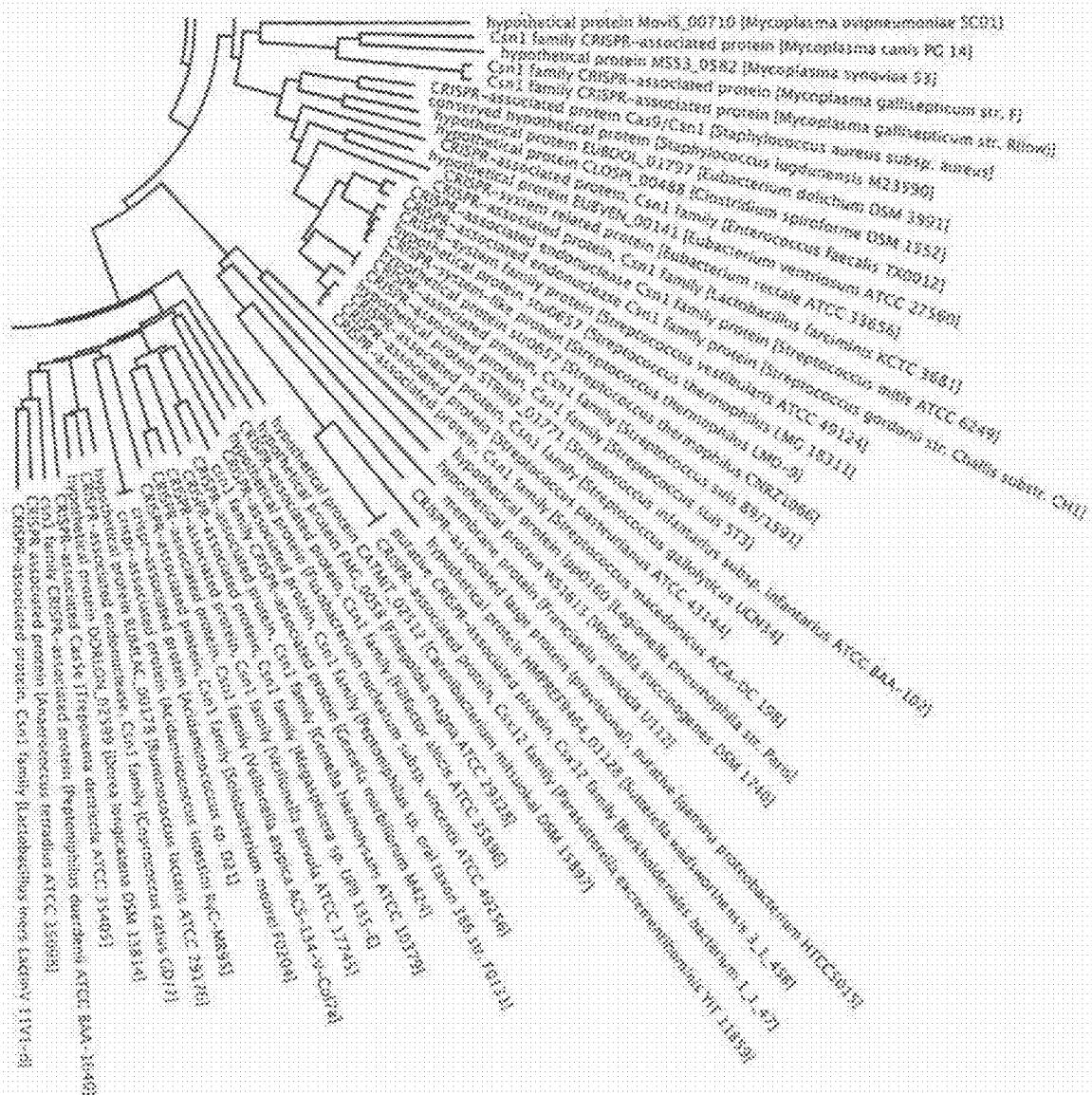
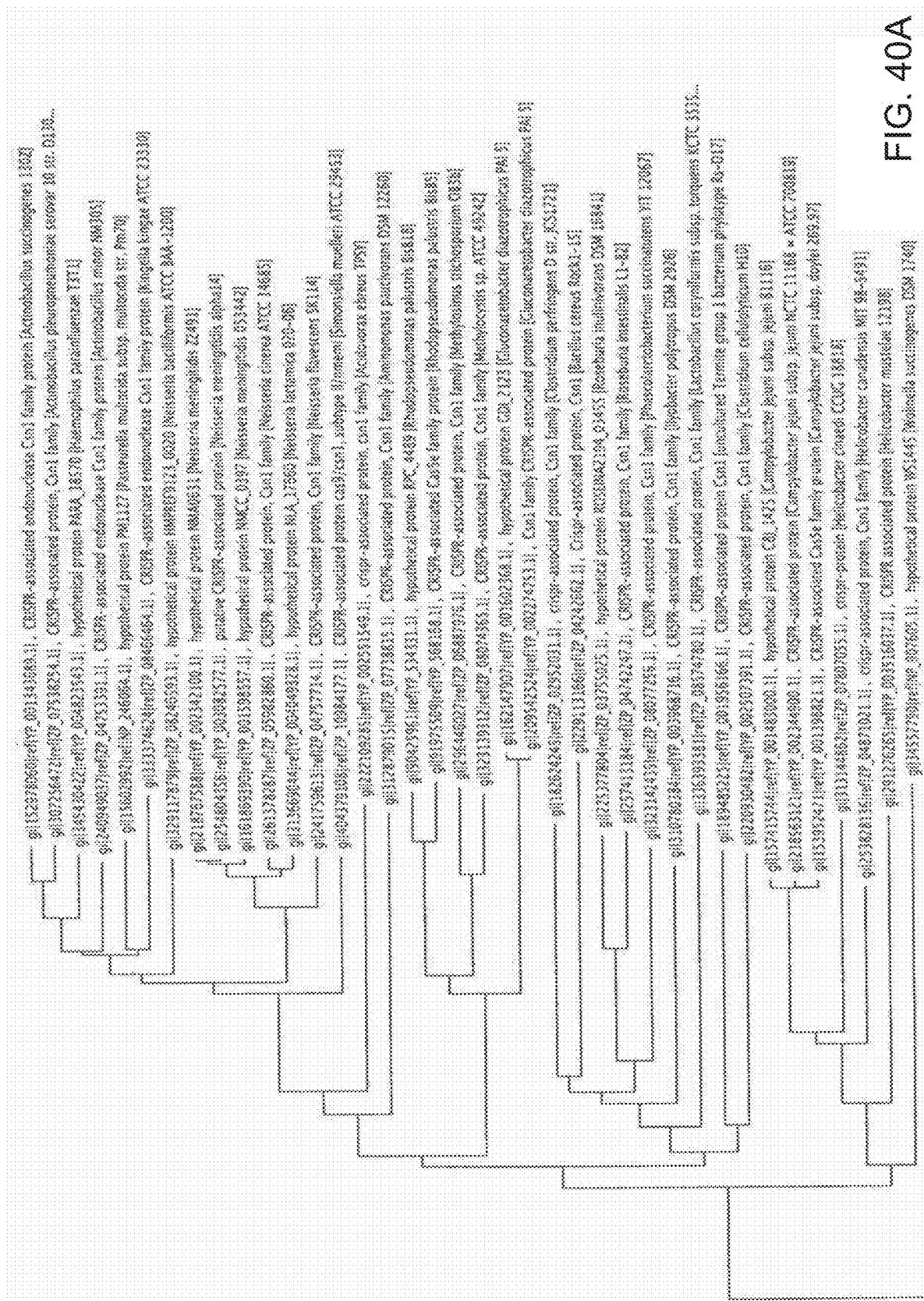
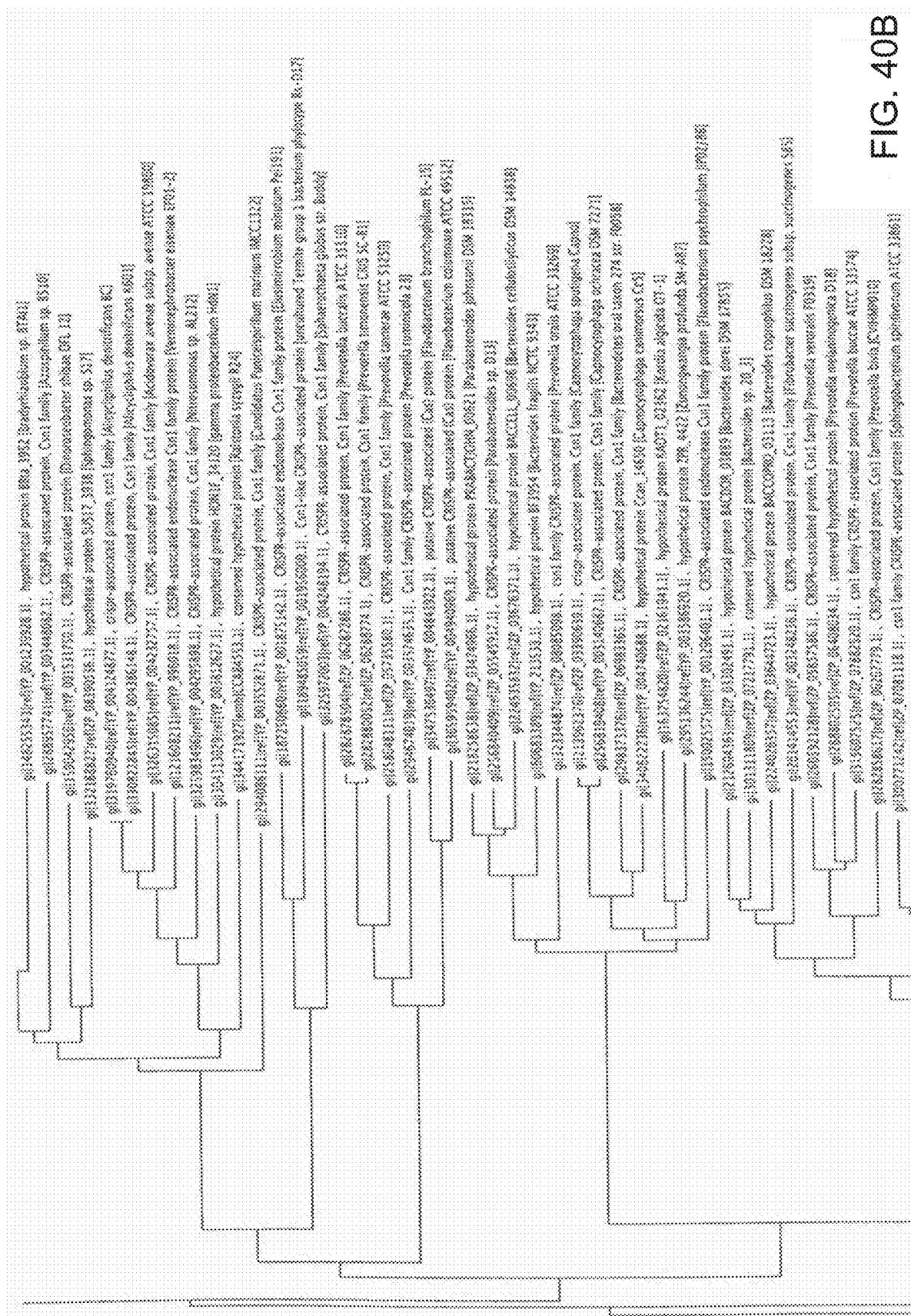


FIG. 39D



EIG. 40A



EIG 40B

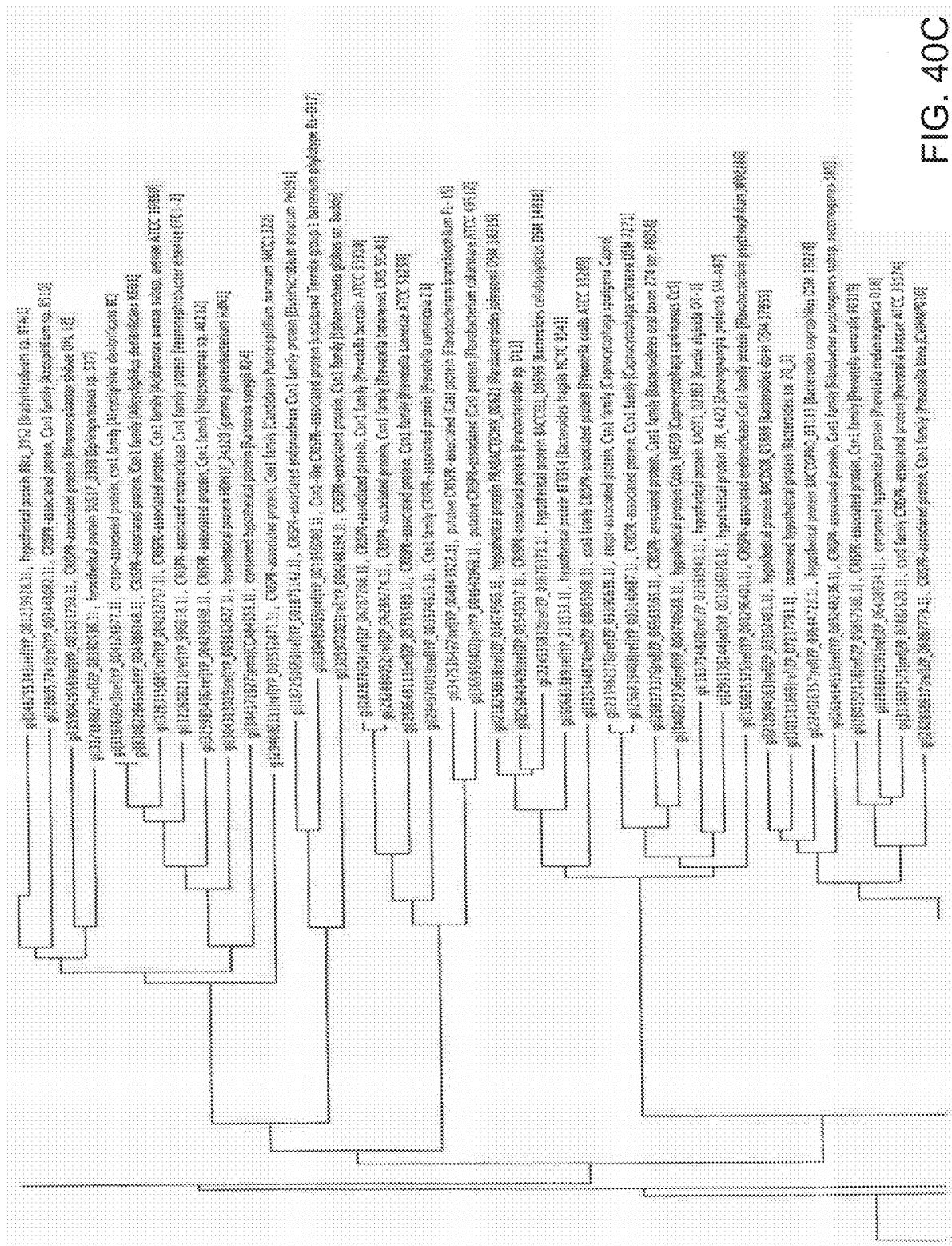


FIG. 40C

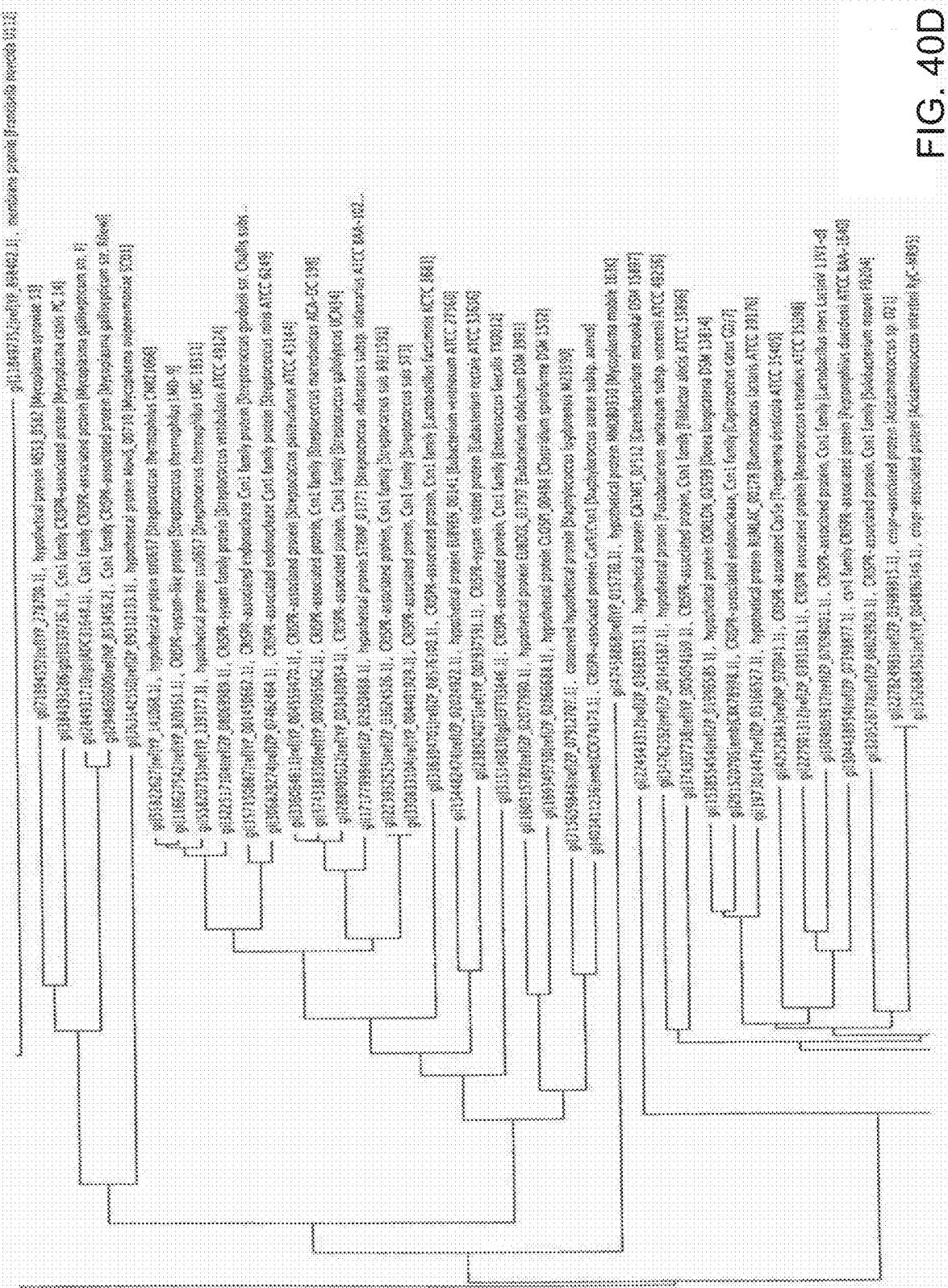
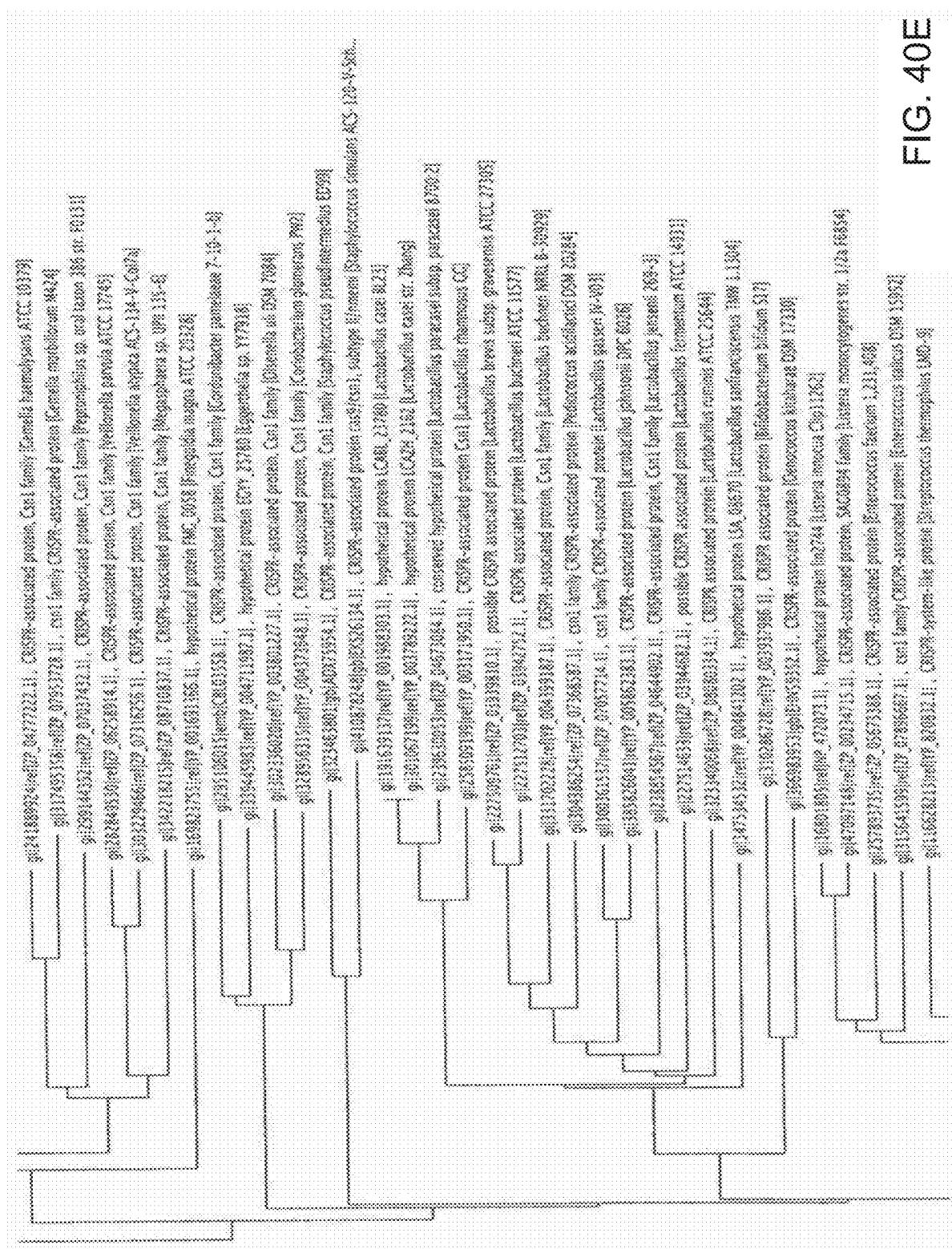


FIG. 40D



EIG. 40E

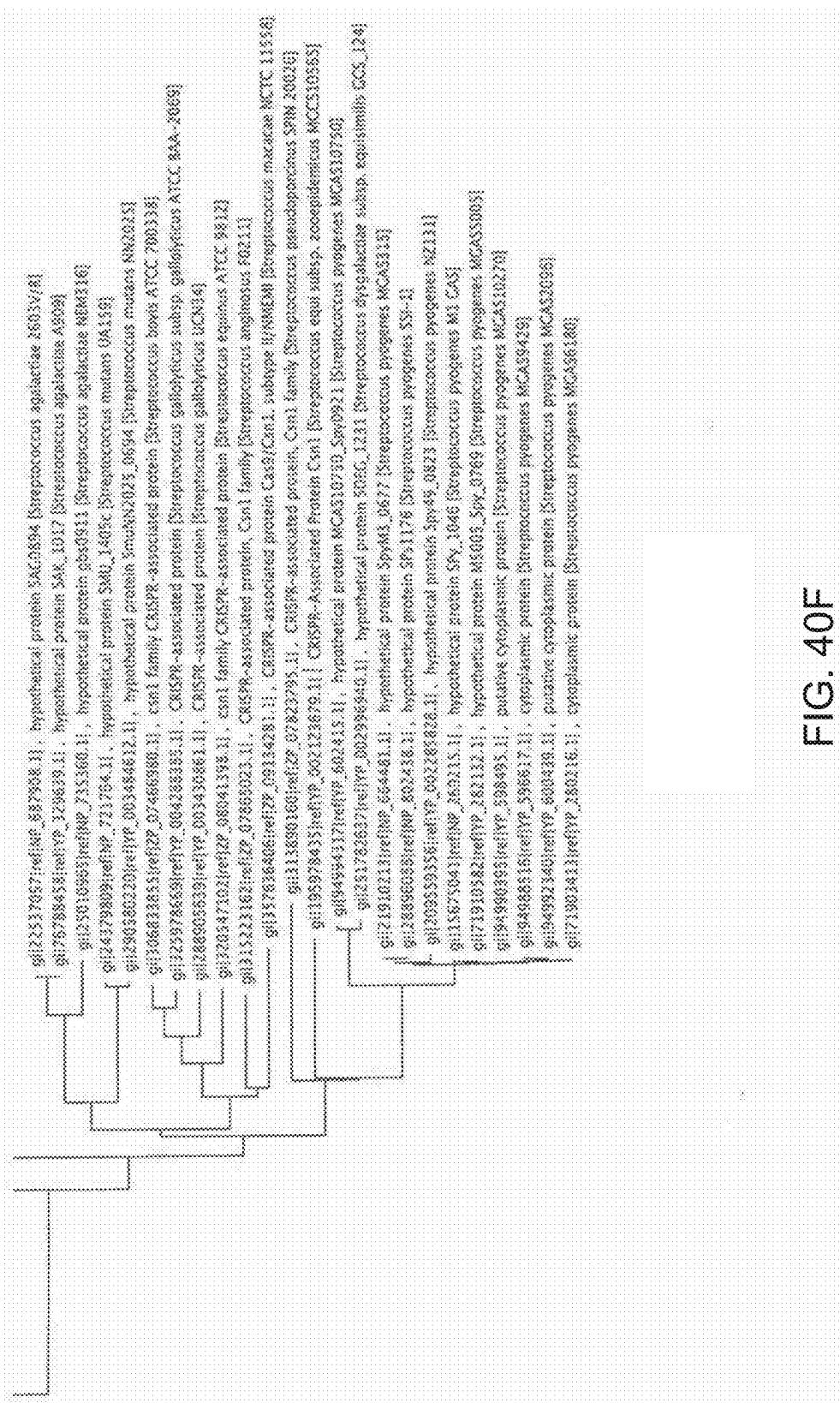


FIG. 40F

SpCas9 mutation positions

hSpCas9

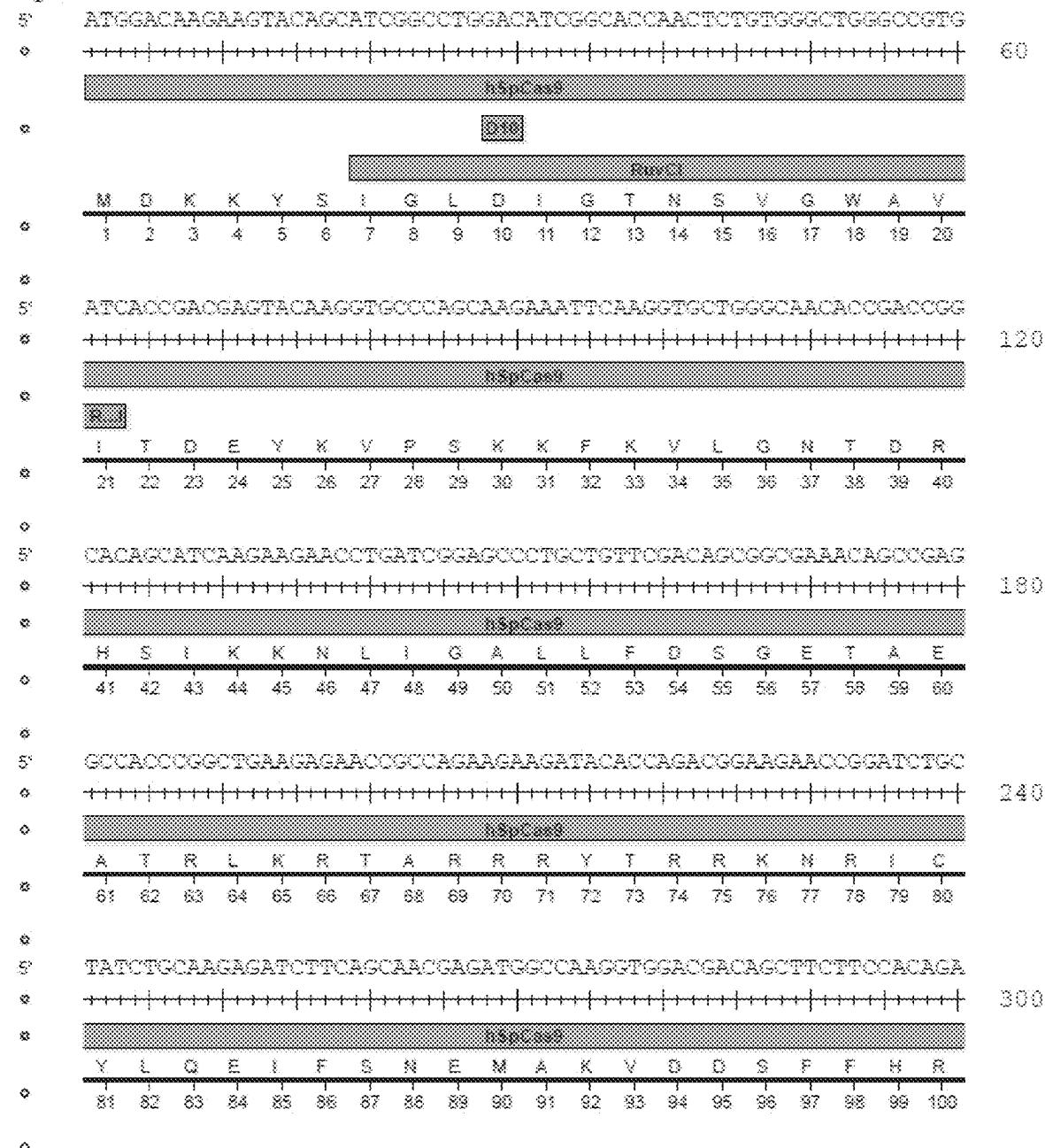


FIG. 41A

FIG. 41B

hSpCas9

5' CGGCTGGAAAAATCTGATCGCCCCAGCTGCCCGCCGAGAAGAAATGCCCTGTTGCCAAC
 * +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 * |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 * |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 * 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' CTGATTGCCCTGAGCCTGGCCTGACCCCGCAACTTCAAGAGCAACTTGACCTGGCCGAG
 * +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 * |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 * |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 * 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' GATGCCAAACTGCAGCTGAGCAAGGACACCTACGAGGACGGACCTGGACAACTGCTGGCC
 * +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 * |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 * |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 * 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' CAGATGGCGACCGAGTACGGCGACCTGTTCTGGCGGCCAAGAACCTGTCCGAAGGUCATC
 * +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 * |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 * |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 * 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' CTGCTGAGCGACATCTGAGACTGAAACAOOGAGATCACCAAGGCCCCCTGAGCGCTCT
 * +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 * |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 * |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 * L L S D I L R Y 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' ATGATCAAGAGATAACGAGGACCCACGGACCTGACCCCTGCTGAAACCTCTCGTCGG
 * +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 * |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 * |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 * 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. 41C

FIG. 41D

FIG. 41E

FIG. 41F

FIG. 41G

FIG. 41H

FIG. 411

FIG. 41J

hSpCas9
3420
 5' CCCAGAAAAGAACCGACTGGGACCCUTAAGAACTAACGGGGGTTTGGACACCCCCACACCGTGGC
 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140
 A R K K Q W D P K N Y G G F D S P T V A

3480
 5' TATTCTGTGCTCGTGGTGGGCCAAAGTGGAAAAGGGCAAGTCGAAGAAACTGAAGAGTG
 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160
 Y S V L V V A K V E K G K S K K L K S V

3540
 5' AAAGAGCTGCTGGCATCACCATCATGGAAAGAACCGCTTOGAGAGAACATCCATCGAC
 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180
 K E L L G I T I M E R S S F E K N P I D

3600
 5' TTCTCTGGAAGCCAAAGGGCTACAAAGAACTGAAAAAGGACCTGATCATCAAAGCTGGCUTAAG
 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
 F L E A K G Y K E V K K D L I I K L P K

3660
 5' TACTCCCTGTTGACCTGGAAAACGGGCGGAAAGAGAAATGCTGGCCTCTGGCGGGGAACTG
 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220
 Y S L F E L E N G R K R M L A S A G E L

3720
 5' CAGAAGGGAAACGAACTGGCCTGGCCTGCATAATATGTAACCTTCCTGTAACCTGGCAGC
 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240
 Q K G N E L A L P S K Y V N F L Y L A S

FIG. 41K

bSpCas9

5' CACTATGAGAACCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGGTCTTTGTGAA
 3780

H Y E K L K G S P E D N E Q K O 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' CAGGCACACAGGACTACCTGGACCCAGATCATGGAGATCAGGCCAGAGTTCTCCAAAGAGACTG
 3840

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' ATCTGGCCGACGCTAACTGGACAAAGTCTCTCCGCTACACAAAGCACCGGGATAAG
 3900

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' CCCATCACAGAGGAGGCGAGATATCATCCACCTGTTACCCCTGACCAAATCTGGGAGCC
 3960

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' CCTGUCGCGCTTCAGTACTTTGACACCACCATGGAAAGAGGGTACACCCAGCAACAAA
 4020

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' GAGGTGCTGGACGCCACCCGTGATGCCAGAGCATCACCGGCGTGTACGAGACACGGATC
 4080

E V L D A T L I H G 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. 41L

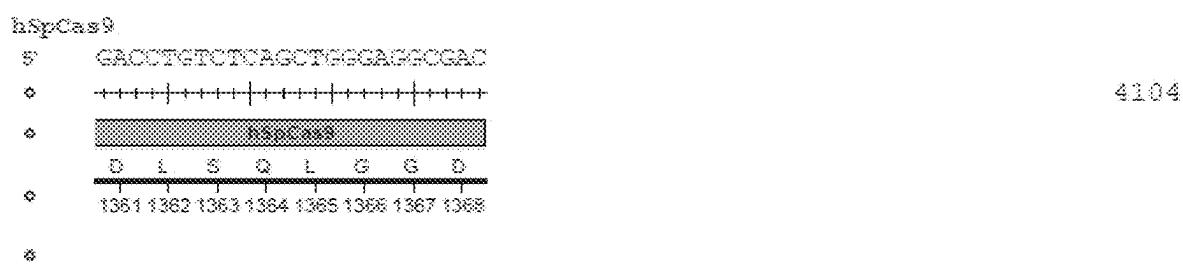


FIG. 41M

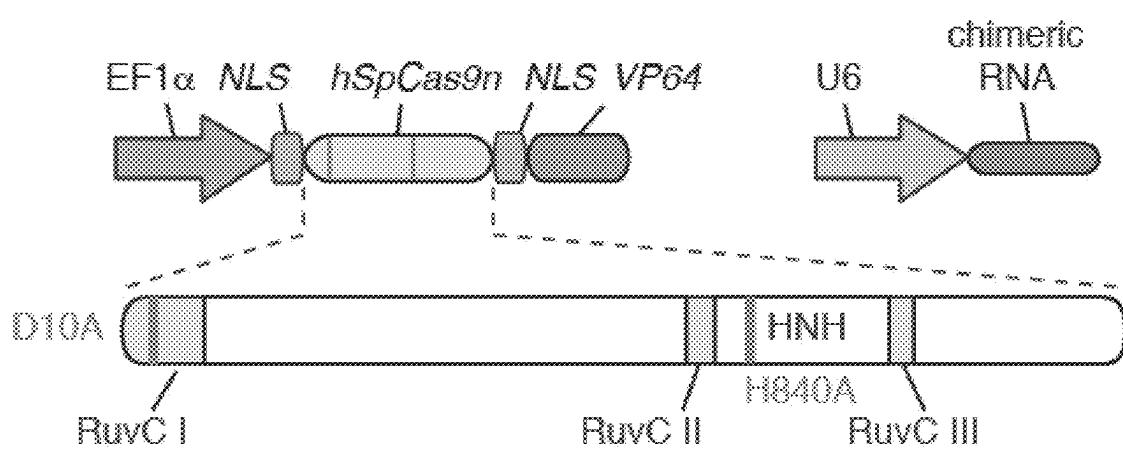


FIG. 42

a

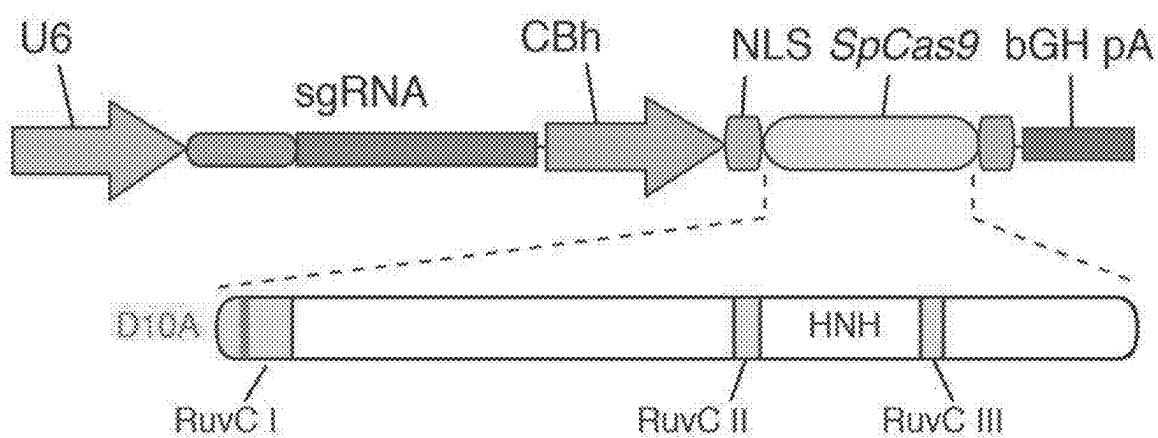


FIG. 43A

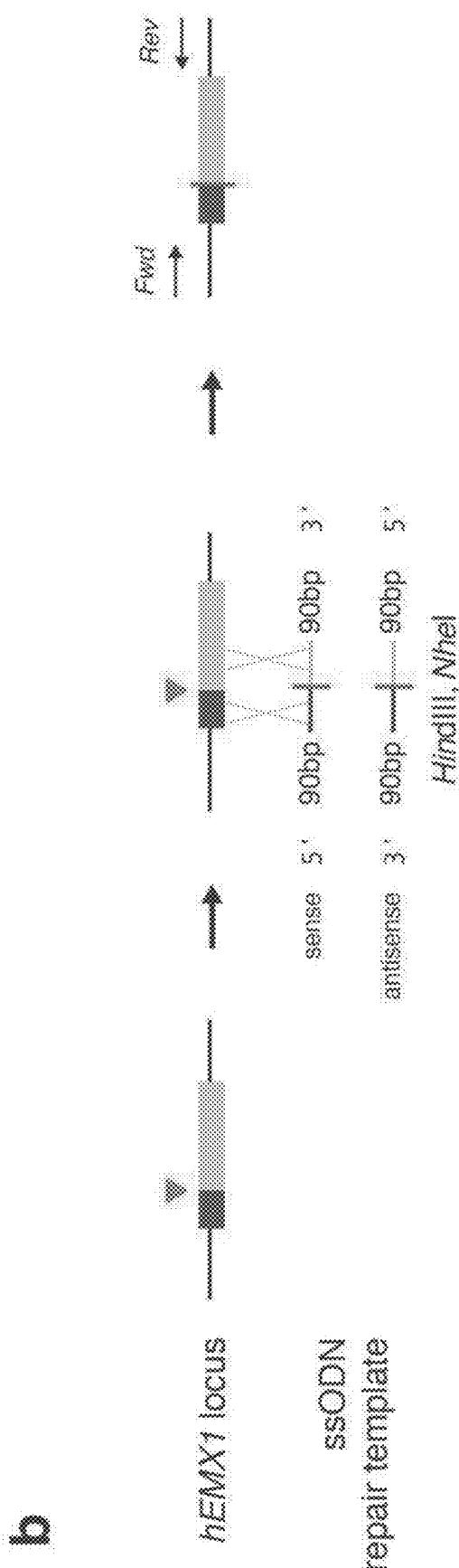


FIG. 43B

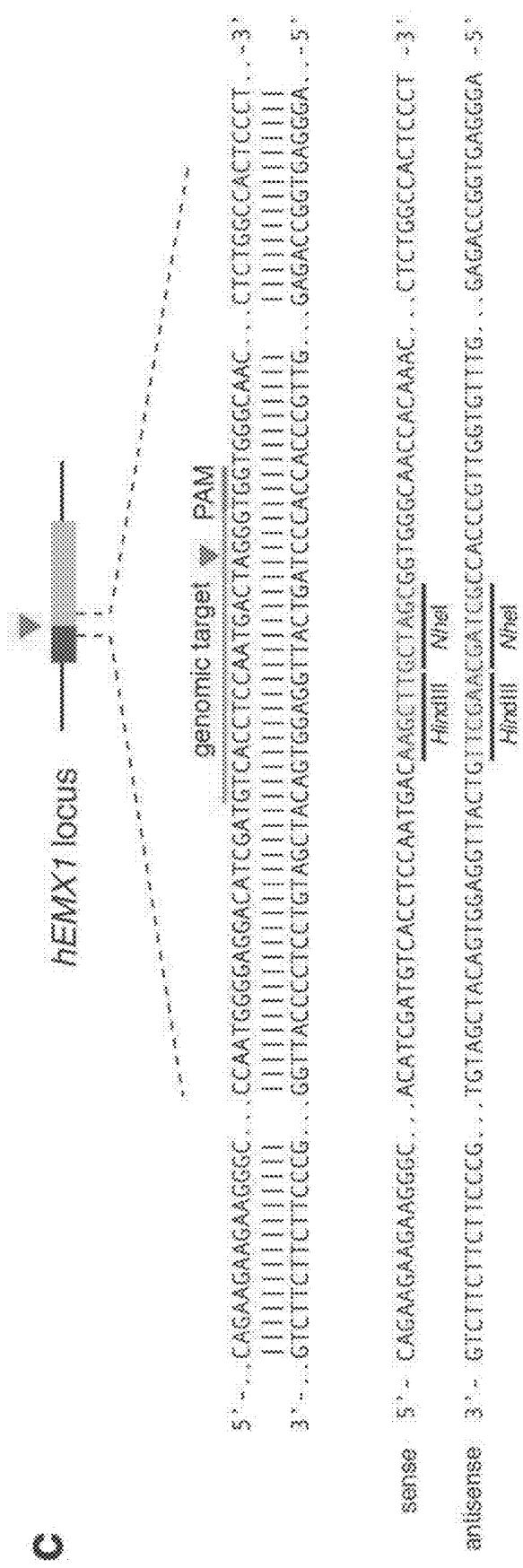


FIG. 43C

d

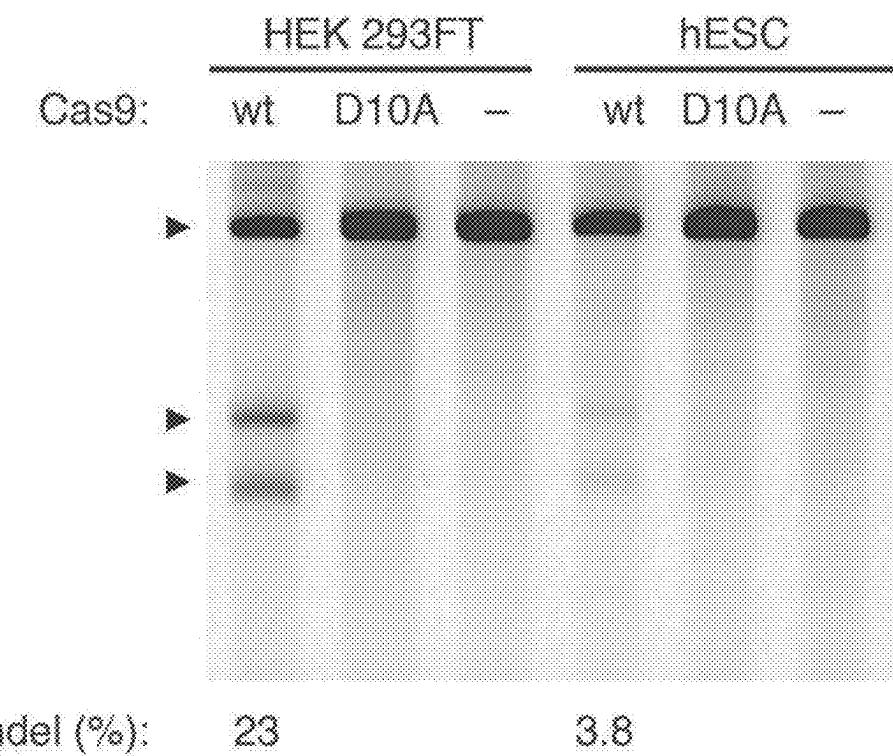


FIG. 43D

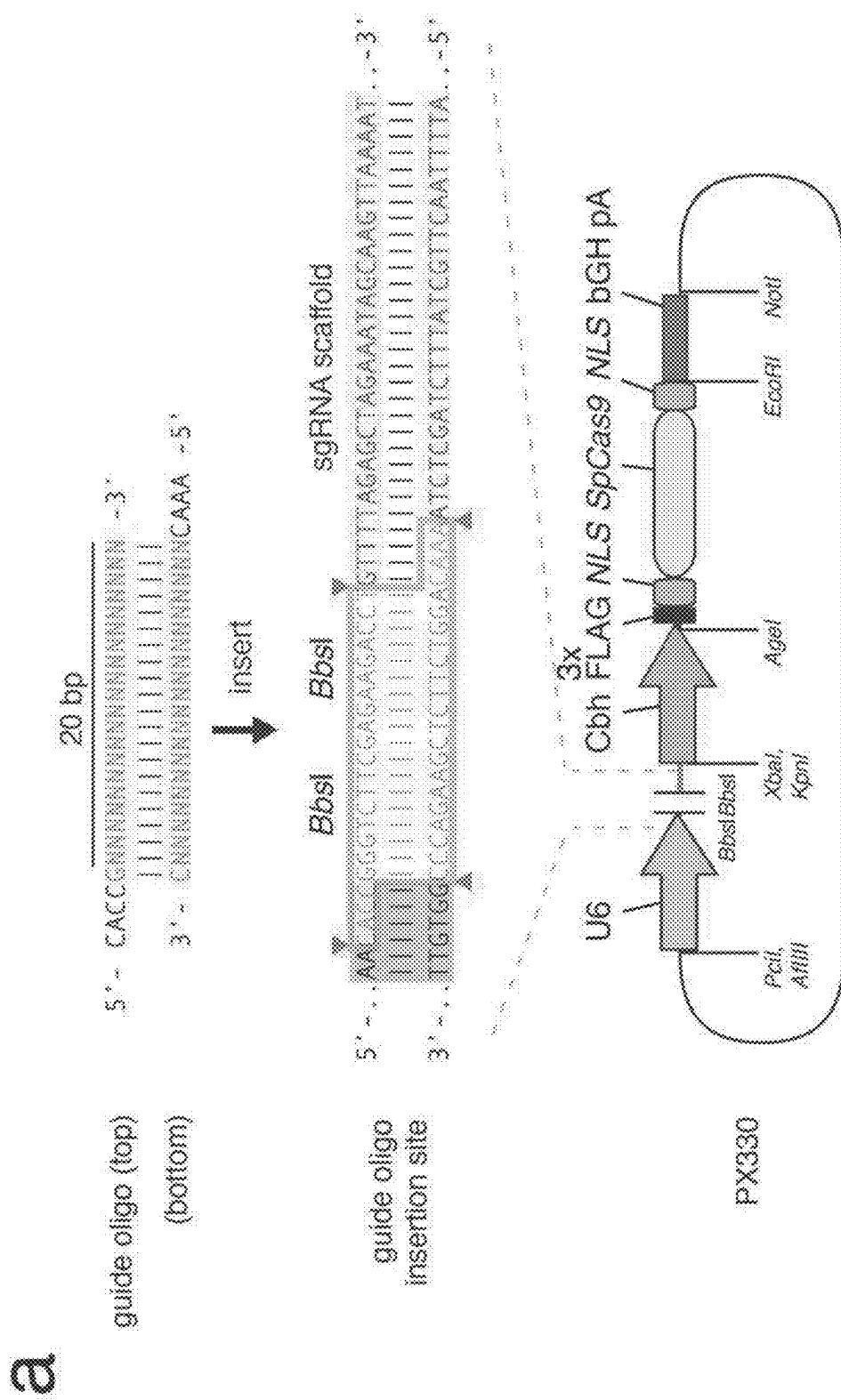


FIG. 44A

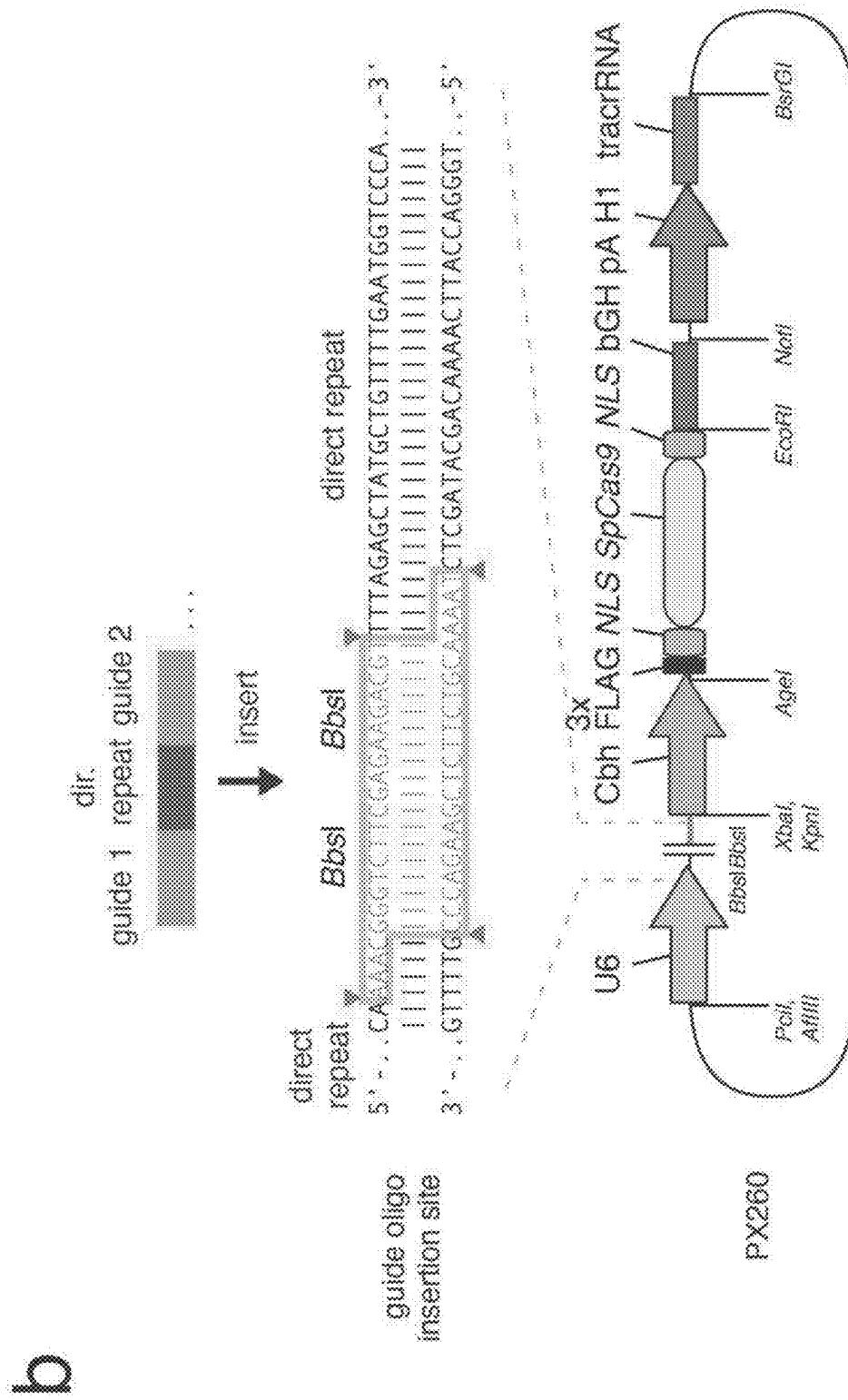


FIG. 44B

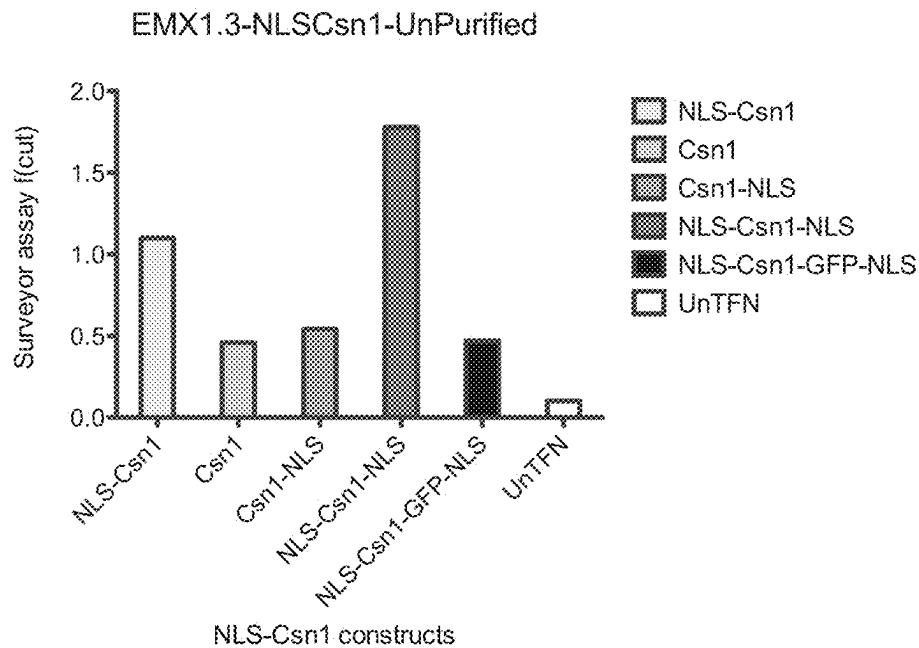


FIG. 45

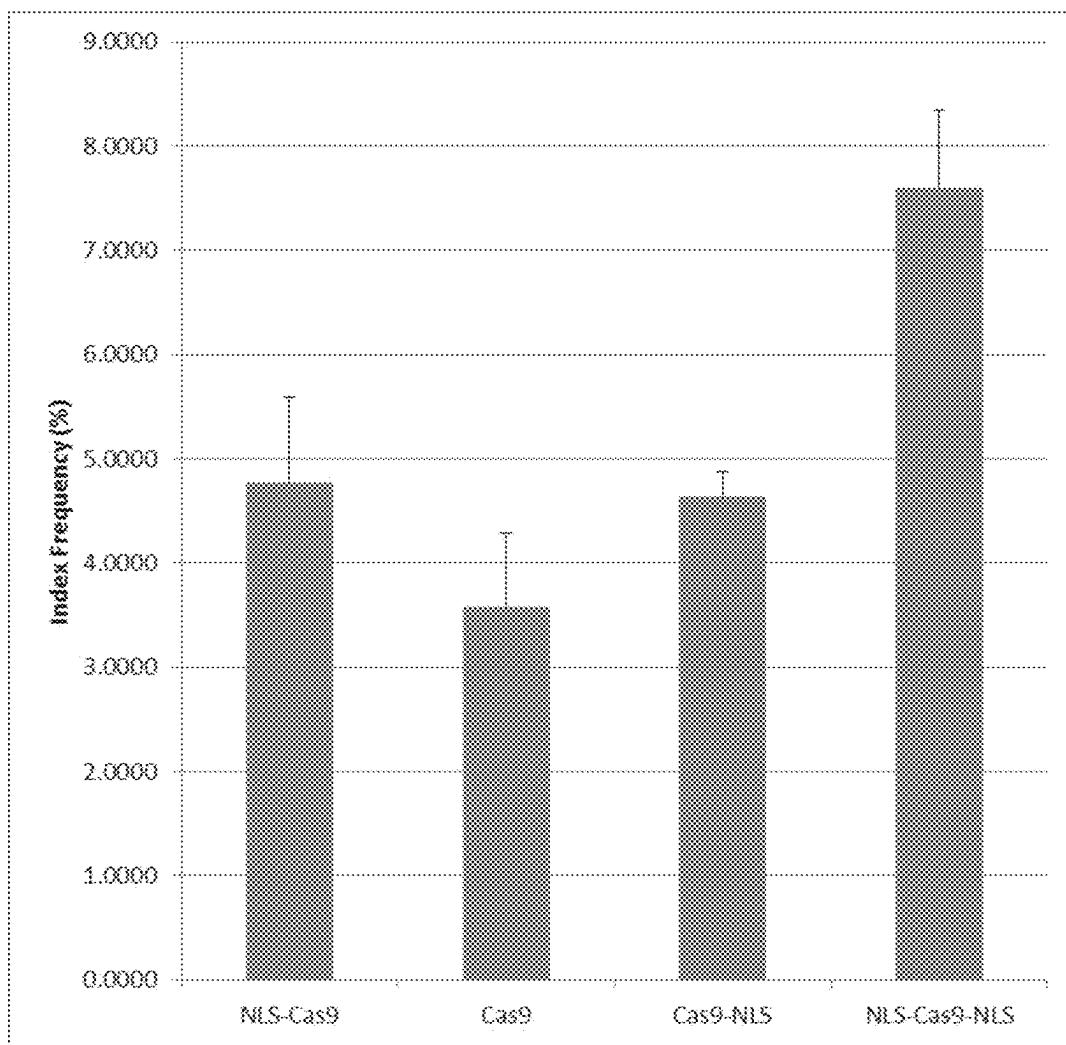


FIG. 46

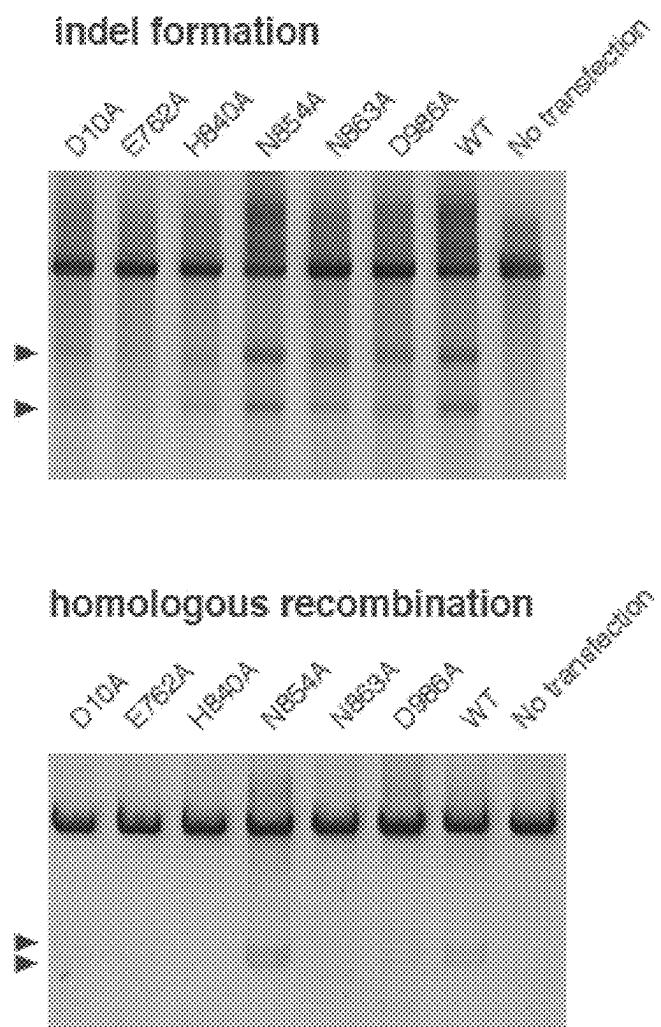


FIG. 47

A

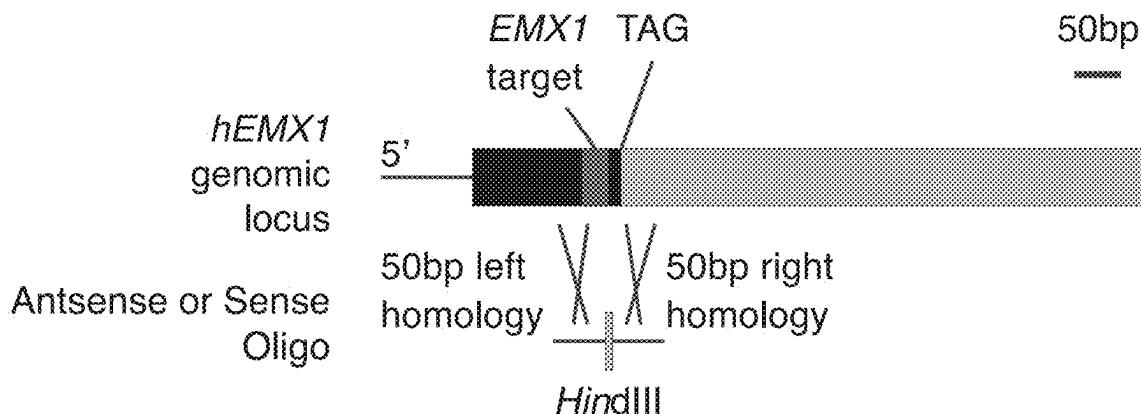


FIG. 48A

B

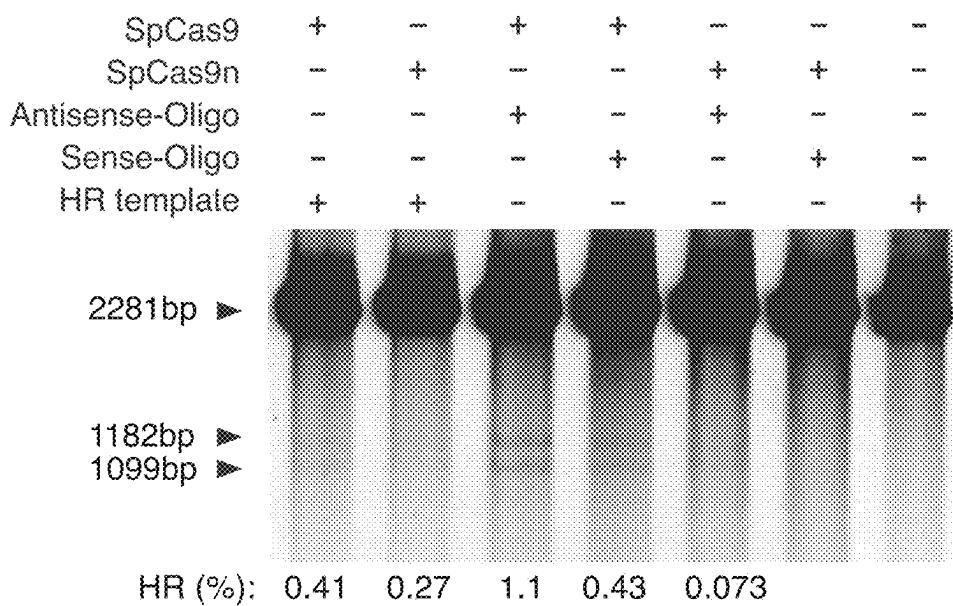


FIG. 48B

Conditional Cas9, Rosa26 targeting vector map

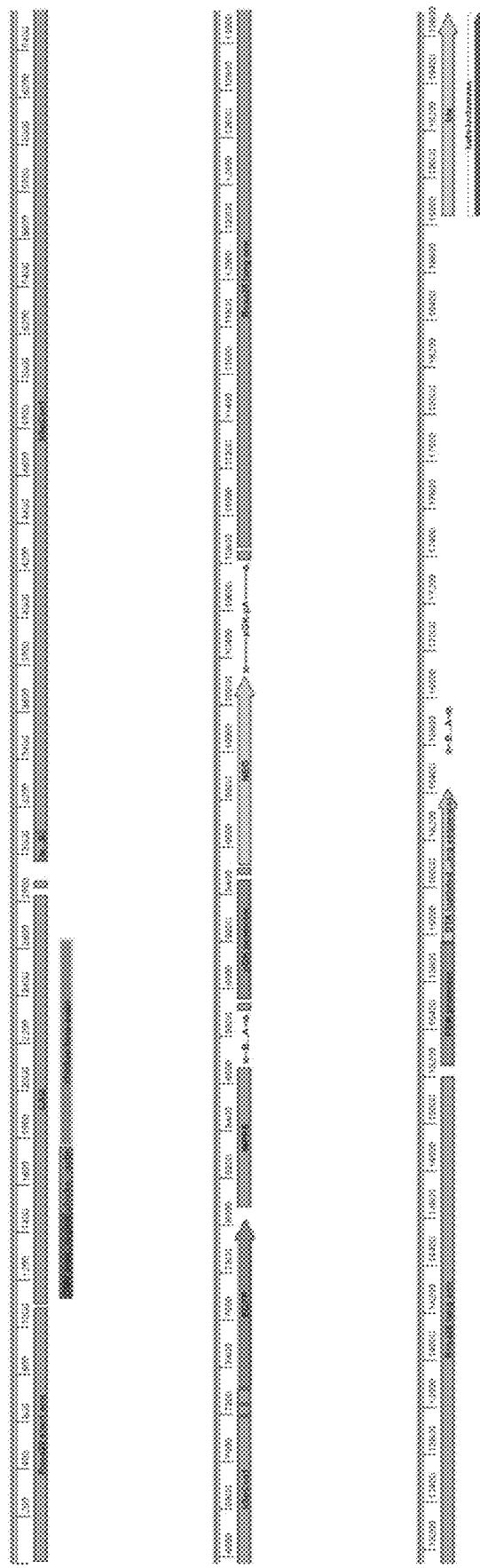


FIG. 49A

Constitutive Cas9, Rosa26 targeting vector map

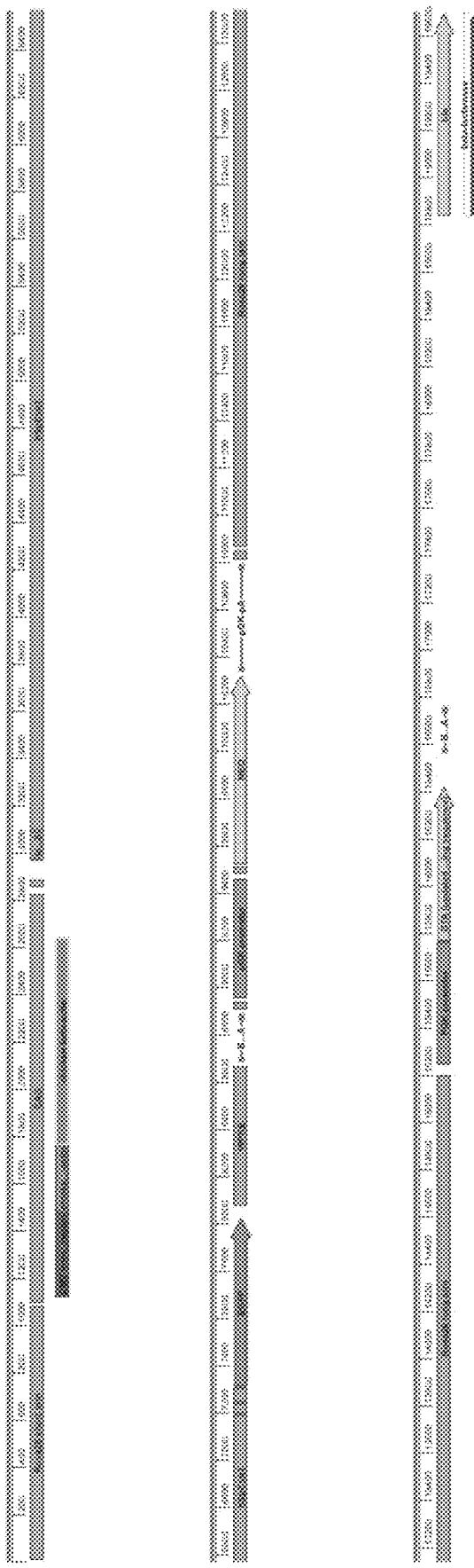


FIG. 49B

pCAG

NLS-Cas9-NLS

ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGA
CGATGACGATAAGATGGCCCCAAAGAAGAAGCAGGGAGGTGGTATCCACGGAGTCC
CAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGC
TGGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGG
CAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTGACA
GCGGCAGAAACAGCCGAGGCCACCCGGTGAAGAGAACCGCCAGAAGAAGATAACAC
CAGACCGGAAGAACCGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCA
AGGTGGACGACAGCTTCTCCACAGACTGGAAGAGTCCTCTGGTGGAAAGAGGAT
AAGAACGACGAGCGGCACCCCATCTCGGCAACATCGTGGACGAGGTGGCCTACCA
CGAGAAAGTACCCCACCATCTACCCACCTGAGAAAGAAACTGGTGGACAGCACCGACA
AGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTCCGGGCC
ACTTCTGATCGAGGGCGACCTGAACCCCCGACAAACAGCGACGTGGACAAAGCTGTTC

FIG. 50A

ATCCAGCTGGTGCAGACCTACAACCAGCTGTTGAGGAAAACCCATCAACGCCAG
CGCGTGGACCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGG
AAAATCTGATGCCAGCTGCCCGCGAGAAGAAGAATGGCCTGTCGGAACCTG
ATTGCCCTGAGCTGGGCTGACCCCCAACTTCAAGAGCAACTCGACCTGGCCGAG
GATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCT
GGCCCAGATCGCGACCACTACGCCGACCTGTTCTGGCCAGAACCTGTCCG
ACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCCGAGATCACCAAGGCCCC
CTGAGCGCCTCATGATCAAGAGATAACGACGAGCACCACCGACCTGACCCCTGCT
GAAAGCTCTCGTGCAGCAGCTGCCCTGAGAAGTACAAAGAGATTTCTCGAC
AGAGCAAGAACGGTACGCCGCTACATTGACGGCGAGCCAGGAAGAGATTC
TACAAGTTCATCAAGCCCATCCTGAAAAGATGGACGGACCCGAGGAAC TGCTCGT
GAAGCTGAACAGAGAGGACCTGCTCGAAGCAGCGGACCTTGACACAACGGCAGC
ATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCCGGCAGGAAGA
TTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTCCG
CATCCCCCTACTACGTGGCCCTCTGCCAGGGAAACAGCAGATTGCCCTGGATGA
CCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTCGAGGAAGTGGTGGACAA
GGCGCTTCCGCCAGAGCTTATCGACGGATGACCAACTTCGATAAGAACCTGC
CCAACGAGAAGGTGCTGCCAAGCACAGCCTGCTGTACGAGTACTCACCGTGTAT
AACGAGCTGACCAAAGTGAATACGTGACCGAGGGAAATGAGAAAGCCGCCCTC
GAGCGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAA
GTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTCGACTC
CGTGGAAATCTCCGGCGTGGAAAGATCGTTCAACGCCCTGGCACATACACG
ATCTGCTGAAAATTATCAAGGACAAGGACTTCTGGACAATGAGGAAAACGAGGAC
ATTCTGGAAGATATCGTGTGACCCCTGACACTGTTGAGGACAGAGAGATGATCGA
GGAACGGCTGAAAACCTATGCCACCTGTCGACGACAAAGTGTGAGCAGCTGA
AGCGGCGGAGATAACCCGGCTGGGAGGCTGAGCCGGAAAGCTGATCAACGGCATC
CGGGACAAGCAGTCCGCAAGACAATCCTGGATTCTGTAAGTCCGACGGCTTCGC
CAACAGAAAATTCATGCAGCTGATCCACGACGACAGCCTGACCTTAAAGAGGACA
TCCAGAAAAGGCCAGGTGCTGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAAT
CTGGCCGGCAGCCCCGCCATTAAAAGAGGGCATCCTGCAGACAGTGAAGGTGG
CGAGCTCGTGAAGTGTGAGGGCGACAAGCCCGAGAACATCGTGTGAAATGG
CCAGAGAGAACCAGACCAACCGAGAAGGGACAGAACAGCCCGAGAGAACATGAA
GCGGATCGAAGAGGGCATCAAAGAGCTGGGAGCCAGATCCTGAAAGAACACCCCC
GTGGAAAACACCCAGCTGCAAGACGAGAAGCTGTACCTGTACTACCTGCA
GCGGGATATGTACGTGGACCAAGGAACCTGGACATCAACCCGCTGTCGACTACGAT

FIG. 50B

GTGGACCATATCGTGCCTCAGAGCTTCTGAAGGACGACTCCATCGACAACAAGGTG
CTGACCAGAAGCGACAAGAACCGGGCAAGAGCGACAAACGTGCCCTCCGAAGAGG
TCGTGAAGAAGATGAAGAACTACTGGCGGAGCTGCTGAACGCCAAGCTGATTACC
CAGAGAAAAGTCGACAACTTGACCAAGGCCAGAGAGAGGCCCTGAGCGAACTGG
ATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCAC
GTGGCACAGATCCTGGACTCCCCGATGAACACTAAGTACGACGAGAATGACAAGCT
GATCCGGGAAGTGAAGTGATCACCCCTGAAGTCCAAGCTGGTGTCCGATTCCGGA
AGGATTCCAGTTTACAAAGTGCAGAGATCAACAACCTACCACACGCCACGAC
GCCTACCTGAACGCCGTGTTGGAAACGCCCTGATCAAAAAGTACCTAAGCTGGA
AAGCGAGTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATGCCA
AGAGCGAGCAGGAAATCGGAAGGCTACGCCAAGTACTTCTTACAGCAACATC
ATGAACCTTCAAGACCGAGATTACCCCTGGCAACGGCGAGATCCGGAAGCGGCC
TCTGATCGAGACAAACGGCAAACCGGGAGATCGTGTGGATAAGGGCCGGATT
TTGCCACCGTGGAAAGTGTGAGCATGCCCAAGTGAATATCGTAAAAAGACC
GAGGTGCAGACAGGCCGCTTCAGCAAAGAGTCTATCCTGCCAAGAGGAACAGCGA
TAAGCTGATGCCAGAAAGAAGGACTGGGACCTAAGAAGTACGGCGGCTCGACA
GCCCAACCGTGGCTATTCTGTGCTGGTGGTGGCAAAGTGGAAAAGGGCAAGTCC
AAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGATCACCATCATGGAAAGAAGCA
GCTTCGAGAAGAATCCCATCGACTTCTGGAAAGCCAAGGGCTACAAAGAAGTGA
AAGGACCTGATCATCAAGCTGCTTAAGTACTCCCTGTCAGCTGGAAAACGGCCG
GAAGAGAAATGCTGGCTCTGCCGCAACTGCAAGAGGAAACGAACCTGGCCCTGC
CCTCCAAATATGTGAACCTCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGT
CCCCCGAGGATAATGAGCAGAAACAGCTGTTGTGGAACAGCACAAGCACTACCTG
GACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGTACCTGGCCACGC
TAATCTGGACAAAGTGTGCTGCCCTACAACAAGCACCCGGATAAGCCCACAGAG
AGCAGGCCGAGAATATCATCCACCTGTTACCCCTGACCAATCTGGGAGCCCTGCCG
CCTTCAAGTACTTGTACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAG
GTGCTGGACGCCACCTGATCCACCAAGAGCATCACCGCCCTGTACGAGACACGGAT
CGACCTGTCTCAGCTGGAGGGACAAAAGGCCGGGCCACGAAAAAGGCCGGC
CAGGCAAAAAGAAAAAG

P2A-EGFP

ggaagcggaggccactaacttctccctgttcaaacaagcaggggatgtcgaaagagaatccggccaGTGAGCAAGGGCGA
GGAGCTGTTCACCGGGGTGGTGCCATCCTGGTCGAGCTGGACGGCGACGTAACG
GCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCT

FIG. 50C

GCACCAACCAGGCAAGCTGCCGTGCCCTGGCCCACCCCTCGTGACCCCTGACCTACG
GCGTGCAGTGCCTCAGCGCTACCCGACCACATGAAGCAGCACGACTCTTCAGT
CCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTCTCAAGGACGACGGC
AACTACAAGACCCGCCGGAGGTGAAGTTCGAGGGCGACACCCCTGGTGAACCGCAT
CGAGCTGAAGGGCATCGACTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGC
ATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGC
CGACCACTACCAGCAGAACACCCCCATCGCGACGGCCCCGTGCTGCTGCCGACA
ACCACTACCTGAGCACCCAGTCGCCCTGAGCAAAGACCCAAACGAGAACGCGAT
CACATGGCTCTGCTGGAGTTCTCGTACCGCCGCCGGATCACTCTCGGCATGGACGAG
CTGTACAAG

WPPE

Cgataatcacccctggattacaaaatttgtgaaaggattgacttgtttaactatgtgtcccttaactatgtggatacgtgttttaatgc
ctttgttatcatgctatgtccctggatgtgttttcatttttccttccttgatataaatccctgggtctgtctttatgaggagttgtgcccggtgtcagg
caacgtggcgtgggtgcactgtgtttgtgtcgaacaaacccccactgggtggggattgecaccacccgtcaactcccttcgggactttcgtt
cccccttcctatggccacggcgaactcatgcggcctgcctggccctgtggacaggggtcgctgtggactgacaattccgt
gtgtgtggggaaatcatgtcccttcgtgtgtccaccctggatttcgtccgggacgtccctgtcaactgtcccttcggc
ctcaatccagggaccccttcggccgtctggcccttcgtccggccttcgtccgtcttcgtccgtccctcagacgagtcggatetc
ecttggggccctcccccgtatcg

bGHpolyA

cgacCTCGACtgtgecttctagttgcagecatctgttgtgccttcctcccgtccttgaccctggaaagggtgccactcccactgt
cctttctaataaaaatgagggaaattgcatgcattgtctgtaggtgtcattctattctgggggggtgggtggggcaggacagcaaggggg
aggatgggaagacaatGgcaggcatg

loxP-SV40polyA \times 3-loxP

ataacttcgtataatgtatctatacgaagtattcgccatgttgcgatcataatggaaacgttgcagatctgcgactcttagaggatctgcgactcttagagg
atcataatcagccntaccacatttgttagaggcttactngctttaaaaacctcccacacetccccctgaacctgaaacataaaaatgaatgc当地
tgttggttgttaacttgttattgcagettataatggttacaataaagaatgcatacataacaaatttcacaataaagcatttttgcattctgt
tgtggtttgtccaaactcatcaatgtatcttatcatgtctggatctgcgactcttagaggatcataatgc当地
gcctttaaaaacctcccacacctccccctgaacctgaaacataaaaatgaatgc当地
aataaaagcaatgc当地
catcacaatattcacaataaageatttttgcattctgttagggtttgtccaaactcatcaatgtatcttatcatgt
ggatctgcgactcttagaggatcataatgc当地
catcacaatatttgttagaggcttactngctttaaaaacctcccacacctccccctgaacctg
aacataaaaatgaatgc当地
aatttgttgttaacttgttattgcagettataatggttacaataaata

FIG. 50D

aageaatgcatacacaatttcacaaaataaaggatcttttactgcattctatgtgggttgtccaaactcateatgttatcatgtctggatccccatcaagctgtatccggaaaccttaataacttgtataatgtatgtatcagaagttat

Rosa26 short homology arm

Rosa26 long homology arm

FIG. 50E

ctgagttcaatccccagcaaccacatgggttcacaaccatctgtaatggatctgtatgccttcttgtgtgtcataaaggaaactgtta
ttcacattaaataaataaaTCCTCCTTCTTCTTCTTTTTTTAAAGAGAATACTGTCTCCAG
TAGAAttTACTGAAGTAATGAAATACTTGTGTTGTTCCAATATGGTAGCCAATAAT
CAAATtACTCTTGTAGCACTGGAAATGTtACCAAGGAACAACTAAATTtATTgAAGTGTaa
CTGTGGACAGAGGAGGCCATAACTGCAGACTGTGAGGATACAGAAAGACCAATGCAGA
CTTAATGTCTTCTTACACTAAGCAATAAAGAAATAAAATGAACCTCTAGTA
TCCTATTGTTAAACTGCTAGCTTACtAACTTTGTGCTTCATCTATAACAAAGCTG
AAAGCTAAGTCTGCAGCCATTACTAAACATGAAAGCAAGTAATGATAATTGGATT
TCAAAAATGTAGGGCCAGAGTTAGCCAGCCAGTGGTGGTGCCTTATGCCtt
AATCCCAGCACTCTGGAGGCAGAGACAGGCAGATCTGAGTTGAGCCCAGCCTG
GtCTACACATCAAGTTCTATCTAGGATAGCCAGGAATAACACACAGAAACCTGTTGG
GGAGGGGGGCTCTGAGATTCTATAAAATTATAATTGAAGCATTCCCTAATGAGCCAC
TATGGATGTGGCTAAATCCGTCTACCTTCTGATGAGATTGGTATTATTTTCTG
TCTCTGCTGTTGGTGGTCTTGACACTGTGGCTTCTTAAAGCCTCCCTGC
CATGTGGTCTCTGTTGCTACTAACCTCCATGGCTAAATGGCATGGCTTTGCC
TTCTAAGGGCAGCTGCTGAGATTGCAAGCCTGATTCCAGGGTGGGAAATC
TTCAACACAAATTGCTTTAAATTTTTAAAAAATGGTTATAATAAA
CCTCATAAAATAGTTATGAGGAGTGAGGTGACTAATATTAAaTGAGTCCCTCCCC
ATAAAAGAGCTATTAGGCTTTGCTTACTAACCTTTAAATGTGGTATC
TTAGAACCAGGGCTTAGAGTTTAGTATAACAGAAACTGTGCACTCGCTTAATCA
GATTTCAGTTCAAATCCAGAGAATCAAATTCTCACAGCCAAAGTCAAATTAA
GAATTCTGACTTTAATGTTAAATTGCTACTGTGAATATAAAATGATAGCTTTCT
GAGGCAGGGTCTCACTATGTATCTGCTGATCTGCAACAAGATATGTAGACTAA
GTTCTGCCTGTTGCTCCTGAATACTAAGGTAAAATGTAGTAATACTTTGGAA
CTGAGGTCAAGATTCTTATAAGGGACACACTAAGGGAGCTGGGTGATAGTTGG
TAAAtgtttaagtgtgaaaactgttaatttatcacgcacactttaaaaaaaaagccaggccgttagagcatgtgtTaaggg
atccctaggactgtgtgacacacaAGAGTAGtTACTTGGCAGGCTCTGGTGGAGAGCATATTCAA
AAAACAAGGCAGACAACCAAGAAACTACAGTtAAGGTACCTGTCTTAAACCACATCT
GCATATACACAGGGATATTAAATATTCAAATAATATTCAAGTTCTCCCC
ATCAAATTGGGACATGGATTCTCCGGTGAATAGGCAGAGTTGGAAACTAAACAAA
TGTGGTTTGATTTGTGAAATTGTTCAAGTGATAGTTAAAGCCCATGAGATAC
AGAACAAAGCTGCTATTGAGGTCTTGGTTtATACTCAGAACACTCTTGGGT
TTCCCTGCACTATCCTGATCATGTGCTAGGCCTACCTTAGGCTGATTGTTGTTCAAAT
aAACTTAAGTTCTGTCAAGGTGATGTATGATTTCATATATCAAGGAAAACATG
TTATATATGTTAAACATITGTACTTAATGTGAAAGTTAGGTCTTGTGGGTT

FIG. 50F

TGATTTtAAfTTCAAAACCTGAGCTAAATAAGTCATTTrACATGTCTTACATTTGGT
GgAATTGTATATTGTGGTTGCAGGCAAGACTCTGTACCTAGTAACCCTaCCTATA
GAGCACTTTGCTGGGTACAAGTCTAGGAGTCAAGCATTCAACCTGAAGTTGAGAC
GTTTGTTAGTGTATACTAGTTATATGTTGGAGGACATGTTATCCAGAAGATATT
AGGACTATTITGACTGGGCTAACCGAATTGATTCTGATTAGCACTGTTAGTGAGCAT
TGAGTGGCCTTAGGCTGAATTggagtacttgtatatactcaaaataatgtggcctttttaaaaggcccttgttattaca
ccctgttttacataattttgttcaaagaaatacttgttggTCTCCTTTGACAACAATAGCATGTTTCAAG
CCATATTITTTTTCCTTTTTTTTTGGTTTTCGAGACAGGGTTCTGTAT
AGCCCTGGCTGTCCCTGGAACTCACTTGTAGACCAGGCTGGCCTCGAACTCAGAAAT
CCGCCTGCCTCTGCCCTTGAGTGCCGGATTAAAGGCCTGCACCACCGCCTGGC
TAAGTTGGATATTGTATATAACTATAACCAACTAACTCCACTGGGTGGATT
TAATTCAAGTCAGTAGTCTTAAGTGGCTTTATTGCCCTeATTAAAATCTACTGTT
ACTCTAACAGAGGCTGGTGGACTAGTGGCACCTAACGCAACTCCTACGGATATACTA
GCAGArTAAGGGTCAGGGATAGAAACTAGTCTAGCGTTTGATACCTACCAAGCTtA
TACTACCTTGTCTGATAGAAATATTtAGGACATCTAGCTT

pPGK-Neo-pPGK-polyA

aattctaccggtaggggaggcgctttcccaaggcagtcggagcatgcgccttagcagccccgtggcacttggcgctacacaagtgg
ccctggcctgcacacatccacatcaccggtaggcgecaacccggctcgctttggggcccttegcgecaectctactctccct
agtccggatgtccccccggccggcagctcgctgtgcaggacgtgacaaatggaaagttagcactgtctcaactgtctgtcagatgg
cagccacggcgtgagcaatggaaagccggtagggcttggggcagccaaatagcactgttgccttgcgttgcggcgtcagggcgtgg
aagggggtgggtccggggggggcgtcagggcggggtcagggcggggcggggcggccgaaggcctccggaggccggcattegc
acgctcaaaagcgeacgtcgccgcgtttcccttccatcattccggcccttgcacccgtcaatgcgcgttagcgaaggcttccat
agaaagtatggacttgcaccatgggateggccattgaacaagatggattgcacgcagggttccggcgtcagccgcaggggcggcgtt
tcggctatgactggcacaacagacaatccggctctgtgcgcgttgcgcgttgcgcgttgcgcgttgcgcgttgcgcgttgcgc
accgcaccttccgggtccctgaatgaactgcaggacgaggcagccggctatgtggctggcactgcacgcagggttgcgcgttgc
gctcgacgttgcactgaagccggaaaggactggactgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
gagaaggatccatcatggctatgcataatgcggcgtgcatacgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
tcgagccgtacttcggatggaaaggcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
ttccgcaggctcaaggcgecatgcggcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
ggccgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
tcggcggcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
aagaagggtgagaacagactacatttgatgaaggatggactacgggggtgggggtgggattagataatgcgttgcgttgcgtt
tttactgaaggcttactattgtttatgataatgtttcatagtt

FIG. 50G

gatatcataattaaacaagcaaaaccaaattaaggccagctcatcttccactcatgtatctatagatctctcgatggatcattgt
tttccttgcattccacttgtggttcaagtactgtggttcaatgttcaatgcctgaagaacgagatcagcagccatgtttccaca
tacacttcatttcattgttttgcctaaatccatcagaaagc

pPGK-DTA

TACCGGGTAGGGGAGGCCTTTCCcAAGGCAGTCTGgAGCATGCGCTTAGCAGCCC
CGCTgGGCACTGGCGCTACACAAGTGGCCTCTGGCCTCGCACACATTCCACATCCA
CCGGTAGGCGCCAACCGGCTCCGTCTTGGTGGCCCTTCGCGCCACCTTCTACTCC
TCCCCTAGTCAGGAAGTCCCCCCCCGCCAGCTCGCGTCGTGcAGGACGTGACA
AAATGGAAGTAGCACGCTCACTAGTCCTGAGATGGACAGCACCGCTGAGCAATG
GAAGCGGGTAGGCCTTGGGCAGCGCCAATAGCAGCTTGCTCCTCGCTTCTG
GGCTCAGAGGCTGGAAAGGGTGGGTCCGGGGCGGGCTCAGGGCGGGCTCAGG
GGCGGGGGCGGGCGCCCGAAGGTCCCTCCGGAGGGCCCGCATTCTGCACGCTTCAAA
GCGCACGTCTGCCCGCTGTTCTCCTCTCCTCATCTCCGGCCTTCGACCTGCAGG
TCCTCGCCATggatctgtatgttgttGattttctaaAtctttgtGatggaaaactttctgtaccacgggaactaacctgtt
atgttagattccattcaaaaaggtaaaaaacggccaaatctgttacacaaggaaattatgacgtatggaaagggttatagtaccgacaa
taaatacgcacgtcgcccataactctgttagataatgaaaaccgcctctggaaaagctggaggcggtcaagtgaegtatccaggactg
acgaagggtctcgactaaaaggtaatgcgaaactattaagaaagagttagtttaagtcteactgaaccgttgcggagcaagtcgga
acggaagagatattcaaaaaggteggtaatgggtgcctcgctgttagtgcacgcctccctegctgagggagatctgcgttgaaatattaa
taactgggaacaggcgaagecgtaagecgtaacttgagataatgttggaaaaccgcgtggaaaacgtggccaagatgegtatgtatgatgata
tggctcaagecctgtgcaggaaatctgtcaggcgaictcttgcaggaaacctacttcgtgtgtgcataattggacaaactacctacag
agatttaaagetctaaggtaaatataaaatttttaagtgtataatgtgttaaactactgttcaattgttgcgtatgtttagattccaaacctatggact
gtgaatgggagcagtggatgcagatcttagagctcgctgatcagccctcactgtgcctcttagttgcagccatctgttgttgcct
ccccctgccttccttgaccctggaagggtgccactccactgtccatataaaaatgagggaaatgcategcattgtgtgatggatgtcat
tctattctgggggggtgggtggggcaggacagcaagggggaggattggaaagacaatgcaggcatg

FIG. 50H

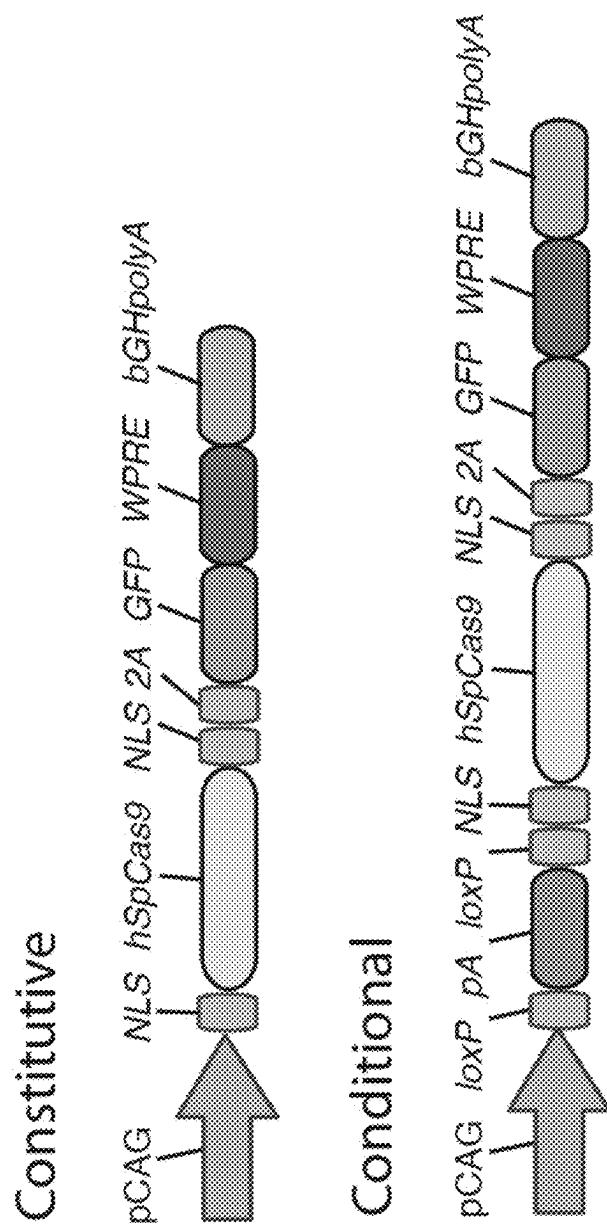


FIG. 51

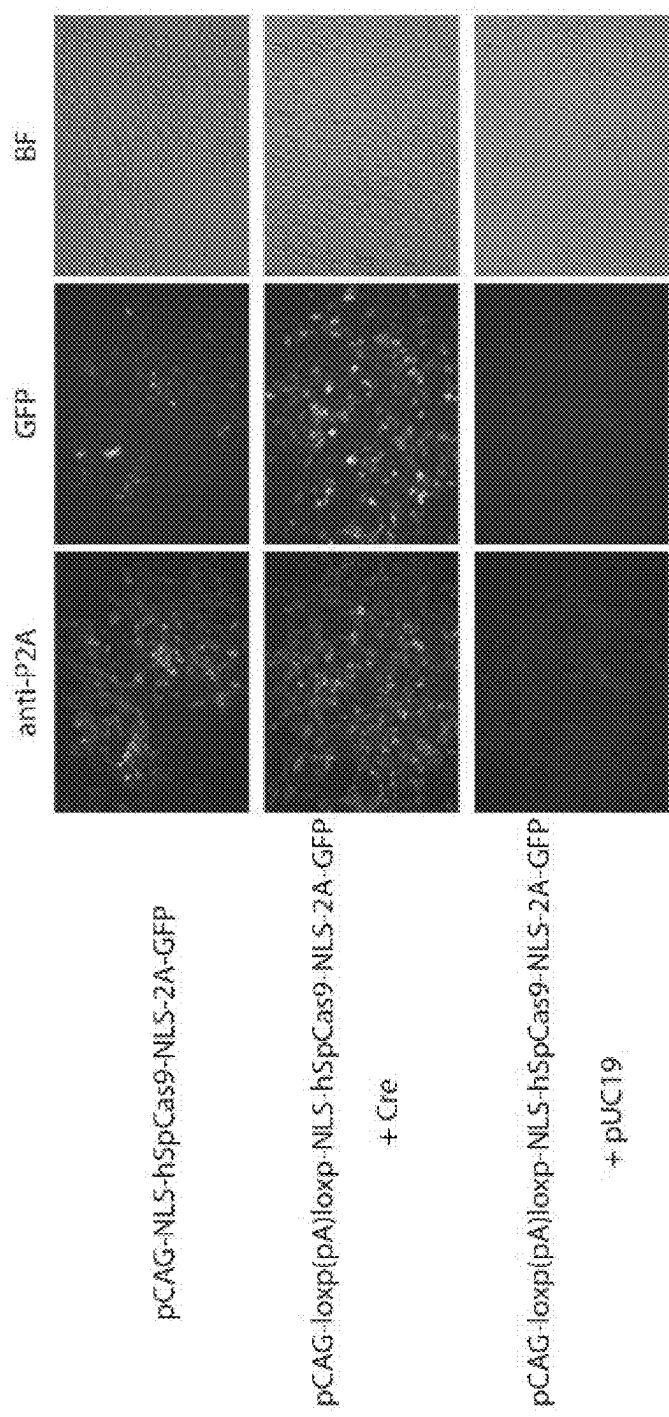


FIG. 52

Validation of Cas9 nuclease activity by Surveyor

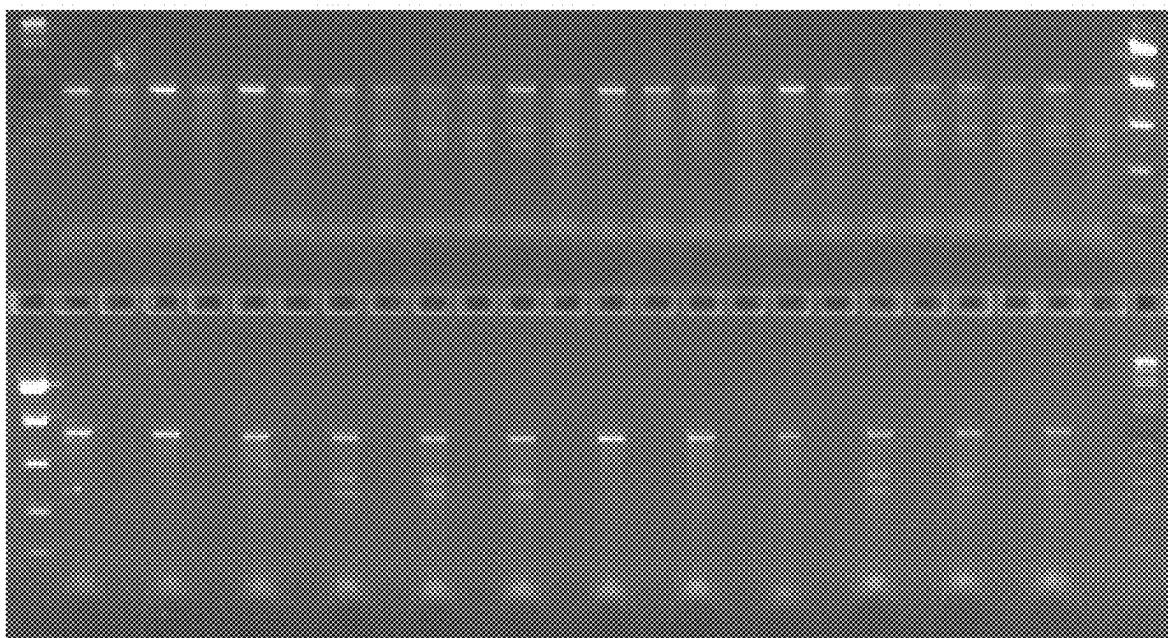


FIG. 53

		Average	StDev
pCAG-loxp(pA)loxp-NLS-hSpCas9-NLS-2A-GFP	Clone 1	32.1	7.1
	Clone 2	27.3	3.5
	Clone 3	35.9	1.4
	Clone 4	39.0	4.7
pCAG-NLS-hSpCas9-NLS-2A-GFP	Clone 1	26.9	1.3
	Clone 2	33.1	2.7

FIG. 54

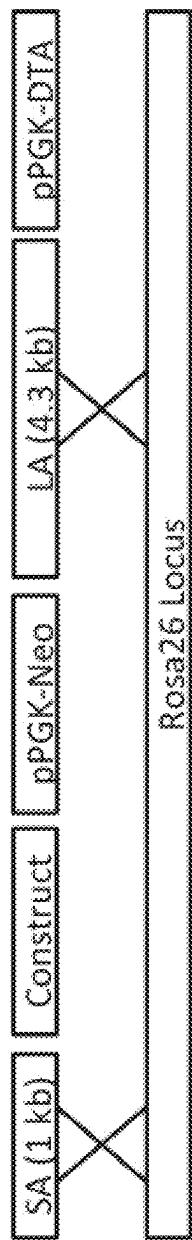


FIG. 55

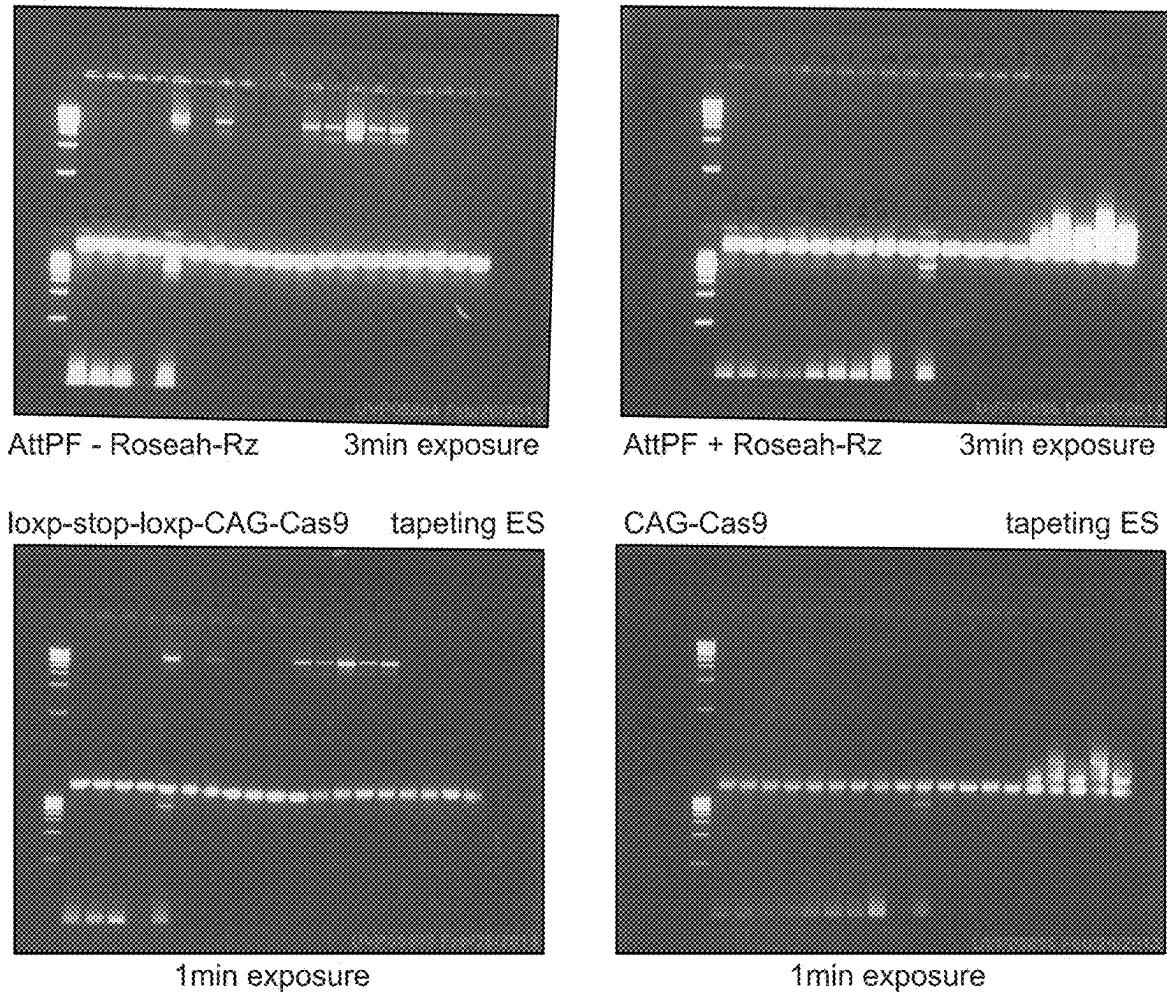


FIG. 56

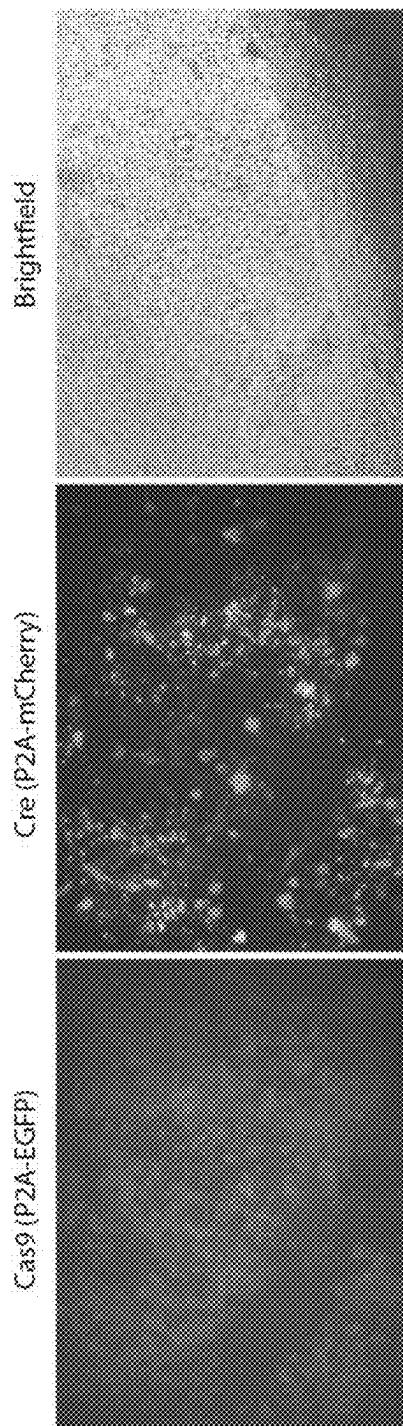


FIG. 57

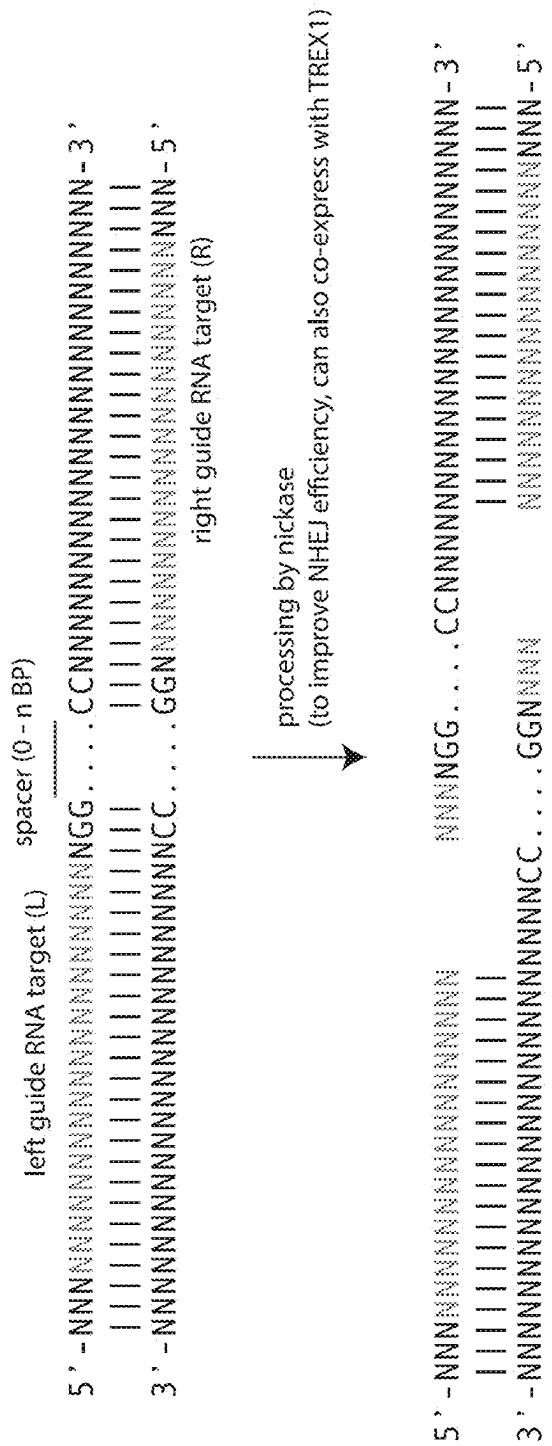


FIG. 58

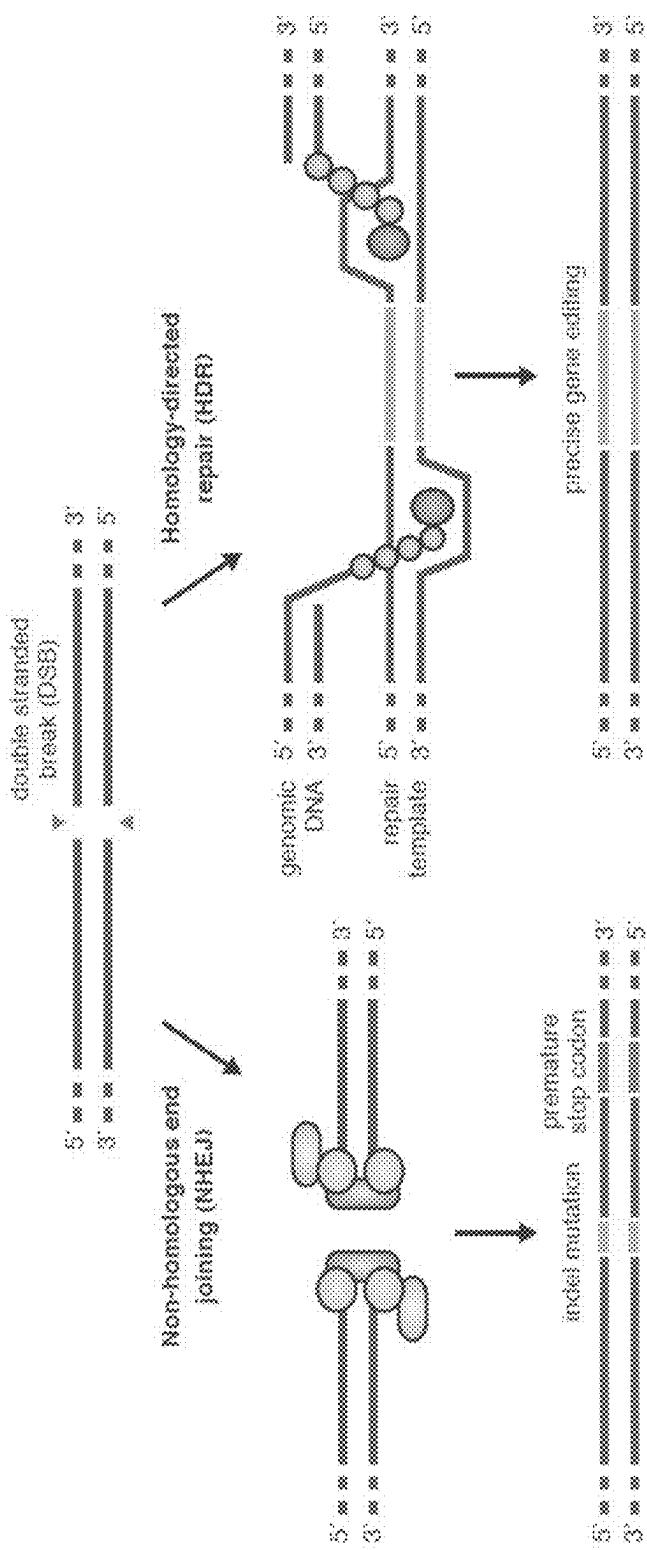


FIG. 59

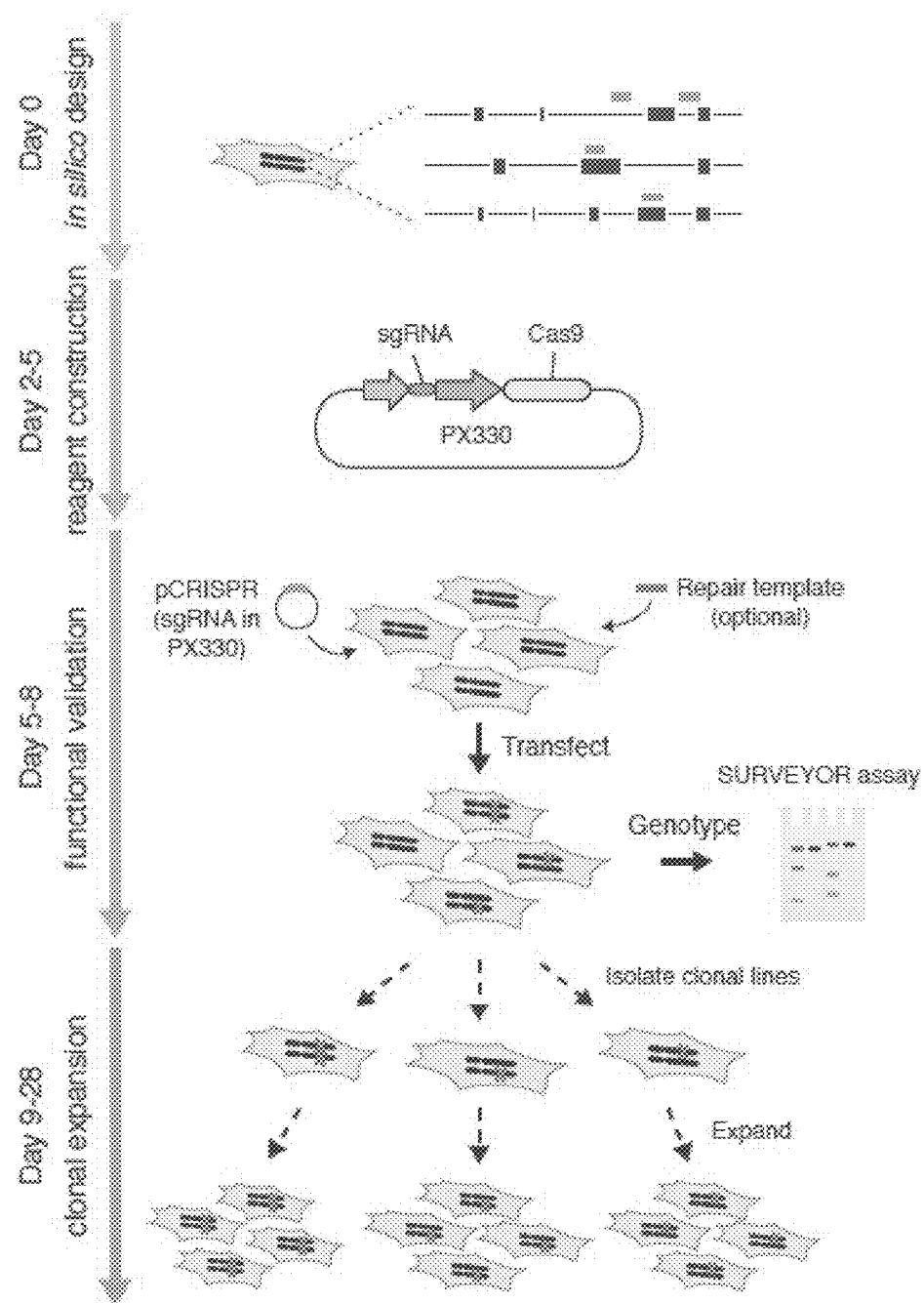


FIG. 60

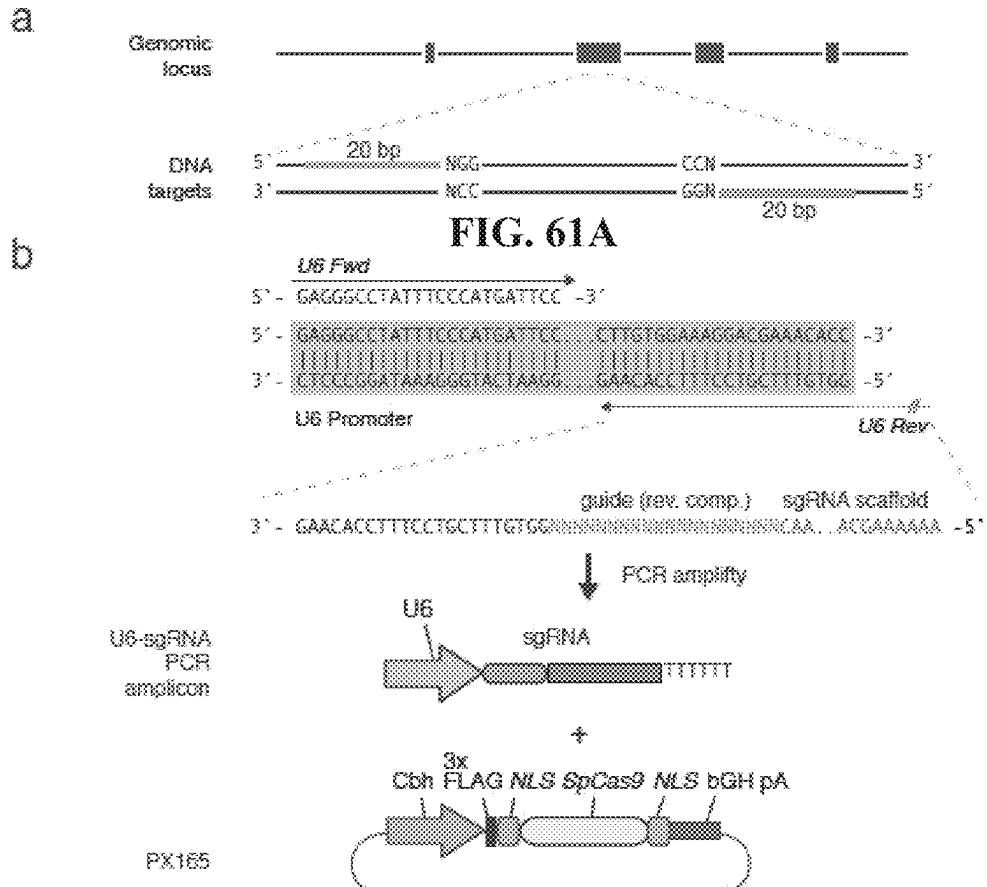


FIG. 61B

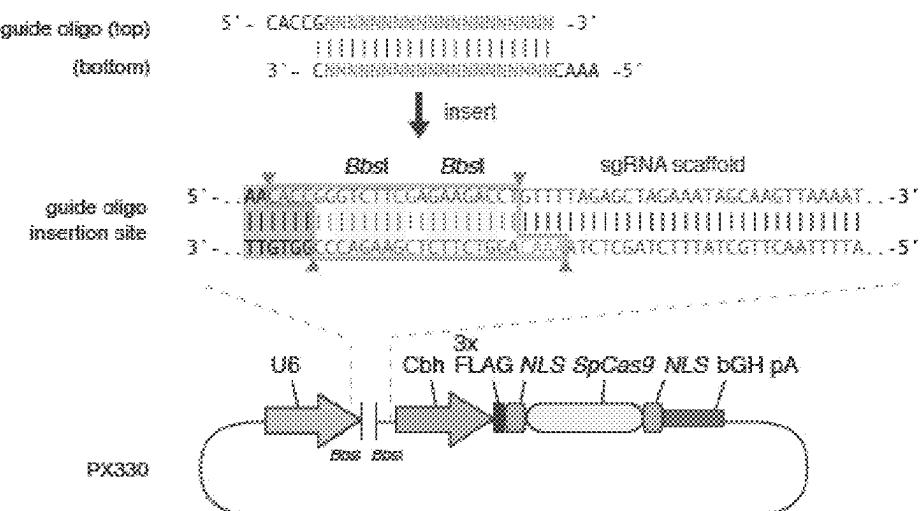


FIG. 61C

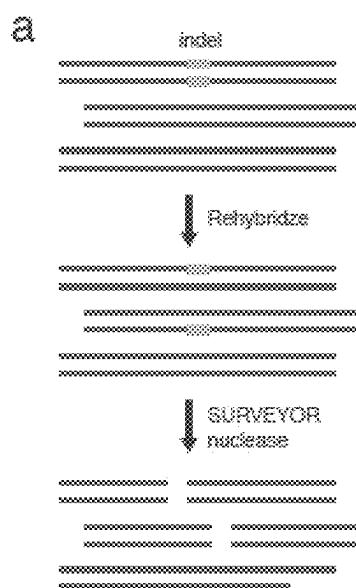


FIG. 62A

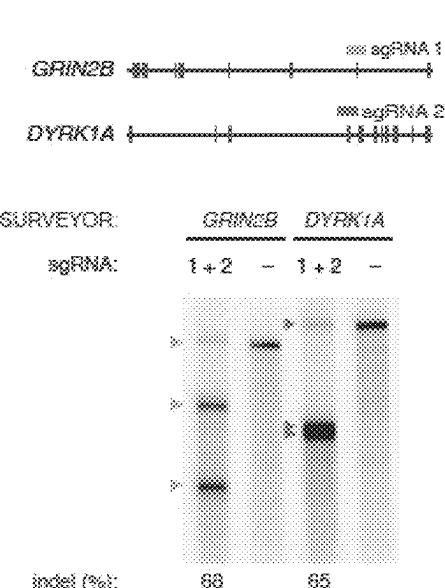


FIG. 62B

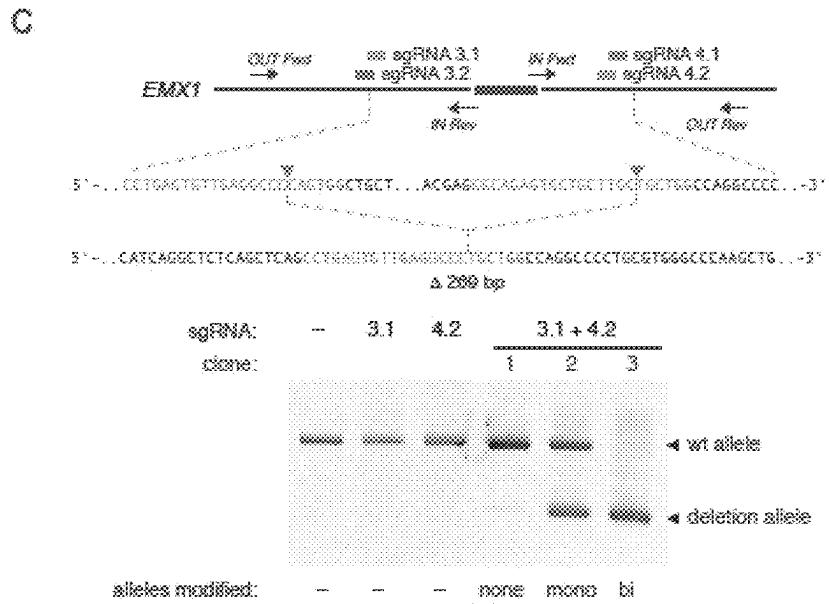


FIG. 62C

d

sgRNA4x	Approx. deletion size (bp)	Number of clones				
		+/+	2/+	2/2	Inversion	None
3.1 + 4.1	282	12	10	3	0	23
3.2 + 4.1	237	20	16	2	0	38
3.1 + 4.2	426	4	14	5	0	23
3.2 + 4.2	239	6	11	4	0	23

FIG. 62D

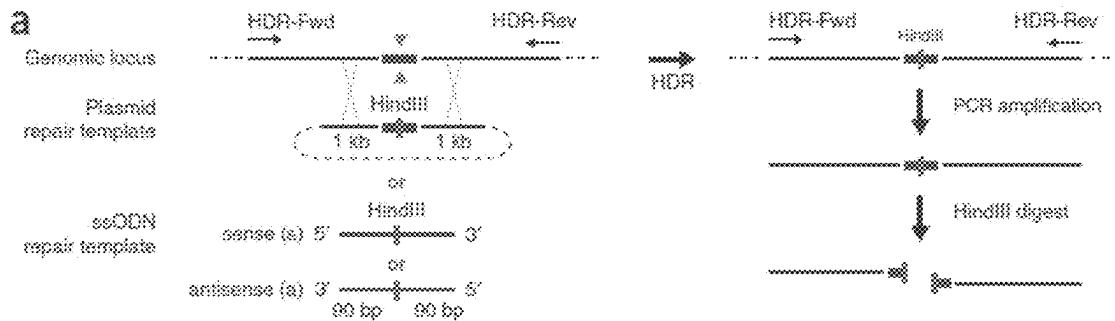


FIG. 63A

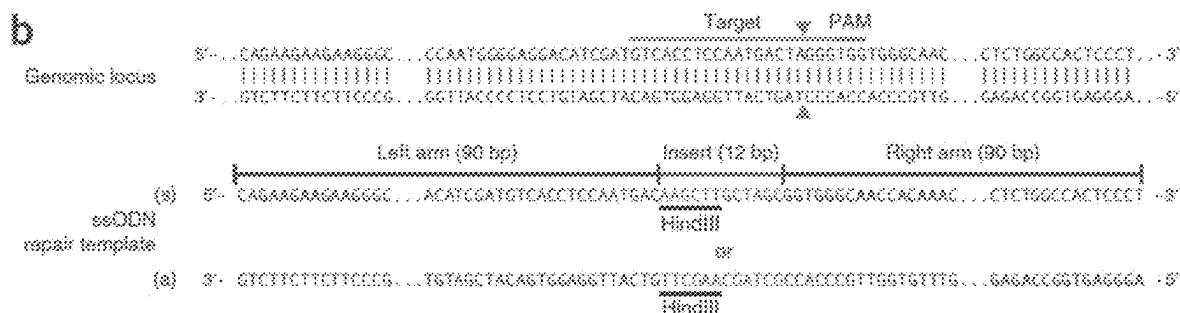


FIG. 63B

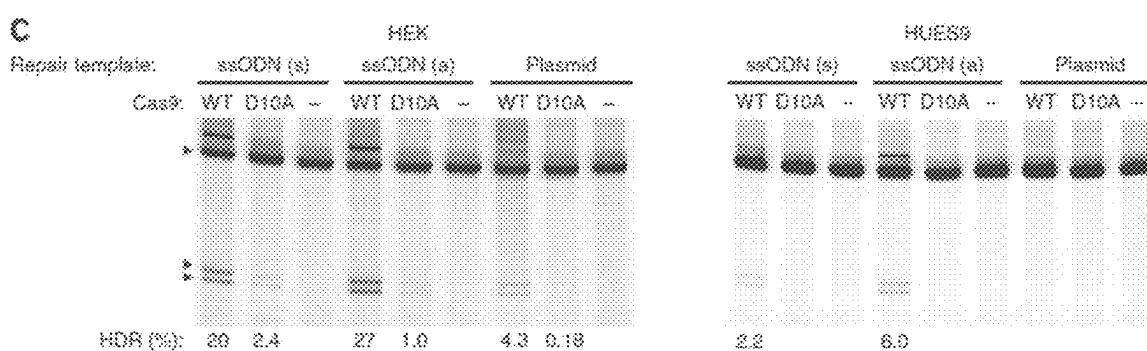


FIG. 63C

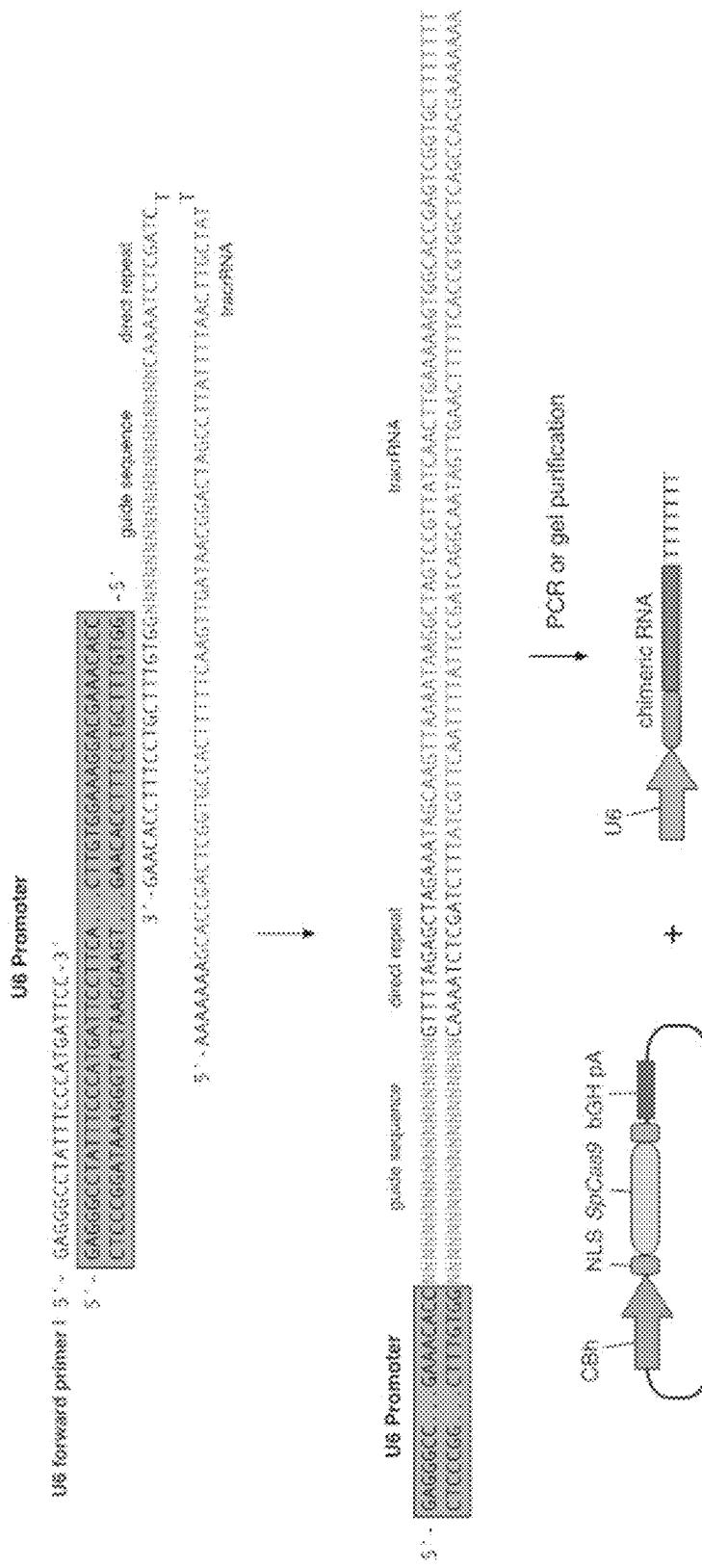


FIG. 64

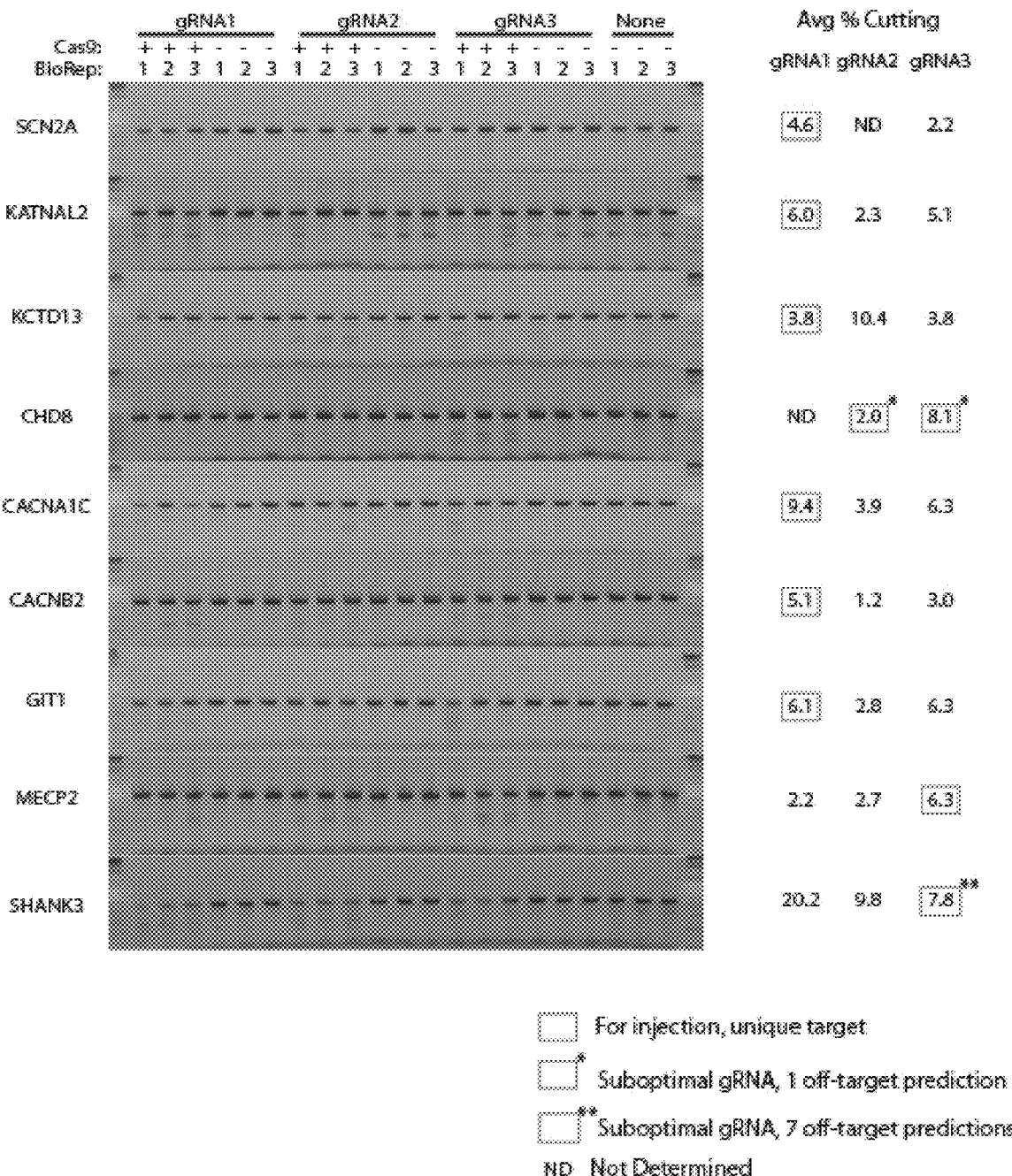
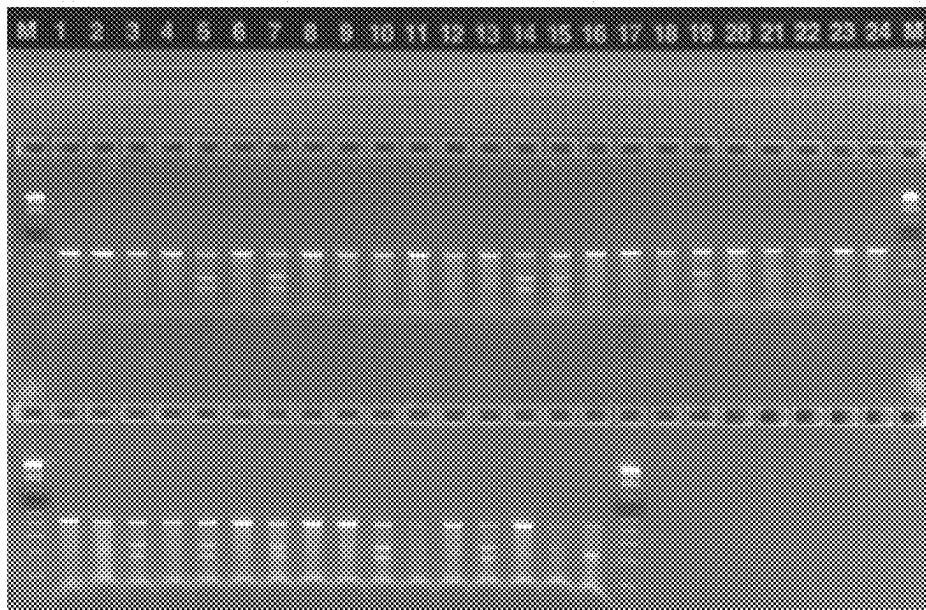


FIG. 65



gRNA sequences for Chd8 targeting:

Chd8.1 - agctgtttactggtcggct
Chd8.2 - aatggatacacacctggcgaa
Chd8.3 - caatggatacacacctggtcga

FIG. 66

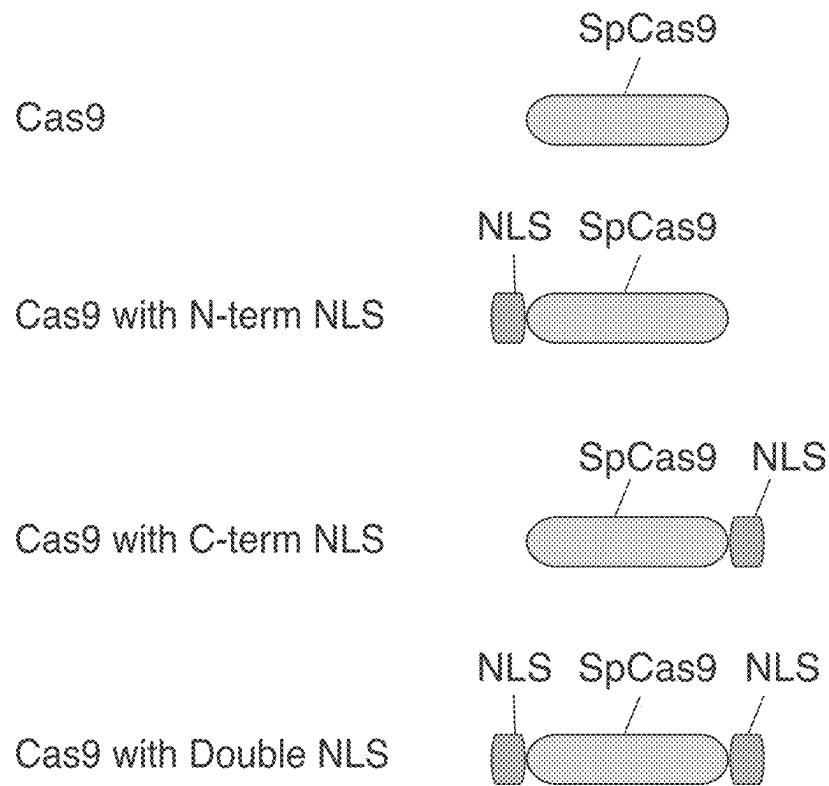


FIG. NLS1

FIG. 67

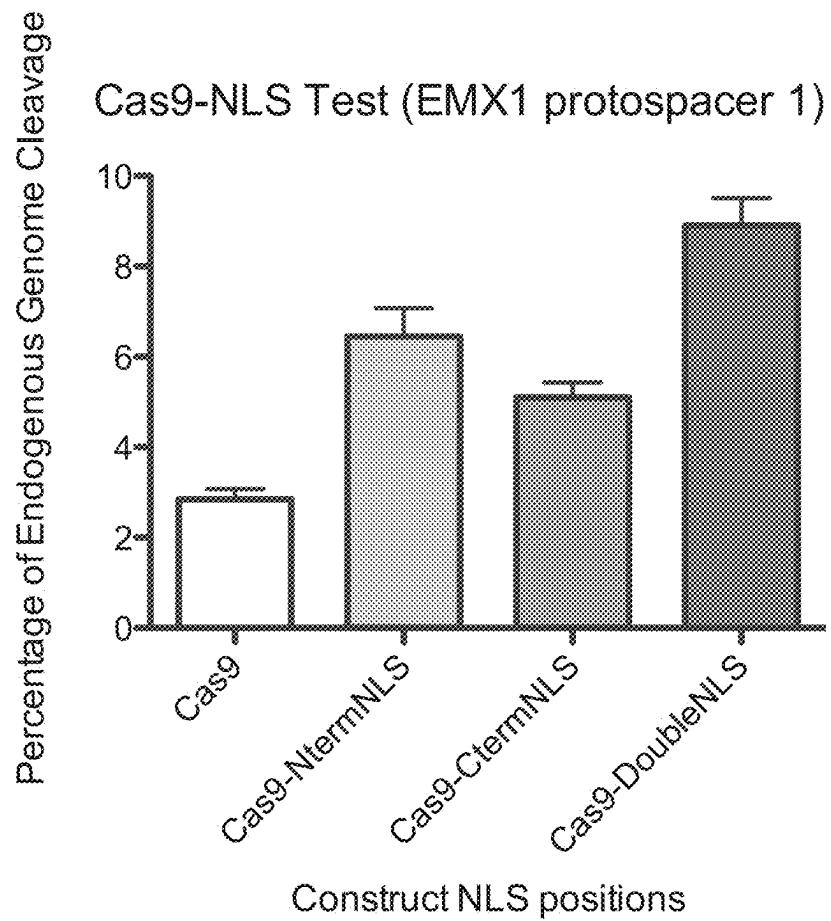


FIG. 68

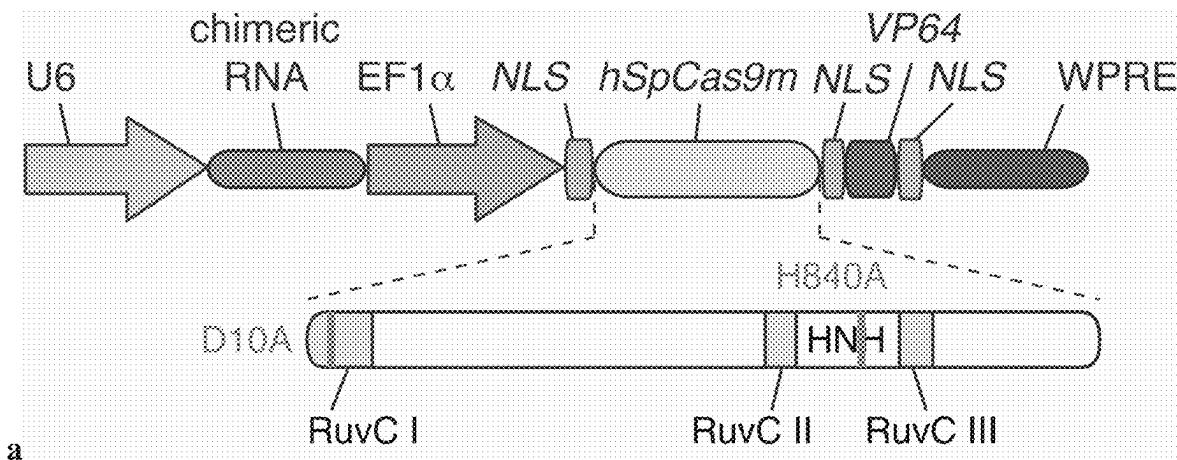


FIG. 69A

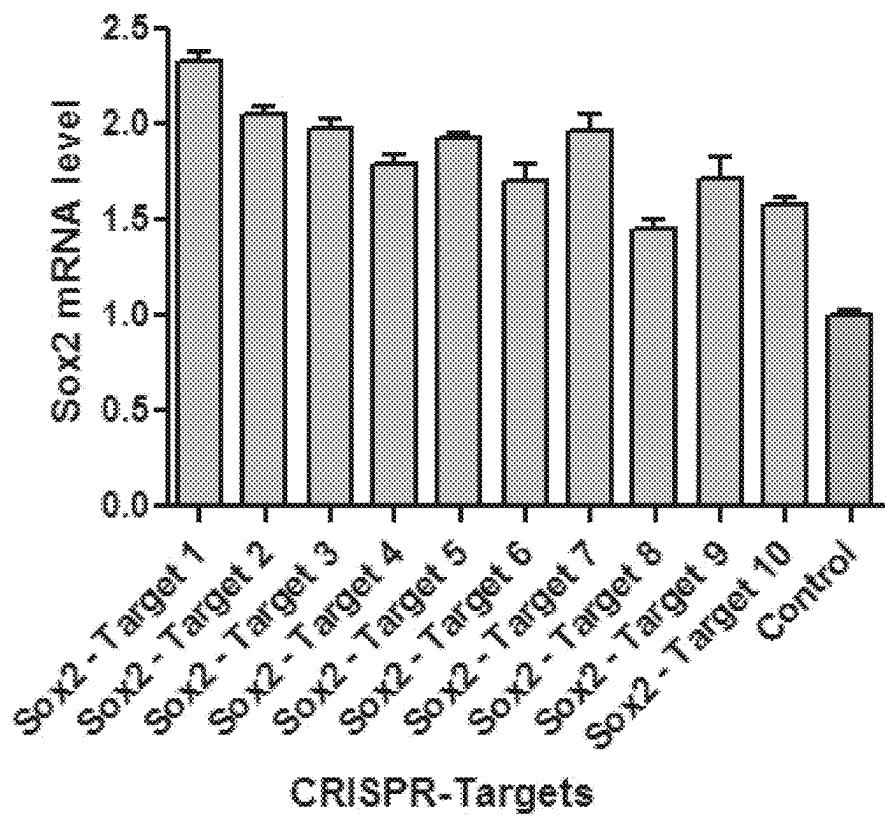


FIG. 69B

NLS architecture optimization for SpCas9

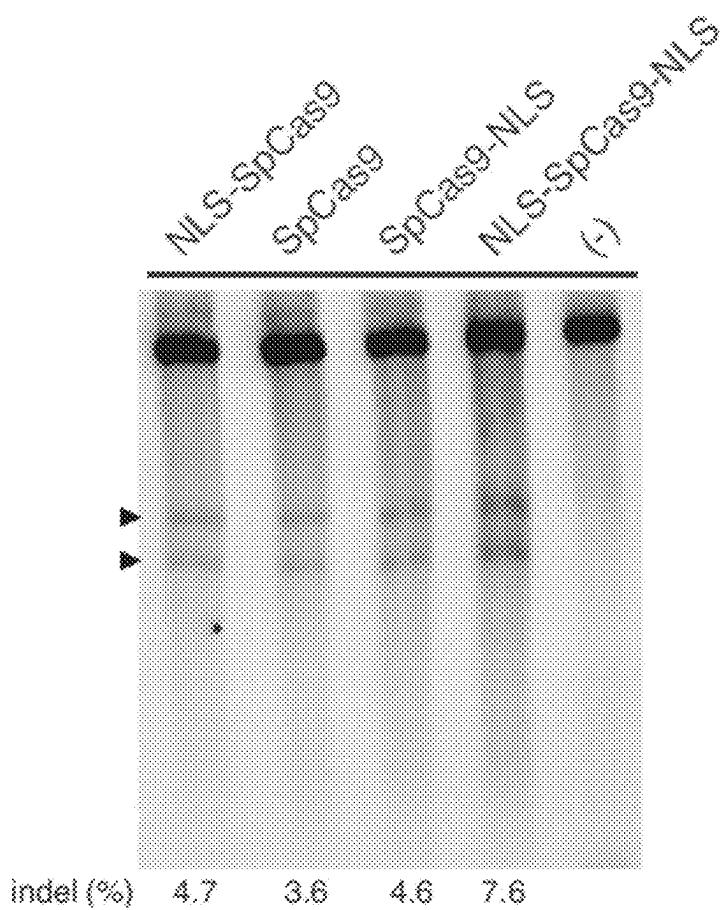
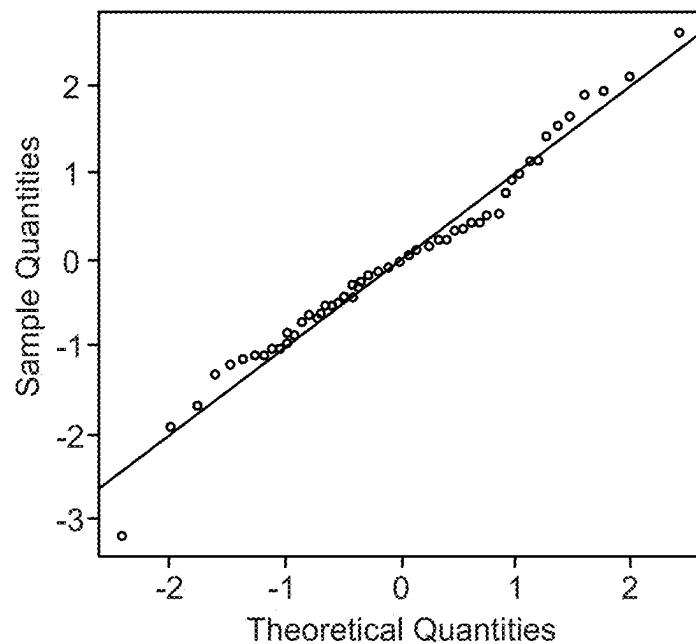


FIG. 70



QQ plot for the NGGNN sequences

FIG. 71

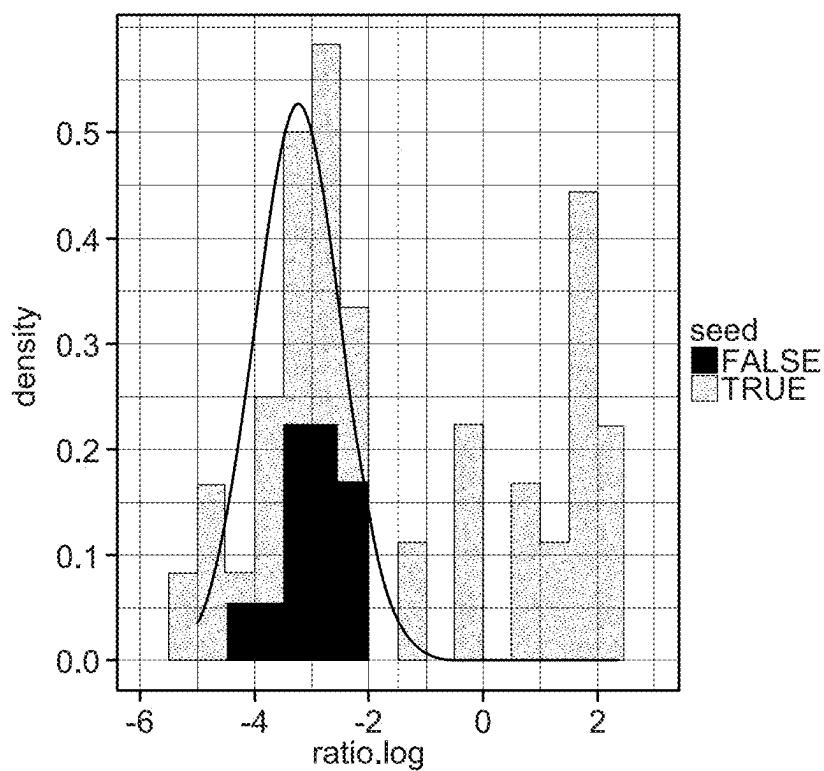


FIG. 72

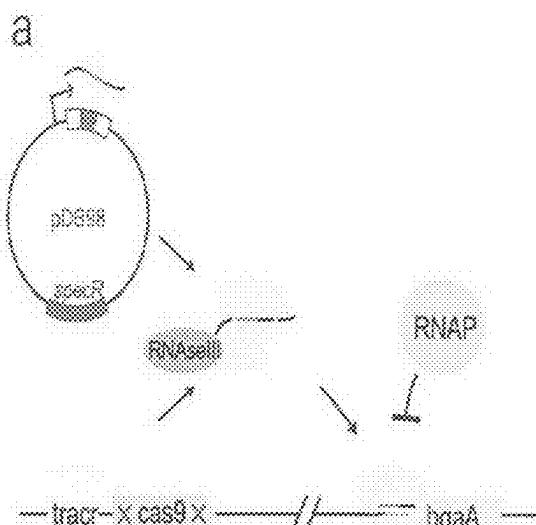


FIG. 73A

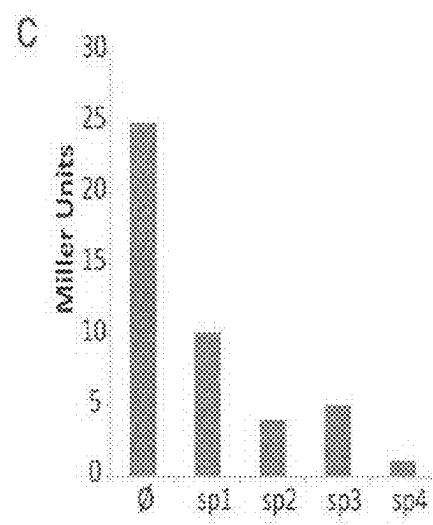


FIG. 73C

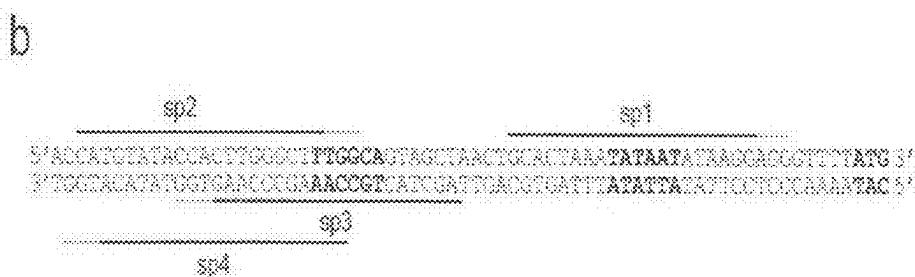


FIG. 73B

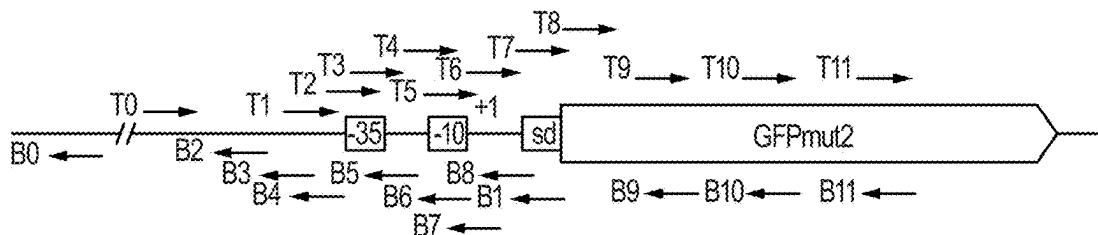


FIG. 74A

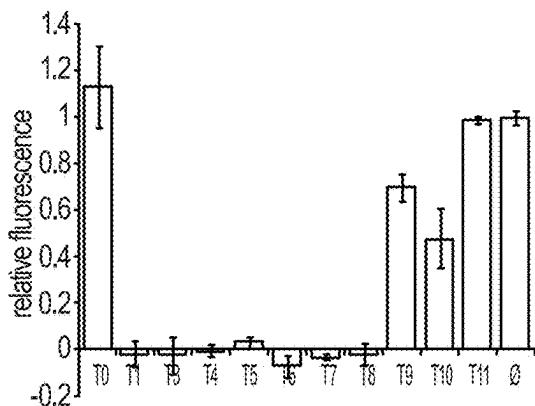


FIG. 74B

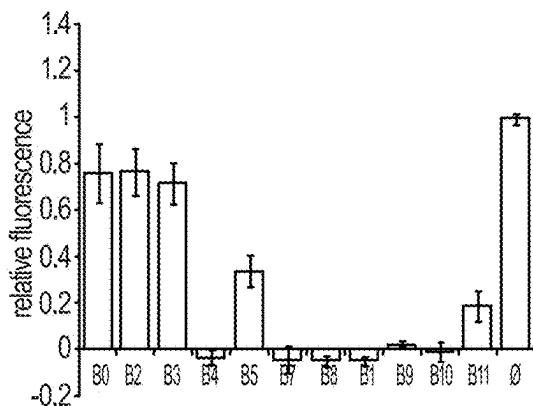


FIG. 74C

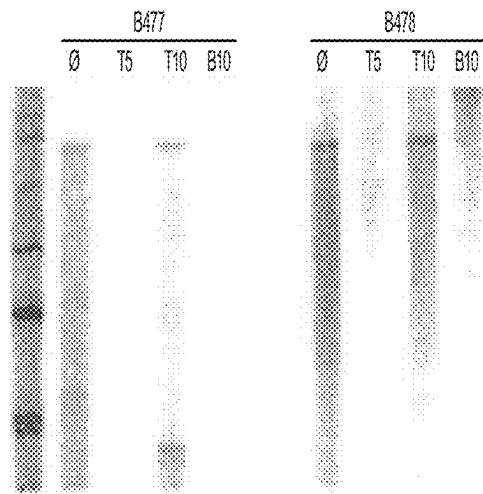


FIG. 74D

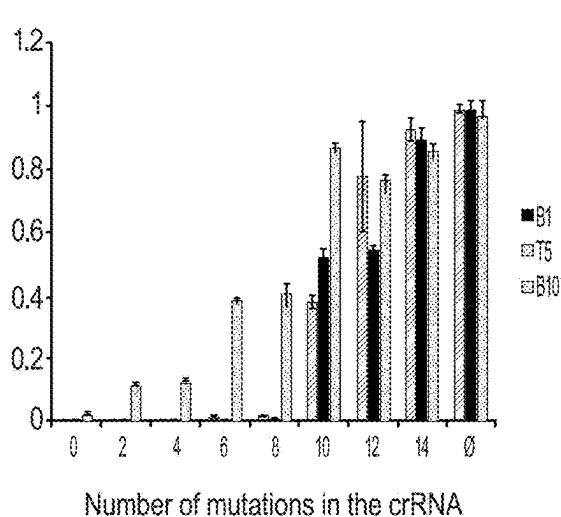


FIG. 74E

CRISPR-CAS COMPONENT SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION

RELATED APPLICATIONS AND INCORPORATION BY REFERENCE

[0001] This application is a continuation of U.S. application Ser. No. 17/034,754, filed Sep. 28, 2020, which is a continuation of U.S. application Ser. No. 15/967,495, filed Apr. 30, 2018, which is a continuation of U.S. application Ser. No. 14/105,035, filed Dec. 12, 2013, and which claims priority to U.S. provisional patent applications 61/736,527, filed on Dec. 12, 2012, 61/748,427, filed on Jan. 2, 2013, 61/768,959, filed Feb. 25, 2013, 61/791,409 filed on Mar. 15, 2013, and 61/835,931, filed on Jun. 17, 2013.

[0002] Reference is made to U.S. provisional patent applications 61/758,468; 61/769,046; 61/802,174; 61/806,375; 61/814,263; 61/819,803 and 61/828,130, each entitled ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION, filed on Jan. 30, 2013; Feb. 25, 2013; Mar. 15, 2013; Mar. 28, 2013; Apr. 20, 2013; May 6, 2013 and May 28, 2013 respectively. Reference is also made to U.S. provisional patent applications 61/835,936, 61/836,127, 61/836,101, 61/836,080, 61/836,123 and 61/835,973 each filed Jun. 17, 2013. Reference is also made to U.S. provisional patent application 61/842,322 and U.S. patent application Ser. No. 14/054,414, each having Broad reference BI-2011/008A, entitled CRISPR-CAS SYSTEMS AND METHODS FOR ALTERING EXPRESSION OF GENE PRODUCTS filed on Jul. 2, 2013 and Oct. 15, 2013 respectively.

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

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0004] This invention was made with government support under Grant No. MH100706, awarded by the National Institutes of Health. The government has certain rights in the invention.

SEQUENCE LISTING

[0005] 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 114203-6222_SL.xml and is 708,469 bytes in size.

FIELD OF THE INVENTION

[0006] The present invention generally relates to systems, methods and compositions used for the control of gene expression involving sequence targeting, such as genome perturbation or gene-editing, that may use vector systems related to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and components thereof.

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] There exists a pressing need for alternative and robust systems and techniques for sequence targeting with a wide array of applications. This invention addresses this need and provides related advantages. The CRISPR/Cas or the CRISPR-Cas system (both terms are used interchangeably throughout this application) 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, in other words the Cas enzyme can be recruited to a specific DNA target using said short RNA molecule. 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 and optimization of these genome engineering tools, which are aspects of the claimed invention.

[0009] In one aspect, the invention provides a vector system comprising one or more vectors. In some embodiments, the system comprises: (a) a first regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting one or more guide sequences upstream of the tracr mate sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein 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; and (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence; wherein components (a) and (b) are located on the same or different vectors of the system. In some embodiments, component (a)

further comprises the tracr sequence downstream of the tracr mate sequence under the control of the first regulatory element. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the system comprises the tracr sequence under the control of a third regulatory element, such as a polymerase III promoter. In some embodiments, the tracr sequence exhibits at least 50%, 60%, 70%, 80%, 90%, 95%, or 99% of sequence complementarity along the length of the tracr mate sequence when optimally aligned. Determining optimal alignment is within the purview of one of skill in the art. For example, there are publically and commercially available alignment algorithms and programs such as, but not limited to, ClustalW, Smith-Waterman in matlab, Bowtie, Geneious, Biopython and SeqMan. In some embodiments, the CRISPR complex comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR complex in a detectable amount in the nucleus of a eukaryotic cell. Without wishing to be bound by theory, it is believed that a nuclear localization sequence is not necessary for CRISPR complex activity in eukaryotes, but that including such sequences enhances activity of the system, especially as to targeting nucleic acid molecules in the nucleus. In some embodiments, the CRISPR enzyme is a type II CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is *S. pneumoniae*, *S. pyogenes*, or *S. thermophilus* Cas9, and may include mutated Cas9 derived from these organisms. The enzyme may be a Cas9 homolog or ortholog. In some embodiments, the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR enzyme lacks DNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 15, 16, 17, 18, 19, 20, 25 nucleotides, or between 10-30, or between 15-25, or between 15-20 nucleotides in length. In general, and throughout this specification, 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). 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.

[0010] 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).

[0011] 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 EFla 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.).

[0012] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[0013] In one aspect, the invention provides a vector comprising a regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising one or more nuclear localization sequences. In some embodiments, said regulatory element drives transcription of the CRISPR enzyme in a eukaryotic cell such that said CRISPR enzyme accumulates in a detectable amount in the nucleus of the eukaryotic cell. In some embodiments, the regulatory element is a polymerase II promoter. In some embodiments, the CRISPR enzyme is a type II CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is *S. pneumoniae*, *S. pyogenes* or *S. thermophilus* Cas9, and may include mutated Cas9 derived from these organisms. In some embodiments, the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR enzyme lacks DNA strand cleavage activity.

[0014] In one aspect, the invention provides a CRISPR enzyme comprising one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the CRISPR enzyme is a type II CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is *S. pneumoniae*, *S. pyogenes* or *S. thermophilus* Cas9, and may include mutated Cas9 derived from these organisms. The enzyme may be a Cas9 homolog or ortholog. In some embodiments, the CRISPR enzyme lacks the ability to cleave one or more strands of a target sequence to which it binds.

[0015] In one aspect, the invention provides a eukaryotic host cell comprising (a) a first regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting one or more guide sequences upstream of the tracr mate sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein 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; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence. In some embodiments, the host cell comprises components (a) and (b). In some embodiments, component (a), component (b), or components (a) and (b) are stably integrated into a genome of the host eukaryotic cell. In some embodiments, component (a) further comprises the tracr sequence downstream of the tracr mate sequence under the control of the first regulatory

element. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the eukaryotic host cell further comprises a third regulatory element, such as a polymerase III promoter, operably linked to said tracr sequence. In some embodiments, the tracr sequence exhibits at least 50%, 60%, 70%, 80%, 90%, 95%, or 99% of sequence complementarity along the length of the tracr mate sequence when optimally aligned. In some embodiments, the CRISPR enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the CRISPR enzyme is a type II CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is *S. pneumoniae*, *S. pyogenes* or *S. thermophilus* Cas9, and may include mutated Cas9 derived from these organisms. The enzyme may be a Cas9 homolog or ortholog. In some embodiments, the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR enzyme lacks DNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 15, 16, 17, 18, 19, 20, 25 nucleotides, or between 10-30, or between 15-25, or between 15-20 nucleotides in length. In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example a mammal. Also, the organism may be an arthropod such as an insect. The organism also may be a plant. Further, the organism may be a fungus.

[0016] In one aspect, the invention provides a kit comprising one or more of the components described herein. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting one or more guide sequences upstream of the tracr mate sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein 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; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence. In some embodiments, the kit comprises components (a) and (b) located on the same or different vectors of the system. In some embodiments, component (a) further comprises the

tracr sequence downstream of the tracr mate sequence under the control of the first regulatory element. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the system further comprises a third regulatory element, such as a polymerase III promoter, operably linked to said tracr sequence. In some embodiments, the tracr sequence exhibits at least 50%, 60%, 70%, 80%, 90%, 95%, or 99% of sequence complementarity along the length of the tracr mate sequence when optimally aligned. In some embodiments, the CRISPR enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the CRISPR enzyme is a type II CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is *S. pneumoniae*, *S. pyogenes* or *S. thermophilus* Cas9, and may include mutated Cas9 derived from these organisms. The enzyme may be a Cas9 homolog or ortholog. In some embodiments, the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR enzyme lacks DNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 15, 16, 17, 18, 19, 20, 25 nucleotides, or between 10-30, or between 15-25, or between 15-20 nucleotides in length.

[0017] 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.

[0018] 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.

[0019] 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 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.

[0020] In one aspect, the invention provides a method for developing a biologically active agent that modulates a cell signaling event associated with a disease gene. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) contacting a test compound with a model cell of any one of the described embodiments; and (b) detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with said mutation in said disease gene, thereby developing said biologically active agent that modulates said cell signaling event associated with said disease gene.

[0021] In one aspect, the invention provides a recombinant polynucleotide comprising a guide sequence upstream of a tracr mate sequence, wherein the guide sequence when expressed directs sequence-specific binding of a CRISPR complex to a corresponding target sequence present in a eukaryotic cell. In some embodiments, the target sequence is a viral sequence present in a eukaryotic cell. In some embodiments, the target sequence is a proto-oncogene or an oncogene.

[0022] 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 a 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.

[0023] Accordingly, it is an object of the invention not to 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.

[0024] 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. These and other embodiments are disclosed or are obvious from and encompassed by, the following detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] 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:

[0026] 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).

[0027] FIGS. 2A-F. FIG. 2A shows an exemplary CRISPR system and a possible mechanism of action. FIG. 2B (left panel) provides an example adaptation of the CRISPR system for expression in eukaryotic cells, and also results of tests assessing nuclear localization and CRISPR activity (right panel). FIG. 2C illustrates mammalian expression of SpCas9 and SpRNase III driven by the constitutive EF1 α promoter and tracrRNA and pre-crRNA array (DR-Spacer-DR) driven by the RNA Pol3 promoter U6; and discloses SEQ ID NOS 279-280, respectively, in order of appearance. FIG. 2D shows results of a surveyor nuclease assay for SpCas9-mediated insertions and deletions. FIG. 2E is a schematic representation of base pairing between target locus and EMX1-targeting crRNA, as well as an example chromatogram of a micro deletion adjacent to the SpCas9 cleavage site. FIG. 2E also discloses SEQ ID NOS 281-283, respectively, in order of appearance. FIG. 2F shows mutated alleles identified from sequencing analysis and discloses SEQ ID NOS 284-288, respectively, in order of appearance.

[0028] FIGS. 3A-C show an exemplary expression cassette for expression of CRISPR system elements in eukaryotic cells (FIG. 3A), predicted structures of example guide sequences (FIG. 3B), and CRISPR system activity as measured in eukaryotic and prokaryotic cells (FIG. 3C) (SEQ ID NOS 289-298, respectively, in order of appearance).

[0029] FIGS. 4A-D show results of an evaluation of SpCas9 specificity for an example target. FIG. 4A shows effect of single-nucleotide mutations in the guide sequence on protospacer cleavage in the mammalian genome, which was analyzed using a series of EMX1-targeting chimeric crRNAs with single point mutations. FIG. 4A also discloses SEQ ID NOS 299, 282 and 300-310, respectively, in order of appearance. FIG. 4B illustrates results of a Surveyor nuclease assay comparing the cleavage efficiency of Cas9 when paired with different mutant chimeric RNAs. FIG. 4C provides a schematic showing of the design of TALENs targeting EMX1 and also discloses SEQ ID NO: 299. FIG. 4D shows a Surveyor gel comparing the efficiency of TALEN and Cas9 (n=3).

[0030] FIGS. 5A-G. FIG. 5A shows an aspartate-to-alanine substitution (D 10A) in the RuvC I catalytic domain of SpCas9 was engineered to convert the nuclease into a nickase. FIG. 5B shows co-expression of EMX1-targeting chimeric crRNA with SpCas9 produced indels in the target site, whereas co-expression with SpCas9n did not (n=3). FIG. 5C provides a schematic illustration of the HR strategy, with relative locations of recombination points and primer

annealing sequences (arrows). FIG. 5D shows that 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), with SpCas9 and SpCas9n mediating similar levels of HR efficiencies. FIG. 5E shows verified HR using Sanger sequencing of genomic amplicons and discloses SEQ ID NO: 311. FIG. 5F shows a schematic design of the crRNA array encoding a pair of EMX1- and PVALB-targeting spacers, and a Surveyor blot showing efficient cleavage at both loci, and also discloses SEQ ID NOS 312-313, respectively, in order of appearance. FIG. 5G shows targeted deletion of larger genomic regions through concurrent DSBs using spacers against two targets within EMX1 spaced by 119 bp, wherein a 1.6% deletion efficacy (3 out of 182 amplicons) was detected. FIG. 5G also discloses SEQ ID NOS 314-318, respectively, in order of appearance.

[0031] FIG. 6 provides a table of protospacer sequences (SEQ ID NOS 33, 32, 31, 322-327, 35, 34 and 330-334, 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).

[0032] FIGS. 7A-C. FIG. 7A shows tracrRNA expression constructs and discloses SEQ ID NOS 335-336, respectively, in order of appearance. FIG. 7B shows a comparison of different tracrRNA transcripts for Cas9-mediated gene targeting in Surveyor assays. FIG. 7C shows results of a Northern blot analysis of total RNA extracted from 293 FT cells transfected various constructs.

[0033] FIG. 8 shows a schematic of a surveyor nuclease assay for detection of double strand break-induced micro- insertions and -deletions.

[0034] FIGS. 9A-B show exemplary bicistronic expression vectors for expression of CRISPR system elements in eukaryotic cells. FIG. 9A discloses SEQ ID NOS 337-339, respectively, in order of appearance. FIG. 9B discloses SEQ ID NOS 340-342, respectively, in order of appearance.

[0035] FIGS. 10A-D show a bacterial plasmid transformation interference assay (FIG. 10D), expression cassettes (FIG. 10A) and plasmids used therein (FIG. 10B), and transformation efficiencies of cells used therein (FIG. 10C). FIG. 10A also discloses SEQ ID NOS 343-345, respectively, in order of appearance.

[0036] FIGS. 11A-C show histograms of distances between adjacent *S. pyogenes* SF370 locus 1 PAM (NGG) (FIG. 11A) and *S. thermophilus* LMD9 locus 2 PAM (NN-GAAW) (FIG. 11B) in the human genome; and distances for each PAM by chromosome (Chr) (FIG. 11C).

[0037] FIGS. 12A-C show an exemplary CRISPR system, an example adaptation for expression in eukaryotic cells, and results of tests assessing CRISPR activity. FIG. 12A provides a Schematic illustration of CRISPR 1 from *S. thermophilus* LMD-9. FIG. 12B illustrates the design of an

expression system for the *S. thermophilus* CRISPR system and discloses SEQ ID NOS 346-347, respectively, in order of appearance. An algorithm generating the colors of each base according to its probability of assuming the predicted secondary structure is indicated by a rainbow scale and reproduced in gray scale in FIG. 12C. FIG. 12C also discloses SEQ ID NO: 348.

[0038] FIGS. 13A-C show exemplary manipulations of a CRISPR system for targeting of genomic loci in mammalian cells. FIG. 13A provides a schematic of the human EMX1 locus showing the location of five protospacers, indicated by the underlined sequences. FIG. 13B provides a schematic of the pre-crRNA/trrRNA complex (top), and a schematic of a chimeric RNA design comprising a 20 bp guide sequence in a hairpin structure (bottom). Results of a Surveyor assay is provided in FIG. 13C. FIG. 13A also discloses SEQ ID NO: 349. FIG. 13B also discloses SEQ ID NOS 350-352, respectively, in order of appearance.

[0039] FIGS. 14A-B. FIG. 14A discloses SEQ ID NO: 353. FIG. 14B shows the results of a Northern blot analysis of crRNA processing in mammalian cells.

[0040] FIGS. 15A-B show an exemplary selection of protospacers in the human PVALB and mouse Th loci. FIG. 15A discloses SEQ ID NO: 354. FIG. 15B discloses SEQ ID NO: 355.

[0041] FIG. 16 shows example protospacer and corresponding PAM sequence targets of the *S. thermophilus* CRISPR system in the human EMX1 locus (SEQ ID NO: 348).

[0042] FIG. 17 provides a table of sequences for primers and probes (SEQ ID NOS 36-39 and 356-363, respectively, in order of appearance) used for Surveyor, RFLP, genomic sequencing, and Northern blot assays.

[0043] FIGS. 18A-C. show exemplary manipulation of a CRISPR system with chimeric RNAs and results of SURVEYOR assays for system activity in eukaryotic cells. FIG. 18A discloses SEQ ID NO: 364, respectively, in order of appearance and illustrates a schematic of a bicistronic expression vector for chimeric RNA and Cas9. Results of Surveyor assays for Cas9-mediated indels at the human EMX1 and PVALB loci are provided in FIG. 18B and FIG. 18C, respectively.

[0044] FIGS. 19A-B show a graphical representation of the results of SURVEYOR assays for CRISPR system activity in eukaryotic cells (SEQ ID NOS 365-443, respectively, in order of appearance). Quantification of the results corresponding to FIGS. 18B and 18C are illustrated by histogram in FIG. 19A and FIG. 19B, respectively.

[0045] FIG. 20 shows an exemplary visualization of some *S. pyogenes* Cas9 target sites in the human genome using the UCSC genome browser.

[0046] FIG. 21 shows predicted secondary structures for exemplary chimeric RNAs comprising a guide sequence, tracr mate sequence, and tracr sequence (SEQ ID NOS 444-463, respectively, in order of appearance).

[0047] FIG. 22 shows exemplary bicistronic expression vectors for expression of CRISPR system elements in eukaryotic cells (SEQ ID NOS 464 and 341-342, respectively, in order of appearance).

[0048] FIGS. 23A-B show that Cas9 nuclease activity against endogenous targets may be exploited for genome editing. FIG. 23A: Concept of genome editing using the CRISPR system. The CRISPR targeting construct directed cleavage of a chromosomal locus and was co-transformed

with an editing template that recombined with the target to prevent cleavage. Kanamycin-resistant transformants that survived CRISPR attack contained modifications introduced by the editing template. tracr, trans-activating CRISPR RNA; aphA-3, kanamycin resistance gene. FIG. 23B: Transformation of crR6M DNA in R6^{8232.5} cells with no editing template, the R6 wild-type srtA or the R6370. 1 editing templates. Recombination of either R6 srtA or R6^{370.1} prevented cleavage by Cas9. Transformation efficiency was calculated as colony forming units (cfu) per µg of crR6M DNA; the mean values with standard deviations from at least three independent experiments are shown. PCR analysis was performed on 8 clones in each transformation. "Un." indicates the unedited srtA locus of strain R6^{8232.5}, "Ed." shows the editing template. R6^{8232.5} and R6^{370.1} targets are distinguished by restriction with EaeI.

[0049] FIGS. 24A-C show analysis of PAM and seed sequences that eliminate Cas9 cleavage. FIG. 24A: PCR products with randomized PAM sequences or randomized seed sequences were transformed in crR6 cells (SEQ ID NOS 465-469, respectively, in order of appearance). These cells expressed Cas9 loaded with a crRNA that targeted a chromosomal region of R6^{8232.5} cells (highlighted in pink) that is absent from the R6 genome. More than 2x10⁵ chloramphenicol-resistant transformants, carrying inactive PAM or seed sequences, were combined for amplification and deep sequencing of the target region. FIG. 24B: Relative proportion of number of reads after transformation of the random PAM constructs in crR6 cells (compared to number of reads in R6 transformants). The relative abundance for each 3-nucleotide PAM sequence is shown. Severely underrepresented sequences (NGG) are shown in red; partially underrepresented one in orange (NAG). FIG. 24C: Relative proportion of number of reads after transformation of the random seed sequence constructs in crR6 cells (compared to number of reads in R6 transformants). The relative abundance of each nucleotide for each position of the first 20 nucleotides of the protospacer sequence is shown (SEQ ID NO: 470). High abundance indicates lack of cleavage by Cas9, i.e. a CRISPR inactivating mutation. The grey line shows the level of the WT sequence. The dotted line represents the level above which a mutation significantly disrupts cleavage (See section "Analysis of deep sequencing data" in Example 5).

[0050] FIGS. 25A-F show introduction of single and multiple mutations using the CRISPR system in *S. pneumoniae*. FIG. 25A: Nucleotide and amino acid sequences of the wild-type and edited (green nucleotides; underlined amino acid residues) bgaA. The protospacer, PAM and restriction sites are shown (SEQ ID NOS 471-475 and 472, respectively, in order of appearance). FIG. 25B: Transformation efficiency of cells transformed with targeting constructs in the presence of an editing template or control. FIG. 25C: PCR analysis for 8 transformants of each editing experiment followed by digestion with BtgZI (R→A) and TseI (NE→AA). Deletion of bgaA was revealed as a smaller PCR product. FIG. 25D: Miller assay to measure the β-galactosidase activity of WT and edited strains. FIG. 25E: For a single-step, double deletion the targeting construct contained two spacers (in this case matching srtA and bgaA) and was co-transformed with two different editing templates. FIG. 25F: PCR analysis for 8 transformants to detect deletions in srtA and bgaA loci. 6/8 transformants contained deletions of both genes.

[0051] FIGS. 26A-D provide mechanisms underlying editing using the CRISPR system. FIG. 26A: A stop codon was introduced in the erythromycin resistance gene ermAM to generate strain JEN53. The wild-type sequence can be restored by targeting the stop codon with the CRISPR: ermAM (stop) construct, and using the ermAM wild-type sequence as an editing template. FIG. 26B: Mutant and wild-type ermAM sequences (SEQ ID NOS 476-479, respectively, in order of appearance). FIG. 26C: Fraction of erythromycin-resistant (erm^R) cfu calculated from total or kanamycin-resistant (kan^R) cfu. FIG. 26D: Fraction of total cells that acquire both the CRISPR construct and the editing template. Co-transformation of the CRISPR targeting construct produced more transformants (t-test, p=0.011). In all cases the values show the mean±s.d. for three independent experiments.

[0052] FIGS. 27A-D illustrate genome editing with the CRISPR system in *E. coli*. FIG. 27A: A kanamycin-resistant plasmid carrying the CRISPR array (pCRISPR) targeting the gene to edit may be transformed in the HME63 recombiner strain containing a chloramphenicol-resistant plasmid harboring cas9 and tracr (pCas9), together with an oligonucleotide specifying the mutation. FIG. 27B: A K42T mutation conferring streptomycin resistance was introduced in the rpsL gene (SEQ ID NOS 480-483, respectively, in order of appearance). FIG. 27C: Fraction of streptomycin-resistant (strep^R) cfu calculated from total or kanamycin-resistant (kan^R) cfu. FIG. 27D: Fraction of total cells that acquire both the pCRISPR plasmid and the editing oligonucleotide. Co-transformation of the pCRISPR targeting plasmid produced more transformants (t-test, p=0.004). In all cases the values showed the mean±s.d. for three independent experiments.

[0053] FIGS. 28A-B illustrate the transformation of crR6 genomic DNA leads to editing of the targeted locus. FIG. 28A: The IS1167 element of *S. pneumoniae* R6 was replaced by the CRISPR01 locus of *S. pyogenes* SF370 to generate crR6 strain. This locus encodes for the Cas9 nuclease, a CRISPR array with six spacers, the tracrRNA that is required for crRNA biogenesis and Cas1, Cas2 and Csn2, proteins not necessary for targeting. Strain crR6M contains a minimal functional CRISPR system without cas1, cas2 and csn2. The aphA-3 gene encodes kanamycin resistance. Protospacers from the streptococcal bacteriophages φ8232.5 and φ370.1 were fused to a chloramphenicol resistance gene (cat) and integrated in the srtA gene of strain R6 to generate strains R68232.5 and R6370.1. FIG. 28B: Left panel: Transformation of crR6 and crR6M genomic DNA in R6^{8232.5} and R6^{370.1}. As a control of cell competence a streptomycin resistant gene was also transformed. Right panel: PCR analysis of 8 R6^{8232.5} transformants with crR6 genomic DNA. Primers that amplify the srtA locus were used for PCR. 7/8 genotyped colonies replaced the R68232.5 srtA locus by the WT locus from the crR6 genomic DNA.

[0054] FIGS. 29A-F provide chromatograms of DNA sequences of edited cells obtained in this study. In all cases the wild-type and mutant protospacer and PAM sequences (or their reverse complement) are indicated. When relevant, the amino acid sequence encoded by the protospacer is provided. For each editing experiment, all strains for which PCR and restriction analysis corroborated the introduction of the desired modification were sequenced. A representative chromatogram is shown. FIG. 29A: Chromatogram for the

introduction of a PAM mutation into the R6^{8232.5} target (FIG. 23d) (SEQ ID NOS 484-485, respectively, in order of appearance).

[0055] FIG. 29B: Chromatograms for the introduction of the R>A and NE>AA mutations into β-galactosidase (bgaA) (FIG. 25c) (SEQ ID NOS 471-475 and 472, respectively, in order of appearance). FIG. 29C: Chromatogram for the introduction of a 6664 bp deletion within bgaA ORF (FIGS. 25c and 25f). The dotted line indicates the limits of the deletion (SEQ ID NOS 486-488, respectively, in order of appearance). FIG. 29D: Chromatogram for the introduction of a 729 bp deletion within srtA ORF (FIG. 25f). The dotted line indicates the limits of the deletion (SEQ ID NOS 489-491, respectively, in order of appearance). FIG. 29E: Chromatograms for the generation of a premature stop codon within ermAM (FIG. 33) (SEQ ID NOS 492-495, respectively, in order of appearance). FIG. 29F: rpsL editing in *E. coli* (FIG. 27) (SEQ ID NOS 480-483, respectively, in order of appearance).

[0056] FIGS. 30A-C illustrate CRISPR immunity against random *S. pneumoniae* targets containing different PAMs. FIG. 30A: Position of the 10 random targets on the *S. pneumoniae* R6 genome. The chosen targets have different PAMs and are on both strands. FIG. 30B: Spacers corresponding to the targets were cloned in a minimal CRISPR array on plasmid pLZ12 and transformed into strain crR6Rc, which supplies the processing and targeting machinery in trans. FIG. 30C: Transformation efficiency of the different plasmids in strain R6 and crR6Rc. No colonies were recovered for the transformation of pDB99-108 (T1-T10) in crR6Rc. The dashed line represents limit of detection of the assay.

[0057] FIG. 31 provides a general scheme for targeted genome editing. To facilitate targeted genome editing, crR6M was further engineered to contain tracrRNA, Cas9 and only one repeat of the CRISPR array followed by kanamycin resistance marker (aphA-3), generating strain crR6Rk. DNA from this strain is used as a template for PCR with primers designed to introduce a new spacer (green box designated with N). The left and right PCRs are assembled using the Gibson method to create the targeting construct. Both the targeting and editing constructs are then transformed into strain crR6Rc, which is a strain equivalent to crR6Rk but has the kanamycin resistance marker replaced by a chloramphenicol resistance marker (cat). About 90% of the kanamycin-resistant transformants contain the desired mutation.

[0058] FIG. 32 illustrates the distribution of distances between PAMs. NGG and CCN that are considered to be valid PAMs. Data is shown for the *S. pneumoniae* R6 genome as well as for a random sequence of the same length and with the same GC-content (39.7%). The dotted line represents the average distance (12) between PAMs in the R6 genome.

[0059] FIGS. 33A-D illustrate CRISPR-mediated editing of the ermAM locus using genomic DNA as targeting construct. To use genomic DNA as targeting construct it is necessary to avoid CRISPR autoimmunity, and therefore a spacer against a sequence not present in the chromosome must be used (in this case the ermAM erythromycin resistance gene). FIG. 33A: Nucleotide and amino acid sequences of the wild-type and mutated (red letters) ermAM gene. The protospacer and PAM sequences are shown (SEQ ID NOS 492-495, respectively, in order of appearance). FIG.

33B: A schematic for CRISPR-mediated editing of the ermAM locus using genomic DNA. A construct carrying an ermAM-targeting spacer (blue box) is made by PCR and Gibson assembly, and transformed into strain crR6Rc, generating strain JEN37. The genomic DNA of JEN37 was then used as a targeting construct, and was co-transformed with the editing template into JEN38, a strain in which the srtA gene was replaced by a wild-type copy of ermAM. Kanamycin-resistant transformants contain the edited genotype (JEN43). FIG. 33C: Number of kanamycin-resistant cells obtained after co-transformation of targeting and editing or control templates. In the presence of the control template 5.4×10^3 cfu/ml were obtained, and 4.3×10^5 cfu/ml when the editing template was used. This difference indicates an editing efficiency of about 99% [$(4.3 \times 10^5 - 5.4 \times 10^3) / 4.3 \times 10^5$]. FIG. 33D: To check for the presence of edited cells seven kanamycin-resistant clones and JEN38 were streaked on agar plates with (erm+) or without (erm-) erythromycin. Only the positive control displayed resistance to erythromycin. The ermAM mut genotype of one of these transformants was also verified by DNA sequencing (FIG. 29e).

[0060] FIGS. 34A-D illustrate sequential introduction of mutations by CRISPR-mediated genome editing. FIG. 34A: A schematic for sequential introduction of mutations by CRISPR-mediated genome editing. First, R6 is engineered to generate crR6Rk. crR6Rk is co-transformed with a srtA-targeting construct fused to cat for chloramphenicol selection of edited cells, along with an editing construct for a ΔsrtA in-frame deletion. Strain crR6 ΔsrtA is generated by selection on chloramphenicol. Subsequently, the ΔsrtA strain is co-transformed with a bgaA-targeting construct fused to aphA-3 for kanamycin selection of edited cells, and an editing construct containing a ΔbgaA in-frame deletion. Finally, the engineered CRISPR locus can be erased from the chromosome by first co-transforming R6 DNA containing the wild-type IS1167 locus and a plasmid carrying a bgaA protospacer (pDB97), and selection on spectinomycin. FIG. 34B: PCR analysis for 8 chloramphenicol (Cam)-resistant transformants to detect the deletion in the srtA locus. FIG. 34C: β-galactosidase activity as measured by Miller assay. In *S. pneumoniae*, this enzyme is anchored to the cell wall by sortase A. Deletion of the srtA gene results in the release of β-galactosidase into the supernatant. ΔbgaA mutants show no activity. FIG. 34D: PCR analysis for 8 spectinomycin (Spec)-resistant transformants to detect the replacement of the CRISPR locus by wild-type IS1167.

[0061] FIGS. 35A-C illustrate the background mutation frequency of CRISPR in *S. pneumoniae*. FIG. 35A: Transformation of the CRISPR::Ø or CRISPR::erm (stop) targeting constructs in JEN53, with or without the ermAM editing template. The difference in kan^R CFU between CRISPR::Ø and CRISPR::erm (stop) indicates that Cas9 cleavage kills non-edited cells. Mutants that escape CRISPR interference in the absence of editing template are observed at a frequency of 3×10^{-3} . FIG. 35B: PCR analysis of the CRISPR locus of escapers shows that 7/8 have a spacer deletion. FIG. 35C: Escaper #2 carries a point mutation in cas9 (SEQ ID NOS 496-499, respectively, in order of appearance).

[0062] FIG. 36 illustrates that the essential elements of the *S. pyogenes* CRISPR locus 1 are reconstituted in *E. coli* using pCas9. The plasmid contained tracrRNA, Cas9, as well as a leader sequence driving the crRNA array. The pCRISPR plasmids contained the leader and the array only. Spacers may be inserted into the crRNA array between BsaI

sites using annealed oligonucleotides (SEQ ID NOS 343, 500 and 127, respectively, in order of appearance). Oligonucleotide design is shown at bottom. pCas9 carried chloramphenicol resistance (CmR) and is based on the low-copy pACYC184 plasmid backbone. pCRISPR is based on the high-copy number pZE21 plasmid. Two plasmids were required because a pCRISPR plasmid containing a spacer targeting the *E. coli* chromosome may not be constructed using this organism as a cloning host if Cas9 is also present (it will kill the host).

[0063] FIG. 37 illustrates CRISPR-directed editing in *E. coli* MG1655. An oligonucleotide (W542) carrying a point mutation that both confers streptomycin resistance and abolishes CRISPR immunity, together with a plasmid targeting rpsL. (pCRISPR::rpsL) or a control plasmid (pCRISPR::Ø) were co-transformed into wild-type *E. coli* strain MG1655 containing pCas9. Transformants were selected on media containing either streptomycin or kanamycin. Dashed line indicates limit of detection of the transformation assay.

[0064] FIGS. 38A-B illustrate the background mutation frequency of CRISPR in *E. coli* HME63. FIG. 38A: Transformation of the pCRISPR::Ø or pCRISPR: rpsL plasmids into HME63 competent cells. Mutants that escape CRISPR interference were observed at a frequency of 2.6×10^{-4} . FIG. 38B: Amplification of the CRISPR array of escapers showed that 8/8 have deleted the spacer.

[0065] FIGS. 39A-D. FIG. 39A shows a portion of a circular depiction of a phylogenetic analysis. FIG. 39B shows another portion of a circular depiction of the phylogenetic analysis. FIG. 39C shows yet another portion of a circular depiction of the phylogenetic analysis. FIG. 39D shows still another portion of a circular depiction of the phylogenetic analysis. The analysis reveals five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids).

[0066] FIGS. 40A-F. FIG. 40A shows a portion of a linear depiction of the phylogenetic analysis. FIG. 40B shows another portion of the linear depiction of the phylogenetic analysis. FIG. 40C shows yet another portion of the linear depiction of the phylogenetic analysis. FIG. 40D shows still another portion of the linear depiction of the phylogenetic analysis. FIG. 40E shows still yet another portion of the linear depiction of the phylogenetic analysis. FIG. 40F shows a last portion of the linear depiction of the phylogenetic analysis. The analysis reveals five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids).

[0067] FIG. 41A-M shows sequences where the mutation points are located within the SpCas9 gene (SEQ ID NOS 501-502, respectively, in order of appearance). FIG. 41A shows a portion of the sequences. FIG. 41B shows a portion of the sequences. FIG. 41C shows a portion of the sequences. FIG. 41D shows a portion of the sequences. FIG. 41E shows a portion of the sequences. FIG. 41F shows a portion of the sequences. FIG. 41G shows a portion of the sequences. FIG. 41H shows a portion of the sequences. FIG. 41I shows a portion of the sequences. FIG. 41J shows a portion of the sequences. FIG. 41K shows a portion of the sequences. FIG. 41L shows a portion of the sequences. FIG. 41M shows a portion of the sequences.

[0068] FIG. 42 shows a schematic construct in which the transcriptional activation domain (VP64) is fused to Cas9 with two mutations in the catalytic domains (D10 and H840).

[0069] FIG. 43A-D shows genome editing via homologous recombination. FIG. 43A: Schematic of SpCas9 nickase, with D10A mutation in the RuvC I catalytic domain. FIG. 43B: 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. FIG. 43C: Sequence of region modified by HR. FIG. 43D: SURVEYOR assay for wildtype (wt) and nickase (D10A) SpCas9-mediated indels at the EMX1 target 1 locus (n 3) (SEQ ID NOS 503-505, 503, 506 and 505, respectively, in order of appearance). Arrows indicate positions of expected fragment sizes.

[0070] FIGS. 44A-B show single vector designs for SpCas9. FIG. 44A discloses SEQ ID NOS 320-321 and 328, respectively, in order of appearance. FIG. 44B discloses SEQ ID NO: 329.

[0071] FIG. 45 shows quantification of cleavage of NLS-Csn1 constructs NLS-Csn1, Csn1, Csn1-NLS, NLS-Csn1-NLS, NLS-Csn1-GFP-NLS and UnTFN.

[0072] FIG. 46 shows index frequency of NLS-Cas9, Cas9, Cas9-NLS and NLS-Cas9-NLS.

[0073] FIG. 47 shows a gel demonstrating that SpCas9 with nickase mutations (individually) do not induce double strand breaks.

[0074] FIG. 48A shows a design of the oligo DNA used as Homologous Recombination (HR) template in this experiment and FIG. 48B shows a comparison of HR efficiency induced by different combinations of Cas9 protein and HR template.

[0075] FIG. 49A shows the Conditional Cas9, Rosa26 targeting vector map.

[0076] FIG. 49B shows the Constitutive Cas9, Rosa26 targeting vector map.

[0077] FIG. 50A-H show the sequences of each element present in the vector maps of FIGS. 49A-B (SEQ ID NOS 507-516, respectively, in order of appearance). FIG. 50A shows a portion of the sequences. FIG. 50B shows a portion of the sequences. FIG. 50C shows a portion of the sequences. FIG. 50D shows a portion of the sequences. FIG. 50E shows a portion of the sequences. FIG. 50F shows a portion of the sequences. FIG. 50G shows a portion of the sequences. FIG. 50H shows a portion of the sequences.

[0078] FIG. 51 shows a schematic of the important elements in the Constitutive and Conditional Cas9 constructs.

[0079] FIG. 52 shows the functional validation of the expression of Constitutive and Conditional Cas9 constructs.

[0080] FIG. 53 shows the validation of Cas9 nuclease activity by Surveyor.

[0081] FIG. 54 shows the quantification of Cas9 nuclease activity.

[0082] FIG. 55 shows construct design and homologous recombination (HR) strategy.

[0083] FIG. 56 shows the genomic PCR genotyping results for the constitutive (Right) and conditional (Left) constructs at two different gel exposure times (top row for 3 min and bottom row for 1 min).

[0084] FIG. 57 shows Cas9 activation in mESCs.

[0085] FIG. 58 shows a schematic of the strategy used to mediate gene knockout via NHEJ using a nickase version of Cas9 along with two guide RNAs.

[0086] FIG. 59 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.

[0087] FIG. 60 shows the timeline and overview of experiments. Steps for reagent design, construction, validation, and cell line expansion. Custom sgRNAs (light blue bars) for each target, as well as genotyping primers, are designed in silico via our online design tool (available at the website genome-engineering.org/tools). sgRNA expression vectors are then cloned into a plasmid containing Cas9 (PX330) and verified via DNA sequencing. Completed plasmids (pCRIS-PRs), and optional repair templates for facilitating homology directed repair, are then transfected into cells and assayed for ability to mediate targeted cleavage. Finally, transfected cells can be clonally expanded to derive isogenic cell lines with defined mutations.

[0088] FIG. 61A-C shows Target selection and reagent preparation. FIG. 61A: For *S. pyogenes* Cas9, 20-bp targets (highlighted in blue) must be followed by 5'-NGG, which can occur in either strand on genomic DNA. We recommend using the online tool described in this protocol in aiding target selection (www.genome-engineering.org/tools). FIG. 61B: Schematic for co-transfection of Cas9 expression plasmid (PX165) and PCR-amplified U6-driven sgRNA expression cassette. Using a U6 promoter-containing PCR template and a fixed forward primer (U6 Fwd), sgRNA-encoding DNA can appended onto the U6 reverse primer (U6 Rev) and synthesized as an extended DNA oligo (Ultragene oligos from IDT). Note the guide sequence (blue N's) in U6 Rev is the reverse complement of the 5'-NGG flanking target sequence (SEQ ID NOS 517 and 517-519, respectively, in order of appearance). FIG. 61C: Schematic for scarless cloning of the guide sequence oligos into a plasmid containing Cas9 and sgRNA scaffold (PX330). The guide oligos (blue N's) contain overhangs for ligation into the pair of BbsI sites on PS330, with the top and bottom strand orientations matching those of the genomic target (i.e. top oligo is the 20-nt sequence preceding 5'-NGG in genomic DNA). Digestion of PX330 with BbsI allows the replacement of the Type IIs restriction sites (blue outline) with direct insertion of annealed oligos. It is worth noting that an extra G was placed before the first base of the guide sequence. Applicants have found that an extra G in front of the guide sequence does not adversely affect targeting efficiency. In cases when the 20-nt guide sequence of choice does not begin with guanine, the extra guanine will ensure the sgRNA is efficiently transcribed by the U6 promoter, which prefers a guanine in the first base of the transcript (SEQ ID NOS 320-321 and 328, respectively, in order of appearance).

[0089] FIG. 62A-D shows the anticipated results for multiplex NHEJ. FIG. 62A: Schematic of the SURVEYOR assay used to determine indel percentage. First, genomic DNA from the heterogeneous population of Cas9-targeted

cells is amplified by PCR. Amplicons are then reannealed slowly to generate heteroduplexes. The reannealed heteroduplexes are cleaved by SURVEYOR nuclease, whereas homoduplexes are left intact. Cas9-mediated cleavage efficiency (% indel) is calculated based on the fraction of cleaved DNA, as determined by integrated intensity of gel bands. FIG. 62B: Two sgRNAs (orange and blue bars) are designed to target the human GRIN2B and DYRK1A loci. SURVEYOR gel shows modification at both loci in transfected cells. Colored arrows indicated expected fragment sizes for each locus. FIG. 62C: A pair of sgRNAs (light blue and green bars) are designed to excise an exon (dark blue) in the human EMX1 locus. Target sequences and PAMs (red) are shown in respective colors, and sites of cleavage indicated by red triangle. Predicted junction is shown below. Individual clones isolated from cell populations transfected with sgRNA 3, 4, or both are assayed by PCR (OUT Fwd, OUT Rev), reflecting a deletion of ~270-bp. Representative clones with no modification (12/23), mono-allelic (10/23), and bi-allelic (1/23) modifications are shown. IN Fwd and IN Rev primers are used to screen for inversion events (FIG. 6d) (SEQ ID NOS 520-522, respectively, in order of appearance). FIG. 62D: Quantification of clonal lines with EMX1 exon deletions. Two pairs of sgRNAs (3.1, 3.2 left-flanking sgRNAs; 4.1, 4.2, right flanking sgRNAs) are used to mediate deletions of variable sizes around one EMX1 exon. Transfected cells are clonally isolated and expanded for genotyping analysis for deletions and inversion events. Of the 105 clones are screened, 51 (49%) and 11 (10%) carrying heterozygous and homozygous deletions, respectively. Approximate deletion sizes are given since junctions may be variable.

[0090] FIG. 63A-C shows the application of ssODNs and targeting vector to mediate HR with both wildtype and nickase mutant of Cas9 in HEK293FT and HUES9 cells with efficiencies ranging from 1.0-27%. FIG. 63A provides a schematic showing of the design of the application of ssODNs and targeting vector to mediate HR. FIG. 63B shows sequences of genomic locus and ssODN repair template, which are SEQ ID NOS 503-505, 503, 506 and 505, respectively, in order of appearance. FIG. 63C shows ssODNs and targeting vector mediate HR with both wild-type and nickase mutant of Cas9 in HEK293FT and HUES9 cells with efficiencies ranging from 1.0-27%.

[0091] FIG. 64 shows a schematic of a PCR-based method for rapid and efficient CRISPR targeting in mammalian cells. A plasmid containing the human RNA polymerase III promoter U6 is PCR-amplified using a U6-specific forward primer and a reverse primer carrying the reverse complement of part of the U6 promoter, the sgRNA (+85) scaffold with guide sequence, and 7 T nucleotides for transcriptional termination. The resulting PCR product is purified and co-delivered with a plasmid carrying Cas9 driven by the CBh promoter (SEQ ID NOS 517, 523, 518 and 524-525, respectively, in order of appearance).

[0092] FIG. 65 shows SURVEYOR Mutation Detection Kit from Transgenomics results for each gRNA and respective controls. A positive SURVEYOR result is one large band corresponding to the genomic PCR and two smaller bands that are the product of the SURVEYOR nuclease making a double-strand break at the site of a mutation. Each gRNA was validated in the mouse cell line, Neuro-N2a, by liposomal transient co-transfection with hSpCas9. 72 hours

post-transfection genomic DNA was purified using QuickExtract DNA from Epicentre. PCR was performed to amplify the locus of interest.

[0093] FIG. 66 shows Surveyor results for 38 live pups (lanes 1-38) 1 dead pup (lane 39) and 1 wild-type pup for comparison (lane 40). Pups 1-19 were injected with gRNA Chd8.2 and pups 20-38 were injected with gRNA Chd8.3. Of the 38 live pups, 13 were positive for a mutation. The one dead pup also had a mutation. There was no mutation detected in the wild-type sample. Genomic PCR sequencing was consistent with the SURVEYOR assay findings (SEQ ID NOS 526-528, respectively, in order of appearance).

[0094] FIG. 67 shows a design of different Cas9 NLS constructs. All Cas9 were the human-codon-optimized version of the Sp Cas9. NLS sequences are linked to the cas9 gene at either N-terminus or C-terminus. All Cas9 variants with different NLS designs were cloned into a backbone vector containing so it is driven by EF1a promoter. On the same vector there is a chimeric RNA targeting human EMX1 locus driven by U6 promoter, together forming a two-component system.

[0095] FIG. 68 shows the efficiency of genomic cleavage induced by Cas9 variants bearing different NLS designs. The percentage indicate the portion of human EMX1 genomic DNA that were cleaved by each construct. All experiments are from 3 biological replicates. n=3, error indicates S.E.M.

[0096] FIG. 69A shows a design of the CRISPR-TF (Transcription Factor) with transcriptional activation activity. The chimeric RNA is expressed by U6 promoter, while a human-codon-optimized, double-mutant version of the Cas9 protein (hSpCas9m), operably linked to triple NLS and a VP64 functional domain is expressed by a EF1a promoter. The double mutations, D10A and H840A, renders the cas9 protein unable to introduce any cleavage but maintained its capacity to bind to target DNA when guided by the chimeric RNA.

[0097] FIG. 69B shows transcriptional activation of the human SOX2 gene with CRISPR-TF system (Chimeric RNA and the Cas9-NLS-VP64 fusion protein). 293 FT cells were transfected with plasmids bearing two components: (1) U6-driven different chimeric RNAs targeting 20-bp sequences within or around the human SOX2 genomic locus, and (2) EF1a-driven hSpCas9m (double mutant)-NLS-VP64 fusion protein. 96 hours post transfection, 293 FT cells were harvested and the level of activation is measured by the induction of mRNA expression using a qRT-PCR assay. All expression levels are normalized against the control group (grey bar), which represents results from cells transfected with the CRISPR-TF backbone plasmid without chimeric RNA. The qRT-PCR probes used for detecting the SOX2 mRNA is Taqman Human Gene Expression Assay (Life Technologies). All experiments represents data from 3 biological replicates, n=3, error bars show s.e.m.

[0098] FIG. 70 depicts NLS architecture optimization for SpCas9.

[0099] FIG. 71 shows a QQ plot for NGGNN sequences.

[0100] FIG. 72 shows a histogram of the data density with fitted normal distribution (black line) and 0.99 quantile (dotted line).

[0101] FIG. 73A-C shows RNA-guided repression of bgaA expression by dgRNA::cas9**. FIG. 73A. The Cas9 protein binds to the tracrRNA, and to the precursor CRISPR RNA which is processed by RNaseIII to form the crRNA. The crRNA directs binding of Cas9 to the bgaA promoter

and represses transcription. FIG. 73B. The targets used to direct Cas9** to the bgaA promoter are represented (SEQ ID NO: 529). Putative -35, -10 as well as the bgaA start codon are in bold. FIG. 73C. Betagalactosidase activity as measure by Miller assay in the absence of targeting and for the four different targets.

[0102] FIG. 74A-E shows characterization of Cas9** mediated repression. FIG. 74A. The gfpmut2 gene and its promoter, including the -35 and -10 signals are represented together with the position of the different target sites used the study. FIG. 74B. Relative fluorescence upon targeting of the coding strand. FIG. 74C. Relative fluorescence upon targeting of the non-coding strand. FIG. 74D. Northern blot with probes B477 and B478 on RNA extracted from T5, T10, B10 or a control strain without a target. FIG. 74E. Effect of an increased number of mutations in the 5' end of the crRNA of B1, T5 and B10.

[0103] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE INVENTION

[0104] 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. 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.

[0105] In aspects of the invention the terms “chimeric RNA”, “chimeric guide RNA”, “guide RNA”, “single guide RNA” and “synthetic guide RNA” are used interchangeably and refer to the polynucleotide sequence comprising the guide sequence, the tracr sequence and the tracr mate sequence. The term “guide sequence” refers to the about 20 bp sequence within the guide RNA that specifies the target site and may be used interchangeably with the terms “guide” or “spacer”. The term “tracr mate sequence” may also be used interchangeably with the term “direct repeat(s)”.

[0106] 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.

[0107] 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.

[0108] 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.

[0109] "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.

[0110] 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.

[0111] "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.

[0112] As used herein, "expression" 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.

[0113] 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.

[0114] The terms "subject," "individual," and "patient" are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. Tissues, cells and their progeny of a biological entity obtained *in vivo* or cultured *in vitro* are also encompassed.

[0115] The terms "therapeutic agent", "therapeutic capable agent" or "treatment agent" are used interchangeably and refer to a molecule or compound that confers some beneficial effect upon administration to a subject. The beneficial effect includes enablement of diagnostic determinations; amelioration of a disease, symptom, disorder, or pathological condition; reducing or preventing the onset of a disease, symptom, disorder or condition; and generally counteracting a disease, symptom, disorder or pathological condition.

[0116] As used herein, "treatment" or "treating," or "palliating" or "ameliorating" are used interchangeably. These terms refer to an approach for obtaining beneficial or desired results including but not limited to a therapeutic benefit and/or a prophylactic benefit. By therapeutic benefit is meant any therapeutically relevant improvement in or effect on one or more diseases, conditions, or symptoms under treatment. For prophylactic benefit, the compositions may be administered to a subject at risk of developing a particular disease, condition, or symptom, or to a subject reporting one or more of the physiological symptoms of a disease, even though the disease, condition, or symptom may not have yet been manifested.

[0117] The term "effective amount" or "therapeutically effective amount" refers to the amount of an agent that is sufficient to effect beneficial or desired results. The therapeutically effective amount may vary depending upon one or more of: the subject and disease condition being treated, the weight and age of the subject, the severity of the disease condition, the manner of administration and the like, which can readily be determined by one of ordinary skill in the art. The term also applies to a dose that will provide an image for detection by any one of the imaging methods described herein. The specific dose may vary depending on one or more of: the particular agent chosen, the dosing regimen to be followed, whether it is administered in combination with other compounds, timing of administration, the tissue to be imaged, and the physical delivery system in which it is carried.

[0118] 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 PRO-

TOCOLS 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) ANTI-BODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R. I. Freshney, ed. (1987)).

[0119] Several aspects of the invention relate to vector systems comprising one or more vectors, or vectors as such. Vectors can be designed for expression of CRISPR transcripts (e.g. nucleic acid transcripts, proteins, or enzymes) in prokaryotic or eukaryotic cells. For example, CRISPR transcripts can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

[0120] Vectors may be introduced and propagated in a prokaryote. 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.

[0121] 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).

[0122] In some embodiments, a vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYEpSec1 (Baldari, et al., 1987. *EMBO J.* 6:229-234), pMFa (Kuijana 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.).

[0123] 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).

[0124] 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.

[0125] In some embodiments, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, et al., 1987. *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8:729-733) and immunoglobulins (Baneiji, et al., 1983. *Cell* 33:729-740; Queen and Baltimore, 1983. *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund, et al., 1985. *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990. *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3:537-546).

[0126] 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 *Haloférax 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*, *Pyrococcus*, *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*.

[0127] 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 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. Full complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization and promote 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. In some embodiments, the target sequence may be within an organelle of a eukaryotic cell, for example, mitochondrion or chloroplast. A sequence or template that may be used for recombination into the targeted locus comprising the target sequences is referred to as an

“editing template” or “editing polynucleotide” or “editing sequence”. In aspects of the invention, an exogenous template polynucleotide may be referred to as an editing template. In an aspect of the invention the recombination is homologous recombination.

[0128] 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, the tracr sequence has sufficient complementarity to a tracr mate sequence to hybridize and participate in formation of a CRISPR complex. As with the target sequence, it is believed that complete complementarity is not needed, provided there is sufficient to be functional. In some embodiments, the tracr sequence has at least 50%, 60%, 70%, 80%, 90%, 95% or 99% of sequence complementarity along the length of the tracr mate sequence when optimally aligned. 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.

[0129] 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.

[0130] 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, homologs thereof, or modified versions thereof. These enzymes are known; for example, the amino acid sequence of *S. pyogenes* Cas9 protein may be found in the SwissProt database under accession number Q99ZW2. In some embodiments, the unmodified CRISPR enzyme has DNA cleavage activity, such as Cas9. In some embodiments the CRISPR enzyme is Cas9, and may be Cas9 from *S. pyogenes* or *S. pneumoniae*. 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. In some embodiments, a Cas9 nickase may be used in combination with guide sequence(es), e.g., two guide sequences, which target respectively sense and antisense strands of the DNA target. This combination allows both strands to be nicked and used to induce NHEJ. Applicants have demonstrated (data not shown) the efficacy of two nickase targets (i.e., sgRNAs targeted at the same location but to different strands

of DNA) in inducing mutagenic NHEJ. A single nickase (Cas9-D10A with a single sgRNA) is unable to induce NHEJ and create indels but Applicants have shown that double nickase (Cas9-D10A and two sgRNAs targeted to different strands at the same location) can do so in human embryonic stem cells (hESCs). The efficiency is about 50% of nuclease (i.e., regular Cas9 without D10 mutation) in hESCs.

[0131] As a further example, two or more catalytic domains of Cas9 (RuvC I, RuvC II, and RuvC III) 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. Other mutations may be useful; where the Cas9 or other CRISPR enzyme is from a species other than *S. pyogenes*, mutations in corresponding amino acids may be made to achieve similar effects.

[0132] 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 primate. 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", 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.

[0133] In some embodiments, a vector encodes a CRISPR enzyme comprising one or more nuclear localization sequences (NLSSs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSSs. 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. Typically, an NLS consists of one or more short sequences of positively charged lysines or arginines exposed on the protein surface, but other types of NLS are known. 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: 1); the NLS from nucleoplasmin (e.g. the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK (SEQ ID NO: 2)); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID NO: 3) or QRQRNELKRSP (SEQ ID NO: 4); the hRNP1A1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAK-PRNQGGY (SEQ ID NO: 5); the sequence RMRIZFKNKGKDAAELRRRVEVSVELRKAKKD-EQILKRRNV (SEQ ID NO: 6) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID NO: 7) and PPKKARED (SEQ ID NO: 8) of the myoma T protein; the sequence PQPKKKPL (SEQ ID NO: 9) of human p53; the sequence SALIKKKKKMAP (SEQ ID NO: 10) of mouse c-abl IV; the sequences DRLRR (SEQ ID NO: 11) and PKQKKRK (SEQ ID NO: 12) of the influenza virus NS1; the sequence RKLKKKKKKL (SEQ ID NO: 13) of the Hepatitis virus delta antigen; the sequence REKKKFLKRR (SEQ ID NO: 14) of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 15) of the human poly (ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARTKK (SEQ ID NO: 16) of the steroid hormone receptors (human) glucocorticoid.

[0134] 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). Examples of detectable markers include fluorescent proteins (such as Green fluorescent proteins, or GFP; RFP; CFP), and epitope tags (HA tag, flag tag, SNAP tag). 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 muta-

tion 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.

[0135] 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, 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.

[0136] 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 MMMMMMMNNNNNNNNNNNNNNXGG (SEQ ID NO: 530) where NNNNNNNNNNNNNNNXGG SEQ ID NOOOOO: 531) (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 MMMMMMMNNNNNNNNNNNNXGG (SEQ ID NO: 532) where NNNNNNNNNNNNNXGG (SEQ ID NO: 533) (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. For the *S. thermophilus* CRISPR1 Cas9, a unique target sequence in a genome may include a Cas9 target site of the form

MMMMMMNNNNNNNNNNXXAGAAW (SEQ ID NO: 17) where NNNNNNNNNNNNNXXAGAAW (SEQ ID NO: 18) (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* CRISPR 1 Cas9 target site of the form MMMMMMMNNNNNNNNNNXXAGAAW (SEQ ID NO: 19) where NNNNNNNNNNNNNXXAGAAW (SEQ ID NO: 20) (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. For the *S. pyogenes* Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMNNNNNNNNNNXGGXG (SEQ ID NO: 534) where NNNNNNNNNNNNNXGGXG (SEQ ID NO: 535) (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 MMMMMMMNNNNNNNNNNXGGXG (SEQ ID NO: 536) where NNNNNNNNNNNXGGXG (SEQ ID NO: 537) (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.

[0137] In some embodiments, a guide sequence is selected to reduce the degree of secondary structure within the guide sequence. Secondary structure 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 PA Carr and GM Church, 2009, *Nature Biotechnology* 27 (12): 1151-62). Further algorithms may be found in U.S. application Ser. No. TBA (attorney docket 44790.11.2022; Broad Reference BI-2013/004A); incorporated herein by reference.

[0138] 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. Example illustrations of optimal alignment between a tracr sequence and a tracr mate sequence are provided in FIGS. 12B and 13B. 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. 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. 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 some embodiments, the single transcript further includes a transcription termination sequence; preferably this is a polyT sequence, for example six T nucleotides. An example illustration of such a hairpin structure is provided in the lower portion of FIG. 13B, where 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 terminator: sequence represents the transcription
(1)
NNNNNNNNNNNNNNNNNNNNNNNNNNNgttttgtactctcaagatt-taGAAAtaaatcttcgcagaactcacaagataaggctt catggcgaaat-caaacacctgtcatttatgcgcagggttgcgttatcaaTTTTTT (SEQ ID NO:
21);
(2)
NNNNNNNNNNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgc-agaagctacaagataaggcttcatgcgcgaaatca acacccctgtcattt-tatggcagggtgtttcggttatcaaTTTTTT (SEQ ID NO: 22); (3)
NNNNNNNNNNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgc-agaagctacaagataaggcttcatgcgcgaaatca acacccctgtcattt-tatggcagggtgtttcggttatcaaTTTTTT (SEQ ID NO: 23); (4)
NNNNNNNNNNNNNNNNNNNNNNNNNNNgttt-tagagctaGAAAtagaagttaaaataaggcttgcgttatcaacttgaaaaa-agtggcacgcaggcttcgttatcaaTTTTTT (SEQ ID NO: 24); (5)
NNNNNNNNNNNNNNNNNNNNNNNNNNNgttt-tagagetaGAAATA Gcaagttaaaataaggcttgcgttatcaacttgaa-aaatgttTTTTTT (SEQ ID NO: 25); and (6)
NNNNNNNNNNNNNNNNNNNNNNNNNNNgttt-tagagetaAAATAGcaagttaaaataaggcttgcgttatcaTTTTT TTT (SEQ ID NO: 26). 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 (such as illustrated in the top portion of FIG. 13B).

[0139] 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, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 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.

[0140] 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), HeRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and auto-fluorescent 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.

[0141] 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 organisms (such as animals, plants, or fungi) 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).

[0142] Methods of non-viral delivery of nucleic acids include lipofection, nucleofection, 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., TransfectamTM and LipofectinTM). 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).

[0143] 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); Blaeze 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).

[0144] The use of RNA or DNA viral based systems for the delivery of nucleic acids takes 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.

[0145] 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 packag-

ing 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); Sommerfelt et al., *J. Virol.* 63:2374-2378 (1989); Miller et al., *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700). In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors may also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., *Virology* 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzychka, *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 & Muzychka, *PNAS* 81:6466-6470 (1984); and Samulski et al., *J. Virol.* 63:03822-3828 (1989).

[0146] 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 w2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing 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 be 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 is more sensitive than AAV. 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.

[0147] 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, Rat1, CVI, 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.

[0148] 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. In certain embodiments, the organism or subject is a plant. In certain embodiments, the organism or subject or plant is algae. Methods for producing transgenic plants and animals are known in the art, and generally begin with a method of cell transfection, such as described herein.

[0149] 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 com-

plex 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.

[0150] 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.

[0151] With recent advances in crop genomics, the ability to use CRISPR-Cas systems to perform efficient and cost effective gene editing and manipulation will allow the rapid selection and comparison of single and multiplexed genetic manipulations to transform such genomes for improved production and enhanced traits. In this regard reference is made to US patents and publications: U.S. Pat. No. 6,603,061-Agrobacterium-Mediated Plant Transformation Method; U.S. Pat. No. 7,868,149-Plant Genome Sequences and Uses Thereof and US 2009/0100536-Transgenic Plants with Enhanced Agronomic Traits, all the contents and disclosure of each of which are herein incorporated by reference in their entirety. In the practice of the invention, the contents and disclosure of Morrell et al "Crop genomics: advances and applications" Nat Rev Genet. 2011 Dec. 29; 13 (2): 85-96 are also herein incorporated by reference in their entirety. In an advantageous embodiment of the invention, the CRISPR/Cas9 system is used to engineer microalgae (Example 15). Accordingly, reference herein to animal cells may also apply, mutatis mutandis, to plant cells unless otherwise apparent.

[0152] 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 plant (including micro-algae), 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 (including micro-algae).

[0153] 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.

[0154] In one aspect, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first 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, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein 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; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence. 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.

[0155] 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.

[0156] 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.

[0157] 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.

[0158] 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.

[0159] 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.

[0160] Examples of disease-associated genes and polynucleotides are 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.

[0161] 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.

[0162] 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 on Dec. 12, 2012 and 61/748,427 filed on Feb. 2, 2013. Such genes, proteins and pathways may be the target polynucleotide of a CRISPR complex.

TABLE A

DISEASE/DISORDERS	GENE(S)
Neoplasia	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; 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	Aber; 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 (Friedreich's Ataxia); ATX3 (Machado-Joseph's Dx); ATXN1 and ATXN2 (spinocerebellar ataxias); DMPK (myotonic dystrophy); Atrophin-1 and Atn1 (DRPLA Dx); CBP (Crb-BP-global instability); VLDLR (Alzheimer's); Atxn7; Atxn10

TABLE A-continued

DISEASE/DISORDERS	GENE(S)
Fragile X Syndrome	FMR2; FXR1; FXR2; mGLUR5
Secretase Related Disorders	APH-1 (alpha and beta); Presenilin (Psen1); nicastrin (Nestn); 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; SORLI1; 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, TAP, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5), Bleeding disorders (TBXA2R, P2RX1, P2X1); Factor H and factor H-like 1 (HF1, CFH, HUS); Factor V and factor VIII (MCFD2); Factor VII deficiency (F7); Factor X deficiency (F10); Factor XI deficiency (F11); Factor XII deficiency (F12, HAF); Factor XIIIa deficiency (F13A1, F13A); Factor XIIIb deficiency (F13B); Fanconi anemia (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANC, 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); Leukocyte 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, CBF2A, 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). 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, SCY45, 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, IPFEX, 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).
Inflammation and immune related diseases and disorders	

TABLE B-continued

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, SC01), Hepatic lipase deficiency (LIPC), Hepatoblastoma, cancer and carcinomas (CTNNB1, PDGFRL, PDGRL, PRLLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; Medullary cystic kidney disease (UMOD, HNF1, FJHN, MCKD2, ADMCKD2); Phenylketonuria (PAH, PKU1, QDPR, DHPR, PTS); Polycystic kidney and hepatic disease (FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PLCD, 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, LGMD2I, 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, PLA2U, URK, ACE, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); Autism (MeCP2, BZRAP1, MDGA2, Sema5A, Neurexin 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); Fragile X Syndrome (FMR2, FXR1, FXR2, mGLUR5); Huntington's disease and disease like disorders (HD, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); Parkinson disease (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK5, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); Rett syndrome (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Synuclein, DJ-1); Schizophrenia (Neuregulin1 (Nrg1), Erb4 (receptor for Neuregulin), Complexin1 (Cpx1), 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, (Nctn), PEN-2, Nos1, Parp1, Nat1, Nat2); Trinucleotide Repeat Disorders (HTT (Huntington's Dx), SBMA/SMAX1/AR (Kennedy's Dx), FXN/X25 (Friedreich's Ataxia), ATX3 (Machado- Joseph's Dx), ATXN1 and ATXN2 (spinocerebellar ataxias), DMPK (myotonic dystrophy), Atrophin-1 and Atn1 (DRPLA Dx), CBP (Crb-BP-global instability), VLDR (Alzheimer's), Atxn7, Atxn10).
Occular diseases and disorders	Age-related macular degeneration (Aber, 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, CRYB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); Corneal clouding and dystrophy (APOA1, TGFB1, 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 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; TSC2D3; MAPK10; NRIP1; KRAS; MAPK13; RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1; PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3; MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP; CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1; ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1; STAT1; IL6; HSP90AA1
Glucocorticoid Receptor 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; GLII; 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

TABLE C-continued

CELLULAR FUNCTION	GENES
Huntington's Disease Signaling	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2; MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SPI1; CAPN2; PIK3CA; HDAC5; CREB1; PRKC1; 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; RAPIA; 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; CTND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK; CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
Integrin Signaling	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAPIA; 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; PDGFR; 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; PMAIPI1; 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; SPI1; 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

TABLE C-continued

CELLULAR FUNCTION	GENES
Xenobiotic Metabolism Signaling	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1; NCOR2; PIK3CA; ARNT; PRKC1; NFKB2; CAMK2A; PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1; ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD; GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL; NEF2L2; 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; TGFBRI; 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
Neuregulin Signaling	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1; MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKC1; 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; TGFBRI; CCND1; GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B; AKT3; SOX2
Insulin Receptor Signaling	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1; PTPN11; AKT2; CBL; PIK3CA; PRKC1; 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; PRKC1; 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; PRKC1; PTK2; FOS; PIK3CB; PIK3C3; MAPK8; IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A; YWHAZ; PNX; RAF1; CASP9; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3; FOXO1; SRF; CTGF; RPS6KB1

TABLE C-continued

CELLULAR FUNCTION	GENES
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; TGFB1; 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; RAPIA; RGS16; MAPK1; GNAS; AKT2; IKBKB; PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB; PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1; IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK; PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3; PRKCA
Inositol Phosphate Metabolism	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6; MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3; MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1; MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK; 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
PDGF 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
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; GNBL1; BCL2L1; MAPK3; ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3; STAT1

TABLE C-continued

CELLULAR FUNCTION	GENES
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; NTSE; TTK; CSNK1A1; BRAF; SGK
Chemokine Signaling	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ; CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13; RHOA; CCR3; SRC; PPP1CC; MAPK14; NOXI1; 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; PPP2RIA; IGF1R; PRKD1; MAPK3; KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA; YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Estrogen Receptor Signaling	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2; SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1; HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP; MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Protein Ubiquitination Pathway	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4; CBL;UBE2I;BTRE;HSPA5;USP7;USP10;FBXW7; USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8; USP1; VHL; HSP90AA1; BIRC3
IL-10 Signaling	TRAF6; CCRL1; ELK1; IKBKB; SP1; FOS; NFkB2; MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF; IKBKKG; 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; TGFBRI1; 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; TGFBRI1; 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; HDAC7; 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

TABLE C-continued

CELLULAR FUNCTION	GENES
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
cAMP-mediated Signaling	RAPIA; 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	SLAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SLAH1 SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Lysine Degradation	ERCC5; ERCC4; XPA; XPC; ERCC1
Nucleotide Excision Repair Pathway	UCHL1; HK2; GCK; GPI; HK1
Starch and Sucrose Metabolism	NQO1; HK2; GCK; HK1
Aminosugars Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
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

TABLE C-continued

CELLULAR FUNCTION	GENES
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
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 Pain	TRPM5; TRPA1 TRPM7; TRPC5; TRPC6; TRPC1; Cnrl; cnr2; Grk2; Trpa1; Pomp; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca; Prkacb; Prkar1a; Prkar2a
Mitochondrial Function	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Developmental	BMP-4; Chordin (Chrd); Noggin (Nog); WNT (Wnt2; Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b; Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenin; Dkk-1; Frizzled related proteins; Otx-2; Gbx2; FGF-8; Reelin; Dab1; unc-86 (Pou4fl or Brn3a); Numb; Reln
Neurology	

[0163] 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 EI, 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.

[0164] 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).

[0165] 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.

[0166] 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.

[0167] 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.

[0168] 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.

[0169] Examples of addiction-related proteins may include ABAT for example.

[0170] 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.

[0171] Examples of cardiovascular diseases associated proteins may include IL1B (interleukin 1, beta), XDH (xanthine dehydrogenase), TP53 (tumor protein p53), PTGIS (prostaglandin 12 (prostacyclin) synthase), MB (myoglobin), IL4 (interleukin 4), ANGPT1 (angiopoietin 1), ABCG8 (ATP-binding cassette, sub-family G (WHITE), member 8), or CTSK (cathepsin K), for example.

[0172] 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 (UBEIC) encoded by the UBA3 gene, for example.

[0173] Examples of proteins associated Autism Spectrum Disorder may include the benzodiazepine 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.

[0174] 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.

[0175] Examples of proteins associated Schizophrenia may include NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXNI, GSK3A, BDNF, DISC1, GSK3B, and combinations thereof.

[0176] 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.

[0177] 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.

[0178] 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.

[0179] 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.

[0180] 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 ADRAID (Alpha-1D adrenergic receptor for Alpha-1D adrenoceptor), for example.

[0181] 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.

[0182] 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.

[0183] 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), TACRI (tachykinin receptor 1), or HTR2c (5-hydroxytryptamine (serotonin) receptor 2C), for example.

[0184] Examples of neurodevelopmental-associated sequences include A2BP1 [ataxin 2-binding protein 1], AADAT [aminoacidate 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.

[0185] Further examples of preferred conditions treatable with the present system include may be selected from: Aicardi-Goutières Syndrome; Alexander Disease; Allan-Herndon-Dudley Syndrome; POLG-Related Disorders; Alpha-Mannosidosis (Type II and III); Alström Syndrome; Angelman; Syndrome; Ataxia-Telangiectasia; Neuronal Ceroid-Lipofuscinoses; Beta-Thalassemia; Bilateral Optic Atrophy and (Infantile) Optic Atrophy Type 1; Retinoblastoma (bilateral); Canavan Disease; Cerebrooculofacioskeletal Syndrome 1 [COFS1]; Cerebrotendinous Xanthomatosis; Cornelia de Lange Syndrome; MAPT-Related Disorders; Genetic Prion Diseases; Dravet Syndrome; Early-Onset Familial Alzheimer Disease; Friedreich Ataxia [FRDA]; Fryns Syndrome; Fucosidosis; Fukuyama Congenital Muscular Dystrophy; Galactosialidosis; Gaucher Disease; Organic Acidemias; Hemophagocytic Lymphohistiocytosis; Hutchinson-Gilford Progeria Syndrome; Mucolipidosis II; Infantile Free Sialic Acid Storage Disease; PLA2G6-Associated Neurodegeneration; Jervell and Lange-Nielsen Syndrome; Junctional Epidermolysis Bullosa; Huntington Disease; Krabbe Disease (Infantile); Mitochondrial DNA-Associated Leigh Syndrome and NARP; Lesch-Nyhan Syndrome; LIS1-Associated Lissencephaly; Lowe Syndrome; Maple Syrup Urine Disease; MECP2 Duplication Syndrome; ATP7A-Related Copper Transport Disorders; LAMA2-Related Muscular Dystrophy; Arylsulfatase A Deficiency; Mucopolysaccharidoses 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; COLIA1/2-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 Chondroplasia 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.

[0186] 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.

EXAMPLES

[0187] 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

[0188] 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) interspersed 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. 2A). 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. 2A). 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

[0189] Human embryonic kidney (HEK) cell line HEK 293 FT (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% CO₂ incubation. Mouse neuro2A (N2A) cell line (ATCC) was maintained with DMEM supplemented with 5% fetal bovine serum (HyClone), 2 mM GlutaMAX (Life Technologies), 100 U/mL penicillin, and 100 µg/mL streptomycin at 37°C. with 5% CO₂.

[0190] HEK 293 FT 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 Lipo-

fectamine 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

[0191] HEK 293 FT 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.

[0192] 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. 8 provides a schematic illustration of this Surveyor assay.

Restriction Fragment Length Polymorphism Assay for Detection of Homologous Recombination

[0193] HEK 293 FT 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

[0194] 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).

Bacterial Plasmid Transformation Interference Assay

[0195] Elements of the *S. pyogenes* CRISPR locus 1 sufficient for CRISPR activity were reconstituted in *E. coli* using pCRISPR plasmid (schematically illustrated in FIG. 10A). pCRISPR contained tracrRNA, SpCas9, and a leader sequence driving the crRNA array. Spacers (also referred to

as "guide sequences") were inserted into the crRNA array between BsaI sites using annealed oligonucleotides, as illustrated. Challenge plasmids used in the interference assay were constructed by inserting the protospacer (also referred to as a "target sequence") sequence along with an adjacent CRISPR motif sequence (PAM) into pUC19 (see FIG. 10B). The challenge plasmid contained ampicillin resistance. FIG. 10C provides a schematic representation of the interference assay. Chemically competent *E. coli* strains already carrying pCRISPR and the appropriate spacer were transformed with the challenge plasmid containing the corresponding protospacer-PAM sequence. pUC19 was used to assess the transformation efficiency of each pCRISPR-carrying competent strain. CRISPR activity resulted in cleavage of the pPSP plasmid carrying the protospacer, precluding ampicillin resistance otherwise conferred by pUC19 lacking the protospacer. FIG. 10D illustrates competence of each pCRISPR-carrying *E. coli* strain used in assays illustrated in FIG. 4C.

RNA Purification

[0196] HEK 293 FT 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

[0197] RNAs were mixed with equal volumes of 2× loading buffer (Ambion), heated to 95° C. for 5 min, chilled on ice for 1 min, and then loaded onto 8% denaturing polyacrylamide gels (SequaGel, National Diagnostics) after pre-running the gel for at least 30 minutes. The samples were electrophoresed for 1.5 hours at 40 W limit. Afterwards, the RNA was transferred to Hybond N+ membrane (GE Healthcare) at 300 mA in a semi-dry transfer apparatus (Bio-rad) at room temperature for 1.5 hours. The RNA was crosslinked to the membrane using autocrosslink button on Stratagene UV Crosslinker the Stratalink (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 [γ -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

[0198] 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. 9). 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 sequences (also referred to herein as "CRISPR motif sequences") were created by ligating hybridized oligos carrying compatible overhangs (Integrated DNA Technology) into BamHI-digested pUC19. Cloning for all constructs was performed in *E. coli* strain JM109 (Zymo Research).

[0199] pCRISPR-carrying cells were made competent using the Z-Competent *E. coli* Transformation Kit and Buffer Set (Zymo Research, T3001) according to manufacturer's instructions. In the transformation assay, 50 μ L aliquots of competent cells carrying pCRISPR were thawed on ice and transformed with 1 ng of spacer plasmid or pUC19 on ice for 30 minutes, followed by 45 second heat shock at 42° C. and 2 minutes on ice. Subsequently, 250 μ L SOC (Invitrogen) was added followed by shaking incubation at 37° C. for 1hr, and 100 μ L of the post-SOC outgrowth was plated onto double selection plates (12.5 μ g/ml chloramphenicol, 100 μ g/ml ampicillin). To obtain cfu/ng of DNA, total colony numbers were multiplied by 3.

[0200] To improve expression of CRISPR components in mammalian cells, two genes from the SF370 locus 1 of *Streptococcus pyogenes* (*S. pyogenes*) were codon-optimized, Cas9 (SpCas9) and RNase III (SpRNase III). To facilitate nuclear localization, a nuclear localization signal (NLS) was included at the amino (N)- or carboxyl (C)-termini of both SpCas9 and SpRNase III (FIG. 2B). To facilitate visualization of protein expression, a fluorescent protein marker was also included at the N- or C-termini of both proteins (FIG. 2B). A version of SpCas9 with an NLS attached to both N- and C-termini (2xNLS-SpCas9) was also generated. Constructs containing NLS-fused SpCas9 and SpRNase III were transfected into 293 FT human embryonic kidney (HEK) cells, and the relative positioning of the NLS to SpCas9 and SpRNase III was found to affect their nuclear localization efficiency. Whereas the C-terminal NLS was sufficient to target SpRNase III to the nucleus, attachment of a single copy of these particular NLS's to either the N- or C-terminus of SpCas9 was unable to achieve adequate nuclear localization in this system. In this example, the C-terminal NLS was that of nucleoplasmin (KRPAATKKAGQAKKK (SEQ ID NO: 2)), and the C-terminal NLS was that of the SV40 large T-antigen (PKKKRKV (SEQ ID NO: 1)). Of the versions of SpCas9 tested, only 2xNLS-SpCas9 exhibited nuclear localization (FIG. 2B).

[0201] The tracrRNA from the CRISPR locus of *S. pyogenes* SF370 has two transcriptional start sites, giving rise to two transcripts of 89-nucleotides (nt) and 171 nt that are subsequently processed into identical 75 nt mature tracrRNAs. The shorter 89 nt tracrRNA was selected for expression in mammalian cells (expression constructs illustrated in FIG. 7A, with functionality as determined by results of the Surveyor assay shown in FIG. 7B). 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. 7C shows results of a Northern blot analysis of total RNA extracted from 293 FT 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 293 FT 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.

[0202] To promote precise transcriptional initiation, the RNA polymerase III-based U6 promoter was selected to drive the expression of tracrRNA (FIG. 2C). 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. 2C). 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. 2C), a key gene in the development of the cerebral cortex.

[0203] 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 293 FT 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. 8) (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. 2D). 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. 2D). 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. 4-7, 12, and 13). 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. 7B).

[0204] FIG. 14 provides an additional Northern blot analysis of crRNA processing in mammalian cells. FIG. 14A 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. 14A. The line indicates the region whose reverse-complement sequence was used to generate Northern blot probes for EMX1 (1) crRNA detection. FIG. 14B shows a Northern blot analysis of total RNA extracted from 293 FT cells transfected with U6 expression constructs carrying DR-EMX1 (1)-DR. Left and right panels are from 293 FT cells transfected without or with SpRNase III respectively. DR-EMX1 (1)-DR was pro-

cessed 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 293 FT 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.

[0205] FIG. 2 illustrates the bacterial CRISPR system described in this example. FIG. 2A 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. 2B 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. 2C 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. 2D 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. 2E 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. 2F 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.

[0206] To further simplify the three-component system, a chimeric crRNA-tracrRNA hybrid design was adapted, where a mature crRNA (comprising a guide sequence) is fused to a partial tracrRNA via a stem-loop to mimic the natural crRNA: tracrRNA duplex (FIG. 3A). To increase co-delivery efficiency, a bicistronic expression vector was created to drive co-expression of a chimeric RNA and SpCas9 in transfected cells (FIGS. 3A and 8). 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. 13B top and bottom). FIG. 9 provides schematic illustrations of bicistronic expression vectors for pre-crRNA array (FIG. 9A) or chimeric crRNA (represented by the short line downstream of the guide sequence insertion site and upstream of the EF1a promoter in FIG. 9B) 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. 9B also shows a partial DR sequence (GTTTTAGAGCTA (SEQ ID NO: 27)) and a partial tracrRNA sequence (TAGCAAGT-TAAAATAAGGCTAGTCGTTTT (SEQ ID NO: 28)). Guide sequences can be inserted between BbsI sites using annealed oligonucleotides. Sequence design for the oligonucleotides are shown below the schematic illustrations in

FIG. 9, 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. 4).

[0207] 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. 15 illustrates the selection of some additional targeted protospacers in human PVALB (FIG. 15A) and mouse Th (FIG. 15B) loci. Schematics of the gene loci and the location of three protospacers within the last exon of each are provided. The underlined sequences include 30 bp of protospacer sequence and 3 bp at the 3' end corresponding to the PAM sequences. Protospacers on the sense and anti-sense strands are indicated above and below the DNA sequences, respectively. A modification rate of 6.3% and 0.75% was achieved for the human PVALB and mouse Th loci respectively, demonstrating the broad applicability of the CRISPR system in modifying different loci across multiple organisms (FIGS. 3B and 6). 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 (FIG. 6).

[0208] FIG. 13 provides a further illustration that SpCas9 can be reprogrammed to target multiple genomic loci in mammalian cells. FIG. 13A provides a schematic of the human EMX1 locus showing the location of five protospacers, indicated by the underlined sequences. FIG. 13B 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. 13C. Each protospacer is targeted using either processed pre-crRNA/trcrRNA complex (crRNA) or chimeric RNA (chiRNA).

[0209] 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 our genome targeting experiment (FIG. 3B) (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.

[0210] FIG. 3 illustrates example expression vectors. FIG. 3A provides a schematic of a bi-cistronic vector for driving the expression of a synthetic crRNA-tracrRNA chimera (chimeric RNA) as well as SpCas9. The chimeric guide RNA contains a 20-bp guide sequence corresponding to the protospacer in the genomic target site. FIG. 3B provides a schematic showing guide sequences targeting the human EMX1, PVALB, and mouse Th loci, as well as their predicted secondary structures. The modification efficiency at each target site is indicated below the RNA secondary structure drawing (EMX1, n=216 amplicon sequencing reads; PVALB, n=224 reads; Th, n=265 reads). The folding algorithm produced an output with each base colored according to its probability of assuming the predicted secondary structure, as indicated by a rainbow scale that is reproduced in FIG. 3B in gray scale. Further vector designs for SpCas9 are shown in FIG. 44, 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. 44b includes a tracrRNA coding sequence linked to an HI promoter.

[0211] To test whether spacers containing secondary structures are able to function in prokaryotic cells where CRISPRs naturally operate, transformation interference of protospacer-bearing plasmids were tested in an *E. coli* strain heterologously expressing the *S. pyogenes* SF370 CRISPR locus 1 (FIG. 10). The CRISPR locus was cloned into a low-copy *E. coli* expression vector and the crRNA array was replaced with a single spacer flanked by a pair of DRs (pCRISPR). *E. coli* strains harboring different pCRISPR plasmids were transformed with challenge plasmids containing the corresponding protospacer and PAM sequences (FIG. 10C). In the bacterial assay, all spacers facilitated efficient CRISPR interference (FIG. 4C). These results suggest that there may be additional factors affecting the efficiency of CRISPR activity in mammalian cells.

[0212] 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. 4A). FIG. 4B 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. 4B). 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. 4C provides a schematic showing the design of TALENs targeting EMX1, and FIG. 4D shows a Surveyor gel comparing the efficiency of TALEN and Cas9 (n=3).

[0213] 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. 5A) (see e.g. Sapranauskas et al., 2011, Nucleic Acids 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. As illustrated in FIG. 5B, 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 293 FT 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. 5C 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. 5D), with SpCas9 and SpCas9n mediating similar levels of HR efficiencies. Applicants further verified HR using Sanger sequencing of genomic amplicons (FIG. 5E). 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.

[0214] Expression constructs mimicking the natural architecture of CRISPR loci with arrayed spacers (FIG. 2A) 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. 4F, 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. 4G) was detected. This demonstrates that the CRISPR system can mediate multiplexed editing within a single genome.

Example 2: CRISPR System Modifications and Alternatives

[0215] 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 Mg²⁺ 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. 11,

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. 12 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. 12A provides a Schematic illustration of CRISPR 1 from *S. thermophilus* LMD-9. FIG. 12B illustrates the design of an expression system for the *S. thermophilus* CRISPR system. Human codon-optimized hStCas9 is expressed using a constitutive EF1 α 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. 12C provides a schematic showing guide sequences targeting the human EMX1 locus as well as their predicted secondary structures. The modification efficiency at each target site is indicated below the RNA secondary structures. The algorithm generating the structures colors each base according to its probability of assuming the predicted secondary structure, which is indicated by a rainbow scale reproduced in FIG. 12C in gray scale. Statistical analysis of cleavage activity across biological replica at these two protospacer sites is also provided in FIG. 6. FIG. 16 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

[0216] 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'-Nx-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'-Nx-NNAGAAW-3' (SEQ ID NO: 29) 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'-Nx-NGGNG-3' both on the input sequence and on the reverse-complement of the input. The value “x” in Nx may be fixed by the program or specified by the user, such as 20.

[0217] 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).

[0218] Further details of methods and algorithms to optimize sequence selection can be found in U.S. application Ser. No. 61/836,080 (attorney docket 44790.11.2022); incorporated herein by reference.

Example 4: Evaluation of Multiple Chimeric crRNA-tracrRNA Hybrids

[0219] 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. 18a 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 GUUUUA-GAGCUA (SEQ ID NO: 30) 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. 18b and 18c, 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. 19a and 19b, corresponding to FIGS. 18b and 18c, 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

proto-spacer ID	genomic target	protospacer sequence (5' to 3')	PAM	strand
1	EMX1	GGACATCGATGTCACCTCCA ATGACTAGGG (SEQ ID NO: 31)	TGG	+

TABLE D-continued

proto-spacer ID	genomic target	protospacer sequence (5' to 3')	PAM	strand
2	EMX1	CATGGAGGTGACATCGATG TCCTCCCCAT (SEQ ID NO: 32)	TGG	-
3	EMX1	GGAAGGGCCTGAGTCCGAGC AGAAAGAGAA (SEQ ID NO: 33)	GGG	+
4	PVALB	GGTGGCGAGAGGGGCCGAGA TTGGGTGTTG (SEQ ID NO: 34)	AGG	+
5	PVALB	ATGCAGGAGGGTGGCGAGAG GGCCGAGAT (SEQ ID NO: 35)	TGG	+

Cell Culture and Transfection

[0220] Human embryonic kidney (HEK) cell line 293 FT (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% CO₂ incubation. 293 FT cells were seeded onto 24-well plates (Corning) 24 hours prior to transfection at a density of 150,000 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 500 ng plasmid was used.

SURVEYOR Assay for Genome Modification

[0221] 293 FT cells were transfected with plasmid DNA as described above. Cells were incubated at 37° C. for 72 hours post-transfection prior to genomic DNA extraction. Genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Epicentre) following the manufacturer's protocol. Briefly, pelleted cells were resuspended in QuickExtract solution and incubated at 65° C. for 15 minutes and 98° C. for 10 minutes. The genomic region flanking the CRISPR target site for each gene was PCR amplified (primers listed in Table E), and products were purified using QiaQuick Spin Column (Qiagen) following the manufacturer's protocol. 400 ng total of the purified PCR products were mixed with 2 µl 10×Taq DNA 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.

TABLE E

primer name	genomic target	primer sequence (5' to 3')
Sp-EMX1-F	EMX1	AAAACCACCCCTCTCTCTGGC (SEQ ID NO: 36)
Sp-EMX1-R	EMX1	GGAGATTGGAGACACGGAGAG (SEQ ID NO: 37)
Sp-PVALB-F	PVALB	CTGGAAAGCCAATGCCCTGAC (SEQ ID NO: 38)
Sp-PVALB-R	PVALB	GGCAGCAAATCCTTGTCCCT (SEQ ID NO: 39)

Computational Identification of Unique CRISPR Target Sites

[0222] To identify unique target sites for the *S. pyogenes* SF370 Cas9 (SpCas9) enzyme in the human, mouse, rat, zebrafish, fruit fly, and *C. elegans* genome, we developed a software package to scan both strands of a DNA sequence and identify all possible SpCas9 target sites. For this example, each SpCas9 target site was operationally defined as a 20 bp sequence followed by an NGG protospacer adjacent motif (PAM) sequence, and we identified all sequences satisfying this 5'-N20-NGG-3' definition on all chromosomes. To prevent non-specific genome editing, after identifying all potential sites, all target sites were filtered based on the number of times they appear in the relevant reference genome. To take advantage of sequence specificity of Cas9 activity conferred by a 'seed' sequence, which can be, for example, approximately 11-12 bp sequence 5' from the PAM sequence, 5'-NNNNNNNNNN-NGG-3' sequences were selected to be unique in the relevant genome. All genomic sequences were downloaded from the UCSC Genome Browser (Human genome hg19, Mouse genome mm9, Rat genome rn5, Zebrafish genome danRer7, *D. melanogaster* genome dm4 and *C. elegans* genome ce10). The full search results are available to browse using UCSC Genome Browser information. An example visualization of some target sites in the human genome is provided in FIG. 21.

[0223] Initially, three sites within the EMX1 locus in human HEK 293 FT 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. 18b and 19a). 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. 18c and 19b).

[0224] For all five targets in the EMX1 and PVALB loci, a consistent increase in genome modification efficiency with increasing tracr sequence length was observed. Without wishing to be bound by any theory, the secondary structure formed by the 3' end of the tracrRNA may play a role in enhancing the rate of CRISPR complex formation. An illustration of predicted secondary structures for each of the chimeric RNAs used in this example is provided in FIG. 21. The secondary structure was predicted using RNAfold using minimum free energy and partition function algorithm. Pseudocolor for each based (reproduced in grayscale) indicates the probability of pairing. Because chiRNAs with longer tracr sequences were able to cleave targets that were not cleaved by native CRISPR crRNA/tracrRNA hybrids, it is possible that chimeric RNA may be loaded onto Cas9 more efficiently than its native hybrid counterpart. To facilitate the application of Cas9 for site-specific genome editing in eukaryotic cells and organisms, all predicted unique target sites for the *S. pyogenes* Cas9 were computationally identified in the human, mouse, rat, zebra fish, *C. elegans*, and *D. melanogaster* genomes. Chimeric RNAs can be designed for Cas9 enzymes from other microbes to expand the target space of CRISPR RNA-programmable nucleases.

[0225] FIG. 22 illustrates an exemplary bicistronic expression vector for expression of chimeric RNA including up to the +85 nucleotide of wild-type tracr RNA sequence, and SpCas9 with nuclear localization sequences. SpCas9 is expressed from a CBh promoter and terminated with the bGH poly A signal (bGH pA). The expanded sequence illustrated immediately below the schematic corresponds to the region surrounding the guide sequence insertion site, and includes, from 5' to 3', 3'-portion of the U6 promoter (first shaded region), BbsI cleavage sites (arrows), partial direct repeat (tracr mate sequence GTTTTAGAGCTA (SEQ ID NO: 27), underlined), loop sequence GAAA, and +85 tracr sequence (underlined sequence following loop sequence). An exemplary guide sequence insert is illustrated below the guide sequence insertion site, with nucleotides of the guide sequence for a selected target represented by an "N".

[0226] Sequences described in the above examples are as follows (polynucleotide sequences are 5' to 3'):

U6-short tracrRNA (*Streptococcus pyogenes* SF370) :

(SEQ ID NO: 40)

```
GAGGGCCTATTCATGATTCTCATATTGCATATACGATAACAGGC  
TGTTAGAGAGATAATTGAATTAATTGACTGTAAACACAAAGATATTAG  
TACAAAATACGTGACGTAGAAAGTAATAATTCTGGTAGTTGCAGTT  
TTAAAATTATGTTAAAATGGACTATCATATGCTTACCGTAACTGAAA  
GTATTCGATTCTGGTTATATCTGTGGAAAGGACGAAACACCG  
GAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCGTTATCA  
ACTTGAAAAAGTGGCACCGAGTCGGTGTTTTTTT
```

(bold = tracr RNA sequence; underline =

terminator sequence)

-continued

U6-long tracrRNA (*Streptococcus pyogenes* SF370) :

(SEQ ID NO: 41)

```
GAGGGCCTATTCATGATTCTCATATTGCATATACGATAACAGGC  
TGTTAGAGAGATAATTGAATTAATTGACTGTAAACACAAAGATATTAG  
TACAAAATACGTGACGTAGAAAGTAATAATTCTGGTAGTTGCAGTT  
TTAAAATTATGTTAAAATGGACTATCATATGCTTACCGTAACTGAAA  
GTATTCGATTCTGGTTATATCTGTGGAAAGGACGAAACACCG  
ACAGCATAGCAAGTTAAAATAAGGCTAGTCGTTATCAACTGAAAAGT  
GGCACCGAGTCGGTGTTTTTTT
```

U6-DR-BbsI backbone-DR (*Streptococcus pyogenes* SF370) :

(SEQ ID NO: 42)

```
GAGGGCCTATTCATGATTCTCATATTGCATATACGATAACAGGC  
TGTTAGAGAGATAATTGAATTAATTGACTGTAAACACAAAGATATTAG  
TACAAAATACGTGACGTAGAAAGTAATAATTCTGGTAGTTGCAGTT  
TTAAAATTATGTTAAAATGGACTATCATATGCTTACCGTAACTGAAA  
GTATTCGATTCTGGTTATATCTGTGGAAAGGACGAAACACCG  
GGTTTAGAGCTATGCTGTTGAATGGCCAAACGGGCTTCGAGAA  
GACGTTTAGAGCTATGCTGTTGAATGGCCAAAC
```

U6-chimeric RNA-BbsI backbone (*Streptococcus pyogenes* SF370) :

(SEQ ID NO: 43)

```
GAGGGCCTATTCATGATTCTCATATTGCATATACGATAACAGGC  
TGTTAGAGAGATAATTGAATTAATTGACTGTAAACACAAAGATATTAG  
TACAAAATACGTGACGTAGAAAGTAATAATTCTGGTAGTTGCAGTT  
TTAAAATTATGTTAAAATGGACTATCATATGCTTACCGTAACTGAAA  
GTATTCGATTCTGGTTATATCTGTGGAAAGGACGAAACACCG  
GGCTCGAGAACCTGTTAGAGCTAGAAATAGCAAGTTAAAATAAG  
GCTAGTCG
```

NLS-SpCas9-EGFP :

(SEQ ID NO: 44)

```
MDYKDHDGDYKDHDIDYKDDDKMAPKKRKVGIHGVPAADKKYSIGLDI  
GTNSVGWAVITDEYKVPSKKFKVLGNTRHSIKKNLIGALLFDSGETAEA  
TRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDK  
KHERHPFPGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHM  
IKPRGHFLIEGDLNPNDVVKLFIQLVQTYNQLFEENPINASGVDAKAI  
LSARLSKSRRLENLIAQLPGEKKNGLPGNLIALSLGLTPNFKNFDLAED  
AKLQLSKDTYDDLDNLQAQIGDQYADLFLAAKNLSDAIISSDILRVNTE  
ITKAPLSASMIKRYDEHHQDLTLKALVRQQLPEKYKEIFFDQSCKNGYAG  
YIDGGASQEEFYKF1KPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIP  
HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR
```

- continued

FAWMTRKSEETITPWNFEEVVDKGASAQSFIERNFTNDKNLPNEKVLPH
 SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQQKAIVDLLFKTNRKTV
 KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEE
 NEDILEDIVLTTLFEDREMIERLKTYAHLFDDKVMKQLKRRRTGWR
 LSRKLINGIRDQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKDQKA
 QVSGQGDSLHEHIANLAGSPAICKGILQTVKVVDELVKVMGRHKPENIVI
 EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL
 YLYYLQNQGRDMYDQELDINRLSDYDVEDHIVPQSFKDDSIDNKVLTRSD
 KNRGKSDNVPSEEVVKMKNYWRQQLNAKLITQRKFNDL
 DKAGFIKRQLVETRQITKHVAQILDLSRMNTKYDENDKLIREVKVITLKS
 LVSDFRKDFQFYKVREINNYHHAHDAYLNAAVGTLAKKYPLESEFVYG
 DYKVDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKP
 LIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVTQGFSKESIL
 PRKNSDKLIARKKDWPCKYGGFDSPVAYSVLVVAKVEGKSKKLKSVK
 ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIILPKYSLFELENGRK
 RMLASAGELQKGNELALPSKYVNFLYASHYEKLKGSPEDNEQKQLFVEQ
 HKHYLDEIIEQISEFSKRVILADANLDKVLAYSINKHRDKPIREQAENIIH
 LFLLTNLGAPAAFKYFDTIDRKRYTSTKEVLDATLIGHQSITGLYETRID
 LSQLGGDAAAVSKGEELFTGVVPIVLDGDVNGHKFSVSGEGEQDATYG
 KLTLKFICTTGKLPVPWPTLVTLTYGVQCFSRYPDHMKQHDFFKSAMPE
 GYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHK
 LEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGD
 GPVLLPDNHYLSTQSAKSMDPNEKRDHMVLLEFVTAAGITLGMDELYK
 SpCas9-EGFP-NLS :
 (SEQ ID NO: 45)
 MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGA
 LLFDGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFTHR
 LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
 LRLIYLALAHMIKFRGHFLIEGDLNPNSDVDFLTIQLVQTYNQLFEE
 INASGVDAKAILSRSRRLNLIQLPGEKKNGLFGNLIALSGLTP
 NFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVQLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRQRQEDFYPFLKDNRKEIKEILTFRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERNFTND
 NLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQQKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKII
 IKDKDFLDNEENEDILEDIVLTTLFEDREMIERLKTYAHLFDDKVMKQ
 LKRRRTGWGRSLRKLINGIRDQSGKTILDFLKSDGFANRNFMQLIHD
 DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKVVDELVK

- continued

VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEH
 PVENTQLQNEKLYLYLQNQGRDMYDQELDINRLSDYDVEDHIVPQSFKLD
 DSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQQLNAKLITQRKFDN
 LTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDLSRMNTKYDENDKL
 IREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAAVGTLAK
 KYPKLESEFVYGDYKVDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE
 ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKT
 VQTGGFSKESILPKRNSDKLIARKKDWPCKYGGFDSPVAYSVLVVAK
 EKGKSKKLKSVKELLGITTERSSFEKNPIDFLEAKGYKEVKKDLIILPK
 KYSLFELENGRKRLMASAGELQKGNELALPSKYVNFLYASHYEKLKGSP
 EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHD
 KPIREQAENIIHLFTLTNLGAPAAFKYFDTIDRKRYTSTKEVLDATL
 QSITGLYETRIDLSQLGGDAAA VSKGEELFTGVVPIVLDGDVNGHKFS
 VSGEGEGDATYGKLTLCFTTGKLPVPWPTLVTTLYGVQCFSRYPDHM
 KQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKPEGDTLVNRIELKG
 DFKEDGNI LGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQL
 ADHYQQNTPIGDGPVLLPDNHYLSTQSAKSMDPNEKRDHMVLLEFVTAAG
 ITLGMDELYKKRPAATKKAGQAKKK
 NLS - SpCas9-EGFP - NLS :
 (SEQ ID NO: 46)
 MDYKDHDGDKDHDIDYKDDDKMAPKKRKVGIGHVPAADKKYSIGLDI
 GTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDGETAEA
 TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFTHRLEESFLVEEDK
 KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHM
 IKFRGHFLIEGDLNPNSDVDFLTIQLVQTYNQLFEEINPINA
 GSVDAKAISARLKSRRLENLIAQLPGEKKNGLFGNLIALSGLTPNPKSNFD
 LAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVQLNREDLLR
 KQRTFDNGSIPHQIHLGELHAILRQRQEDFYPFLKDNRKEIKEILTFRIPY
 YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERNFTND
 NLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQQKAIVD
 LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKII
 IKDKDFLDNEENEDILEDIVLTTLFEDREMIERLKTYAHLFDDKVMKQ
 LKRRRTGWGRSLRKLINGIRDQSGKTILDFLKSDGFANRNFMQLIHD
 DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKVVDELVK

- continued
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVTALIKKYPKLESEFVYG
DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNNFKTEITLANGEIRKRP
LIETNGETGEIVWDKGKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL
PKRNSDKLIARKKDWPDKYGGFDSPVTAVSVLVAKEKGSKKLKSVK
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDIIKLPKYSLFELNGRK
RMLASAGELQGNELALPSKYVNFLYLAHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLAYSNKHRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIITGLYETRID
LSQLGGDAAAVSKGEELFTGVVPIVLDGDVNGHKFSVSGEGEGDATYG
KLTLCFCTTGKLPVPWPTLVTLTYGVQCFSRYPDHMKQHDFFKSAMPE
GYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHK
LEYNNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGD
GPVLLPDNHYLSTQSALKDPNEKRDHMVLLFVTAAGITLGMDELYKKR
PAATKKAGQAKKK

NLS - SpCas9 - NLS :

(SEQ ID NO: 47)

MDYKDHDG DYKDHDIDYKDDDKMAPKKRKVGIGHGVPAA DKKYSIGLDI
GTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDGETAE
TRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKPYTIYHLRKKLVDSTDKADLRLIYLALAHM
IKFRGHFLIEGDLNPNDNSDVKLFIFIQLVQTYNQLFEENPINASGVDAKAI
LSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSGLTPNPKSNFDLAED
AKLQLSKDTYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG
YIDGGASQEEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIP
HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR
FAWMTRKSEETITPWNFEVVVDKGASAQS FIERMTNF DKNLPNEKVL PKH
SLLYEYFTVYNELTKV KYVTEGMRKPAFLSGEOKKAIVD LFLK TNRKTV
KQLKEDYFKKIECFDSVEISGVEDRFNASL GTYH DLLK I KDKFLDNEE
NEDILEDIVLT LTFEDREMI EERLK TYAHL FDDKVMQ KLR RYTG WGR
LSRKLINGIRDQSGKTILD FLKSDGFANRNF MQLIH DDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELV KVMGRHK PENIVI
EMARENQTTQKGQKNSRERMKRIE EGI KELGSQ I LKEHPVENT QLQNEKL
YI LYQLQNGR DMV DQED INRL SLDYD VHIV PQSFLK DDSID NKV LTRSD
KNRGKSDN VPSEEV VKKMKN YWRQ LNAKLITQRKP DNLTKAER GGLSEL
DKAGFIK QL VETRQ ITKHVAQ I L DSRM NT K YDEND K LIREV KVITL KSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVTALIKKYPKLESEFVYG
DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNNFKTEITLANGEIRKRP
LIETNGETGEIVWDKGKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL
PKRNSDKLIARKKDWPDKYGGFDSPVTAVSVLVAKEKGSKKLKSVK

- continued
ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDIIKLPKYSLFELNGRK
RMLASAGELQGNELALPSKYVNFLYLAHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLAYSNKHRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIITGLYETRID
LSQLGGD KRPAATKKAGQAKKK

NLS - mCherry - SpRNase3 :

(SEQ ID NO: 48)

MFLFLSLSTSFLSSRTLVSKGEEDNM A I KEFMRFKVHMEGSVNGHEFEI
E GEGEGRPYEGTQTAKLKVTKGGLPFAWDILSPQFM YGSKAYV KHPADI
PDYLKLSFPEGFKWERVMNFEDGGVVT VTQDSSLQDG E FIYKV KLRG TNF
PSDGPVMQKKTMGWEASSER M PEGDALKGEIKQRLKLKDGGHYDAEVKT
TYAKAKPVQVLPAGYVNVIKLDITSHNEDYTIVEQYERAEGRHSTGGMDEL
YKGSKQLEELLSTSFDI QFNDL TLLETAFTH TSYANEHRLLN VSHNERLE
FLGDAV LQI I SEYLF AKYPKK TEGDM SKLRS MIV REESLAG FSRFC SF
AYIKLGKGEEKSGGRRDTI LGDLF EAFLG ALL DKG IDA VRRFL KQV MI
PQVEKG NFERV KDYK TCL QEF LQTK GDVA IDYQ VISEKG PAHAK QF EV SI
VVNGAVL SKGLG KSKK LAE QDAK NAL AQL SE

SpRNase3 - mCherry - NLS :

(SEQ ID NO: 49)

MKQLEELLSTSFDI QFNDL TLLETAFTH TSYANEHRLLN VSHNERLE F L G
DAVLQI I SEYLF AKYPKK TEGDM SKLRS MIV REESLAG FSRFC SF DAY I
KLKGKGEEKSGGRRDTI LGDLF EAFLG ALL DKG IDA VRRFL KQV MI PQV
EKG NFERV KDYK TCL QEF LQTK GDVA IDYQ VISEKG PAHAK QF EV SI VVN
GAVL SKGLG KSKK LAE QDAK NAL AQL SE VGS VSK GEED NM A I KEF MR
KV HME GS VNG HE FEI E GEGEGRPYEGTQTAKLKVTKGGLPFAWDILSPQ
F MYGSKAYV KHPADI PDYLKLSFPEGFKWERVMNFEDGGVVT VTQDSSLQ
DGE F IYKV KLRG TNF PSDGP VMQKKT MGWEASSER M PEGDALKGEIKQ
LKLKDGGHYDAEVK TTYKAKKPVQVLPAGYVNVIKLDITSHNEDYTIVEQY
ERAEGRHSTGGMDELYKKRPAATKKAGQAKKK

NLS - SpCas9n - NLS (the D10A nickase mutation is lowercase) :

(SEQ ID NO: 50)

MDYKDHDG DYKDHDIDYKDDDKMAPKKRKVGIGHGVPAA DKKYSIGLa I
GTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDGETAE
TRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKPYTIYHLRKKLVDSTDKADLRLIYLALAHM
IKFRGHFLIEGDLNPNDNSDVKLFIFIQLVQTYNQLFEENPINASGVDAKAI
LSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSGLTPNPKSNFDLAED
AKLQLSKDTYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG
YIDGGASQEEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIP
HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR
FAWMTRKSEETITPWNFEVVVDKGASAQS FIERMTNF DKNLPNEKVL PKH

-continued

SLLYEFYFTVYNELTKVKYVTEGMRKPAFLSGEQQKAIVDLLFKTNRKVT
KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKEFLDNEE
NEDILEDIVLTTLFEDREMIIEERLKTyahLFDDKVMQQLKRRRTGWGR
LSRKLINGIRDQSGKTILDFLKSDFANRNPMLIHDDSLTFKEDIQKA
QVSGQGDSDLHEHIANLAGSPAIIKGQILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL
YLYYLQNQGRDMYVDQELDINRLSDYDHDIVPQSFLKDDSDNQVLTQD
KNRGKSDNVPEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
LVSDFRKDQFYKVREINNNYHHAHDAYLNAAVGTALIKKPYLESEFVY
DYKVDVORKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP
LIETNGETGEIVWDKGRCDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL
PKRNSDKLIARKKDWDPKYGGFDSPTVAYSVLVAKVEKGSKKLKSVK
ELLGITIMERSSEFKNPIDPLEAKGYKEVKDKLIIKLPKYSLFELENGK
RMLASAGELQKGNELALPSKYVNFLYASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSiTGLYETRID
LSQLGGDKRPAATKKAGQAKKKK

hEMX1-HR Template-HindII-NheI:

(SEQ ID NO: 51)

GAATGCTGCCCTAGACCCGCTTCCCTCCCTGTCTGTCTGGCACAGGAG
AATGAGGTCTCACTGGTGGATTTCGGACTACCCCTGAGGAGCTGGCACCTG
AGGGACAAGGCCCCACCTGCCTAGCTCCAGCCTCTGATGAGGGGTGGG
AGAGAGCTACATGAGGTTGCTAAGAAAGCCTCCCTGAAGGAGACCACAC
AGTGTGTGAGGTTGGAGTCTCTAGCAGCGGGTCTGTGCCCTCAGGGATA
GTCTGGCTGTCCAGGCACTGCTCTGATATAAACACCCACCTCTAGTTAT
GAAACCATGCCATTCTGCCTCTGTATGGAAAAGAGCATGGGGTGGC
CCGTGGGGTGGTGTCCACTTAGGCCCTGTGGGAGATCATGGGAACCCAC
GCAGTGGGTATAGGCTCTCTCATTTACTACACATCCACTGTGAAG
AAGCGATTATGATCTCTCTAGAAACTCGTAGAGTCCCATGTCTGCCG
GCTTCAGAGCCTGCACTCTCCACCTGGCTGGCTTGCTGGGCTAG
AGGAGCTAGGATGCACAGCAGCTGTGACCCCTTGTGAGAGGAACAG
GAAAACCACCCCTCTCTGGCCCACTGTGCTCTTCCTGCCCTGCCAT
CCCCCTCTGTGAATGTTAGACCCATGGGAGCAGCTGGTCAAGAGGGACCC
CGGCCTGGGCCCTAACCTATGTAGCCTAGCTTCCCCTGCCCTGCCAT
CAGCTCAGCTGAGTGTGAGGGCCCACTGGCTGCTCTGGGGCTCTG
AGTTTCTCATCTGTGCCCTCCCTCCCTGGCCAGGTGAAGGTGTGGTTC
CAGAACCGGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCC
TGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGGCCA
TTGCCACGAAGCAGGCCATTGGGGAGGACATCGATGTCACCTCCAATGAC

-continued

-continued

VGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPOTFEK
VIEPILENPNKQINEKGKEVCPNPFKLKYKEEHGYIRKYSKKGNPPEIKS
LKYYDSKLGHNHIDITPKDSNNKVVLQSVPWRADVYFNKTGKYEILGLK
YADLQFEGTGTGTYKISQEYKNDIKKKEGVDSDFSEFKFTLYKNDLLLKVDT
ETKEQQLFRFLSRTMPQKHYVELKPYDKQKFEFGGEALIKVGLNVANSGQ
CKKGLGKSNISIYKVRTDVLGNQHIIKNEGDPKPLDFKRPAATKKAGQAK
KKK

U6-St_tracrRNA(7-97) :

(SEQ ID NO: 53)

GAGGGCCTATTCATGATTCCCTCATATTGCATATACGATAACAAGGC
TGTTAGAGAGATAATTGGAATTAAATTGACTGTAAACACAAAGATATTAG
TACAAAATACGTGACGTAGAAGTAATAATTCTGGTAGTTGCAGTT
TTAAAATTATGTTTAAATGGACTATCATATGCTTACCGTAACCTGAAA
GTATTCGATTTCTGGCTTATATCTTGAGGAAAGGACGAAACACCG
TTACTTAAATCTGCAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCA
ACACCCCTGTCATTTATGGCAGGGTGTTCGTTATTTAA

U6-DR-spacer-DR (*S. pyogenes* SF370)

(SEQ ID NO: 54)

gagggcctatcccattgttccatattgcatacgtacaaggc
tgtagagagataatttgaattttactgtaaacacaaagatattag
tacaaaatacgtgacgtagaaagaataatttctggtagttcagtt
ttaaaattatgtttaaatggactatcatatgcttaccgttaacttgaaa
gtatccgatttatggattatatcttgaaaaggacgaaacacccgg
tttagagctatgtgtttgaatggccaaacNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNtttagagctatgtgtttgaatggccaaaa
CTTTTTTT

(lower case underline = direct repeat;

N = guide sequence; bold = terminator)

Chimeric RNA containing +48 tracr RNA
(*S. pyogenes* SF370)

(SEQ ID NO: 55)

gagggcctatcccattgttccatattgcatacgtacaaggc
tgtagagagataatttgaattttactgtaaacacaaagatattag
tacaaaatacgtgacgtagaaagaataatttctggtagttcagtt
ttaaaattatgtttaaatggactatcatatgcttaccgttaacttgaaa
gtatccgatttatggattatatcttgaaaaggacgaaacacccN
NNNNNNNNNNNNNNNNtttagagctagaaatagcaagttaaaataagg

aggctagccgTTTTTT

(N = guide sequence; first underline =
tracr mate sequence; second underline =
tracr sequence; bold = terminator)

-continued

Chimeric RNA containing +54 tracr RNA
(*S. pyogenes* SF370)

(SEQ ID NO: 56)

gagggcctatcccattgttccatattgcatacgtacaaggc
tgtagagagataatttgaattttactgtaaacacaaagatattag
tacaaaatacgtgacgtagaaagaataatttgcgttaccgttaacttgaaa
tttcgatttatggattatatcttgaaaaggacgaaacacccNNNN
NNNNNNNNNNNNNNNNtttagagctagaaatagcaagttaaaataagg

ctagccgttatcaTTTTTT

(N = guide sequence; first underline =
tracr mate sequence; second underline =
tracr sequence; bold = terminator)

Chimeric RNA containing +67 tracr RNA
(*S. pyogenes* SF370)

(SEQ ID NO: 57)

gagggcctatcccattgttccatattgcatacgtacaaggc
tgtagagagataatttgaattttactgtaaacacaaagatattag
tacaaaatacgtgacgtagaaagaataatttgcgttaccgttaacttgaaa
tttcgatttatggattatatcttgaaaaggacgaaacacccNNNN
NNNNNNNNNNNNNNNNtttagagctagaaatagcaagttaaaataagg

ctagccgttatcaacttgaaaaagt**T**TTTTTT

(N = guide sequence; first underline =
tracr mate sequence; second underline =
tracr sequence; bold = terminator)

Chimeric RNA containing +85 tracr RNA
(*S. pyogenes* SF370)

(SEQ ID NO: 58)

gagggcctatcccattgttccatattgcatacgtacaaggc
tgtagagagataatttgaattttactgtaaacacaaagatattag
tacaaaatacgtgacgtagaaagaataatttgcgttaccgttaacttgaaa
tttcgatttatggattatatcttgaaaaggacgaaacacccNNNN
NNNNNNNNNNNNNNNNtttagagctagaaatagcaagttaaaataagg

ctagccgttatcaacttgaaaaagtggcaccgagtegg**T**TTTTTT

(N = guide sequence; first underline =
tracr mate sequence; second underline =
tracr sequence; bold = terminator)

CBh-NLS-SpCas9-NLS

(SEQ ID NO: 59)

CGTTACATAACTTACGGTAAATGGCCGCCTGGCTGACCGCCAAACGACC
CCCGCCCATGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATA
GGGACTTCCATTGACGTCAATGGTGGAGTATTACGGTAAACTGCCA

- continued

CTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCTATTGACG
 TCAATGACGGTAAATGGCCCGCTGGATTATGCCAGTACATGACCTA
 TGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC
 CATGGTCAGGTGAGGCCACGTTCTGCTTCACTCTCCCCATCTCCCC
 CCTCCCCACCCCAATTGTATTATTATTATTATTGTGCA
 GCGATGGGGGGGGGGGGGGGGGGGGGGCGCGGCCAGGGGGGGGGCG
 GGGCGAGGGCGGGCGGGCGAGGGAGGGTGCAGGGAGGGAGCGCAATC
 AGAGCGCGCGCTCCGAAAGTTCTTTATGGCGAGGGGGGGCGCG
 CGGCCCTATAAAAGCGAAGCGCGGGGGAGTCGCTCGACGCT
 GCCTCGCCCCGTGCCCGCTCCGCCGCGCTCGCGCCGCCCG
 CTCTGACTGACCCTGTTACTCCACAGGTGAGGGGGGGACGGCCCTC
 TCCTCCGGCTGTAATTAGCTGAGCAAGAGGTAAAGGTTAAGGGATGGT
 TGGTTGGTGGGTATTAATGTTAATTACCTGGAGCACCTGCCCTGAAATC
 ACTTTTTTCAGGTTGGaccgggaccATGACTATAAGGACACGAC
GGAGACTACAAGGATCATGATATTGATTACAAGACGATGACGATAAGAT
GGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCCG
ACAAGAAGTACAGCATGGCTGGACATCGCACCAACTCTGGCTGG
GCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCT
GGCAACACCGACCGGACAGCATCAAGAAGAACCTGATCGGAGCCCTG
TGTCGACAGGGGAAACAGCCGAGGCCACCCGGCTGAAGAGAACGCC
AGAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGAT
CTTCAGCAACGAGATGGCAAGGTGGACGACAGCTTCTCACAGACTGG
AAGAGTCCTCTGGTGAAGAGGATAAGAACACGAGCGGACCCCATC
TTGGCAACATGTGGACGAGGTGGCTTACACGAGAAGTACCCACCAT
CTACCACTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCACCTG
GGCTGATCTATGGCCCTGGCCACATGATCAAGTCTGGGGCCACTTC
CTGATCGAGGGGACCTGAACCCGACAACAGCGACGTGGACAAGCTGTT
CATCCAGCTGGTCAGACCTACAACCAGCTTGGAGGAAAACCCCATCA
ACGCCAGCGGCTGGACGCCAGGACATCCTGCTGCCAGACTGAGCAAG
AGCAGACGGCTGGAAAATCTGATGCCAGCTGCCAGGGCGAGAACAGA
TGGCCTTCTGGCAACCTGATTGCCCTGAGCTGGGCTGACCCCAACT
TCAAGAGCAACTCGACCTGGCGAGGATGCCAAACTGAGCTGAGCAAG
GACACCTACGACGACCTGGACAACCTGCTGGCCAGATCGCGACCA
GTACGCCACCTGTTCTGGCCCAAGAACCTGTCCGACGCCATCCTG
TGAGCGACATCTGAGAGTGAACACCGAGATACCAAGGCCCCCTGAGC
GCCTCTATGATCAAGAGATAACGAGCACCCAGGACCTGACCTGCT
GAAAGCTCTCGTGGCAGCAGCTGCCAGAGTACAAAGAGATTTCT

- continued

TCGACCAGAGCAAGAACGGCTACGCCCTACATTGACGGGGAGCCAGC
CAGGAAGAGTTCTACAATTCTCATCAAGCCCATCTGGAAAAGATGGACGG
CACCGAGGAACCTGCTCGTAAGCTGAACAGAGAGGACCTGCTGCGGAAGC
AGCGGACCTCGACAAACGGCAGCATCCCCACAGATCCACCTGGAGAG
CTGCACGCCATTCTGCGCGCAGGAAGATTTCACCATCCCTGAAGGA
CAACCGGGAAAAGATCGAGAAGATCTGACCTTCCGATCCCTACTACG
TGGGCCCTCTGCCAGGGAAACAGCAGATTGCCTGGATGACCAGAAAAG
AGCGAGGAACCATCACCCCTGAACTTCGAGGAAGTGGTGGACAAGGG
CGCTCCGCCAGAGCTCATCGAGCGGATGACCAACTTCGATAAGAAC
TGCCCAACGAGAAGGTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTC
ACCGTGATAACGAGCTGACCAAAGTGAATACGTGACCGAGGGAAATGAG
AAAGCCGCCCTCTGAGCGGGAGCAGAAAAAGGCATGTGGACCTG
TGTTCAAGACCAACCGAAAGTGCACGAGCTGAAAGAGGACTAC
TTCAAGAAAATCGAGTCTGACTCCGTGAAATCTCGGTGAAAGA
TCGGTTCAACGCCCTGGCACATACCACGATCTGCTGAAATTATCA
AGGACAAGGACTCCCTGGACAATGAGGAAACGAGGACATTCTGAAAGAT
ATCGTGACTGCCCTGACACTGTTGAGGACAGAGAGATGATCGAGGAACG
GCTGAAAACCTATGCCACCTGTCGACGACAAAGTGTGATGAGCAGCTGA
AGCGGGAGATACACCGCTGGGAGGCTGAGCCGAAGCTGATCAAC
GGCATCCGGACAAGCAGTCCGCAAGACAATCCTGGATTCCTGAAAGTC
CGACGGCTCGCCAACAGAAACTTCATGCACTGATCCACGACGACAGCC
TGACCTTAAAGAGGACATCCAGAAAGCCCAGTGTCCGGCAGGGCAG
AGCCTGACGAGCACATTGCAATCTGCCGGCAGCCCGCATTAAAGAA
GGGCATCTGAGACAGTGAAGGTGGACGAGCTGTTGAAAGTGTG
GCCGGACAAGCCGAGAACATCGTGTGAAATGGCCAGAGAGAACAG
ACCAACCCAGAAGGGACAGAACAGCCGAGAGAATGAGCAGTCGA
AGAGGGCATCAAAGAGCTGGGAGGCCAGATCTGAAAGAACACCCCTGG
AAAACACCCAGCTGAGAACGAGAAGCTGTACCTGTACTACCTGAGAAT
GGCGGGATATGATCGTGGACAGGAACGGACATCAACCGGCTGTCGA
CTACGATGTGGACCATATCGTGCCTCAGAGCTTCTGAAAGGACGACTCCA
TCGACAACAAGGTGCTGACGAGCAGCGACAAGAACCGGGCAAGCGAC
AACGTGCCCTCGAAGAGGTGAGAAGAGATGAGAACACTACTGGCGCA
GCTGCTGAAACGCAAGCTGATTACCCAGAGAAAGTCTGACAACTCTGACCA
AGGCCGAGAGAGGGCGGCTGAGCGAAGTGGATAAGGCCGCTCATCAAG
AGACAGCTGGTGGAAAACCCGGCAGATCACAAAGCACGAGTGGCACAGATCCT
GGACTCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGG
AAAGTGAAGTGTGATCACCTGAGTCAAGCTGAGTGGTGTCCGATTCCGGAAG
GATTTCCAGTTTACAAGTGCAGCAGATCAACAACTACCAACACGCCA
CGACGCCACCTGAACGCCGCTGAGGGAAACGCCCTGATCAAAAGTAC

-continued

```

CTAACGCTGAAAGCGAGTCGTACGGCAGACTACAAGGTGTACGACGTG  

CGGAAGATGATGCCAAGAGCGAGCAGGAATCGCAAGGCTACGCCAA  

GTACTTCTACAGCAACATCATGAACTTTCAAGACCGAGATTACCC  

TGGCCAACGGCAGATCGGAAGCGGCCTCTGATCGAGACAAACGGCAGA  

ACCGGGGAGATCGTGTGGATAAGGGCGGGATTTGCCACCGTGCAGGAA  

AGTGCTGAGCATGCCCAAGTGAATATCGTAAAAAGACCGAGGTGCAGA  

CAGCGGCTTCAGCAAAGAGTCTATCCTGCCAAGAGGAACAGCGATAAAG  

CTGATGCCAGAAAAGAGGACTGGGACCTAAGAAGTACGGCGCTTCGA  

CAGCCCCACCGTGGCTATTCTGTGCTGGTGGCTGGCAAAGTGGAAAAGG  

GCAAGTCCAAGAAACTGAAGAGTGTAAAGAGACTGCTGGGATCACCATC  

ATGGAAAGAACAGCTTCGAGAAGAACATCCCCTCGACTTCTGGAAGCCAA  

GGGCTACAAAGAAGTAAAAAGACCTGATCATCAAGCTGCCAAAGTACT  

CCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAAATGCTGGCCTCTGCCGC  

GAACTCAGAACAGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAAC  

CCTGTACCTGGCAGCCACTATGAGAAGCTGAAGGGCTCCCCGAGGATA  

ATGAGCAGAACAGCTGTTGTGGAACAGCACAAAGCAACTACCTGGACGAG  

ATCATCGAGCAGATCAGCGAGTCTCCAAAGAGAGTGTACCTGGCGACGC  

TAATCTGGACAAAGTGTGTCGCCCTACAACAAAGCACCGGGATAAGCCCA  

TCAGAGAGCAGGCCGAGAATATCATCCACCTGTTACCCCTGACCAATCTG  

GGAGCCCCCTGCCCTCAAGTACTTGACACCACTCGACCGGAAGAG  

GTACACCAGCACAAAGAGGTGCTGGACGCCACCCCTGATCACCAGAGCA  

TCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCAG  

TTCTTTCTTAGCTGACCGACTTCTTAGTAGCAGCAGGACGCTTAA  

A  

(underline = NLS-hSpCas9-NLS)

```

Example Chimeric RNA for *S. thermophilus* LMD-9 CRISPR1 Cas9 (with PAM of NNAGAAW)

```

(SEQ ID NO: 21)  

NNNNNNNNNNNNNNNNNNgttttgcatacgatttaGAAAtaaa  

tcttgcaagactacaagataaggcttcatgcccaaatcaacaccctgt  

catttatggcagggtttcgatTTTTTT  

(N = guide sequence; first underline = tracr mate sequence; second underline = tracr sequence; bold = terminator)

```

Example Chimeric RNA for *S. thermophilus* LMD-9 CRISPR1 Cas9 (with PAM of NNAGAAW)

```

(SEQ ID NO: 22)  

NNNNNNNNNNNNNNNNNNgttttgcatacgattcaGAAAtgcagaagcta  

caaagataaggcttcatgcccaaatcaacaccctgtcatttatggcagg  

gtgtttcgatTTTTTT  

(N = guide sequence; first underline = tracr mate sequence; second underline = tracr sequence; bold = terminator)

```

Example chimeric RNA for *S. thermophilus* LMD-9 CRISPR1 Cas9 (with PAM of NNAGAAW)

```

(SEQ ID NO: 23)  

NNNNNNNNNNNNNNNNNNgttttgcatacgattcaGAAAtgcagaagcta  

caaagataaggcttcatgcccaaatcaacaccctgtcatttatggcagg  

gtgtTTTTTT (N = guide sequence;  

(N = guide sequence; first underline = tracr mate sequence; second underline = tracr sequence; bold = terminator)

```

Example Chimeric RNA for *S. thermophilus* LMD-9 CRISPR1 Cas9 (with PAM of NNAGAAW)

```

(SEQ ID NO: 60)  

NNNNNNNNNNNNNNNNNNgttttgcatacgatttaGAAAtaaa  

tcttgcaagactacaagataaggcttcatgcccaaatcaacaccctgt  

catttatggcagggtttcgatTTTTTT  

(N = guide sequence; first underline = tracr mate sequence; second underline = tracr sequence; bold = terminator)

```

Example Chimeric RNA for *S. thermophilus* LMD-9 CRISPR1 Cas9 (with PAM of NNAGAAW)

```

(SEQ ID NO: 61)  

NNNNNNNNNNNNNNNNgttttgcatacgattcaGAAAtgcagaagcta  

caaagataaggcttcatgcccaaatcaacaccctgtcatttatggcagg  

gtgtttcgatTTTTTT  

(N = guide sequence; first underline = tracr mate sequence; second underline = tracr sequence; bold = terminator)

```

Example Chimeric RNA for *S. thermophilus* LMD-9 CRISPR1 Cas9 (with PAM of NNAGAAW)

```
(SEQ ID NO: 62)
NNNNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagcta
caaagataaggcattatgccaaatcaacaccctgtcatttatggcagg
gtgtTTTTTT (N = guide sequence;
first underline = tracr mate sequence; second
underline = tracr sequence; bold = terminator)
```

Example Chimeric RNA for *S. thermophilus* LMD-9 CRISPR1 Cas9 (with PAM of NNAGAAW)

```
(SEQ ID NO: 63)
NNNNNNNNNNNNNNNNNNgttattgtactctcaagatthaGAAAtaaa
tcttcgagaagctacaatgataaggcattatgccaaatcaacaccctgt
cattttatggcagggtttcgatTTTTTT
(N = guide sequence; first underline =
tracr mate sequence; second underline = tracr
sequence; bold = terminator)
```

Example chimeric RNA for *S. thermophilus* LMD-9 CRISPR1 Cas9 (with PAM of NNAGAAW)

```
(SEQ ID NO: 64)
NNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagcta
caatgataaggcattatgccaaatcaacaccctgtcatttatggcagg
gttttcgttatttaaTTTTTT
(N = guide sequence; first underline = tracr mate
sequence; second underline = tracr sequence;
bold = terminator)
```

Example Chimeric RNA for *S. thermophilus* LMD-9 CRISPR1 Cas9 (with PAM of NNAGAAW)

```
(SEQ ID NO: 65)
NNNNNNNNNNNNNNNNNNgttattgtactctcaGAAAtgcagaagcta
caatgataaggcattatgccaaatcaacaccctgtcatttatggcagg
gtgtTTTTTT (N = guide sequence;
first underline = tracr mate sequence; second
underline = tracr sequence; bold = terminator)
```

Example Chimeric RNA for *S. thermophilus* LMD-9 CRISPR3 Cas9 (with PAM of NGGNG)

```
(SEQ ID NO: 66)
NNNNNNNNNNNNNNNNNNgttttagagctgtgGAAAcacagcgaggta
aaataaggcttagccgtactcaacttgaaaagggtggcaccgattcggtg
tTTTTTT (N = guide sequence;
first underline = tracr mate sequence; second
underline = tracr sequence; bold = terminator)
```

Codon-optimized version of Cas9 from *S. thermophilus* LMD-9 CRISPR3 locus (with an NLS at both 5' and 3' ends)

```
(SEQ ID NO: 67)
ATGAAAAGGCCGGCGGCCACGAAAAAGGCCGGCAGGCCAAAAAGAAAAA
GACCAAGCCCTACAGCATGGCCTGGACATGGCACCAATAGCGTGGGCT
GGCCGCTGACCAGCACAACTACAAGGTGCCAGCAAGAAAATGAAGGTG
CTGGCAACACCTCCAAGAAGTACATCAAGAAAAACCTGCTGGGCTGCT
GCTGTCGACAGCGGCATTACAGCCGAGGGCAGACGGCTGAAGAGAACCG
CCAGACGGCGGTACACCCGGCGAGAACAGAACATCCTGTATCTGCAAGAG
ATCTTCAGCACCGAGATGGCTACCCCTGGACAGCAGCCTCTCCAGCGCT
GGACGACAGCTTCTGGTGCCTGACGACAAGCGGGACAGCAAGTACCCCA
TCTTCGGCAACCTGGTGAAGAGAACGGCTTACACGACGAGTCCCCACC
ATCTTACACCTGAGAAAGTACCTGGCCGACAGCACCAAGAAGGCCGACCT
GAGACTGGTGTACTGGCCCTGGCCACATGATCAAGTACCGGGCCACT
TCCTGATCGAGGGCGAGTTAACAGCAAGAACAGACATCCAGAAGAAC
TTCCAGGACTTCCTGGACACCTACAAACGCCATCTCGAGAGCGACCTGTC
CCTGGAAAACAGCAAGCAGCTGGAAAGAGATCGTGAAGGACAAGATCAGCA
AGCTGGAAAAGAAGGCCGACAGCACCAAGAAGGCCGACCT
AGCGGAATCTTCAGCGAGTTCTGAAGCTGATCGTGGCAACCAGGCCGA
CTTCAGAAAAGTGTCTAACCTGGACAGGAAAGCCAGCTGCACCTCAGCA
AAGAGAGCTACGACGAGGACCTGGAAACCTGCTGGATATATCGGCAC
GACTACAGCGACGTGTTCTGAAGGCCAGAGCTGTACGACGCTATCCT
GCTGAGCGCTTCTGACCGTGACCGACAACGAGACAGAGGCCACTGA
GCAGCGCCATGATTAAGCGGTACAACGAGCACAAAGAGGATCTGGCTCTG
CTGAAAGAGTACATCCGGAACATCAGCCTGAAAACCTACAATGAGGTGTT
CAAGGACGACACCAAGAACGGTACGCCGGCTACATCGACGGCAAGACCA
ACCAGGAAGATTCTATGTGTACCTGAAGAAGCTGCTGGCCAGTCAG
GGGGCCGACTACTTCTGGAAAAAAATCGACCGCGAGGATTTCTGCGGAA
GCAGCGGACCTCGACAAACGGCAGCATCCCTACCGATCCATCTGCAGG
AAATCGGGCCATCCTGGACAAGCAGGCCAAGTTCTACCCATTCTGGCC
AAGAACAAAGAGCGGATCGAGAAGATCTGACCTTCCGATCCCTTACTA
CGTGGGCCCCCTGGCAGAGGCAACAGCGATTTCGCTGGCATCCGGA
```

- continued

```

ACCGCAATGAGAAGATCACCCCTGGAACCTTGAGGACGTGATCGACAAA
GAGTCCAGCAGGCCAGGGCCTTCATCAACCGGATGACCAGCTCGACCTGTA
CCTGCCGAGGAAAGGTGCTGCCAACGACAGCCTGCTGTACGAGACAT
TCAATGTGTATAACGAGCTGACCAAAGTGCAGTTATCGCGAGTCTATG
CGGGACTACAGTCTGGACTCCAAGCAGAAAAGGACATCGTGGGCT
GTACTTCAGGACAAGCGGAAAGTGAACGATAAGGACATCATCGAGTACC
TGCACGCCATCTACGGCTACGATGGCATCGAGCTGAAGGGCATCGAGAAG
CAGTTCAACTCCAGCCTGAGCACATACCACGACCTGCTGAACATTATCAA
CGACAAAGAATTCTGGACGACTCCAGCAACGAGGCCATCGAAGAGA
TCATCCACACCTGACCATCTTGAGGACCGCAGATGATCAAGCAGCGG
CTGAGCAAGTTCGAGAACATCTCGACAAGAGCGTGTGAAAAAGCTGAG
CAGACGGCACTACACCGCTGGGCAAGCTGAGCAGCAAGCTGATCAACG
GCATCCGGACGAGAAAGTCCGCAACACAATCTGGACTACCTGATCGAC
GACGGCATCAGCAACCGGAACTTCATCGAGCTGATCCACGACGACGCCCT
GAGCTTCAAGAAGAAGATCCAGAAGGCCAGATCATGGGGACGAGGACA
AGGGCAACATCAAAGAAGTCGTGAAGTCCCTGCCGGCAGCCCCGCCATC
AAGAAGGGAAATCTCGAGACATCAAGATCGTGGACGAGCTGTGAAAGT
GATGGCGGCAGAAAGCCGAGAGCATCGTGGTGGAAATGGTAGAGAGA
ACCAGTACACCAATCAGGGCAAGAGCAACAGCAGCAGAGACTGAAGAGA
CTGGAAAAGTCCCTGAAAGAGCTGGCAGCAAGATTCTGAAAGAGAAATAT
CCCTGCCAAGCTGCTCAAGATCGACAACAAACGCCCTGAGAACGACCGC
TGTACCTGTACTACCTGAGAATGGCAAGGACATGTATACAGGGACGAC
CTGGATATCGACCGCCTGAGCAACTACGACATCGACCATAATTATCCCCA
GGCCTTCCTGAAAGACAACAGCATTGACAACAAAAGTGTGGTGTCCCTCG
CCAGCAACCGCGCAAGTCCGATGATGTGCCAGCCTGGAAGTCGTGAAA
AAGAGAAAGACCTCTGGTATCAGCTGCTGAAAAGCAAGCTGATTAGCCA
GAGGAAGTTCGACAACCTGACCAAGGCCAGAGGAGGGCCTGAGCCCTG
AAGATAAGGCCGCTTCATCCAGAGACAGCTGGTGGAAACCCGGCAGATC
ACCAAGCACGTGGCCAGACTGTGGATGAGAAAGTTAACAAACAAGAGGA
CGAGAACAAACGGGGCGTGGGACCGTGAAGATCATCACCTGAAGTCCA
CCCTGGTGTCCCAGTTCCGGAGGGACTTCGAGCTGTATAAAAGTGCACGAG
ATCAATGACTTCAACCACGCCACGACGCCACCTGAATGCCGTGGTGGC
TTCGCCCTGCTGAAGAAGTACCCCTAACGCTGGAAACCCGAGTTCGTGTACG
GCGACTACCCCAAGTACAACCTCTCAGAGAGCGGAAGTCCGCCACCGAG
AAGGTGTACTTCTACTCCAACATCATGAATATCTTAAAGAAGTCCATCTC
CCTGGCCGATGGCAGAGTGATCGAGCGGCCCTGATCGAAGTGAACGAAG
AGACAGCGCAGAGCGTGTGGAAACAAAGAAAGCAGACCTGGCCACCGTGG
CGGGTGTGAGTTATCTCAAGTGAATGTCGTGAAGAAGGTGGAAGAAC
GAACCAACGGCCTGGATCGGGCAAGGCCAAGGGCCTGTTCAACGCCAAC

```

- continued

```

TGTCCAGCAAGCTTAAGCCAACTCCAAACGAGAAATCTCGTGGGGCCAAA
GAGTACCTGGACCCCTAACAGAAGTACGGCGGATACGCCGCATCTCCAATAG
CTTCACCGTGCTCGTAAGGGCACAATCGAGAAGGGCCTAACAAAAAGA
TCACAAACGCTGCTGGATTTCAGGGGATCTCTATCTGGACGGATCAAC
TACCGGAAGGATAAGCTGAACCTCTGCTGGAAAAGGCTAACAGGACAT
TGAGCTGATTATCGAGCTGCTAACGACTCCCTGTTCAACTGAGCGACG
GCTCCAGACGGATGCTGGCCTCCATCTGTCCACCAACAACAAGCGGGC
GAGATCCACAAGGAAACCAGATCTTCTGAGCCAGAAAATTGTGAAACT
GCTGTACACGCAAGGGATCTCCACACCATCAATGAGAACCAACCGGA
AATACGTGAAAACCACAAGAAAGAGTTGAGGAACCTGTTACTACATC
CTGGAGTTCAACGAGAACTATGTGGGAGCCAAGAAGAACGGCAAATGCT
GAACTCCGCCCTCCAGAGCTGGCAGAACACAGCATTGAGCGACTGTGCA
GCTCCTTCATCGGCCCTACCGCAGCGAGCGGAAGGGACTGTTGAGCTG
ACCTCCAGAGGCTCTGCCCGACTTGTAGTTCTGGAGGTGAAGATCCC
CCGGTACAGAGACTACACCCCTCTAGTCTGCTGAAGGACGCCACCCCTGA
TCCACCAAGACGTGACCGCCCTGTACGAAACCCGGATCGACCTGGCTAAG
CTGGCGAGGGAAAGCGCTCTGCTACTAACGAAAGCTGGTCAAGCTAA
GAAAAAGAAATAA

```

Example 5: RNA-Guided Editing of Bacterial Genomes Using CRISPR-Cas Systems

[0227] Applicants used the CRISPR-associated endonuclease Cas9 to introduce precise mutations in the genomes of *Streptococcus pneumoniae* and *Escherichia coli*. The approach relied on Cas9-directed cleavage at the targeted site to kill unmutated cells and circumvented the need for selectable markers or counter-selection systems. Cas9 specificity was reprogrammed by changing the sequence of short CRISPR RNA (crRNA) to make single- and multi-nucleotide changes carried on editing templates. Simultaneous use of two crRNAs enabled multiplex mutagenesis. In *S. pneumoniae*, nearly 100% of cells that survived Cas9 cleavage contained the desired mutation, and 65% when used in combination with recombineering in *E. coli*. Applicants exhaustively analyzed Cas9 target requirements to define the range of targetable sequences and showed strategies for editing sites that do not meet these requirements, suggesting the versatility of this technique for bacterial genome engineering.

[0228] The understanding of gene function depends on the possibility of altering DNA sequences within the cell in a controlled fashion. Site-specific mutagenesis in eukaryotes is achieved by the use of sequence-specific nucleases that promote homologous recombination of a template DNA containing the mutation of interest. Zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and homing meganucleases can be programmed to cleave genomes in specific locations, but these approaches require engineering of new enzymes for each target sequence. In prokaryotic organisms, mutagenesis methods either introduce a selection marker in the edited locus or

require a two-step process that includes a counter-selection system. More recently, phage recombination proteins have been used for recombineering, a technique that promotes homologous recombination of linear DNA or oligonucleotides. However, because there is no selection of mutations, recombineering efficiency can be relatively low (0.1-10% for point mutations down to Oct. 5, 2010-6 for larger modifications), in many cases requiring the screening of a large number of colonies. Therefore new technologies that are affordable, easy to use and efficient are still in need for the genetic engineering of both eukaryotic and prokaryotic organisms.

[0229] Recent work on the CRISPR (clustered, regularly interspaced, short palindromic repeats) adaptive immune system of prokaryotes has led to the identification of nucleases whose sequence specificity is programmed by small RNAs. CRISPR loci are composed of a series of repeats separated by 'spacer' sequences that match the genomes of bacteriophages and other mobile genetic elements. The repeat-spacer array is transcribed as a long precursor and processed within repeat sequences to generate small crRNA that specify the target sequences (also known as protospacers) cleaved by CRISPR systems. Essential for cleavage is the presence of a sequence motif immediately downstream of the target region, known as the protospacer-adjacent motif (PAM). CRISPR-associated (cas) genes usually flank the repeat-spacer array and encode the enzymatic machinery responsible for crRNA biogenesis and targeting. Cas9 is a dsDNA endonuclease that uses a crRNA guide to specify the site of cleavage. Loading of the crRNA guide onto Cas9 occurs during the processing of the crRNA precursor and requires a small RNA antisense to the precursor, the tracrRNA, and RNase III. In contrast to genome editing with ZFNs or TALENs, changing Cas9 target specificity does not require protein engineering but only the design of the short crRNA guide.

[0230] Applicants recently showed in *S. pneumoniae* that the introduction of a CRISPR system targeting a chromosomal locus leads to the killing of the transformed cells. It was observed that occasional survivors contained mutations in the target region, suggesting that Cas9 dsDNA endonuclease activity against endogenous targets could be used for genome editing. Applicants showed that marker-less mutations can be introduced through the transformation of a template DNA fragment that will recombine in the genome and eliminate Cas9 target recognition. Directing the specificity of Cas9 with several different crRNAs allows for the introduction of multiple mutations at the same time. Applicants also characterized in detail the sequence requirements for Cas9 targeting and show that the approach can be combined with recombineering for genome editing in *E. coli*.

Results: Genome Editing by Cas9 Cleavage of a Chromosomal Target

[0231] *S. pneumoniae* strain crR6 contains a Cas9-based CRISPR system that cleaves a target sequence present in the bacteriophage ϕ 8232.5. This target was integrated into the srtA chromosomal locus of a second strain R6^{8232.5}. An altered target sequence containing a mutation in the PAM region was integrated into the srtA locus of a third strain R6^{370.1}, rendering this strain 'immune' to CRISPR cleavage (FIG. 28a). Applicants transformed R6^{8232.5} and R6^{370.1} cells with genomic DNA from crR6 cells, expecting that

successful transformation of R6^{8232.5} cells should lead to cleavage of the target locus and cell death. Contrary to this expectation, Applicants isolated R6^{8232.5} transformants, albeit with approximately 10-fold less efficiency than R6^{370.1} transformants (FIG. 28b). Genetic analysis of eight R6^{8232.5} transformants (FIG. 28) revealed that the great majority are the product of a double recombination event that eliminates the toxicity of Cas9 targeting by replacing the ϕ 8232.5 target with the crR6 genome's wild-type srtA locus, which does not contain the protospacer required for Cas9 recognition. These results were proof that the concurrent introduction of a CRISPR system targeting a genomic locus (the targeting construct) together with a template for recombination into the targeted locus (the editing template) led to targeted genome editing (FIG. 23a).

[0232] To create a simplified system for genome editing, Applicants modified the CRISPR locus in strain crR6 by deleting cas1, cas2 and csn2, genes which have been shown to be dispensable for CRISPR targeting, yielding strain crR6M (FIG. 28a). This strain retained the same properties of crR6 (FIG. 28b). To increase the efficiency of Cas9-based editing and demonstrate that a template DNA of choice can be used to control the mutation introduced, Applicants co-transformed R6^{8232.5} cells with PCR products of the wild-type srtA gene or the mutant R6^{370.1} target, either of which should be resistant to cleavage by Cas9. This resulted in a 5- to 10-fold increase of the frequency of transformation compared with genomic crR6 DNA alone (FIG. 23b). The efficiency of editing was also substantially increased, with 8/8 transformants tested containing a wild-type srtA copy and 7/8 containing the PAM mutation present in the R6^{370.1} target (FIG. 23b and FIG. 29a). Taken together, these results showed the potential of genome editing assisted by Cas9.

[0233] Analysis of Cas9 target requirements: To introduce specific changes in the genome, one must use an editing template carrying mutations that abolish Cas9-mediated cleavage, thereby preventing cell death. This is easy to achieve when the deletion of the target or its replacement by another sequence (gene insertion) is sought. When the goal is to produce gene fusions or to generate single-nucleotide mutations, the abolishment of Cas9 nuclease activity will only be possible by introducing mutations in the editing template that alter either the PAM or the protospacer sequences. To determine the constraints of CRISPR-mediated editing, Applicants performed an exhaustive analysis of PAM and protospacer mutations that abrogate CRISPR targeting.

[0234] Previous studies proposed that *S. pyogenes* Cas9 requires an NGG PAM immediately downstream of the protospacer. However, because only a very limited number of PAM-inactivating mutations have been described so far, Applicants conducted a systematic analysis to find all 5-nucleotide sequences following the protospacer that eliminate CRISPR cleavage. Applicants used randomized oligonucleotides to generate all possible 1,024 PAM sequences in a heterogeneous PCR product that was transformed into crR6 or R6 cells. Constructs carrying functional PAMs were expected to be recognized and destroyed by Cas9 in crR6 but not R6 cells (FIG. 24a). More than 2×10^5 colonies were pooled together to extract DNA for use as template for the co-amplification of all targets. PCR products were deep sequenced and found to contain all 1,024 sequences, with coverage ranging from 5 to 42,472 reads (See section "Analysis of deep sequencing data"). The functionality of

each PAM was estimated by the relative proportion of its reads in the crR6 sample over the R6 sample. Analysis of the first three bases of the PAM, averaging over the two last bases, clearly showed that the NGG pattern was under-represented in crR6 transformants (FIG. 24b). Furthermore, the next two bases had no detectable effect on the NGG PAM (See section “Analysis of deep sequencing data”), demonstrating that the NGGN sequence was sufficient to license Cas9 activity. Partial targeting was observed for NAG PAM sequences (FIG. 24b). Also the NNGGN pattern partially inactivated CRISPR targeting (Table G), indicating that the NGG motif can still be recognized by Cas9 with reduced efficiency when shifted by 1 bp. These data shed light onto the molecular mechanism of Cas9 target recognition, and they revealed that NGG (or CCN on the complementary strand) sequences are sufficient for Cas9 targeting and that NGG to NAG or NNGGN mutations in the editing template should be avoided. Owing to the high frequency of these tri-nucleotide sequences (once every 8 bp), this means that almost any position of the genome can be edited. Indeed, Applicants tested ten randomly chosen targets carrying various PAMs and all were found to be functional (FIG. 30).

[0235] Another way to disrupt Cas9-mediated cleavage is to introduce mutations in the protospacer region of the editing template. It is known that point mutations within the ‘seed sequence’ (the 8 to 10 protospacer nucleotides immediately adjacent to the PAM) can abolish cleavage by CRISPR nucleases. However, the exact length of this region is not known, and it is unclear whether mutations to any nucleotide in the seed can disrupt Cas9 target recognition. Applicants followed the same deep sequencing approach described above to randomize the entire protospacer sequence involved in base pair contacts with the crRNA and to determine all sequences that disrupt targeting. Each position of the 20 matching nucleotides (14) in the spe1 target present in R6^{8232.5} cells (FIG. 23a) was randomized and transformed into crR6 and R6 cells (FIG. 24a). Consistent with the presence of a seed sequence, only mutations in the 12 nucleotides immediately upstream of the PAM abrogated cleavage by Cas9 (FIG. 24c). However, different mutations displayed markedly different effects. The distal (from the PAM) positions of the seed (12 to 7) tolerated most mutations and only one particular base substitution abrogated targeting. In contrast, mutations to any nucleotide in the proximal positions (6 to 1, except 3) eliminated Cas9 activity, although at different levels for each particular substitution. At position 3, only two substitutions affected CRISPR activity and with different strength. Applicants concluded that, although seed sequence mutations can prevent CRISPR targeting, there are restrictions regarding the nucleotide changes that can be made in each position of the seed. Moreover, these restrictions can most likely vary for different spacer sequences. Therefore Applicants believe that mutations in the PAM sequence, if possible, should be the preferred editing strategy. Alternatively, multiple mutations in the seed sequence may be introduced to prevent Cas9 nuclease activity.

[0236] Cas9-mediated genome editing in *S. pneumoniae*: To develop a rapid and efficient method for targeted genome editing, Applicants engineered strain crR6Rk, a strain in which spacers can be easily introduced by PCR (FIG. 33). Applicants decided to edit the β-galactosidase (bgaA) gene of *S. pneumoniae*, whose activity can be easily measured. Applicants introduced alanine substitutions of amino acids

in the active site of this enzyme: R481A (R→A) and N563A,E564A (NE→AA) mutations. To illustrate different editing strategies, Applicants designed mutations of both the PAM sequence and the protospacer seed. In both cases the same targeting construct with a crRNA complementary to a region of the β-galactosidase gene that is adjacent to a TGG PAM sequence (CCA in the complementary strand, FIG. 26) was used. The R→A editing template created a three-nucleotide mismatch on the protospacer seed sequence (CGT to GCA, also introducing a BtgZI restriction site). In the NE→AA editing template Applicants simultaneously introduced a synonymous mutation that created an inactive PAM (TGG to TTG) along with mutations that are 218 nt downstream of the protospacer region (AAT GAA to GCT GCA, also generating a Tsel restriction site). This last editing strategy demonstrated the possibility of using a remote PAM to make mutations in places where a proper target may be hard to choose. For example, although the *S. pneumoniae* R6 genome, which has a 39.7% GC content, contains on average one PAM motif every 12 bp, some PAM motifs are separated by up to 194 bp (FIG. 33). In addition Applicants designed a ΔbgaA in-frame deletion of 6,664 bp. In all three cases, co-transformation of the targeting and editing templates produced 10-times more kanamycin-resistant cells than co-transformation with a control editing template containing wild-type bgaA sequences (FIG. 25b). Applicants genotyped 24 transformants (8 for each editing experiment) and found that all but one incorporated the desired change (FIG. 25c). DNA sequencing also confirmed not only the presence of the introduced mutations but also the absence of secondary mutations in the target region (FIG. 29b,c). Finally, Applicants measured β-galactosidase activity to confirm that all edited cells displayed the expected phenotype (FIG. 25d).

[0237] Cas9-mediated editing can also be used to generate multiple mutations for the study of biological pathways. Applicants decided to illustrate this for the sortase-dependent pathway that anchors surface proteins to the envelope of Gram-positive bacteria. Applicants introduced a sortase deletion by co-transformation of a chloramphenicol-resistant targeting construct and a ΔsrtA editing template (FIG. 33a,b), followed by a ΔbgaA deletion using a kanamycin-resistant targeting construct that replaced the previous one. In *S. pneumoniae*, β-galactosidase is covalently linked to the cell wall by sortase. Therefore, deletion of srtA results in the release of the surface protein into the supernatant, whereas the double deletion has no detectable β-galactosidase activity (FIG. 34c). Such a sequential selection can be iterated as many times as required to generate multiple mutations.

[0238] These two mutations may also be introduced at the same time. Applicants designed a targeting construct containing two spacers, one matching srtA and the other matching bgaA, and co-transformed it with both editing templates at the same time (FIG. 25e). Genetic analysis of transformants showed that editing occurred in 6/8 cases (FIG. 25f). Notably, the remaining two clones each contained either a ΔsrtA or a ΔbgaA deletion, suggesting the possibility of performing combinatorial mutagenesis using Cas9. Finally, to eliminate the CRISPR sequences, Applicants introduced a plasmid containing the bgaA target and a spectinomycin resistance gene along with genomic DNA from the wild-type strain R6. Spectinomycin-resistant transformants that retain the plasmid eliminated the CRISPR sequences (FIG. 34a,d).

[0239] Mechanism and efficiency of editing: To understand the mechanisms underlying genome editing with Cas9, Applicants designed an experiment in which the editing efficiency was measured independently of Cas9 cleavage. Applicants integrated the ermAM erythromycin resistance gene in the srtA locus, and introduced a premature stop codon using Cas9-mediated editing (FIG. 33). The resulting strain (JEN53) contains an ermAM (stop) allele and is sensitive to erythromycin. This strain may be used to assess the efficiency at which the ermAM gene is repaired by measuring the fraction of cells that restore antibiotic resistance with or without the use of Cas9 cleavage. JEN53 was transformed with an editing template that restores the wild-type allele, together with either a kanamycin-resistant CRISPR construct targeting the ermAM (stop) allele (CRISPR::ermAM (stop)) or a control construct without a spacer (CRISPR::Ø) (FIG. 26a,b). In the absence of kanamycin selection, the fraction of edited colonies was on the order of 10^{-2} (erythromycin-resistant cfu/total cfu) (FIG. 26c), representing the baseline frequency of recombination without Cas9-mediated selection against unedited cells. However, if kanamycin selection was applied and the control CRISPR construct was co-transformed, the fraction of edited colonies increased to about 10-1 (kanamycin- and erythromycin-resistant cfu/kanamycin-resistant cfu) (FIG. 26c). This result shows that selection for the recombination of the CRISPR locus co-selected for recombination in the ermAM locus independently of Cas9 cleavage of the genome, suggesting that a subpopulation of cells is more prone to transformation and/or recombination. Transformation of the CRISPR::ermAM (stop) construct followed by kanamycin selection resulted in an increase of the fraction of erythromycin-resistant, edited cells to 99% (FIG. 26c). To determine if this increase is caused by the killing of non-edited cells, Applicants compared the kanamycin-resistant colony forming units (cfu) obtained after co-transformation of JEN53 cells with the CRISPR::ermAM (stop) or CRISPR::Ø constructs.

[0240] Applicants counted 5.3 times less kanamycin-resistant colonies after transformation of the ermAM (stop) construct ($2.5 \times 10^4 / 4.7 \times 10^3$, FIG. 35a), a result that suggests that indeed targeting of a chromosomal locus by Cas9 leads to the killing of non-edited cells. Finally, because the introduction of dsDNA breaks in the bacterial chromosome is known to trigger repair mechanisms that increase the rate of recombination of the damaged DNA, Applicants investigated whether cleavage by Cas9 induces recombination of the editing template. Applicants counted 2.2 times more colonies after co-transformation with the CRISPR::erm (stop) construct than with the CRISPR::Ø construct (FIG. 26d), indicating that there was a modest induction of recombination. Taken together, these results showed that co-selection of transformable cells, induction of recombination by Cas9-mediated cleavage and selection against non-edited cells, each contributed to the high efficiency of genome editing in *S. pneumoniae*.

[0241] As cleavage of the genome by Cas9 should kill non-edited cells, one would not expect to recover any cells that received the kanamycin resistance-containing Cas9 cassette but not the editing template. However, in the absence of the editing template Applicants recovered many kanamycin-resistant colonies after transformation of the CRISPR::ermAM (stop) construct (FIG. 35a). These cells that 'escape' CRISPR-induced death produced a background that determined a limit of the method. This background fre-

quency may be calculated as the ratio of CRISPR: ermAM (stop)/CRISPR::Ø cfu, 2.6×10^{-3} ($7.1 \times 10^1 / 2.7 \times 10^4$) in this experiment, meaning that if the recombination frequency of the editing template is less than this value, CRISPR selection may not efficiently recover the desired mutants above the background. To understand the origin of these cells, Applicants genotyped 8 background colonies and found that 7 contained deletions of the targeting spacer (FIG. 35b) and one harbored a presumably inactivating mutation in Cas9 (FIG. 35c).

[0242] Genome editing with Cas9 in *E. coli*: The activation of Cas9 targeting through the chromosomal integration of a CRISPR-Cas system is only possible in organisms that are highly recombinogenic. To develop a more general method that is applicable to other microbes, Applicants decided to perform genome editing in *E. coli* using a plasmid-based CRISPR-Cas system. Two plasmids were constructed: a pCas9 plasmid carrying the tracrRNA, Cas9 and a chloramphenicol resistance cassette (FIG. 36), and a pCRISPR kanamycin-resistant plasmid carrying the array of CRISPR spacers. To measure the efficiency of editing independently of CRISPR selection, Applicants sought to introduce an A to C transversion in the rpsL gene that confers streptomycin resistance. Applicants constructed a pCRISPR: LrpsL plasmid harboring a spacer that would guide Cas9 cleavage of the wild-type, but not the mutant rpsL allele (FIG. 27b). The pCas9 plasmid was first introduced into *E. coli* MG1655 and the resulting strain was co-transformed with the pCRISPR::rpsL plasmid and W542, an editing oligonucleotide containing the A to C mutation. streptomycin-resistant colonies after transformation of the pCRISPR: rpsL plasmid were only recovered, suggesting that Cas9 cleavage induces recombination of the oligonucleotide (FIG. 37). However, the number of streptomycin-resistant colonies was two orders of magnitude lower than the number of kanamycin-resistant colonies, which are presumably cells that escape cleavage by Cas9. Therefore, in these conditions, cleavage by Cas9 facilitated the introduction of the mutation, but with an efficiency that was not enough to select the mutant cells above the background of 'escapers'.

[0243] To improve the efficiency of genome editing in *E. coli*, Applicants applied their CRISPR system with recombineering, using Cas9-induced cell death to select for the desired mutations. The pCas9 plasmid was introduced into the recombineering strain HME63 (31), which contains the Gam, Exo and Beta functions of the -red phage. The resulting strain was co-transformed with the pCRISPR::rpsL plasmid (or a pCRISPR::Ø control) and the W542 oligonucleotide (FIG. 27a). The recombineering efficiency was 5.3×10^{-5} , calculated as the fraction of total cells that become streptomycin-resistant when the control plasmid was used (FIG. 27c). In contrast, transformation with the pCRISPR: rpsL plasmid increased the percentage of mutant cells to $65 \pm 14\%$ (FIGS. 27c and 29f). Applicants observed that the number of cfu was reduced by about three orders of magnitude after transformation of the pCRISPR::rpsL plasmid than the control plasmid ($4.8 \times 10^5 / 5.3 \times 10^2$, FIG. 38a), suggesting that selection results from CRISPR-induced death of non-edited cells. To measure the rate at which Cas9 cleavage was inactivated, an important parameter of Applicants' method, Applicants transformed cells with either pCRISPR:: rpsL or the control plasmid without the W542 editing oligonucleotide (FIG. 38a). This background of CRISPR 'escapers', measured as the ratio of pCRISPR::rpsL/

pCRISPR::Ø cfu, was 2.5×10^{-4} ($1.2 \times 10^2 / 4.8 \times 10^5$). Genotyping eight of these escapers revealed that in all cases there was a deletion of the targeting spacer (FIG. 38b). This background was higher than the recombineering efficiency of the rpsL mutation, 5.3×10^{-5} , which suggested that to obtain 65% of edited cells, Cas9 cleavage must induce oligonucleotide recombination. To confirm this, Applicants compared the number of kanamycin- and streptomycin-resistant cfu after transformation of pCRISPR: rpsL or pCRISPR::Ø (FIG. 27d). As in the case for *S. pneumoniae*, Applicants observed a modest induction of recombination, about 6.7 fold ($2.0 \times 10^{-4} / 3.0 \times 10^{-5}$). Taken together, these results indicated that the CRISPR system provided a method for selecting mutations introduced by recombineering.

[0244] Applicants showed that CRISPR-Cas systems may be used for targeted genome editing in bacteria by the co-introduction of a targeting construct that killed wild-type cells and an editing template that both eliminated CRISPR cleavage and introduced the desired mutations. Different types of mutations (insertions, deletions or scar-less single-nucleotide substitutions) may be generated. Multiple mutations may be introduced at the same time. The specificity and versatility of editing using the CRISPR system relied on several unique properties of the Cas9 endonuclease: (i) its target specificity may be programmed with a small RNA, without the need for enzyme engineering, (ii) target specificity was very high, determined by a 20 bp RNA-DNA interaction with low probability of non-target recognition, (iii) almost any sequence may be targeted, the only requirement being the presence of an adjacent NGG sequence, (iv) almost any mutation in the NGG sequence, as well as mutations in the seed sequence of the protospacer, eliminates targeting.

[0245] Applicants showed that genome engineering using the CRISPR system worked not only in highly recombinogenic bacteria such as *S. pneumoniae*, but also in *E. coli*. Results in *E. coli* suggested that the method may be applicable to other microorganisms for which plasmids may be introduced. In *E. coli*, the approach complements recombineering of mutagenic oligonucleotides. To use this methodology in microbes where recombineering is not a possible, the host homologous recombination machinery may be used by providing the editing template on a plasmid. In addition, because accumulated evidence indicates that CRISPR-mediated cleavage of the chromosome leads to cell death in many bacteria and archaea, it is possible to envision the use of endogenous CRISPR-Cas systems for editing purposes.

[0246] In both *S. pneumoniae* and *E. coli*, Applicants observed that although editing was facilitated by a co-selection of transformable cells and a small induction of recombination at the target site by Cas9 cleavage, the mechanism that contributed the most to editing was the selection against non-edited cells. Therefore the major limitation of the method was the presence of a background of cells that escape CRISPR-induced cell death and lack the desired mutation. Applicants showed that these 'escapers' arose primarily through the deletion of the targeting spacer, presumably after the recombination of the repeat sequences that flank the targeting spacer. Future improvements may focus on the engineering of flanking sequences that can still support the biogenesis of functional crRNAs but that are sufficiently different from one another to eliminate recombination. Alternatively, the direct transformation of chimeric crRNAs may be explored. In the particular case of *E. coli*,

the construction of the CRISPR-Cas system was not possible if this organism was also used as a cloning host. Applicants solved this issue by placing Cas9 and the tracrRNA on a different plasmid than the CRISPR array. The engineering of an inducible system may also circumvent this limitation.

[0247] Although new DNA synthesis technologies provide the ability to cost-effectively create any sequence with a high throughput, it remains a challenge to integrate synthetic DNA in living cells to create functional genomes. Recently, the co-selection MAGE strategy was shown to improve the mutation efficiency of recombineering by selecting a sub-population of cells that has an increased probability to achieve recombination at or around a given locus. In this method, the introduction of selectable mutations is used to increase the chances of generating nearby non-selectable mutations. As opposed to the indirect selection provided by this strategy, the use of the CRISPR system makes it possible to directly select for the desired mutation and to recover it with a high efficiency. These technologies add to the toolbox of genetic engineers, and together with DNA synthesis, they may substantially advance both the ability to decipher gene function and to manipulate organisms for biotechnological purposes. Two other studies also relate to CRISPR-assisted engineering of mammalian genomes. It is expected that these crRNA-directed genome editing technologies may be broadly useful in the basic and medical sciences.

[0248] Strains and culture conditions. *S. pneumoniae* strain R6 was provided by Dr. Alexander Tomasz. Strain crR6 was generated in a previous study. Liquid cultures of *S. pneumoniae* were grown in THYE medium (30 g/l Todd-Hewitt agar, 5 g/l yeast extract). Cells were plated on tryptic soy agar (TSA) supplemented with 5% defibrinated sheep blood. When appropriate, antibiotics were added as follows: kanamycin (400 µg/ml), chloramphenicol (5 µg/ml), erythromycin (1 µg/ml) streptomycin (100 µg/ml) or spectinomycin (100 µg/ml). Measurements of β-galactosidase activity were made using the Miller assay as previously described.

[0249] *E. coli* strains MG1655 and HME63 (derived from MG1655, Δ(argF-lac) U169 λ cI857 Δcro-bioA galK tyr 145 UAG mutS<>amp) (31) were provided by Jeff Roberts and Donald Court, respectively. Liquid cultures of *E. coli* were grown in LB medium (Difco). When appropriate, antibiotics were added as follows: chloramphenicol (25 µg/ml), kanamycin (25 µg/ml) and streptomycin (50 µg/ml).

[0250] *S. pneumoniae* transformation. Competent cells were prepared as described previously (23). For all genome editing transformations, cells were gently thawed on ice and resuspended in 10 volumes of M2 medium supplemented with 100 ng/ml of competence-stimulating peptide CSP1 (40), and followed by addition of editing constructs (editing constructs were added to cells at a final concentration between 0.7 ng/µl to 2.5 µg/µl). Cells were incubated 20 min at 37° C. before the addition of 2 µl of targeting constructs and then incubated 40 min at 37° C. Serial dilutions of cells were plated on the appropriate medium to determine the colony forming units (cfu) count.

[0251] *E. coli* Lambda-red recombineering. Strain HME63 was used for all recombineering experiments. Recombineering cells were prepared and handled according to a previously published protocol (6). Briefly, a 2 ml overnight culture (LB medium) inoculated from a single colony obtained from a plate was grown at 30° C. The

overnight culture was diluted 100-fold and grown at 30° C. with shaking (200 rpm) until the OD600 is from 0.4-0.5 (approximately 3 hrs). For Lambda-red induction, the culture was transferred to a 42° C. water bath to shake at 200 rpm for 15 min. Immediately after induction, the culture was swirled in an ice-water slurry and chilled on ice for 5-10 min. Cells were then washed and aliquoted according to the protocol. For electro-transformation, 50 µl of cells were mixed with 1 mM of salt-free oligos (IDT) or 100-150 ng of plasmid DNA (prepared by QIAprep Spin Miniprep Kit, Qiagen). Cells were electroporated using 1 mm Gene Pulser cuvette (Bio-rad) at 1.8 kV and were immediately resuspended in 1 ml of room temperature LB medium. Cells were recovered at 30° C. for 1-2 hrs before being plated on LB agar with appropriate antibiotic resistance and incubated at 32° C. overnight.

[0252] Preparation of *S. pneumoniae* genomic DNA. For transformation purposes, *S. pneumoniae* genomic DNA was extracted using the Wizard Genomic DNA Purification Kit, following instructions provided by the manufacturer (Promega). For genotyping purposes, 700 µl of overnight *S. pneumoniae* cultures were pelleted, resuspended in 60 µl of lysozyme solution (2 mg/ml) and incubated 30 min at 37° C. The genomic DNA was extracted using QIAprep Spin Miniprep Kit (Qiagen).

[0253] Strain construction. All primers used in this study are provided in Table G. To generate *S. pneumoniae* crR6M, an intermediate strain, LAM226, was made. In this strain the *aphA-3* gene (providing kanamycin resistance) adjacent to the CRISPR array of *S. pneumoniae* crR6 strain was replaced by a *cat* gene (providing chloramphenicol resistance). Briefly, crR6 genomic DNA was amplified using primers L448/L444 and L447/L481, respectively. The *cat* gene was amplified from plasmid pC194 using primers L445/L446. Each PCR product was gel-purified and all three were fused by SOEing PCR with primers L448/L481. The resulting PCR product was transformed into competent *S. pneumoniae* crR6 cells and chloramphenicol-resistant transformants were selected. To generate *S. pneumoniae* crR6M, *S. pneumoniae* crR6 genomic DNA was amplified by PCR using primers L409/L488 and L448/L481, respectively. Each PCR product was gel-purified and they were fused by SOEing PCR with primers L409/L481. The resulting PCR product was transformed into competent *S. pneumoniae* LAM226 cells and kanamycin-resistant transformants were selected.

[0254] To generate *S. pneumoniae* crR6Rc, *S. pneumoniae* crR6M genomic DNA was amplified by PCR using primers L430/W286, and *S. pneumoniae* LAM226 genomic DNA was amplified by PCR using primers W288/L481. Each PCR product was gel-purified and they were fused by SOEing PCR with primers L430/L481. The resulting PCR product was transformed into competent *S. pneumoniae* crR6M cells and chloramphenicol-resistant transformants were selected.

[0255] To generate *S. pneumoniae* crR6Rk, *S. pneumoniae* crR6M genomic DNA was amplified by PCR using primers L430/W286 and W287/L481, respectively. Each PCR product was gel-purified and they were fused by SOEing PCR with primers L430/L481. The resulting PCR product was transformed into competent *S. pneumoniae* crR6Rc cells and kanamycin-resistant transformants were selected.

[0256] To generate JEN37, *S. pneumoniae* crR6Rk genomic DNA was amplified by PCR using primers L430/W356 and W357/L481, respectively. Each PCR product was

gel-purified and they were fused by SOEing PCR with primers L430/L481. The resulting PCR product was transformed into competent *S. pneumoniae* crR6Rc cells and kanamycin-resistant transformants were selected.

[0257] To generate JEN38, R6 genomic DNA was amplified using primers L422/L461 and L459/L426, respectively. The *ermAM* gene (specifying erythromycin resistance) was amplified from plasmid pFW15 43 using primers L457/L458. Each PCR product was gel-purified and all three were fused by SOEing PCR with primers L422/L426. The resulting PCR product was transformed into competent *S. pneumoniae* crR6Rc cells and erythromycin-resistant transformants were selected.

[0258] *S. pneumoniae* JEN53 was generated in two steps. First JEN43 was constructed as illustrated in FIG. 33. JEN53 was generated by transforming genomic DNA of JEN25 into competent JEN43 cells and selecting on both chloramphenicol and erythromycin.

[0259] To generate *S. pneumoniae* JEN62, *S. pneumoniae* crR6Rk genomic DNA was amplified by PCR using primers W256/W365 and W366/L403, respectively. Each PCR product was purified and ligated by Gibson assembly. The assembly product was transformed into competent *S. pneumoniae* crR6Rc cells and kanamycin-resistant transformants were selected.

[0260] Plasmid construction. pDB97 was constructed through phosphorylation and annealing of oligonucleotides B296/B297, followed by ligation in pLZ12spec digested by EcoRI/BamHI. Applicants fully sequenced pLZ12spec and deposited its sequence in genebank (accession: KC112384).

[0261] pDB98 was obtained after cloning the CRISPR leader sequence was cloned together with a repeat-spacer-repeat unit into pLZ12spec. This was achieved through amplification of crR6Rc DNA with primers B298/B320 and B299/B321, followed by SOEing PCR of both products and cloning in pLZ12spec with restriction sites BamHI/EcoRI. In this way the spacer sequence in pDB98 was engineered to contain two BsaI restriction sites in opposite directions that allow for the scar-less cloning of new spacers.

[0262] pDB99 to pDB108 were constructed by annealing of oligonucleotides B300/B301 (pDB99), B302/B303 (pDB100), B304/B305 (pDB101), B306/B307 (pDB102), B308/B309 (pDB10³), B310/B311 (pDB104), B312/B313 (pDB105), B314/B315 (pDB106), B315/B317 (pDB107), B318/B319 (pDB108), followed by ligation in pDB98 cut by BsaI.

[0263] The pCas9 plasmid was constructed as follow. Essential CRISPR elements were amplified from *Streptococcus pyogenes* SF370 genomic DNA with flanking homology arms for Gibson Assembly. The tracrRNA and Cas9 were amplified with oligos HC008 and HC010. The leader and CRISPR sequences were amplified HC011/HC014 and HC015/HC009, so that two BsaI type IIS sites were introduced in between two direct repeats to facilitate easy insertion of spacers.

[0264] pCRISPR was constructed by subcloning the pCas9 CRISPR array in pZE21-MCS1 through amplification with oligos B298+B299 and restriction with EcoRI and BamHI. The rpsL targeting spacer was cloned by annealing of oligos B352+B353 and cloning in the BsaI cut pCRISPR giving pCRISPR::rpsL.

[0265] Generation of targeting and editing constructs. Targeting constructs used for genome editing were made by Gibson assembly of Left PCRs and Right PCRs (Table G). Editing constructs were made by SOEing PCR fusing PCR

products A (PCR A), PCR products B (PCR B) and PCR products C (PCR C) when applicable (Table G). The CRISPR::Ø and CRISPR::ermAM (stop) targeting constructs were generated by PCR amplification of JEN62 and crR6 genomic DNA respectively, with oligos L409 and L481.

[0266] Generation of targets with randomized PAM or protospacer sequences. The 5 nucleotides following the spacer 1 target were randomized through amplification of R6^{8232,5} genomic DNA with primers W377/L426. This PCR product was then assembled with the cat gene and the srtA upstream region that were amplified from the same template with primers L422/W376. 80 ng of the assembled DNA was used to transform strains R6 and crR6. Samples for the randomized targets were prepared using the following primers: B280-B290/L426 to randomize bases 1-10 of the target and B269-B278/L426 to randomize bases 10-20. Primers L422/B268 and L422/B279 were used to amplify the cat gene and srtA upstream region to be assembled with the first and last 10 PCR products respectively. The assembled constructs were pooled together and 30 ng was transformed in R6 and crR6. After transformation, cells were plated on chloramphenicol selection. For each sample more than 2×10^5 cells were pooled together in 1 ml of THYE and genomic DNA was extracted with the Promega Wizard kit. Primers B250/B251 were used to amplify the target region. PCR products were tagged and run on one Illumina MiSeq paired-end lane using 300 cycles.

Analysis of Deep Sequencing Data.

[0267] Randomized PAM: For the randomized PAM experiment 3,429,406 reads were obtained for crR6 and 3,253,998 for R6. It is expected that only half of them will correspond to the PAM-target while the other half will sequence the other end of the PCR product. 1,623,008 of the crR6 reads and 1,537, 131 of the R6 reads carried an error-free target sequence. The occurrence of each possible PAM among these reads is shown in supplementary file. To estimate the functionality of a PAM, its relative proportion in the crR6 sample over the R6 sample was computed and is denoted r_{ijklm} where I,j,k,l,m are one of the 4 possible bases. The following statistical model was constructed:

$$\log(r_{ijklm}) = \mu + b_2 i + b_3 j + b_4 k + b_2 b_3 i_j + b_3 b_4 j_k + \epsilon_{ijklm}$$

[0268] where ϵ is the residual error, b_2 is the effect of the 2nd base of the PAM, b_3 of the third, b_4 of the fourth, $b_2 b_3$ is the interaction between the second and third bases, $b_3 b_4$ between the third and fourth bases. An analysis of variance was performed:

Anova Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
b3	3	151.693	50.564	601.8450	<2.2e-16 ***
b2	3	90.521	30.174	359.1454	<2.2e-16 ***
b4	3	1.881	0.627	7.4623	6.070e-05 ***
b3:b2	9	228.940	25.438	302.7738	<2.2e-16 ***
b3:b4	9	3.010	0.334	3.9809	5.227e-05 ***
Residuals	996	83.680	0.084		

[0269] When added to this model, b_1 or b_5 do not appear to be significant and other interactions than the ones included can also be discarded. The model choice was made through successive comparisons of more or less complete models using the anova method in R. Tukey's honest sig-

nificance test was used to determine if pairwise differences between effects are significant.

[0270] NGGNN Patterns are Significantly Different from all Other Patterns and Carry the Strongest Effect (See Table Below).

[0271] In order to show that positions 1, 4 or 5 do not affect the NGGNN pattern Applicants looked at these sequences only. Their effect appears to be normally distributed (see QQ plot in FIG. 71), and model comparisons using the anova method in R shows that the null model is the best one, i.e. there is no significant role of b_1 , b_4 and b_5 .

Model Comparison Using the Anova Method in R for the NGGNN Sequences

Model 1: ratio.log ~1					
Model 2: ratio.log ~b1 + b4 + b5					
Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	63	14.579			
2	54	11.295	9	3.2836	1.7443 0.1013

Partial Interference of NAGNN and NNGGN Patterns

[0272] NAGNN patterns are significantly different from all other patterns but carry a much smaller effect than NGGNN (see Tukey's honest significance test below).

[0273] Finally, NTGGN and NCGGN patterns are similar and show significantly more CRISPR interference than NTGHN and NCGHN patterns (where H is A,T or C), as shown by a bonferroni adjusted pairwise student-test.

Pairwise Comparisons of the Effect of b_4 on NYGNN Sequences Using t Tests with Pooled SD

Data: b4			
	A	C	G
C	1.00	—	—
G	9.2e-05	2.4e-06	—
T	0.31	1.00	1.2e-08

[0274] Taken together, these results allow concluding that NNGGN patterns in general produce either a complete interference in the case of NGGNN, or a partial interference in the case of NAGGN, NTGGN or NCGGN.

Tukey Multiple Comparisons of Means: 95% Family-Wise Confidence Level

	diff	lwr	upr	p adj
\$b2:b3				
G:G-A:A	-2.76475	-2.94075	-2.58875	<1E-07
G:G-C:A	-2.79911	-2.97511	-2.62311	<1E-07
G:G-T:A	-2.7809	-2.9569	-2.6049	<1E-07
G:G-A:C	-2.81643	-2.99244	-2.64043	<1E-07
G:G-C:C	-2.77903	-2.95504	-2.60303	<1E-07
G:G-G:C	-2.64867	-2.82468	-2.47267	<1E-07
G:G-T:C	-2.79718	-2.97319	-2.62118	<1E-07
G:G-A:G	-2.67068	-2.84668	-2.49468	<1E-07
G:G-C:G	-2.73525	-2.91125	-2.55925	<1E-07
G:G-T:G	-2.7976	-2.62159	-2.9736	<1E-07
G:G-A:T	-2.76727	-2.59127	-2.94328	<1E-07

-continued

	diff	lwr	upr	p adj
G:G:C:T	-2.84114	-2.66513	-3.01714	<1E-07
G:G:G:T	-2.76409	-2.58809	-2.94009	<1E-07
G:G:T:T	-2.76781	-2.59181	-2.94381	<1E-07
G:G:G:A	-2.13964	-2.31565	-1.96364	<1E-07
G:A:A:A	-0.62511	-0.80111	-0.4491	<1E-07
G:A:C:A	-0.65947	-0.83547	-0.48346	<1E-07
G:A:T:A	-0.64126	-0.46525	-0.81726	<1E-07
G:A:C:C	-0.67679	-0.50078	-0.85279	<1E-07
G:A:C:C	-0.63939	-0.46339	-0.81539	<1E-07
G:A:G:C	-0.50903	-0.33303	-0.68503	<1E-07
G:A:T:C	-0.65754	-0.48154	-0.83354	<1E-07
G:A:A:G	-0.53104	-0.35503	-0.70704	<1E-07
G:A:C:G	-0.59561	-0.4196	-0.77161	<1E-07
G:A:T:G	-0.65795	-0.48195	-0.83396	<1E-07
G:A:A:T	-0.62763	-0.45163	-0.80363	<1E-07
G:A:C:T	-0.70149	-0.52549	-0.8775	<1E-07
G:A:G:T	-0.62445	-0.44844	-0.80045	<1E-07
G:A:T:T	-0.62817	-0.45216	-0.80417	<1E-07
	\$b3:b4			

Randomized Target

[0275] For the randomized target experiment 540,726 reads were obtained for crR6 and 753,570 for R6. As before, only half of the reads are expected to sequence the interesting end of the PCR product. After filtering for reads that carry a target that is error-free or with a single point mutation, 217,656 and 353,141 reads remained for crR6 and R6 respectively. The relative proportion of each mutant in the crR6 sample over the R6 sample was computed (FIG. 24c). All mutations outside of the seed sequence (13-20 bases away from the PAM) show full interference. Those sequences were used as a reference to determine if other mutations inside the seed sequence can be said to significantly disrupt interference. A normal distribution was fitted to these sequences using the fitdistr function of the MASS R package. The 0.99 quantile of the fitted distribution is shown as a dotted line in FIG. 24c. FIG. 72 shows a histogram of the data density with fitted normal distribution (black line) and 0.99 quantile (dotted line).

TABLE F

Relative abundance of PAM sequences in the crR6/R6 samples averaged over bases 1 and 5.								
3rd position								
		A	C	G	T			
2nd position	A	AAA 1.04	ACA 1.12	AGA 0.73	ATA 1.10	A 4th position		
	AAC 1.07	ACC 1.04	AGC 0.64	ATC 0.97	C 4th position			
	AAG 1.00	ACG 1.09	AGG 0.61	ATG 1.07	G			
	AAT 0.98	ACT 1.02	AGT 0.65	ATT 1.01	T			
C	CAA 1.05	CCA 1.05	CGA 0.99	CTA 1.07	A			
	CAC 1.04	CCC 1.02	CGC 1.08	CTC 1.04	C			
	CAG 1.08	CCG 1.08	CGG 0.81	CTG 1.05	G			
	CAT 1.13	CCT 1.05	CGT 1.07	CTT 1.08	T			
G	GAA 0.97	GCA 1.05	GGA 0.08	GTA 0.99	A			
	GAC 0.92	GCC 1.00	GGC 0.05	GTC 1.15	C			
	GAG 0.96	GCG 0.98	GGG 0.07	GTG 0.98	G			
	GAT 0.98	GCT 0.99	GGT 0.06	GTT 1.05	T			
T	TAA 1.08	TCA 1.16	TGA 1.05	TTA 1.14	A			
	TAC 1.00	TCC 1.08	TGC 1.08	TTC 1.05	C			
	TAG 1.02	TCG 1.11	TGG 0.77	TTG 1.01	G			
	TAT 1.01	TCT 1.12	TGT 1.21	TTT 1.02	T			

TABLE G

Primers used in this study (SEQ ID NOS 68-183, respectively, in order of appearance).	
Primer	Sequence 5'-3'
B217	TCCTAGCAGGATTTCTGATATTACTGTCACGTTTAGAGCTATGCTGTTTGA
B218	GTGACAGTAATATCAGAAATCCTGCTAGGAGTTGGGACCATTCAAACAGC
B229	GGGTTCAAGTCCTTGCTAGCAAGAG
B230	GCCAATGAAACGGGAACCCCTGGTC
B250	NNNNNGACGAGGCAATGGCTGAAATC
B251	NNNNNTATTTGGCTCATATTTGCTG
B255	CTTTACACCAATCGCTGCAACAGAC

TABLE G-continued

Primers used in this study (SEQ ID NOS 68-183, respectively, in order of appearance).	
Primer	Sequence 5'-3'
B256	CAAAATTCTAGTCTTCTTCGCTTTCCCCATAAAACCCCTCTTA
B257	AGGGTTTATGGGAAAGGCAAAGAAGACTAGAAATTTGATACC
B258	CTTACGGTGCATAAAAGTCATTCC
B269	TGGCTCGATTCAGCCATTGC
B270	CTTGACGAGGCAATGGCTGAAATCGAGCCAANAAAGCGCAAG
B271	CTTGACGAGGCAATGGCTGAAATCGAGCCAANAAAGCGCAAG
B272	CTTGACGAGGCAATGGCTGAAATCGAGCCAANAAAGCGCAAG
B273	CTTGACGAGGCAATGGGTGAAATCGAGCCAANAAAGCGCAAG
B274	CTTGACGAGGCAATGGGTGAAATCGAGCCAANAAAGCGCAAG
B275	CTTGACGAGGCAATGGGTGAAATCGAGCCAANAAAGCGNAAG
B276	CTTGACGAGGCAATGGGTGAAATCGAGCCAANAAAGCNCAAG
B277	CTTGACGAGGCAATGGGTGAAATCGAGCCAANAAAGCGNAAG
B278	CTTGACCAGCCAATGGCTGAAATCGAGCCAANAAAGCGCNAGAAG
B279	CCGCTTTTGGCTCGATTCAG
B280	CAATGGCTGAAATCGAGCCAANAAAGCGCANGAAGAAATC
B281	CAATGGCTGAAATCGAGCCAANAAAGCGCAANAAAGAAATC
B282	CAATGGCTGAAATCGAGCCAANAAAGCGCAAGNAGAAATC
B283	CAATGGCTGAAATCGAGCCAANAAAGCGCAAGANGAAATC
B284	CAATGGCTGAAATCGAGCCAANAAAGCGCAAGAANAAATC
B285	CAATGGCTGAAATCGAGCCAANAAAGCGCAAGAAGNAATCAACC
B286	CAATGGCTGAAATCGAGCCAANAAAGCGCAAGAAGANATCAACC
B287	CAATGGCTGAAATCGAGCCAANAAAGCGCAAGAAGAANTCAACC
B288	CAATGGCTGAAATCGAGCCAANAAAGCGCAAGAAGAAANCAACC
B289	CAATGGCTGAAATCGAGCCAANAAAGCGCAAGAAGAAATNAACCAGC
B290	CAATGGCTGAAATCGAGCCAANAAAGCGCAAGAAGAAATCNACCAGC
B296	gaccctCCCATCCGTACAACCCACAACCTGg
B297	aattcCAGGGTTGTGGTTGTACGGATGGAg
B298	CATGGATCCTATTCCTTAATAACTAAAAATATGG
B299	CATGAATTCAACTCAACAAGTCTCAGTGTGCTG
B300	AAACATTTCTCCATTAGGAAAAGGATGCTG
B301	AAAACAGCATCTTTCTAAATGGAGAAAAAT
B302	AAACCTTAAATCAGTCACAAATAGCAGCAAAATTG
B303	AAAACAATTGCTGCTATTGTGACTGATTTAAG
B304	AAACTTTCATCATACGACCAATCTGCTTTATTTG
B305	AAAACAAATAAGCAGATTGGTCGTATGATGAAAA
B306	AAACTCGTCCAGAAGTTATCGTAAAGAAATCGAG

TABLE G-continued

Primers used in this study (SEQ ID NOS 68-183, respectively, in order of appearance).	
Primer	Sequence 5'-3'
B307	AAAACTCGATTCTTTACGATAACTTCTGGACGA
B308	AAACAATCTCTCCAAGGTTCCCTAAAAATCTCG
B309	AAAACAGAGATTAAAGGAAACCTTGGAGAGATT
B310	AAACGCCATCGTCAGGAAGAAGCTATGCTTGAGTG
B311	AAAACACTCAAGCATAGCTTCTCCTGACGATGGC
B312	AAACATCTCTATACCTATTGAAATTCTTTGTATG
B313	AAAACATACAAAGAAATTCAATAAGTATAGAGAT
B314	AAACTAGCTGTGATAGTCGCAAACCGAGCCTTCG
B315	AAAACGAAGGCTGGTTTGCGGACTATCACAGCTA
B316	AAACATCGGAAGGTCGAGCAAGTAATTATCTTTG
B317	AAAACAAAAGATAATTACTTGCTCGACCTTCCGAT
B318	AAACAAGATGGTATCGCAAAGTAAGTGACAATAAG
B319	AAAACTTATTGTCACTTACTTGCAGTACCATCTT
B320	GAGACCTTGAGCTTCCGAGACTGGTCTCAGTTGGGACCATTCAAAACAG
B321	TGAGACCAGTCTCGGAAGCTCAAAGGTCTCGTTTAGAGCTATGCTGTTTG
B352	aaacTACTTTACGCAGCGCGAGTTCGGTTTTTg
B353	aaaacAAAAAACGAACTCCGCCTGCGTAAAGTA
HC008_SP	ATGCCGGTACTGCCGGGCCTTGTGGGATTACGAAATCATCCTG
HC009_SP	GTGACTGGCGATGCTGTGGAAATGGACGATCACACTACTCTTCTT
HC010_SP	TTAAGAAAATAATCTTCATCTAAATATACTTCAGTCACCTCTAGCTGAC
HC011_SP	ATTGATTTGAGTCAGCTAGGAGGTGACTGAAGTATATTTAGATGAAG
HC014_SP	GAGACCTTGAGCTTCCGAGACTGGTCTCAGTTGGGACCATTCAAAAC ACCATAGCTAAAAACCTCGTAGACTATTTTGTC
HC015_SP	GAGACCAGTCTCGGAAGCTAAAGGTCTCGTTTAGAGCTATGCTGTTT GAATGGTCCAAAACCTCAGCACACTGAGACTTG
L403	AGTCATCCCAGCAACAAATGG
L409	CGTGGTAAATCGGATAACGTTCCAAGTGAAG
L422	Tgtctttttcacaaacaagg
L426	AAGCCAAAGTTGGCACCAACC
L430	GTAGCTTATTCACTCCTAGTGG
L444	CGTTTGTGAACTAATGGGTGCAAATTACGAATCTCTCCTGACG
L445	CGTCAGGAGGAGATTGTAATTGCACCCATTAGTCAACAAACG
L446	GATATTATGGAGCCTATTTGTGGTTTTAGGCATAAAACTATATG
L447	CATATAGTTATGCCTAAAAACCCACAAAAATAGGCTCCATAATATC
L448	ATTATTTCTTAATAACTAAAAATATGG
L457	CGTgtacaatttgtacgtacggc

TABLE G-continued

Primers used in this study (SEQ ID NOS 68-183, respectively, in order of appearance).	
Primer	Sequence 5'-3'
L458	GCACCGGTGATCACTAGTCCTAGG
L459	cctaggactagtatcacccgtGCAAATATGAGCCAAATAATATAT
L461	GCCGTACGCTAGCAATTGTACACGTTGTTGAACTAATGGGTGC
L481	TTCAAATTTCCCATTGATTCTCC
L488	CCATATTTTAGTTATTAAGAAATAATACCAGCCATCAGTCACCTCC
W256	AGACGATTCAATAGACAATAAGG
W286	GTTTGGGACCATTCAAAAACAGCATAGCTCTAAAACCTCGTAGAC
W287	GCTATGCTGTTTGAATGGTCCAAAACGCACCCATTAGTTCAACAAACG
W288	GCTATGCTGTTTGAATGGTCCAAAACGCACCCATTAGTTCAACAAACG
W326	AATTCTTTCTTCATCATCGGTC
W327	AAGAAAAGAATGAAGATTGTTCATG
W341	GGTACTAATCAAAATAGTGAGGAGG
W354	GTTTTCAAAATCTGCGGTTGCG
W355	AAAAATTGAAAAATGGTGGAAACAC
W356	ATTCGTAAACGGTATCGGTTCTTTAAAGTTGGGACCATTCAAAACAGC
W357	TTTAAAGAAACCGATACCGTTACGAAATGTTAGAGCTATGCTGTTGA
W365	AAACGGTATCGGTTCTTTAAATTCAATTGTTGGGACCATTCAAAACAGC
W366	AATTGAATTAAAAGAAACCGATACCGTTGTTAGAGCTATGCTGTTGA
W370	GTTCCTTAAACCAAAACGGTATCGGTTCTTTAAATTCAATTGTTGGGACCATTCAAAACAGC
W371	GAAACCGATACCGTTGGTTAACGAAACAGGTAAAGGGCATTAAAC
W376	CGATTCAGCATTGCCTCGTC
W377	GCCTTGACGAGGCAATGGCTGAAATCGNNNNAAAAGCGCAAGAAGAAATCAAC
W391	TCCGTACAACCCACAACCTGCTAGTGAGCGTTGGGACCATTCAAAACAGC
W392	GCTCACTAGCAGGGTTGGGTGACGGAGTTAGAGCTATGCTGTTGA
W393	TTGTTGCCACTCTCCTTCTTC
W397	CAGGGTTGTTGGGTGCGATGGAGTTAACCTCCATCTCC
W398	GGGAGTTAACTCCATCGCAACAACCCACAACCTGCTAGTG
W403	GTGGTATCTATCGTGATGTGAC
W404	TTACCGAAACGGAATTATCTGC
W405	AAAGCTAGAGTCCGCAATTGG
W431	GTGGGTTGACGGATTGAGTTAACCTCCATCTCCTTC
W432	GATGGGAGTTAACTCAATCCGTACAACCCACAACCTG
W433	GCTTCACCTATTGCAGCACCAATTGACCACATGAAGATAG
W434	GTGGTCAATTGGTGCTGCAATAGGTGAAGCTAATGGTGATG
W463	CTGATTGTATTAATTGAGACATTATGCTCACCTC
W464	GCATAATGTCTAAAAATTAATACAAATCAGTGAATCATG

TABLE G-continued

Primers used in this study (SEQ ID NOS 68-183, respectively, in order of appearance).	
Primer	Sequence 5'-3'
W465	GTTTTGGGACCATTCAAAACAGCATAGCTCTAAAACGTGACAGTAATATCAG
W466	GTTTTAGAGCTATGCTGTTTGAATGGTCCAAAACGCTCACTAGCAGGGTTG
W542	ATACTTTACGCAGCGCGAGTCGGTTTgTAGGAGTGGTAGTATAACACGAGTACAT

TABLE H

Design of targeting and editing constructs used in this study (SEQ ID NOS 184, 184, 184, 185, and 186, respectively, in order of appearance).						
Targeting Constructs						
Edition	Template DNA	Left PCR	Right PCR	Spacer sequence	PAM	
bgaA R>A	crR6Rk	W256/W391	W392/L403	GCTCACTAGCAGGGTTGTGGTTGTACGGA	TGG	
bgaA NE>AA	crR6Rk	W256/W391	W392/L403	GCTCACTAGCAGGGTTGTGGTTGTACGGA	TGG	
ΔbgaA	crR6Rk	W256/W391	W392/L403	GCTCACTAGCAGGGTTGTGGTTGTACGGA	TGG	
ΔsrtA	crR6Rc	W256/B218	B217/L403	TCCTAGCAGGATTCTGATATTACTGTCAC	TGG	
ermB Stop	crR6Rk	W256/W356	W357/L403	TTTAAAGAACCGATAACGTTACGAAAT	TGG	
ΔsrtA ΔbgaA JEN51 (for Left PCR) and JEN52 (for Right PCR)		W256/W465	W466/W403	same as the ones used for ΔsrtA TGG and ΔbgaA		
Editing Constructs						
Edition	Template DNA	PCR A	PCR B	PCR C	SOEing PCR	Primers used to verify name of resulting strains edited genotype
bgaA R>A	R6	W403/W397	W398/W404	N/A	W403/W404	JEN56 W403/W404
bgaA NE>AA	R6	W403/W431	W432/W433	W434/W404	W403/W404	JEN60 W403/W404
ΔbgaA	R6	B255/B256	B257/B258	N/A	B255/B258	JEN52 W393/W405
ΔsrtA	R6	B230/W463	W464/B229	N/A	B230/B229	JEN51 W422/W426
ermB Stop	JEN38	L422/W370	W371/L426	N/A	L422/L426	JEN43 L457/L458
ΔsrtA ΔbgaA		same as the ones used for ΔsrtA and ΔbgaA			JEN64	same as the ones used for ΔsrtA and ΔbgaA

Example 6: Optimization of the Guide RNA for *Streptococcus pyogenes* Cas9 (Referred to as SpCas9)

[0276] Applicants mutated the tracrRNA and direct repeat sequences, or mutated the chimeric guide RNA to enhance the RNAs in cells.

[0277] The optimization is based on the observation that there were stretches of thymines (Ts) in the tracrRNA and guide RNA, which might lead to early transcription termination by the pol 3 promoter. Therefore Applicants gener-

ated the following optimized sequences. Optimized tracrRNA and corresponding optimized direct repeat are presented in pairs.

Optimized tracrRNA 1 (Mutation Underlined):

(SEQ ID NO: 187)
GGAACCATTCAtAACAGCATAGCAAGTTtAATAAGGCTAGTCGTTATC
AACTTGAAAAAGTGGCACCGAGTCGGTGT

Optimized Direct Repeat 1 (Mutation Underlined):

(SEQ ID NO: 188)
GTTaTAGAGCTATGCTGTTaTGAATGGTCCCAAAAC

Optimized tracrRNA 2 (Mutation Underlined):

(SEQ ID NO: 189)
GGAACCATTCAatACAGCATAGCAAGTTAAatATAAGGCTAGTCGGTTATC
AACTTGAAAAAGTGGCACCGAGTCGGTCTTTT

Optimized Direct Repeat 2 (Mutation Underlined):

(SEQ ID NO: 190)
GTaTTAGAGCTATGCTGTaTTGAATGGTCCCAAAAC

[0278] Applicants also optimized the chimeric guideRNA for optimal activity in eukaryotic cells.

Original Guide RNA:

(SEQ ID NO: 191)
NNNNNNNNNNNNNNNNNNNNNNNNNTTTAGAGCTAGAAATAGCAAGTTAAAA
TAAGGCTAGTCGGTTATCAACTTGAAAAGTGGCACCGAGTCGGTCTTT
TTTTT

Optimized Chimeric Guide RNA Sequence 1:

(SEQ ID NO: 192)
NNNNNNNNNNNNNNNNNNNNNNNTTTTAGAGCTAGAAATAGCAAGTTAA
TAAGGCTAGTCGGTTATCAACTTGAAAAGTGGCACCGAGTCGGTCTTT
TTTTT

Optimized Chimeric Guide RNA Sequence 2:

(SEQ ID NO: 193)
NNNNNNNNNNNNNNNNNNNNNNNTTTAGAGCTATGCTGTTGGAAACAA
AACAGCATAGCAAGTTAAATAAGGCTAGTCGGTTATCAACTTGAAAAAA
GTGGCACCGAGTCGGTCTTTTTT

Optimized Chimeric Guide RNA Sequence 3:

(SEQ ID NO: 194)
NNNNNNNNNNNNNNNNNNNTTTTAGAGCTATGCTGTTATTGGAAACAA
TACAGCATAGCAAGTTAATATAAGGCTAGTCGGTTATCAACTTGAAAAAA
GTGGCACCGAGTCGGTCTTTTTT

[0279] Applicants showed that optimized chimeric guide RNA works better as indicated in FIG. 3. The experiment was conducted by co-transfected 293 FT cells with Cas9 and a U6-guide RNA DNA cassette to express one of the four RNA forms shown above. The target of the guide RNA is the same target site in the human Emx1 locus: "GTCACCTCCAATGACTAGGG (SEQ ID NO: 195)"

Example 7: Optimization of *Streptococcus* Thermophiles LMD-9 CRISPR1 Cas9 (Referred to as St1Cas9)

[0280] Applicants designed guide chimeric RNAs as shown in FIG. 4.

[0281] The St1Cas9 guide RNAs can undergo the same type of optimization as for SpCas9 guide RNAs, by breaking the stretches of poly thymines (Ts)

Example 8: Cas9 Diversity and Mutations

[0282] 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 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 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.

[0283] 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. 39 and 40A-F).

[0284] In this example, Applicants show that the following mutations can convert SpCas9 into a nicking enzyme: D10A, E762A, H840A, N854A, N863A, D986A.

[0285] Applicants provide sequences showing where the mutation points are located within the SpCas9 gene (FIG. 41). Applicants also show that the nickases are still able to mediate homologous recombination (Assay indicated in FIG. 2). Furthermore, Applicants show that SpCas9 with these mutations (individually) do not induce double strand break (FIG. 47).

Example 9: Supplement to DNA Targeting Specificity of the RNA-Guided Cas9 Nuclease

Cell Culture and Transfection

[0286] Human embryonic kidney (HEK) cell line 293 FT (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% CO₂ incubation.

[0287] 293 FT cells were seeded either onto 6-well plates, 24-well plates, or 96-well plates (Corning) 24 hours prior to transfection. Cells were transfected using Lipofectamine 2000 (Life Technologies) at 80-90% confluence following the manufacturer's recommended protocol. For each well of a 6-well plate, a total of 1 µg of Cas9+sgRNA plasmid was used. For each well of a 24-well plate, a total of 500 ng Cas9+sgRNA plasmid was used unless otherwise indicated. For each well of a 96-well plate, 65 ng of Cas9 plasmid was used at a 1:1 molar ratio to the U6-sgRNA PCR product.

[0288] Human embryonic stem cell line HUES9 (Harvard Stem Cell Institute core) was maintained in feeder-free conditions on GelTrex (Life Technologies) in mTesR medium (Stemcell Technologies) supplemented with 100 µg/ml Normocin (InvivoGen). HUES9 cells were transfected with Amaxa P3 Primary Cell 4-D Nucleofector Kit (Lonza) following the manufacturer's protocol.

SURVEYOR Nuclease Assay for Genome Modification

[0289] 293 FT cells were transfected with plasmid DNA as described above. Cells were incubated at 37° C. for 72 hours post-transfection prior to genomic DNA extraction. Genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Epicentre) following the manufacturer's protocol. Briefly, pelleted cells were resuspended in QuickExtract solution and incubated at 65° C. for 15 minutes and 98° C. for 10 minutes.

[0290] The genomic region flanking the CRISPR target site for each gene was PCR amplified (primers listed in Tables J and K), and products were purified using QiaQuick Spin Column (Qiagen) following the manufacturer's protocol. 400 ng total of the purified PCR products were mixed with 2 µl 10×Taq DNA 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.

Northern Blot Analysis of tracrRNA Expression in Human Cells

[0291] Northern blots were performed as previously described¹. Briefly, RNAs were heated to 95° C. for 5 min before loading on 8% denaturing polyacrylamide gels (SequaGel, National Diagnostics). Afterwards, RNA was transferred to a pre-hybridized Hybond N+ membrane (GE Healthcare) and crosslinked with Stratagene UV Crosslinker (Stratagene). Probes were labeled with [γ -32P] ATP (Perkin Elmer) with T4 polynucleotide kinase (New England Biolabs). After washing, membrane was exposed to phosphor screen for one hour and scanned with phosphorimager (Typhoon).

Bisulfite Sequencing to Assess DNA Methylation Status

[0292] HEK 293 FT cells were transfected with Cas9 as described above. Genomic DNA was isolated with the

DNeasy Blood & Tissue Kit (Qiagen) and bisulfite converted with EZ DNA Methylation-Lightning Kit (Zymo Research). Bisulfite PCR was conducted using KAPA2G Robust HotStart DNA Polymerase (KAPA Biosystems) with primers designed using the Bisulfite Primer Seeker (Zymo Research, Tables J and K). Resulting PCR amplicons were gel-purified, digested with EcoRI and HindIII, and ligated into a pUC19 backbone prior to transformation. Individual clones were then Sanger sequenced to assess DNA methylation status.

In Vitro Transcription and Cleavage Assay

[0293] HEK 293 FT cells were transfected with Cas9 as described above. Whole cell lysates were then prepared with a lysis buffer (20 mM HEPES, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT, 5% glycerol, 0.1% Triton X-100) supplemented with Protease Inhibitor Cocktail (Roche). T7-driven sgRNA was in vitro transcribed using custom oligos (Example 10) and HiScribe T7 In Vitro Transcription Kit (NEB), following the manufacturer's recommended protocol. To prepare methylated target sites, pUC19 plasmid was methylated by M.SssI and then linearized by NheI. The in vitro cleavage assay was performed as follows: for a 20 µL cleavage reaction, 10 µL of cell lysate with incubated with 2 µL cleavage buffer (100 mM HEPES, 500 mM KCl, 25 mM MgCl₂, 5 mM DTT, 25% glycerol), the in vitro transcribed RNA, and 300 ng pUC19 plasmid DNA.

Deep Sequencing to Assess Targeting Specificity

[0294] HEK 293 FT cells plated in 96-well plates were transfected with Cas9 plasmid DNA and single guide RNA (sgRNA) PCR cassette 72 hours prior to genomic DNA extraction (FIG. 72). The genomic region flanking the CRISPR target site for each gene was amplified (FIG. 74, FIG. 80, (Example 10) by a fusion PCR method to attach the Illumina P5 adapters as well as unique sample-specific barcodes to the target amplicons (schematic described in FIG. 73). PCR products were purified using EconoSpin 96-well Filter Plates (Epoch Life Sciences) following the manufacturer's recommended protocol.

[0295] Barcoded and purified DNA samples were quantified by Quant-iT PicoGreen dsDNA Assay Kit or Qubit 2.0 Fluorometer (Life Technologies) and pooled in an equimolar ratio. Sequencing libraries were then deep sequenced with the Illumina MiSeq Personal Sequencer (Life Technologies).

Sequencing Data Analysis and Indel Detection

[0296] MiSeq reads were filtered by requiring an average Phred quality (Q score) of at least 23, as well as perfect sequence matches to barcodes and amplicon forward primers. Reads from on- and off-target loci were analyzed by first performing Smith-Waterman alignments against amplicon sequences that included 50 nucleotides upstream and downstream of the target site (a total of 120 bp). Alignments, meanwhile, were analyzed for indels from 5 nucleotides upstream to 5 nucleotides downstream of the target site (a total of 30 bp). Analyzed target regions were discarded if part of their alignment fell outside the MiSeq read itself, or if matched base-pairs comprised less than 85% of their total length.

[0297] Negative controls for each sample provided a gauge for the inclusion or exclusion of indels as putative cutting events. For each sample, an indel was counted only

if its quality score exceeded $\mu - \sigma$, where μ was the mean quality-score of the negative control corresponding to that sample and σ was the standard deviation of same. This yielded whole target-region indel rates for both negative controls and their corresponding samples. Using the negative control's per-target-region-per-read error rate, q , the sample's observed indel count n , and its read-count R , a maximum-likelihood estimate for the fraction of reads having target-regions with true-indels, p , was derived by applying a binomial error model, as follows.

[0298] Letting the (unknown) number of reads in a sample having target regions incorrectly counted as having at least 1 indel be E , we can write (without making any assumptions about the number of true indels)

$$\text{Prob}(E | p) = \binom{R(1-p)}{E} q^E (1-q)^{R(1-p)-E}$$

[0299] since $R(1-p)$ is the number of reads having target-regions with no true indels. Meanwhile, because the number of reads observed to have indels is n , $n=E+Rp$, in other words the number of reads having target-regions with errors but no true indels plus the number of reads whose target-regions correctly have indels. We can then re-write the above

$$\text{Prob}(E | p) = \text{Prob}(n = E + Rp | p) = \binom{R(1-p)}{n-Rp} q^{n-Rp} (1-q)^{R-n}$$

[0300] Taking all values of the frequency of target-regions with true-indels p to be equally probable a priori, $\text{Prob}(n|p) \propto \text{Prob}(p|n)$. The maximum-likelihood estimate (MLE) for the frequency of target regions with true-indels was therefore set as the value of p that maximized $\text{Prob}(n|p)$. This was evaluated numerically.

[0301] In order to place error bounds on the true-indel read frequencies in the sequencing libraries themselves, Wilson score intervals (2) were calculated for each sample, given the MLE-estimate for true-indel target-regions, Rp , and the number of reads R . Explicitly, the lower bound l and upper bound u were calculated as

$$l = \left(Rp + \frac{z^2}{2} - z\sqrt{Rp(1-p) + z^2/4} \right) / (R + z^2)$$

$$u = \left(Rp + \frac{z^2}{2} + z\sqrt{Rp(1-p) + z^2/4} \right) / (R + z^2)$$

[0302] where z , the standard score for the confidence required in normal distribution of variance 1, was set to 1.96, meaning a confidence of 95%. The maximum upper bounds and minimum lower bounds for each biological replicate are listed in FIGS. 80-83.

qRT-PCR Analysis of Relative Cas9 and sgRNA Expression
[0303] 293 FT cells plated in 24-well plates were transfected as described above. 72 hours post-transfection, total RNA was harvested with miRNeasy Micro Kit (Qiagen). Reverse-strand synthesis for sgRNAs was performed with qScript Flex cDNA kit (VWR) and custom first-strand synthesis primers (Tables J and K). qPCR analysis was performed with Fast SYBR Green Master Mix (Life Tech-

nologies) and custom primers (Tables J and K), using GAPDH as an endogenous control. Relative quantification was calculated by the $\Delta\Delta CT$ method.

Table I

Target site sequences. Tested target sites for <i>S. pyogenes</i> type II CRISPR system with the requisite PAM. Cells were transfected with Cas9 and either crRNA-tracrRNA or chimeric sgRNA for each target.			
Target site ID	genomic target	Target site sequence (5' to 3')	PAM strand
1	EMX1	GTCACCTCCAATGACTAGGG (SEQ ID NO: 319)	TGG +
2	EMX1	GACATCGATGTCCTCCCAT (SEQ ID NO: 196)	TGG -
3	EMX1	GAGTCCCGAGCAGAAGAAAGAA (SEQ ID NO: 197)	GGG +
6	EMX1	GCGCCACCGGTTGATGTGAT (SEQ ID NO: 198)	GGG -
10	EMX1	GGGCCACAGATGAGAAACTC (SEQ ID NO: 199)	AGG -
11	EMX1	GTACAAACGGCAGAACGCTGG (SEQ ID NO: 200)	AGG +
12	EMX1	GGCAGAACGCTGGAGGGAGGA (SEQ ID NO: 201)	GGG +
13	EMX1	GGAGCCCTTCTCTCTCTGCT (SEQ ID NO: 202)	CGG -
14	EMX1	GGGCAACCACAAACCCACGA (SEQ ID NO: 203)	GGG +
15	EMX1	GCTCCCATTACATCAACCGG (SEQ ID NO: 204)	TGG +
16	EMX1	GTGGCGCATGGCACGAAGC (SEQ ID NO: 205)	AGG +
17	EMX1	GGCAGAGTGCTGCTGCTGC (SEQ ID NO: 206)	TGG +
18	EMX1	GCCCCCTGCGTGGGCCAACGC (SEQ ID NO: 207)	TGG +
19	EMX1	GACTGGCCAGAGTCCAGCTT (SEQ ID NO: 208)	GGG -
20	EMX1	GGCCTCCCCAAAGCCTGGCC (SEQ ID NO: 209)	AGG -
4	PVALB	GGGGCCGAGATTGGGTGTT (SEQ ID NO: 210)	AGG +
5	PVALB	GTGGCGAGAGGGGCCGAGAT (SEQ ID NO: 211)	TGG +
1	SERPINB5	GAGTGCCGCCGAGGCCGGGC (SEQ ID NO: 212)	GGG +
2	SERPINB5	GGAGTGCCGCCGAGGCCGGGG (SEQ ID NO: 213)	CGG +
3	SERPINB5	GGAGAGGAGTGCCGCCGAGG (SEQ ID NO: 214)	CGG +

TABLE J

Primer sequences		
SURVEYOR assay		
primer name	genomic target	primer sequence (5' to 3')
Sp-EMX1-F1	EMX1	AAAACCACCCCTCTCTCTGGC (SEQ ID NO: 36)
Sp-EMX1-R1	EMX1	GGAGATTGGAGACACGGAGAG (SEQ ID NO: 37)
Sp-EMX1-F2	EMX1	CCATCCCCCTCTGTGAATGT (SEQ ID NO: 215)
Sp-EMX1-R2	EMX1	GGAGATTGGAGACACGGAGA (SEQ ID NO: 216)
Sp-PVALB-F	PVALB	CTGAAAGCCAATCGCTGAC (SEQ ID NO: 38)
Sp-PVALB-R	PVALB	GGCAGCAAACCTCCTGTGCCT (SEQ ID NO: 39)
qRT-PCR for Cas9 and sgRNA expression		
primer name	primer sequence (5' to 3')	
sgRNA reverse-strand synthesis	AAGCACCGACTCGGTGCCAC (SEQ ID NO: 217)	
EMX1.1 sgRNA qPCR F	TCACCTCCAATGACTAGGGG (SEQ ID NO: 218)	
EMX1.1 sgRNA qPCR R	CAAGTTGATAACGGACTAGCCT (SEQ ID NO: 219)	
EMX1.3 sgRNA qPCR F	AGTCCGAGCAGAAGAAGAAGTTT (SEQ ID NO: 220)	
EMX1.3 sgRNA qPCR R	TTTCAAGTTGATAACGGACTAGCCT (SEQ ID NO: 221)	
Cas9 qPCR F	AAACAGCAGATTGGCTGGAA (SEQ ID NO: 222)	
Cas9 qPCR R	TCATCCGCTCGATGAAGCTC (SEQ ID NO: 223)	
GAPDH qPCR F	TCCAAAATCAAGTGGGGCGA (SEQ ID NO: 224)	
GAPDH qPCR R	TGATGACCCTTTGGCTCCC (SEQ ID NO: 225)	
Bisulfite PCR and sequencing		
primer name	primer sequence (5' to 3')	
Bisulfite PCR F (SERPINB5 locus)	GAGGAATTCTTTTTGTTYGAAATATGT TGGAGGTTTTGGAAAG (SEQ ID NO: 226)	
Bisulfite PCR R (SERPINB5 locus)	GAGAACGCTTAAATAAAAAACRACAATAC TCAACCCAACAAC (SEQ ID NO: 227)	
pUC19 sequencing	CAGGAAACAGCTATGAC (SEQ ID NO: 228)	

Table K

Sequences for primers to test sgRNA architecture. Primers hybridize to the reverse strand of the U6 promoter unless otherwise indicated. The U6 priming site is in italics, the guide sequence is indicated as a stretch of Ns, the direct repeat sequence is highlighted in bold, and the tracrRNA sequence underlined. The secondary structure of each sgRNA architecture is shown in FIG. 43.	
primer name	primer sequence (5' to 3')
U6-Forward	GCCTCTAGAGGTACCTGAGGGCTATTCATGATTCC (SEQ ID NO: 229)
I: sgRNA (DR + 12, tracrRNA + 85)	<u>ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTTC</u> <u>AAGTTGATAACGGACTAGCCTTATTTAACTTGCTATTT</u> <u>CTAGCTCTAAACCNNNNNNNNNNNNNNNNNNGTGT</u> <u>TCGTCCTTTCCACAAG</u> (SEQ ID NO: 230)
II: sgRNA (DR + 12, tracrRNA + 85) mut2	<u>ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTTC</u> <u>AAGTTGATAACGGACTAGCCTTATTTAACTTGCTATTT</u> <u>CTAGCTCTAAATACNNNNNNNNNNNNNNNNNGTGT</u> <u>TCGTCCTTTCCACAAG</u> (SEQ ID NO: 231)
III: sgRNA (DR + 22, tracrRNA + 85)	<u>ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTTC</u> <u>AAGTTGATAACGGACTAGCCTTATTTAACTTGCTATGC</u> <u>TGTTTGTTCCAATACAGCATAGCTCTAAACNNNNNN</u> <u>NNNNNNNNNNNNNNNGTGTTCGTCCTTTCCACAAG</u> (SEQ ID NO: 232)
IV: sgRNA (DR + 22, tracrRNA + 85) mut4	<u>ACCTCTAGAAAAAAAGCACCGACTCGGTGCCACTTTTC</u> <u>AAGTTGATAACGGACTAGCCTTATTTAACTTGCTATGC</u> <u>TGTATTGTTCCAATACAGCATAGCTCTAAACNNNNNN</u> <u>NNNNNNNNNNNNNNNGTGTTCGTCCTTTCCACAAG</u> (SEQ ID NO: 233)

Table L

Target sites with alternate PAMs for testing PAM specificity of Cas9. All target sites for PAM specificity testing are found within the human EMX1 locus.

Target site sequence (5' to 3')	PAM
AGGCCCCAGTGGCTGCTCT	NAA
ACATCAACCGGTGGCGCAT	NAT
AAGGTGTGGTCCAGAACCGG	NAC
CCATCACATCAACCGGTGG	NAG
AAACGGCAGAACGCTGGAGG	NTA
GGCAGAACGCTGGAGGAGGA	NTT
GGTGTGGTCCAGAACCGG	NTC
AACCGGAGGACAAAGTACA	NTG
TTCCAGAACCGGAGGACAA	NCA
GTGTGGTCCAGAACCGG	NCT
TCCAGAACCGGAGGACAA	NCC
CAGAACGCTGGAGGAGGA	NCG
CATCAACCGGTGGCGATT	NGA
GCAGAACGCTGGAGGAGGA	NGT

Table L-continued

Target sites with alternate PAMs for testing PAM specificity of Cas9. All target sites for PAM specificity testing are found within the human EMX1 locus.	
Target site sequence (5' to 3')	PAM
CCTCCCTCCCTGGCCCAGG (SEQ ID NO: 248)	NGC
TCATCTGTGCCCTCCCTC (SEQ ID NO: 249)	NAA
GGGAGGACATCGATGTCAC (SEQ ID NO: 250)	NAT
CAAACGGCAGAACGCTGGAG (SEQ ID NO: 251)	NAC
GGGTGGGCAACCACAAACC (SEQ ID NO: 252)	NAG
GGTGGGCAACCACAAACCC (SEQ ID NO: 253)	NTA
GGCTCCCATCACATCAACC (SEQ ID NO: 254)	NTT
GAAGGGCCTGAGTCCGAGC (SEQ ID NO: 255)	NTC
CAACCGGTGGCGCATGGC (SEQ ID NO: 256)	NTG

Table L-continued

Target site sequence (5' to 3')	PAM
AGGAGGAAGGGCCTGAGTC (SEQ ID NO: 257)	NCA
AGCTGGAGGAGGAAGGCC (SEQ ID NO: 258)	NCT
GCATTGCCACGAAGCAGGC (SEQ ID NO: 259)	NCC
ATTGCCACGAAGCAGGCCA (SEQ ID NO: 260)	NCG
AGAACCGGAGGACAAGTA (SEQ ID NO: 261)	NGA
TCAACCGGTGGCGATTGC (SEQ ID NO: 262)	NGT
GAAGCTGGAGGAGGAAGGG (SEQ ID NO: 263)	NGC

Example 10: Supplementary Sequences

[0304] All sequences are in the 5' to 3' direction. For U6 transcription, the string of underlined Ts serve as the transcriptional terminator.

TTTTT
(tracrRNA sequence in bold)

aaaaaa **NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN** gttttagatgtatgtg
ttttgaatggtoccaaaa**TTTTTTTT**
(direct repeat sequence is highlighted in gray and the
guide sequence is in bold Ns)

-continued

TTTTAATTATTTGTGCAGCGATGGGGCGGGGGGGGGGGGGGGCGCGCCAG
GCGGGCGGGCGGGCGAGGGCGGGCGGGCGAGGCGAGAGGTGCGCGGC
AGCCAATCAGAGCGCGCGCTCGAAAGTTCTTTATGGCGAGGC GGCGCG
CGCGGGCCCTATAAAAAGCGAAGCGCGCGGGCGGGAGTCGCTGCGACGCTGC
CTTCGCCCCGTGCCCGCTCCGCCGCCCTCGCGCCGCCGCCGGCTTGAC
TGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGCCCTCTCCGGCTG
TAATTAGCTGAGCAAGAGGTAAGGGTTAAGGGATGGTTGGTGGTGGGTATTA
ATGTTAATTACCTGGAGCACCTGCCTGAAATCACTTTTCAGGTTGGaccgg
tgccaccATGGACTATAAGGACCACGCGAGACTACAAGGATCATGATATTGAT
TACAAAGACGATGACGATAAGATGGCCCAAAGAAGAACGGAAGGTGGTATCC
ACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAA
CTCTGTGGCTGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTG
AAGGTGCTGGCAACACCGACCGGACAGCATCAAGAAGAACCTGATCGGAGCCC
TGCTGTTGACAGCGGGAAAAGCGGAGGCCACCCGGCTGAAAGAGAACGCCAG
AAGAAGATACACCAAGCAGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGC
AACGAGATGGCCAAGGTGGACGACAGCTTCTCCACAGACTGGAAGAGTCCTTCC
TGGTGGAAAGAGGATAAGAAGCAGAGCGGGCACCCATCTCGGCAACATCGTGG
CGAGGTGGCTACCACGAGAAGTACCCACCATCTACCACTGAGAAAGAAACTG
GTGGACAGCACCAGAACGGCCACCTGCGGCTGATCTATCTGGCCCTGGCCACA
TGATCAAGTCCGGGGCACTTCTGATCGAGGGCGACCTGAACCCGACAACAG
CGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCGACTGTTGAG
GAAAACCCATCAACGCCAGCGCGTGGACGCCAAGGCCATCTGTCTGCCAG
ACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCAGCTGCCGGCGAGA
AGAAGAATGGCCTGTCGGCAACCTGATTGCCCTGAGCCTGGCCTGACCCCC
AACTTCAGAGACAACCTCGACCTGGCCAGGGATGCCAAACTGAGCTGAGCAA
GGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGCGACCAGT
ACGCCGACCTGTTCTGGCCGCAAGAACCTGTCGACGCCATCTGTGAGC
GACATCCTGAGAGTGAAACACCGAGATCACCAAGGCCCCCTGAGCGCCTCTAT
GATCAAGAGATAACGACGAGCACCCAGGACCTGACCCCTGCTGAAAGCTCTCG
TGGGCGAGCAGCTGCCCTGAGAAGTACAAGAGATTTCTCGACCAGAGCAAG
AACGGCTACGCCGGCTACATTGACGGCGAGCCAGCCAGGAAGAGTTCTACAA
GTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACCTGCTGAG
AGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTCGACAACCGCAG
CATCCCCCACCAGATCCACCTGGAGAGCTGACGCCATTCTGCGGGCGAGG
AAGATTTTACCCATTCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTG
ACCTTCCGCATCCCTACTACGTGGCCCTCTGGCAGGGAAACAGCAGATT
CGCCTGGATGACAGAAAGAGCAGGAAACCATCACCCCTGGAACCTCGAGG
AAAGTGGTGGACAAGGGCGCTTCCGCCAGAGCTTACGAGGGATGACCAAC
TTCGATAAGAACCTGCCAACGAGAAGGTGCTGCCAAGCAGCAGCTGCTGTA

-continued

CGAGTACTTCACCGTGATAACGAGCTGACCAAAAGTGAAATACGTGACCGAGG
GAATGAGAAAAGCCGCCTTCCTGAGCGCGAGCAGAAAAAGGCCATCGTGGAC
CTGCTGTTCAAGACCAACCGAAAAGTGACCGTGAAGCAGCTGAAAGAGGACTA
CTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGGTGAAAGATC
GGTTCAACGCCTCCCTGGCACATACCACGATCTGCTGAAATTATCAAGGAC
AAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGT
GACCCGTGACACTGTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCT
ATGCCAACCTGTTGACGACAAAGTGATGAGAAGCAGCTGAAGCGGGAGATAC
ACCGGCTGGGCAAGGGCTGAGCCGAAGCTGATCAACGGCATCCGGACAAGC
AGTCCGGCAAGACAATCCTGGATTCCTGAAGTCCGACGGCTCGCCAACAGA
AACTTCATGCAGCTGATCCACGACAGCCTGACCTTAAAGAGGACATCCA
GAAAGCCCAGGTGTCGGCCAGGGCGATAGCTGACGAGCACATTGCCAATC
TGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTG
GACGAGCTCGTAAAGTGATGGCCGGCACAGCCGAGAACATCGTGTGAG
AATGGCCAGAGGAAACGACCAACCCAGAAGGGACAGAAGAACAGCCGCGAG
AGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGCAGCCAGATCCTGAA
AGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAACGCTGTACCTGTACT
ACCTGCAGAATGGCGGGATATGTACGTGGACCAGGAACGGACATCAACCGG
CTGTCGCACTACGATGTGGACCATATCGTCGCTCAGAGCTTCTGAAGGACGA
CTCCATCGACAACAAGGTGCTGACCGAGCGACAAGAACCGGGCAAGAGCG
ACAACGTGCCCTCGAAGAGGTGTGAAGAAGATGAAGAACTACTGGCGGAG
CTGCTGAACGCCAAGCTGATTACCCAGAGAAAAGTCGACAATCTGACCAAGC
CGAGAGAGGGCGCTGAGCGAACTGGATAAGGGCGCTTCATCAAGAGACAG
CTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCG
GATGAAACACTAAAGTACGACGAGAATGACAAGCTGATCCGGAAAGTAAAGTGA
TCACCCCTGAAGTCCAAGCTGGTCCGATTCGGAAAGGATTCCAGTTTACA
AAAGTGCAGAGATCAACAACCTACCAACCGCCACGACGCCAACCTGACG
GTCGTGGAACCGCCCTGATCAAAAGTACCTTAAGCTGAAAGCGAGTTCGT
GTACGGCACTACAAGGTGTACGACGTGCGGAAGATGATGCCAAGAGCGAGC
AGGAAATCGGCAGGGTACCGCCAAAGTACTTCTTACAGCAACATCATGAACT
TTTCAGAACCGAGATTACCCCTGGCCAACCGGGAGATCCGGAAGCGGGCTCTG
ATCGAGACAAACGGCGAAACCGGGAGATCGTGTGGATAAGGGCGGGATT
TTGCCACCGTGGGAAAGTGTGAGCATGCCAAGCTGAATATCGTAAAAAG
ACCGAGGTGCAGACAGCGGCTTCAGCAAAGAGTCTATCCTGCCAAGAGGAA
CAGCGATAAGCTGATGCCAGAAAAGAGGACTGGACCTAAGAAGTACGGCG
GCTTCGACAGCCCCACCGTGGCTATTCTGTGCTGGTGGTGGCAAAGTGGAA
AAGGGCAAGTCCAAGAAAAGTGAAGAGTGTGAAAGAGCTGCTGGGATCACCAC
CATGGAAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAG
GCTACAAAGAAGTAAAAAGGACCTGATCATCAAGCTGCCAAGTACTCCCTG

-continued

```

TTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCA
GAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACCTCCTGTACCTGG
CCAGCCACTATGAGAAGCTGAAGGGCTCCCCGAGGATAATGAGCAGAACAG
CTGTTTGAAACAGCACAAAGCAACTACCTGGACGAGATCATCGAGCAGATCAG
CGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGTG
CCGCCTACAACAAGCACCGGGATAAGCCCCTGAGAGCAGGGCGAGAACATC
ATCCACCTGTTACCCGTACCAATCTGGAGCCCCTGCCGCTTCAGTACTTT
GACACCACCATCGACCGGAAGAGGTACACCAGCACAAAGAGGTGCTGGACGC
CACCTGATCCACCAGAGCATCACCGCCTGTACGAGACACGGATCGACCTGT
CTCAGCTGGGAGGCGACTTCTTTCTTAGCTTGACCAGCTTCTTAGTAGCA

GCAGGACGCTTTAA
(NLS-hSpCas9-NLS is highlighted in bold)

> Sequencing amplicon for EMX1 guides 1.1, 1.14, 1.17
(SEQ ID NO: 264)
CCAATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGCAACCCACAA
CCACGAGGGCAGAGTGCTGCTGCTGGCCAGGCCCTGCGTGGGCCAAG
CTGGACTCTGGCCAC

> Sequencing amplicon for EMX1 guides 1.2, 1.16
(SEQ ID NO: 265)
CGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGATTGCCACGA
AGCAGGCCAATGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGCAA
CCACAAACCCACGAG

> Sequencing amplicon for EMX1 guides 1.3, 1.13, 1.15
(SEQ ID NO: 266)
GGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCTGAGTCCGAG
CAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGATTGCCACGAAGCA
GGCCAATGGGGAGGACATCGAT

> Sequencing amplicon for EMX1 guides 1.6
(SEQ ID NO: 267)
AGAAGCTGGAGGAGGAAGGGCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAT
CACATCAACCGGTGGCGATTGCCACGAAGCAGGCCAATGGGAGGACATCGA
TGTACCTCCAATGACTAGGGTGG

> Sequencing amplicon for EMX1 guides 1.10
(SEQ ID NO: 268)
CCTCAGTCTCCCATCAGGCTCTCAGCTCAGGCTGAGTGTGAGGCCCTGGCCAG
CTGCTCTGGGGCCTCTGAGTTCTCATCTGTGCCCTCCCTGGCCAG
GTGAAGGTGTGGTTCCA

> Sequencing amplicon for EMX1 guides 1.11, 1.12
(SEQ ID NO: 269)
TCATCTGTGCCCTCCCTCCCTGGCCCAGGTGAAGGTGTGGTTCCAGAACCGG
AGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCTGAGTCCGAGCA
GAAGAAGAAGGGCTCCCATCACA

```

-continued

```
> Sequencing amplicon for EMX1 guides 1.18, 1.19
(SEQ ID NO: 270)
CTCCAATGACTAGGGTGGCAACCACAAACCCACGAGGGCAGAGTCGTGCTGGTTG

CTGCTGGCCAGGCCCTGCCTGGGCCAAGCTGGACTCTGCCACTCCCTGGC

CAGGCTTGGGAGGCCTGGAGT

> Sequencing amplicon for EMX1 guides 1.20
(SEQ ID NO: 271)
CTGCTTGCTGCTGGCCAGGCCCTGCCTGGGCCAAGCTGGACTCTGCCACTGGCCACT

CCCTGGCCAGGCTTGGGAGGCCTGGAGTCATGGCCCACAGGGCTTGAAGC

CCGGGGCGCCATTGACAGAG

> T7 promoter F primer for annealing with target strand
(SEQ ID NO: 272)
GAAATTAATACGACTCACTATAGGG

> oligo containing pUC19 target site 1 for methylation
(T7 reverse)
(SEQ ID NO: 273)
AAAAAAAGACCGACTCGGTGCCACTTTCAAGTTGATAACGGACTAGCCTTA

TTTAACTTGCTATTCTAGCTAAACAAACGACGAGCGTACACCCCTAA

TAGTGAGTCGTATTAATTTC

> oligo containing pUC19 target site 2 for methylation
(T7 reverse)
(SEQ ID NO: 274)
AAAAAAAGACCGACTCGGTGCCACTTTCAAGTTGATAACGGACTAGCCTTAT

TTTAACTTGCTATTCTAGCTAAACACGAAACAAATTAGACTGGACCTATA

GTGAGTCGTATTAATTTC
```

Example 11: Oligo-Mediated Cas9-Induced Homologous Recombination

[0305] The oligo homologous recombination test is a comparison of efficiency across different Cas9 variants and different HR template (oligo vs. plasmid).

[0306] 293 FT cells were used. SpCas9=Wildtype Cas9 and SpCas9n=nickase Cas9 (D10A). The chimeric RNA target is the same EMX1 Protospacer Target 1 as in Examples 5, 9 and 10 and oligos synthesized by IDT using PAGE purification.

[0307] FIG. 44 depicts a design of the oligo DNA used as Homologous Recombination (HR) template in this experiment. Long oligos contain 100 bp homology to the EMX1 locus and a HindIII restriction site. 293 FT cells were co-transfected with: first, a plasmid containing a chimeric RNA targeting human EMX1 locus and wild-type cas9 protein, and second, the oligo DNA as HR template. Samples are from 293 FT cells collected 96 hours post transfection with Lipofectamine 2000. All products were amplified with an EMX1 HR Primer, gel purified, followed by digestion with HindIII to detect the efficiency of integration of HR template into the human genome.

[0308] FIGS. 45 and 46 depict a comparison of HR efficiency induced by different combination of Cas9 protein and HR template. The Cas9 construct used were either wild-type Cas9 or the nickase version of Cas9 (Cas9n). The HR template used were: antisense oligo DNA (Antisense-Oligo in above figure), or sense oligo DNA (Sense-Oligo in above figure), or plasmid HR template (HR template in above figure). The sense/anti-sense definition is that the actively-transcribed strand with sequence corresponding to

the transcribed mRNA is defined as the sense strand of genome. HR Efficiency is shown as percentage of HindIII digestion band as against all genomic PCR amplified product (bottom numbers).

Example 12: Autistic Mouse

[0309] Recent large-scale sequencing initiatives have produced a large number of genes associated with disease. Discovering the genes is only the beginning in understanding what the gene does and how it leads to a diseased phenotype. Current technologies and approaches to study candidate genes are slow and laborious. The gold standards, gene targeting and genetic knockouts, require a significant investment in time and resources, both monetary and in terms of research personnel. Applicants set out to utilize the hSpCas9 nuclease to target many genes and do so with higher efficiency and lower turnaround compared to any other technology. Because of the high efficiency of hSpCas9 Applicants can do RNA injection into mouse zygotes and immediately get genome-modified animals without the need to do any preliminary gene targeting in mESCs.

[0310] Chromodomain helicase DNA binding protein 8 (CHD8) is a pivotal gene involved in early vertebrate development and morphogenesis. Mice lacking CHD8 die during embryonic development. Mutations in the CHD8 gene have been associated with autism spectrum disorder in humans. This association was made in three different papers published simultaneously in Nature. The same three studies identified a plethora of genes associated with autism spectrum disorder. Applicants' aim was to create knockout mice for the four genes that were found in all papers, Chd8,

Katnal2, Kctd13, and Scn2a. In addition, Applicants chose two other genes associated with autism spectrum disorder, schizophrenia, and ADHD, GIT1, CACNA1C, and CACNB2. And finally, as a positive control Applicants decide to target MeCP2.

[0311] For each gene Applicants designed three gRNAs that would likely knockout the gene. A knockout would occur after the hSpCas9 nuclease makes a double strand break and the error prone DNA repair pathway, non-homologous end joining, corrects the break, creating a mutation. The most likely result is a frameshift mutation that would knockout the gene. The targeting strategy involved finding proto-spacers in the exons of the gene that had a PAM sequence, NGG, and was unique in the genome. Preference was given to proto-spacers in the first exon, which would be most deleterious to the gene.

[0312] Each gRNA was validated in the mouse cell line, Neuro-N2a, by liposomal transient co-transfection with hSp-Cas9. 72 hours post-transfection genomic DNA was purified using QuickExtract DNA from Epicentre. PCR was performed to amplify the locus of interest. Subsequently the SURVEYOR Mutation Detection Kit from Transgenomics was followed. The SURVEYOR results for each gRNA and respective controls are shown in Figure A1. A positive SURVEYOR result is one large band corresponding to the genomic PCR and two smaller bands that are the product of the SURVEYOR nuclease making a double-strand break at the site of a mutation. The average cutting efficiency of each gRNA was also determined for each gRNA. The gRNA that was chosen for injection was the highest efficiency gRNA that was the most unique within the genome.

[0313] RNA (hSpCas9+gRNA RNA) was injected into the pronucleus of a zygote and later transplanted into a foster mother. Mothers were allowed to go full term and pups were sampled by tail snip 10 days postnatal. DNA was extracted and used as a template for PCR, which was then processed by SURVEYOR. Additionally, PCR products were sent for sequencing. Animals that were detected as being positive in either the SURVEYOR assay or PCR sequencing would have their genomic PCR products cloned into a pUC19 vector and sequenced to determine putative mutations from each allele.

[0314] So far, mice pups from the Chd8 targeting experiment have been fully processed up to the point of allele sequencing. The Surveyor results for 38 live pups (lanes 1-38) 1 dead pup (lane 39) and 1 wild-type pup for comparison (lane 40) are shown in Figure A2. Pups 1-19 were injected with gRNA Chd8.2 and pups 20-38 were injected with gRNA Chd8.3. Of the 38 live pups, 13 were positive for a mutation. The one dead pup also had a mutation. There was no mutation detected in the wild-type sample. Genomic PCR sequencing was consistent with the SURVEYOR assay findings.

Example 13: CRISPR Cas-Mediated Transcriptional Modulation

[0315] FIG. 67 depicts a design of the CRISPR-TF (Transcription Factor) with transcriptional activation activity. The chimeric RNA is expressed by U6 promoter, while a human-codon-optimized, double-mutant version of the Cas9 protein (hSpCas9m), operably linked to triple NLS and a VP64 functional domain is expressed by a EF1a promoter. The double mutations, D10A and H840A, renders the cas9 protein unable to introduce any cleavage but maintained its capacity to bind to target DNA when guided by the chimeric RNA.

[0316] FIG. 68 depicts transcriptional activation of the human SOX2 gene with CRISPR-TF system (Chimeric RNA and the Cas9-NLS-VP64 fusion protein). 293 FT cells were transfected with plasmids bearing two components: (1) U6-driven different chimeric RNAs targeting 20-bp sequences within or around the human SOX2 genomic locus, and (2) EF1a-driven hSpCas9m (double mutant)-NLS-VP64 fusion protein. 96 hours post transfection, 293 FT cells were harvested and the level of activation is measured by the induction of mRNA expression using a qRT-PCR assay. All expression levels are normalized against the control group (grey bar), which represents results from cells transfected with the CRISPR-TF backbone plasmid without chimeric RNA. The qRT-PCR probes used for detecting the SOX2 mRNA is Taqman Human Gene Expression Assay (Life Technologies). All experiments represents data from 3 biological replicates, n=3, error bars show s.e.m.

Example 14: NLS: Cas9 NLS

[0317] 293 FT cells were transfected with plasmid containing two components: (1) EF1a promoter driving the expression of Cas9 (wild-type human-codon-optimized Sp Cas9) with different NLS designs (2) U6 promoter driving the same chimeric RNA targeting human EMX1 locus.

[0318] Cells were collect at 72 h time point post transfection, and then extracted with 50 µl of the QuickExtract genomic DNA extraction solution following manufacturer's protocol. Target EMX1 genomic DNA were PCR amplified and then Gel-purify with 1% agarose gel. Genomic PCR product were re-anneal and subjected to the Surveyor assay following manufacturer's protocol. The genomic cleavage efficiency of different constructs were measured using SDS-PAGE on a 4-12% TBE-PAGE gel (Life Technologies), analyzed and quantified with ImageLab (Bio-rad) software, all following manufacturer's protocol.

[0319] FIG. 69 depicts a design of different Cas9 NLS constructs. All Cas9 were the human-codon-optimized version of the Sp Cas9. NLS sequences are linked to the cas9 gene at either N-terminus or C-terminus. All Cas9 variants with different NLS designs were cloned into a backbone vector containing so it is driven by EF1a promoter. On the same vector there is a chimeric RNA targeting human EMX1 locus driven by U6 promoter, together forming a two-component system.

TABLE M

Cas9 NLS Design Test Results. Quantification of genomic cleavage of different cas9-nls constructs by surveyor assay.

	Percentage Genome Cleavage as measured by Surveyor assay				
	Biological Replicate 1 (%)	Biological Replicate 2 (%)	Biological Replicate 3 (%)	Average (%)	Error (S.E.M., standard error of the mean)
Cas9 (No NLS)	2.50	3.30	2.73	2.84	0.24
Cas9 with N-term NLS	7.61	6.29	5.46	6.45	0.63
Cas9 with C-term NLS	5.75	4.86	4.70	5.10	0.33
Cas9 with Double (N-term and C-term) NLS	9.08	9.85	7.78	8.90	0.60

[0320] FIG. 70 depicts the efficiency of genomic cleavage induced by Cas9 variants bearing different NLS designs. The percentage indicate the portion of human EMX1 genomic DNA that were cleaved by each construct. All experiments are from 3 biological replicates. n=3, error indicates S.E.M.

Example 15: Engineering of Microalgae Using Cas9

Methods of Delivering Cas9

[0321] 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.

[0322] 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.

[0323] 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.

[0324] For Homologous recombination, Applicants provide an additional homology directed repair template.

[0325] Sequence for a cassette driving the expression of Cas9 under the control of beta-2 tubulin promoter, followed by the 3' UTR of Cop1.

(SEQ ID NO: 275)
TCTTTCTTGCCTATGACACTTCCAGCAAAAGTAGGGCGGGCTGGAGAC
GGCTTCCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCGAAGCTCCTTC
GGGGCTGCATGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTAAATAGCCAG
GCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAGCCATATTCAAACACCTAGA
TCACTACCACCTCTACACAGGCCACTCGAGCTGTGATCGCACTCCGTAAGGGGGC
GCTCTTCTCTCGTTTCAGTCACAACCCGCAAACATGTACCCATACGATGTTCCA
GATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCGACAAGAAGTACAG
CATCGGGCTGGACATCGGCACCAAACCTGTGGCTGGCCGTGATCACCAGCAGGAGT
ACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGCACAGCATC
AAGAAGAACCTGATCGGAGCCCTGCTGTCAGCAGCGCGAAACAGCCGAGGCCAC
CCGGCTGAAGAGAACGCCAGAAGAAGATAACCCAGACGGAAGAACGGATCTGCT
ATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCAC
AGACTGGAAGAGTCCTTCCTGGTGGAAAGAGGATAAGAAGCAGCAGCGGCCACCCAT
CTTCGGCAACATCGTGGACGAGGTGGCTTACACGAGAAAGTACCCACCATCTACC
ACCTGAGAAAGAAAATGGTGACAGCACCGACAAGGCCACCTGGCTGATCTAT
CTGGCCCTGGCCCACATGATCAAGTTCCGGGCCACTTCTGATCGAGGGCGACCTG
AACCCCCACAAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAA
CCAGCTGTTGAGGAAAACCCATCAACGCCAGCGCGTGGACGCCAAGGCCATCC

- continued

TGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCAGCTGCC
GGCGAGAAGAAGAAATGGCCTGTTCGGAACCTGATTGCCCTGAGCCTGGGCCTGAC
CCCCAACTTCAAGAGCAACTTGCACCTGGCGAGGATGCCAACTGCAGCTGAGCA
AGGACACCTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCACCAGTAC
GCCGACCTGTTCTGGCCGCAAGAACCTGTCCGACGCCATCTGCTGAGCGACATC
CTGAGAGTGAAACACCGAGATACCCAAGGCCCCCTGAGCGCCTCTATGATCAAGAG
ATACGACGAGCACCAACCAGGACCTGACCCCTGCTGAAAGCTCTCGTGCAGCAGCAG
TGCCTGAGAAGTACAAGAGATTTCTCGACCAGAGCAAGAACGGCTACGCCGC
TACATTGACGGGGAGCCAGCAGGAAGAGTTCTACAAGTTCATCAAGCCCACATC
GGAAAAGATGGACGGCACCGAGGAAGTCTGCTGTGAAGCTGAACAGAGAGGACCTGC
TGCAGAACGAGCGGACCTTCGACAACGGCAGCATCCCCACCAGATCCACCTGGGA
GAGCTGCACGCCATTCTGGCGGGCAGGAAGATTTCACCCATTCTGAAGGACAA
CCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGCCCTC
TGGCCAGGGGAAACAGCAGATTGCGCTGGATGACCAGAAAGAGCGAGGAAACCATC
ACCCCTGGAACCTCGAGGAAGTGGTGACAAGGGCGTCCGCCCCAGAGCTTCAT
CGAGCGGATGACCAACTTCGATAAGAACCTGCCAACGAGAAGGTGCTGCCAAC
ACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATAC
GTGACCGAGGGAAATGAGAAAGCCGCCTCCTGAGCGCGAGCAGAAAAAGGCCA
TCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAG
GACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCTGGAAGAT
CGGTTCAACGCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAA
GGACTTCCCTGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGTGACCC
TGACACTGTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCAC
CTGTTGACGACAAAGTGATGAGACGAGCTGAAGCGGGAGATAACCGCTGGG
CAGGCTGAGCGGAAAGCTGATCAACGGCATCCGGACAAGCAGTCCGGCAAGACA
ATCCTGGATTCCCTGAAGTCCGACGGCTCGCCAACAGAAACTTATGAGCTGATC
CACGACGACAGCCTGACCTTAAAGAGGACATCCAGAAAGCCAGGTGCTGGCCA
GGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCGGCAGCCCCGCCATTAAAGA
AGGGCATCCTGCAGACAGTGAAGGTGGACGAGCTGTAAGTGATGGGGCG
CACAAAGCCCGAGAACATCGTGTGAAATGCCAGAGAGAACAGACCCAGA
AGGGACAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGA
GCTGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGAGAAC
AGAAGCTGTACCTGACTACCTGAGAATGGCGGGATATGTAACGTGGACCAGGAA
CTGGACATCAACCGCTGTCGACTACGATGTGGACCATACTGCTCAGAGCTT
CTGAAGGACGACTCCATCGACAACAAGGTGCTGACCGAGCGACAAGAACCGG
GCAAGAGCGACAACGTGCCCTCGAAGAGGTGCTGAAGAAGATGAAGAACTACTGG
CGCGAGCTGCTGAACGCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAA
GCCCGAGAGAGCGGGCTGAGCGAACTGGATAAGGCCGGCTCATCAAGAGACAG
CTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCGGAT

- continued

```

GAACACTAAGTACGACGAGAAATGACAAGCTGATCCGGAAAGTGAAAGTGATCACCC
TGAAGTCCAAGCTGGTGTCCGATTCCGGAAGGATTCCAGTTTACAAAGTGCACG
AGATCAACAACCTACCAACCACGCCACGACGCCACCTGAACGCCGTGGAAACC
GCCCTGATCAAAAAGTACCTAACGCTGGAAAGCGAGTCGTGACGGCAGTACAA
GGTGTACGACGTGCGGAAGATGATGCCAAGAGCGAGCAGGAAATCGGCAAGGCT
ACCGCCAAGTACTTCTTCTACAGCAACATCATGAACCTTTCAAGACCGAGATTACC
CTGGCCAACGGCGAGATCCGGAAGCGGCCCTGTGATCGAGACAAACGGCGAAACCGG
GGAGATCGTGTGGATAAGGGCGGGATTGCCACCGTGCAGAAAGTGTGAGCA
TGCCCAAGTGAATATCGTGAAGAAAGCCGAGGTGCAGACAGGCGCTTCAGCAAA
GAGTCTATCCTGCCAAGAGAACAGCGATAAGCTGATGCCAGAAAGAAGGACTG
GGACCTAAGAAGTACGGCGCTTCGACAGCCCCACCGTGGCTATTCTGTGCTGGT
GGTGGCAAAGTGGAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTG
CTGGGGATCACCACATGGAAAGAACAGCAGCTCGAGAAGAAATCCCATCGACTTCT
GGAAGCCAAGGGCTACAAAAGAAGTGAAGAAAGGACCTGATCATCAAGCTGCCAAGT
ACTCCCTGTTCGAGCTGGAAACCGGCCCTGCCCTCCAATATGTGAACTTCTGTACCTG
CTGCAGAAGGGAAACGAACCTGGCCCTGCCCTGAGAATGCTGCCCTGCCGCGAA
GCCAGCCACTATGAGAAGCTGAAGGGCTCCCCGAGGATAATGAGCAGAAACAGCT
GTTTGTGGAACAGCACAAGCAACTACCTGGACGAGATCATCGAGCAGATCAGCGAGT
TCTCCAAGAGAGTGTGATCTGGCGACGCTAATCTGGACAAAGTGTGCTGCCCTACA
ACAAGCACCGGATAAGCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTT
ACCTGACCAATCTGGAGCCCTGCCCTTAAGTACTTGTGACACCACATCGAC
CGGAAGAGGTACACCAGCACCAGAGGTGCTGGACGCCACCTGATCCACCAAGAG
CATCACCGCCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGAGGCGACAGCC
CCAAGAAGAAGAGAAAGGTGGAGGCCAGCTAAGGATCCGGCAAGACTGGCCCCGC
TTGGCAACGCAACAGTGAGCCCCCTCCAGTGTGTTGGGATGTGACTATGTATTG
GTGTGTTGGCAACGGGCAACCCGAACAGATTGATACCCGCCCTGGCATTTCTGT
CAGAATGTAACGTCAGTTGATGGTACT

```

[0326] Sequence for a cassette driving the expression of T7 polymerase under the control of beta-2 tubulin promoter, followed by the 3' UTR of Cop1:

```

(SEQ ID NO: 276)
TCTTTCTTGCCTATGACACTTCCAGCAAAAGGTAGG
GGGGCTCGAGACGGCTCCCGGCCGCTGCATGCAAC
ACCGATGATGCTCGACCCCCCGAACGCTCCTCGGGG
CTGCATGGCGCTCCGATGCCGCTCCAGGGCGAGCGC
TGTAAATAGCCAGGCCCCGATTGCAAAGACATTA
TAGCGAGCTACCAAGCCATATTCAAAACACCTAGATC
ACTACCACTCTACACAGGCCACTCGAGCTTGTGATC
GCACCTCGCTAAGGGGGCCCTTCCCTTCGTTTC

```

- continued

```

AGTCACAACCGCAAAatgcctaagaagaagaggaa
ggtaaacacgattaacatcgctaagaacgacttctct
gacatcgaaactggctgtatccgttcaacactctgg
ctgaccattacggtgagcgtttagctcgcaacagt
ggcccttgagcatgagttacgagatgggtgaagca
cgcttccgcaagatgtttgagcgtcaacttaagctg
gtgaggttgcggataacgctggccaagcctctcat
cactaccctactccctaagatgttgcacgcaccaac
gactggttgaggaagtgaaagctaaagcgcggcaagc
ccccacagcctccagttcctgcaagaaatcaagcc

```

-continued

```

ggaagccgtagcgtacatcaccattaagaccactctg
gcttgctaaccagtgtgacaatacaaccgttcagg
ctgttagcaagcgcaatcggtcgcccattgaggacga
ggctcgcttcggctgtatccgtgaccttgcggtaag
caactcaagaaaaacgttgaggaaacaactcaacaagc
gcgttagggcacgttacaagaaaagcattatgcgtt
tgtcgaggctgacatgtctctaagggtctactcggt
ggcgaggcgtggcttcgtggcataaggaagactcta
ttcatgttaggatgtcgcatcgagatgttcattgt
gtcaaccggaatggtagttacaccggccaaatgtc
ggcgttaggtcaagactctgagactatcgaaactcg
cacctgaatacgctgaggctatcgcaaccgtgcagg
tgcgtggctggcatctccatgttccaaccttgc
gtatgtccctctaagccgtggactggcattactggc
gtggctattgggtaacggctgtgtcccttggcgtt
ggtgctactcacagaagaaagactgtgcgtac
gaagacgttacatgcgtgagggttacaaagcgat
acattgcgaaaacaccgcattggaaatcaacaagaa
agtccttagggcgccaaacgttaatccacaaatggaa
cattgtccggcgaggacatccctgcgttgcgttgc
aagaactcccgatgaaaccggaaagacatcgatggaa
tcctgaggcttacccgtggaaacgtgtgtccgttgc
gctgttacccgtggaaacggctcgcaagtctcgcc
gtatcagccttgcgttgcgttgcgttgcgttgc
gtttgttaaccataaggccatctgggtttcccttacaa
atggactggcgccgtgttacccgtgtgtcaatgt
tcaaccggcaaggtaacgatgtaccaaggactgt
tacgcgtggcgaaaggtaaccaatcggttgcgttgc
tactactggctgaaaatccacccgttgcgttgcgtt
gtgtcgacaagggttcgttccctgtgggtggctgt
cattgaggaaaaccacccgttgcgttgcgttgcgtt
aagtctccactgggaaacacttgggtggctgtggcgtt
attctccgttgcgttccctgtgggtggctgtggcgtt
cgctgggttacccgttgcgttgcgttgcgttgcgtt
tccctccgtggcggttgcgttgcgttgcgttgcgtt
tccagcacttcccgatgtccgttgcgttgcgttgcgtt
tggtcgcgcgttaacttgcgttgcgttgcgttgcgtt
caggacatctacccgttgcgttgcgttgcgttgcgtt
agattctacaaggccatcggttgcgttgcgttgcgtt

```

-continued

```

cgaagtagtaccgtgaccgttgcgttgcgttgcgtt
atctctgagaaatgtcaagctggactaaggcactgg
ctggtaatggctggcttacgggttactcgcaatgt
gactaagegttgcgttgcgttgcgttgcgttgcgtt
aaagagttcggttccgttacaatgtgtggaaagata
ccattcagccagcatgttgcgttgcgttgcgttgcgtt
gttcaactcagccgttgcgttgcgttgcgttgcgtt
aagctgatttggatctgtgtgggttgcgttgcgttgcgtt
ctgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
taagctgtggctgtgtgggttgcgttgcgttgcgttgcgtt
ggagagattttcgcaagcggttgcgttgcgttgcgttgcgtt
taactcctgtgtgttccctgtgtggcaggaaataca
gaagcctattcagacgcgttgcgttgcgttgcgttgcgtt
ggtagtcccgatccatgttgcgttgcgttgcgttgcgtt
aagatagcggatgttgcgttgcgttgcgttgcgttgcgtt
tatcgctccatgttgcgttgcgttgcgttgcgttgcgtt
caccctcgtaagactgtgtgtggcacacgagaatgtt
acggaaatcgaaatctttgcactgttgcgttgcgttgcgtt
cggtacgatcccgatccatgttgcgttgcgttgcgttgcgtt
gcagtagcgttgcgttgcgttgcgttgcgttgcgttgcgtt
gtgtatgtactggctgttgcgttgcgttgcgttgcgtt
ccagttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
cttccggctaaaggtaacttgcgttgcgttgcgttgcgtt
tagagtcggacttgcgttgcgttgcgttgcgttgcgtt
GACTGGCCCCCGCTTGGCAACGCAACAGTGAGCCCCCTC
CCTAGTGTGTTGGGATGTGACTATGTATTGTGTG
TTGGCCAACGGGTCACCCGAACAGATTGATAACCGC
CTTGGCATTTCTGTCAAGATGTAACGTCAGTTGATG
GTACT

```

[0327] Sequence of guide RNA driven by the T7 promoter (T7 promoter, Ns represent targeting sequence):

(SEQ ID NO: 277)
gaaat**TAATACGACTCACTAT**NNNNNNNNNNNNNNNNNN
gttttagagcttGAAAtagcaagttaaaataaggctgttgcgtt
ttatcaacttgcgttgcgttgcgttgcgttgcgttgcgtt

Gene Delivery:

[0328] *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.

[0329] Also, Applicants generate a line of *Chlamydomonas reinhardtii* that expresses Cas9 constitutively. This can be done by using pChlamy1 (linearized using PvU1) 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: 278)
 TGC GGTTATTCACACCGCATCAGGTGGCACTTTGGAAATGTGCGGG
 AACCCCTATTGTTTCTAAATACATTCAAATATGTATCCGCTCATGAGATT
 ATCAAAAAGGATCTCACCTAGATCCTTAAATTAAAAATGAAGTTAAATCAAT
 CTAAAGTATAATGAGTAAACTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGC
 ACCTATCTCAGCGATCTGCTATTGCTCATCCATAGTGGCCTGACTCCCCGTCGT
 GTAGATAACTACGATACGGGGCTTACCATCTGGCCCCAGTGTGCAATGATACC
 GCGAGACCACGCTCACGGCTCCAGATTATCAGCAATAAACAGCCAGCCGGAAG
 GGCGAGCGCAGAAGTGGTCTGCAACTTATCCGCTCCATCCAGTCTATTAAATTG
 TTGCGGGAAAGCTAGAGTAAGTAGTTGCGCAGTTAATAGTTGCGAACGTTGTC
 CATTGCTACAGGCATGGTGTACGCTCGTGTGGTATGGCTTCACTCAGCTCC
 GGTTCCAACGATCAAGGGAGTTACATGATCCCCATGGTGTGCAAAAAGCGGTT
 AGCTCCTTCGGTCCTCGATGTTGTCAGAAGTAAGTGGCCGAGTGTATCACTC
 ATGGTTATGGCAGCACTGATAATTCTCTACTGTCATGCCATCGTAAGATGCTTT
 CTGTGACTGGTGGACTCAACCAAGTCATTCTGAGAATAGTGTATGGCCGACCGA
 GTTGCTCTGGCCGGCGTCAATACGGGATAATACCGCAGCACATAGCAGAACTTAA
 AAGTGCTCATGGAAACGTTCTGGGGCGAAACTCTCAAGGATCTTACCGC
 TGTTGAGATCCAGTTCGATGTAACCCACTCGTCGACCCAACTGATCTCAGCATCTT
 TTACTTCAACCAGCGTTCTGGGTGAGAAAAACAGGAAGGCAAAATGCCGAAAAA
 AGGGATAAGGGCGACACGGGAAATGTTGAATACTCATACTCTCCTTTCAATATT
 ATTGAAGCATTATCAGGGTATTGTCATGACCAAAATCCCTAACGTGAGTTTC
 GTTCCACTGAGCGTCAGACCCCGTAGAAAGATCAAAGGATCTTGTGAGATCCTT
 TTTTCTGCGCTAATCTGCTTGCAAAACAAAAAACCCACCGCTACCAGCGTGGT
 TTGTTGCGCGATCAAGAGCTACCAACTTTTCCGAAGGTAACGTTGAGATCCTT
 AGCGCAGATACCAAAACTGTTCTAGTGTAGCCGTAGTTAGGCCACCACTCAA
 GAAACTCTGTTAGCACCGCTACATACCTCGCTGCTAATCTGTTACCGATGGCTGTT
 GCCAGTGGCGATAAGTCGTCTTACCGGGTGGACTCAAGACGATAGTTACCGGAT
 AAGGGCGAGCGCGTGGCTGAACGGGGGGTCTGCAACAGCCAGCTGGAGCG
 AACGACCTACACCGAACTGAGATACTACAGCGTGAGCTATGAGAAAGGCCACGC
 TTCCCGAAGGGAGAAAGGGGACAGGTATCCGGTAAGCGGCAGGGCGAACAGG
 AGAGCGCACGAGGGAGCTCCAGGGGAAACGCCCTGGTATCTTATAGTCTGTC
 GTTTCGCCACCTCTGACTTGAGCGTCGATTTGTGATGCTCGTCAAGGGGGCGGA
 GCCTATGGAAAAACGCCAGCAACCGCCCTTTACGGTTCTGGCTTTGCTGGC
 CTTTGCTCACATGTTCTCTGCTTACCCCTGATTCTGTTGAGATGCTGCTGG
 CGCCTTGAGTGAGCTGATAACCGCTGCCAGCGAACGACCGAGCGCAGCGAGTC
 AGTGAGCGAGGAAGCGGTGCGCTGAGGTTGACATGATTGGTGCATGTTGATGA

-continued

AGCTACAGGACTGATTGGCGGGCTATGAGGGCGGGGAAGCTCTGGAAGGGCCGC
GATGGGGCGCGCGCGTCCAGAAGGCGCCATACGGCCCGCTGGCGCACCCATCCG
GTATAAAAGCCCGCACCCGAACGGTGACCTCCTTCAGCGACAAACGAGCAC
TTATACATACGCGACTATTCTCCGCTATACATAACCACCTAGCTAGCTTAAGATCC
CATCAAGCTTGATGCCGGCGCGAGAAGGAGCGCAGCAAACCCAGGATGATGT
TTGATGGGGTATTGAGCACTTGCAACCCCTATCCGAAGCCCCCTGGCCACAAAG
GCTAGGCACCATGCAAGCAGTCGATGCAGCCCTGGAGCGGTGCCCTCTGAT
AAACCGGCAGGGGCCTATGTTTTACTTTTACAAGAGAAGTCACTAACATC
TTAAAATGGCCAGGTGAGTCGACGAGCAAGCCGGGATCAGGCAGCGTGCCTG
AGATTTGACTTGCAACGCCGCATTGTGTCGACGAAGGCTTTGGCTCCTTGCT
GTCTCAAGCAGCATCTAACCTCGCTGCCGTTCCATTGAGGAGATTGAGGTA
CCATGTACCCATACGATGTTCCAGATTACGCTTCGCGAAGAAAAAGCGCAAGGTC
GAAGCGTCCGACAAGAAGTACAGCATGGCCTGGACATCGCACCAACTCTGTGG
CTGGCCGCTGATACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGG
GCAACACCGACCGGACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTGAC
AGCGGCAGAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CCAGACGGAAGAACCGGATCTGCTATCGCAAGAGATCTTCAGCAACGAGATGGC
AAAGGTGGACGACAGCTCTTCCACAGACTGGAAGAGTCCTCTGGTGAAGAGGA
TAAGAACGACGAGCGGACCCATCTCGGCAACATCGTGACGAGGTGGCTTAC
ACGAGAAAGTACCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGAC
AAGGCGACCTCGGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTGGGGC
CACTTCCCTGATCGAGGGCGACCTGAACCCGACAACAGCGACGTGGACAAGCTGTT
CATCCAGCTGGTGCAGCCTACAACCCAGCTGTCGAGGAAACCCCATCAACGCCA
GCGGCGTGGACGCCAAGGCATCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTG
GAAAATGATGCCAGCTGCCGGGAGAAGAAGAATGGCTGTTGGCAACCT
GATTGCCCTGAGCCTGGCCTGACCCCAACTTCAAGAGCAACTCGACCTGGCA
GGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGC
TGGCCCAAGATCGGCACCGACTGCCGACCTGTTCTGGCCGCAAGAACCTGTCCG
ACGCCATCCTGAGCGACATCCTGAGAGTGAACACCGAGATCAGCAAGGCCCC
CTGAGCGCCCTATGATCAAGAGATAACGACGAGCACCACCGAGCTGACCCCTGCT
GAAAGCTCTCGTGGCAGCAGCTGCTGAGAAGTACAAGAGATTTCTCGACC
AGAGCAAGAACGGCTACGCCGCTACATTGACGGCGGAGCCAGGAAGAGTTC
TACAAGTTCATCAAGCCCATCTGGAAAAGATGGACGGCACCGAGGAACCTGCTCG
GAAGCTGAACAGAGAGGACCTGCTGGGAAGCAGCGGACCTTCGACAACGGCAGC
ATCCCCCACCAGATCCACCTGGGAGAGCTGACGCCATTCTGCCGGCGCAGGAAGA
TTTTTACCCATTCTGAAGGACAACCGGAAAAGATCGAGAAGATCCTGACCTCCG
CATCCCCCTACTACGTTGGCCCTCTGGCAGGGAAACAGCAGATTGCGCTGGATGA
CCAGAAAGAGCGAGGAAACCATCACCCCTGGAACCTCGAGGAAGTGGTGGACAA
GGCGCTTCCGCCAGAGCTTACGAGCGGATGACCAACTTCGATAAGAACCTGC

-continued

CCACAGAGAAGGTGCTGCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGAT
AACGAGCTGACCAAAGTGAATACGTGACCGAGGGATGAGAAAGCCCGCCTCCT
GAGCGGCAGCAGAAAAAGGCCATCGGGACCTGCTGTTCAAGACCAACCGGAAA
GTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTCGACTC
CGTGGAAATCTCCGGCGTGGAGATCGGTTCAACGCCTCCCTGGGACATACCACG
ATCTGCTGAAATTATCAAGGACAAGGACTTCTGGACAATGAGGAAACGAGGAC
ATTCTGGAAGATATCGTGTGACCTGACACTGTTGAGGACAGAGAGATGATCGA
GGAACGGCTGAAAACCTATGCCACCTGTTGACGACAAAGTGTGATGAGCAGCTGA
AGCAGCGAGATAACCGGCTGGGCAGGCTGAGCCGGAAAGCTGATCAACGGCATC
CGGGACAAGCAGTCCGCAAGACAATCTGGATTCTGAAAGTCCGACGGCTTCGC
CAACAGAAACTTCAATGCAGCTGATCCACGACAGCCTGACCTTAAAGAGGACA
TCCAGAAAGCCCAGGTGTCCGCCAGGGCAGAGCTGCACGAGCACATTCCAAT
CTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGAGACAGTGAAGGTGGTGG
CGAGCTGTTGAAAGTGTGGCCGGACAAGCCGAGAACATCGTGTGAAATGG
CCAGAGAGAACAGACCAACCCAGAAGGGACAGAAGAACAGCCGAGAGAACATGG
GCGGATCGAAGAGGGCATCAAAGAGCTGGCAGCCAGATCCTGAAAGAACACCC
GTGGAAACACCCAGCTGCAGAACAGAGCTGTACCTGTACTACCTGAGAATGG
GCGGGATATGTACGTGGACCAGGAACGGACATCAACCGGCTGTCCGACTACGATG
TGGACCATATCGTGCCTCAGAGCTTCTGAAGGACACTCCATGACAACAAGGTGC
TGACCCAGAACGGACAAGAACCGGGCAAGAGCGACACGTGCCCTCGAAGAGGT
CGTGAAGAAGATGAAGAACTACTGGCGCAGCTGCTGAACGCCAAGCTGATTACCC
AGAGAAAGTTCGACAATCTGACCAAGGCCAGAGAGGGCGGCTGAGCGAACTGGAT
AAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGT
GGCACAGATCCTGGACTCCCGATGAACACTAAGTACGACGAGAACAGCTGA
TCCGGGAAGTGAAGTGTACCGTCAAGCTGGTGGGATTTCCGGAAAG
GATTTCAGTTTACAAAGTGCAGGAGATCAACAACTACCAACGCCACGACGCC
TACCTGAACGCCGTGTGGAACCGCCCTGATCAAAAGTACCTAAGCTGGAAAG
CGAGTTCGTGTACGGGACTACAAGGTGTACGACGTGCGGAAGATGATGCCAAGA
GCGAGCAGGAATCGGCAAGGCTACGCCAAGTACTTCTTACAGCAACATCATG
AACTTTTCAAGACCGAGATTACCCCTGCCAACGGGAGATCCGGAGCGGCTCT
GATCGAGACAAACGGCGAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTG
CCACCGTGCAGAACAGTGTGAGCATGCCCAAGTGAATATCGTAAAAAGACCGAG
GTGCAGACAGGGCTTCAGCAAAGAGCTATCCTGCCAAGAGAACAGCGATAA
GCTGATGCCAGAAAGAAGGACTGGGACCTAAGAAGTACGGCGCTCGACAGCC
CCACCGTGGCCTATTCTGTGCTGGTGGCCAAAGTGGAAAGGGCAAGTCCAAG
AAACTGAAGAGTGTGAAAGAGCTGCTGGGATCACCACATGAAAGAACAGCAGCTT
CGAGAAGAACCCATCGACTTCTGGAAGCCAAGGGTACAAAGAACAGTAAAAAGG
ACCTGATCATCAAGCTGCCATAAGTACTCCCTGTTGAGCTGAGAACCGGCCAG
AAGATGCTGGCCTCTGCCGGCAACTGCAGAAGGGAAACGAACGGCCCTGCCCTC

-continued

CAAATATGTGAACCTCCTGTACCTGGCAGCCACTATGAGAAGGCTGAAGGGCTCCCC
CGAGGATAATGAGCAGAACAGCTGTTGTGAAACAGCACAAGCACTACCTGGACG
AGATCATCGAGCAGATCAGCAGTCTCCAAGAGAGTGATCCTGGCCGACGCTAAT
CTGGACAAAGTGTCTGCCCTACAACAAGCACGGGATAAGCCATCAGAGAGCA
GGCGGAGAATATCATCCACCTGTTACCCCTGACCAATCTGGAGCCCTGCCGCTT
CAAGTACTTTGACACCACATCGACCGGAAGAGGTACACCAGCACAAAGAGTGC
TGGACGCCACCTGATCCACAGAGCATCACCCGCTGTACGAGACACGGATCGAC
CTGTCAGCTGGGAGGCAGACGCCAAGAAGAAGAGAAAGGTGGAGGCCAGCT
AACATATGATTGAAATGTCTTCTTGCGCTATGACACTTCAGCAAAGGTAGGGCG
GGCTGCGAGACGGCTTCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCG
AAAGCTCTCGGGGCTGCATGGCGCTCCGATGCCGCTCCAGGGGAGCGCTGTTA
AATAGCCAGGCCCGATTGCAAAGACATTATAGCAGGCTACCAAAGCCATATTCA
AACACCTAGATCACTACCAACTCTACACAGGCCACTCGAGCTGTGATCGCACTCCG
CTAAGGGGCGCCCTTCCCTTCGTTTAGTCAGTCACAACCCGCAAACATGACACAAGA
ATCCCTGTTACTCTCGACCGTATTGATTGGATGATTCCCTACGCGAGCCTGCGGAA
CGACCAAGGAATTCTGGAGGTGAGTCGACGAGCAAGCCGGGGATCAGGAGCGT
GCTTCGAGATTGACTTGCACGCCGATTGTTGACGAGGCTTTGGCTCCCT
GTCGCTGCTCAAGCAGCATCTAACCTGCGTCGCCGTTCCATTGCAAGCCGCTGG
CCCGCCGAGCCCTGGAGGAGCTGGGCTGCCGGTGCCGCCGGTGCTGCCGGGCC
GGCGAGACACCAACCCCGTACTGGCGGAGGCCCTCGCGTGGAGTGGAGGGCTACCGG
CGCCGAGGACTGGTGGGCTGGAGAGGCCCTCGCGTGGAGTGGAGGGCTACCGG
TCCTGGCGGACGCCCGGTGCCGGTGCCGGCTCCCTCGGCCGGCAGCTGCCG
CCGGCACGGAGCCTGCCGTGCCCTACCTGGTGTGAGGCCGGATGACCGGACCC
ACCTGGCGGTCCCGATGGACGGACGACCGAACCGGAACGCCGTGCTGCCCTGGC
CCGCGAACCTGGCGGGTGCTGGCGCTGACAGGGTGCCGCTGACCGGGAAACA
CCGTGCTACCCCCCATTCCGAGGTCTTCCCGAACCTGCTGCCGGAACGCCGGCG
CGACCGCTGAGGACCAACCGCGGGTGGGCTACCTCTGCCCGGTGCTGGACCGC
CTGGAGGACTGGCTGCCGGACGGTGGACACGCTGCTGCCGCCGGAACCCGGTT
CGTCCACGGGACCTGACGGACCAACATCTCGTGGACCTGGCGGACCGAGG
TCACCGGGATCGTCGACTTCACCGACGTCTATGCGGGAGACTCCGCTACAGCCTGG
TGCACACTGCATCTCAACGCCCTCCGGGGCGACCGGAGATCTGGCGCTGCTCG
ACGGGGCGAGTGGAAAGCGGACCGAGGACTTCGCCCGCAACTGCTGCCCTCACC
TTCCTGCACGACTTCGAGGTGTTGAGGAGACCCGCTGGATCTCCGGCTTCAAC
GATCCGGAGGAACGGCGAGTTCTGGGGCGCCGGACACGCCCGCG
CTGATAAGGATCGGCAAGACTGGCCCGCTGGCAACGCAACAGTGAGCCCTC
CTAGTGTGTTGGGATGTGACTATGTATTGCTGTTGGCAACGGTCAACCCGA
ACAGATTGATAACCCGCTTGGCATTCTGTCAGAATGTAACGTCAGTTGATGGTACT

[0330] For all modified *Chlamydomonas reinhardtii* cells, Applicants used PCR, SURVEYOR nuclease assay, and DNA sequencing to verify successful modification.

Example 16: Use of Cas9 as a Transcriptional Repressor in Bacteria

[0331] The ability to artificially control transcription is essential both to the study of gene function and to the construction of synthetic gene networks with desired properties. Applicants describe here the use of the RNA-guided Cas9 protein as a programmable transcriptional repressor.

[0332] Applicants have previously demonstrated how the Cas9 protein of *Streptococcus pyogenes* SF370 can be used to direct genome editing in *Streptococcus pneumoniae*. In this study Applicants engineered the crR6Rk strain containing a minimal CRISPR system, consisting of cas9, the tracrRNA and a repeat. The D10A-H840 mutations were introduced into cas9 in this strain, giving strain crR6Rk**. Four spacers targeting different positions of the bgaA β -galactosidase gene promoter were cloned in the CRISPR array carried by the previously described pDB98 plasmid. Applicants observed a X to Y fold reduction in β -galactosidase activity depending on the targeted position, demonstrating the potential of Cas9 as a programmable repressor (FIG. 73).

[0333] To achieve Cas9** repression in *Escherichia coli* a green fluorescence protein (GFP) reporter plasmid (pDB127) was constructed to express the gfpmut2 gene from a constitutive promoter. The promoter was designed to carry several NPP PAMs on both strands, to measure the effect of Cas9** binding at various positions. Applicants introduced the D10A-H840 mutations into pCas9, a plasmid described carrying the tracrRNA, cas9 and a minimal CRISPR array designed for the easy cloning of new spacers. Twenty-two different spacers were designed to target different regions of the gfpmut2 promoter and open reading frame. An approximately 20-fold reduction of fluorescence was observed upon targeting regions overlapping or adjacent to the -35 and -10 promoter elements and to the Shine-Dalgarno sequence. Targets on both strands showed similar repression levels. These results suggest that the binding of Cas9** to any position of the promoter region prevents transcription initiation, presumably through steric inhibition of RNAP binding.

[0334] To determine whether Cas9** could prevent transcription elongation, Applicants directed it to the reading frame of gfpmut2. A reduction in fluorescence was observed both when the coding and non-coding strands were targeted, suggesting that Cas9 binding is actually strong enough to represent an obstacle to the running RNAP. However, while a 40% reduction in expression was observed when the coding strand was the target, a 20-fold reduction was observed for the non-coding strand (FIG. 21b, compare T9, T10 and T11 to B9, B10 and B11). To directly determine the effects of Cas9** binding on transcription, Applicants extracted RNA from strains carrying either the T5, T10, B10 or a control construct that does not target pDB127 and subjected it to Northern blot analysis using either a probe binding before (B477) or after (B510) the B10 and T10 target sites. Consistent with Applicants' fluorescence methods, no gfpmut2 transcription was detected when Cas9** was directed to the promoter region (T5 target) and a transcription was observed after the targeting of the T10 region. Interestingly, a smaller transcript was observed with the B477 probe. This band corresponds to the expected size

of a transcript that would be interrupted by Cas9**, and is a direct indication of a transcriptional termination caused by dgRNA::Cas9** binding to the coding strand. Surprisingly, Applicants detected no transcript when the non-coding strand was targeted (B10). Since Cas9** binding to the B10 region is unlikely to interfere with transcription initiation, this result suggests that the mRNA was degraded. DgRNA::Cas9 was shown to bind ssRNA in vitro. Applicants speculate that binding may trigger degradation of the mRNA by host nucleases. Indeed, ribosome stalling can induce cleavage on the translated mRNA in *E. coli*.

[0335] Some applications require a precise tuning gene expression rather than its complete repression. Applicants sought to achieve intermediate repression levels through the introduction of mismatches that will weaken the crRNA/target interactions. Applicants created a series of spacers based on the B1, T5 and B10 constructs with increasing numbers of mutations in the 5' end of the crRNA. Up to 8 mutations in B1 and T5 did not affect the repression level, and a progressive increased in fluorescence was observed for additional mutations.

[0336] The observed repression with only an 8 nt match between the crRNA and its target raises the question of off-targeting effects of the use of Cas9** as a transcriptional regulator. Since a good PAM (NGG) is also required for Cas9 binding, the number of nucleotides to match to obtain some level of repression is 10. A 10 nt match occurs randomly once every ~1 Mbp, and such sites are thus likely to be found even in small bacterial genomes. However, to effectively repress transcription, such site needs to be in the promoter region of gene, which makes off-targeting much less likely. Applicants also showed that gene expression can be affected if the non-coding strand of a gene is targeted. For this to happen, a random target would have to be in the right orientation, but such events relatively more likely to happen. As a matter of fact, during the course of this study Applicants were unable to construct one of the designed spacer on pCas9**. Applicants later found this spacer showed a 12 bp match next to a good PAM in the essential murC gene. Such off-targeting could easily be avoided by a systematic blast of the designed spacers.

[0337] Aspects of the invention are further described in the following numbered paragraphs:

[0338] 1. A vector system comprising one or more vectors, wherein the system comprises

[0339] a. a first regulatory element operably linked to a traer mate sequence and one or more insertion sites for inserting a guide sequence upstream of the traer mate sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the traer mate sequence that is hybridized to the traer sequence; and

[0340] b. a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence;

[0341] wherein components (a) and (b) are located on the same or different vectors of the system.

[0342] 2. The vector system of paragraph 1, wherein component (a) further comprises the traer sequence downstream of the traer mate sequence under the control of the first regulatory element.

- [0343] 3. The vector system of paragraph 1, wherein component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell.
- [0344] 4. The vector system of paragraph 1, wherein the system comprises the traer sequence under the control of a third regulatory element.
- [0345] 5. The vector system of paragraph 1, wherein the traer sequence exhibits at least 50% of sequence complementarity along the length of the traer mate sequence when optimally aligned.
- [0346] 6. The vector system of paragraph 1, wherein the CRISPR enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell.
- [0347] 7. The vector system of paragraph 1, wherein the CRISPR enzyme is a type II CRISPR system enzyme.
- [0348] 8. The vector system of paragraph 1, wherein the CRISPR enzyme is a Cas9 enzyme.
- [0349] 9. The vector system of paragraph 1, wherein the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell.
- [0350] 10. The vector system of paragraph 1, wherein the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence.
- [0351] 11. The vector system of paragraph 1, wherein the CRISPR enzyme lacks DNA strand cleavage activity.
- [0352] 12. The vector system of paragraph 1, wherein the first regulatory element is a polymerase III promoter.
- [0353] 13. The vector system of paragraph 1, wherein the second regulatory element is a polymerase II promoter.
- [0354] 14. The vector system of paragraph 4, wherein the third regulatory element is a polymerase III promoter.
- [0355] 15. The vector system of paragraph 1, wherein the guide sequence is at least 15 nucleotides in length.
- [0356] 16. The vector system of paragraph 1, wherein fewer than 50% of the nucleotides of the guide sequence participate in self-complementary base-pairing when optimally folded.
- [0357] 17. A vector comprising a regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising one or more nuclear localization sequences, wherein said regulatory element drives transcription of the CRISPR enzyme in a eukaryotic cell such that said CRISPR enzyme accumulates in a detectable amount in the nucleus of the eukaryotic cell.
- [0358] 18. The vector of paragraph 17, wherein said regulatory element is a polymerase II promoter.
- [0359] 19. The vector of paragraph 17, wherein said CRISPR enzyme is a type IICRISPR system enzyme.
- [0360] 20. The vector of paragraph 17, wherein said CRISPR enzyme is a Cas9 enzyme.
- [0361] 21. The vector of paragraph 17, wherein said CRISPR enzyme lacks the ability to cleave one or more strands of a target sequence to which it binds.
- [0362] 22. A CRISPR enzyme comprising one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell.
- [0363] 23. The CRISPR enzyme of paragraph 22, wherein said CRISPR enzyme is a type IICRISPR system enzyme.

- [0364] 24. The CRISPR enzyme of paragraph 22, wherein said CRISPR enzyme is a Cas9 enzyme.
- [0365] 25. The CRISPR enzyme of paragraph 22, wherein said CRISPR enzyme lacks the ability to cleave one or more strands of a target sequence to which it binds.
- [0366] 26. A eukaryotic host cell comprising:
- [0367] a. a first regulatory element operably linked to a traer mate sequence and one or more insertion sites for inserting a guide sequence upstream of the traer mate sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the traer mate sequence that is hybridized to the traer sequence; and/or
- [0368] b. a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence.
- [0369] 27. The eukaryotic host cell of paragraph 26, wherein said host cell comprises components (a) and (b).
- [0370] 28. The eukaryotic host cell of paragraph 26, wherein component (a), component (b), or components (a) and (b) are stably integrated into a genome of the host eukaryotic cell.
- [0371] 29. The eukaryotic host cell of paragraph 26, wherein component (a) further comprises the traer sequence downstream of the traer mate sequence under the control of the first regulatory element.
- [0372] 30. The eukaryotic host cell of paragraph 26, wherein component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell.
- [0373] 31. The eukaryotic host cell of paragraph 26, further comprising a third regulatory element operably linked to said traer sequence.
- [0374] 32. The eukaryotic host cell of paragraph 26, wherein the traer sequence exhibits at least 50% of sequence complementarity along the length of the traer mate sequence when optimally aligned.
- [0375] 33. The eukaryotic host cell of paragraph 26, wherein the CRISPR enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable mount in the nucleus of a eukaryotic cell.
- [0376] 34. The eukaryotic host cell of paragraph 26, wherein the CRISPR enzyme is a type II CRISPR system enzyme.
- [0377] 35. The eukaryotic host cell of paragraph 26, wherein the CRISPR enzyme is a Cas9 enzyme.
- [0378] 36. The eukaryotic host cell of paragraph 26, wherein the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell.
- [0379] 37. The eukaryotic host cell of paragraph 26, wherein the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence.
- [0380] 38. The eukaryotic host cell of paragraph 26, wherein the CRISPR enzyme lacks DNA strand cleavage activity.

- [0381] 39. The eukaryotic host cell of paragraph 26, wherein the first regulatory element is a polymerase III promoter.
- [0382] 40. The eukaryotic host cell of paragraph 26, wherein the second regulatory element is a polymerase II promoter.
- [0383] 41. The eukaryotic host cell of paragraph 31, wherein the third regulatory element is a polymerase III promoter.
- [0384] 42. The eukaryotic host cell of paragraph 26, wherein the guide sequence is at least 15 nucleotides in length.
- [0385] 43. The eukaryotic host cell of paragraph 26, wherein fewer than 50% of the nucleotides of the guide sequence participate in self-complementary base-pairing when optimally folded.
- [0386] 44. A non-human animal comprising a eukaryotic host cell of any one of paragraphs 26-43.
- [0387] 45. A kit comprising a vector system and instructions for using said kit, the vector system comprising:
- [0388] a. a first regulatory element operably linked to a traer mate sequence and one or more insertion sites for inserting a guide sequence upstream of the traer mate sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the traer mate sequence that is hybridized to the traer sequence; and/or
- [0389] b. a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence.
- [0390] 46. The kit of paragraph 45, wherein said kit comprises components (a) and (b) located on the same or different vectors of the system.
- [0391] 47. The kit of paragraph 45, wherein component (a) further comprises the traer sequence downstream of the traer mate sequence under the control of the first regulatory element.
- [0392] 48. The kit of paragraph 45, wherein component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell.
- [0393] 49. The kit of paragraph 45, wherein the system comprises the traer sequence under the control of a third regulatory element.
- [0394] 50. The kit of paragraph 45, wherein the traer sequence exhibits at least 50% of sequence complementarity along the length of the traer mate sequence when optimally aligned.
- [0395] 51. The kit of paragraph 45, wherein the CRISPR enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable mount in the nucleus of a eukaryotic cell.
- [0396] 52. The kit of paragraph 45, wherein the CRISPR enzyme is a type II CRISPR system enzyme.
- [0397] 53. The kit of paragraph 45, wherein the CRISPR enzyme is a Cas9 enzyme.
- [0398] 54. The kit of paragraph 45, wherein the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell.
- [0399] 55. The kit of paragraph 45, wherein the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence.
- [0400] 56. The kit of paragraph 45, wherein the CRISPR enzyme lacks DNA strand cleavage activity.
- [0401] 57. The kit of paragraph 45, wherein the first regulatory element is a polymerase III promoter.
- [0402] 58. The kit of paragraph 45, wherein the second regulatory element is a polymerase II promoter.
- [0403] 59. The kit of paragraph 49, wherein the third regulatory element is a polymerase III promoter.
- [0404] 60. The kit of paragraph 45, wherein the guide sequence is at least 15 nucleotides in length.
- [0405] 61. The kit of paragraph 45, wherein fewer than 50% of the nucleotides of the guide sequence participate in self-complementary base-pairing when optimally folded.
- [0406] 62. A computer system for selecting a candidate target sequence within a nucleic acid sequence in a eukaryotic cell for targeting by a CRISPR complex, the system comprising:
- [0407] a. a memory unit configured to receive and/or store said nucleic acid sequence; and
- [0408] b. one or more processors alone or in combination programmed to (i) locate a CRISPR motif sequence within said nucleic acid sequence, and (ii) select a sequence adjacent to said located CRISPR motif sequence as the candidate target sequence to which the CRISPR complex binds.
- [0409] 63. The computer system of paragraph 62, wherein said locating step comprises identifying a CRISPR motif sequence located less than about 500 nucleotides away from said target sequence.
- [0410] 64. The computer system of paragraph 62, wherein said candidate target sequence is at least 10 nucleotides in length.
- [0411] 65. The computer system of paragraph 62, wherein the nucleotide at the 3' end of the candidate target sequence is located no more than about 10 nucleotides upstream of the CRISPR motif sequence.
- [0412] 66. The computer system of paragraph 62, wherein the nucleic acid sequence in the eukaryotic cell is endogenous to the eukaryotic genome.
- [0413] 67. The computer system of claim 62, wherein the nucleic acid sequence in the eukaryotic cell is exogenous to the eukaryotic genome.
- [0414] 68. A computer-readable medium comprising codes that, upon execution by one or more processors, implements a method of selecting a candidate target sequence within a nucleic acid sequence in a eukaryotic cell for targeting by a CRISPR complex, said method comprising: (a) locating a CRISPR motif sequence within said nucleic acid sequence, and (b) selecting a sequence adjacent to said located CRISPR motif sequence as the candidate target sequence to which the CRISPR complex binds.
- [0415] 69. The computer-readable medium of paragraph 68, wherein said locating comprises locating a CRISPR motif sequence that is less than about 500 nucleotides away from said target sequence.
- [0416] 70. The computer-readable of paragraph 68, wherein said candidate target sequence is at least 10 nucleotides in length.

[0417] 71. The computer-readable of paragraph 68, wherein the nucleotide at the 3' end of the candidate target sequence is located no more than about 10 nucleotides upstream of the CRISPR motif sequence.

[0418] 72. The computer-readable of paragraph 68, wherein the nucleic acid sequence in the eukaryotic cell is endogenous the eukaryotic genome.

[0419] 73. The computer-readable of paragraph 68, wherein the nucleic acid sequence in the eukaryotic cell is exogenous to the eukaryotic genome.

[0420] 74. A method of modifying a target polynucleotide in a eukaryotic cell, the method comprising 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 traer mate sequence which in tum hybridizes to a traer sequence.

[0421] 75. The method of paragraph 74, wherein said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme.

[0422] 76. The method of paragraph 74, wherein said cleavage results in decreased transcription of a target gene.

[0423] 77. The method of paragraph 74, further comprising 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.

[0424] 78. The method of paragraph 77, wherein said mutation results in one or more amino acid changes in a protein expressed from a gene comprising the target sequence.

[0425] 79. The method of paragraph 74, further comprising 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 traer mate sequence, and the traer sequence.

[0426] 80. The method of paragraph 79, wherein said vectors are delivered to the eukaryotic cell in a subject.

[0427] 81. The method of paragraph 74, wherein said modifying takes place in said eukaryotic cell in a cell culture.

[0428] 82. The method of paragraph 74, further comprising isolating said eukaryotic cell from a subject prior to said modifying.

[0429] 83. The method of paragraph 82, further comprising returning said eukaryotic cell and/or cells derived therefrom to said subject.

[0430] 84. A method of modifying expression of a polynucleotide in a eukaryotic cell, the method comprising: 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 traer mate sequence which in tum hybridizes to a traer sequence.

[0431] 85. The method of paragraph 74, further comprising 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 traer mate sequence, and the traer sequence.

[0432] 86. A method of generating a model eukaryotic cell comprising a mutated disease gene, the method comprising:

[0433] 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 traer mate sequence, and a traer sequence; and

[0434] 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 traer mate sequence that is hybridized to the traer sequence, thereby generating a model eukaryotic cell comprising a mutated disease gene.

[0435] 87. The method of paragraph 86, wherein said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme.

[0436] 88. The method of paragraph 86, wherein said cleavage results in decreased transcription of a target gene.

[0437] 89. The method of paragraph 86, further comprising 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.

[0438] 90. The method of paragraph 89, wherein said mutation results in one or more amino acid changes in a protein expressed from a gene comprising the target sequence.

[0439] 91. A method of developing a biologically active agent that modulates a cell signaling event associated with a disease gene, comprising:

[0440] a. contacting a test compound with a model cell of any one of paragraphs 86-90; and

[0441] b. detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with said mutation in said disease gene, thereby developing said biologically active agent that modulates said cell signaling event associated with said disease gene.

[0442] 92. A recombinant polynucleotide comprising a guide sequence upstream of a traer mate sequence, wherein the guide sequence when expressed directs sequence-specific binding of a CRISPR complex to a corresponding target sequence present in a eukaryotic cell.

[0443] 93. The recombinant polynucleotide of paragraph 89, wherein the target sequence is a viral sequence present in a eukaryotic cell.

[0444] 94. The recombinant polynucleotide of paragraph 89, wherein the target sequence is a proto-oncogene or an oncogene.

[0445] 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. It is intended that the following claims

define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

REFERENCES

- [0446] 1. 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).
- [0447] 2. Bogdanove, A. J. & Voytas, D. F. TAL effectors: customizable proteins for DNA targeting. *Science* 333, 1843-1846 (2011).
- [0448] 3. Stoddard, B. L. Homing endonuclease structure and function. *Q. Rev. Biophys.* 38, 49-95 (2005).
- [0449] 4. Bae, T. & Schneewind, O. Allelic replacement in *Staphylococcus aureus* with inducible counter-selection. *Plasmid* 55, 58-63 (2006).
- [0450] 5. Sung, C. K., Li, H., Claverys, J. P. & Morrison, D. A. An rpsL cassette, janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. *Appl. Environ. Microbiol.* 67, 5190-5196 (2001).
- [0451] 6. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* 4, 206-223 (2009).
- [0452] 7. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821 (2012).
- [0453] 8. Deveau, H., Garneau, J. E. & Moineau, S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* 64, 475-493 (2010).
- [0454] 9. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167-170 (2010).
- [0455] 10. Terns, M. P. & Terns, R. M. CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* 14, 321-327 (2011).
- [0456] 11. van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M. & Brouns, S. J. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends. Biochem. Sci.* 34, 401-407 (2009).
- [0457] 12. Brouns, S. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960-964 (2008).
- [0458] 13. Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* 22, 3489-3496 (2008).
- [0459] 14. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602-607 (2011).
- [0460] 15. Hatoum-Aslan, A., Maniv, I. & Marraffini, L. A. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc. Natl. Acad. Sci. U.S.A.* 108, 21218-21222 (2011).
- [0461] 16. Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329, 1355-1358 (2010).
- [0462] 17. Deveau, H. et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1390-1400 (2008).
- [0463] 18. 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. U.S.A.* (2012).
- [0464] 19. Makarova, K. S., Aravind, L., Wolf, Y. I. & Koonin, E. V. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct.* 6, 38 (2011).
- [0465] 20. Barrangou, R. RNA-mediated programmable DNA cleavage. *Nat. Biotechnol.* 30, 836-838 (2012).
- [0466] 21. Brouns, S. Molecular biology. A Swiss army knife of immunity. *Science* 337, 808-809 (2012).
- [0467] 22. Carroll, D. A CRISPR Approach to Gene Targeting. *Mol. Ther.* 20, 1658-1660 (2012).
- [0468] 23. Bikard, D., Hatoum-Aslan, A., Mucida, D. & Marraffini, L. A. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* 12, 177-186 (2012).
- [0469] 24. Sapranauskas, R. et al. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* (2011).
- [0470] 25. Semenova, E. et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.* (2011).
- [0471] 26. Wiedenheft, B. et al. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U.S.A.* (2011).
- [0472] 27. Zahner, D. & Hakenbeck, R. The *Streptococcus pneumoniae* beta-galactosidase is a surface protein. *J. Bacteriol.* 182, 5919-5921 (2000).
- [0473] 28. Marraffini, L. A., Dedent, A. C. & Schneewind, O. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* 70, 192-221 (2006).
- [0474] 29. Motamedi, M. R., Szigety, S. K. & Rosenberg, S. M. Double-strand-break repair recombination in *Escherichia coli*: physical evidence for a DNA replication mechanism in vivo. *Genes Dev.* 13, 2889-2903 (1999).
- [0475] 30. Hosaka, T. et al. The novel mutation K87E in ribosomal protein S12 enhances protein synthesis activity during the late growth phase in *Escherichia coli*. *Mol. Genet. Genomics* 271, 317-324 (2004).
- [0476] 31. Costantino, N. & Court, D. L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. U.S.A.* 100, 15748-15753 (2003).
- [0477] 32. Edgar, R. & Qimron, U. The *Escherichia coli* CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. *J. Bacteriol.* 192, 6291-6294 (2010).
- [0478] 33. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463, 568-571 (2010).
- [0479] 34. Fischer, S. et al. An archaeal immune system can detect multiple Protospacer Adjacent Motifs (PAMs) to target invader DNA. *J. Biol. Chem.* 287, 33351-33363 (2012).
- [0480] 35. Gudbergsdottir, S. et al. Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* 79, 35-49 (2011).

- [0481] 36. Wang, H. H. et al. Genome-scale promoter engineering by coselection MAGE. *Nat Methods* 9, 591-593 (2012).
- [0482] 37. Cong, L. et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* In press (2013).
- [0483] 38. Mali, P. et al. RNA-Guided Human Genome Engineering via Cas9. *Science* In press (2013).
- [0484] 39. Hoskins, J. et al. Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J. Bacteriol.* 183, 5709-5717 (2001).
- [0485] 40. Havarstein, L. S., Coomaraswamy, G. & Morrison, D. A. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. U.S.A.* 92, 11140-11144 (1995).
- [0486] 41. Horinouchi, S. & Weisblum, B. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol.* 150, 815-825 (1982).
- [0487] 42. Horton, R. M. In Vitro Recombination and Mutagenesis of DNA: SOEing Together Tailor-Made Genes. *Methods Mol. Biol.* 15, 251-261 (1993).
- [0488] 43. Podbielski, A., Spellerberg, B., Woischnik, M., Pohl, B. & Luttkien, R. Novel series of plasmid vectors for gene inactivation and expression analysis in group A streptococci (GAS). *Gene* 177, 137-147 (1996).
- [0489] 44. Husmann, L. K., Scott, J. R., Lindahl, G. & Stenberg, L. Expression of the Arp protein, a member of the M protein family, is not sufficient to inhibit phagocytosis of *Streptococcus pyogenes*. *Infection and immunity* 63, 345-348 (1995).
- [0490] 45. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6, 343-345 (2009).

SEQUENCE LISTING

```

Sequence total quantity: 537
SEQ ID NO: 1      moltype = AA  length = 7
FEATURE          Location/Qualifiers
source           1..7
mol_type = protein
organism = Simian virus 40

SEQUENCE: 1
PKKKRKV

7

SEQ ID NO: 2      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: 2
KRPAATKKAG QAKKKK

16

SEQ ID NO: 3      moltype = AA  length = 9
FEATURE          Location/Qualifiers
REGION          1..9
note = Description of Unknown: C-myc NLS sequence
1..9
source           1..9
mol_type = protein
organism = unidentified

SEQUENCE: 3
PAAKRVKLQ

9

SEQ ID NO: 4      moltype = AA  length = 11
FEATURE          Location/Qualifiers
REGION          1..11
note = Description of Unknown: C-myc NLS sequence
1..11
source           1..11
mol_type = protein
organism = unidentified

SEQUENCE: 4
QRRRNELKRS P

11

SEQ ID NO: 5      moltype = AA  length = 38
FEATURE          Location/Qualifiers
source           1..38
mol_type = protein
organism = Homo sapiens

SEQUENCE: 5
NQSSSNFGPMK GGNFGGGRSSG PYGGGGQYFA KPRNQGGY

38

SEQ ID NO: 6      moltype = AA  length = 42
FEATURE          Location/Qualifiers
REGION          1..42

```

-continued

```

note = Description of Unknown: IBB domain from
      importin-alpha sequence
source          mol_type = protein
                organism = unidentified
SEQUENCE: 6
RMRIZFKNKG KDTAELRRRR VEVSVELRKA KKDEQILKRR NV           42

SEQ ID NO: 7          moltype = AA length = 8
FEATURE          Location/Qualifiers
REGION           1..8
note = Description of Unknown: Myoma T protein sequence
source          mol_type = protein
                organism = unidentified
SEQUENCE: 7
VSRKRPRP          8

SEQ ID NO: 8          moltype = AA length = 8
FEATURE          Location/Qualifiers
REGION           1..8
note = Description of Unknown: Myoma T protein sequence
source          mol_type = protein
                organism = unidentified
SEQUENCE: 8
PPKKARED          8

SEQ ID NO: 9          moltype = AA length = 8
FEATURE          Location/Qualifiers
source           1..8
mol_type = protein
                organism = Homo sapiens
SEQUENCE: 9
PQPKKKPL          8

SEQ ID NO: 10         moltype = AA length = 12
FEATURE          Location/Qualifiers
source           1..12
mol_type = protein
                organism = Mus musculus
SEQUENCE: 10
SALIKKKKKM AP        12

SEQ ID NO: 11         moltype = AA length = 5
FEATURE          Location/Qualifiers
source           1..5
mol_type = protein
                organism = Influenza virus
SEQUENCE: 11
DRLRR             5

SEQ ID NO: 12         moltype = AA length = 7
FEATURE          Location/Qualifiers
source           1..7
mol_type = protein
                organism = Influenza virus
SEQUENCE: 12
PKQKKRK            7

SEQ ID NO: 13         moltype = AA length = 10
FEATURE          Location/Qualifiers
source           1..10
mol_type = protein
                organism = Hepatitis delta virus
SEQUENCE: 13
RKLKKKIKKL          10

SEQ ID NO: 14         moltype = AA length = 10
FEATURE          Location/Qualifiers
source           1..10
mol_type = protein
                organism = Mus musculus
SEQUENCE: 14
REKKKFLKRR          10

```

-continued

```

SEQ ID NO: 15      moltype = AA  length = 20
FEATURE          Location/Qualifiers
source           1..20
mol_type = protein
organism = Homo sapiens
SEQUENCE: 15
KRKGDEVGDGV DEVAKKKSKK                                         20

SEQ ID NO: 16      moltype = AA  length = 17
FEATURE          Location/Qualifiers
source           1..17
mol_type = protein
organism = Homo sapiens
SEQUENCE: 16
RKCLQAGMNL EARKTKK                                         17

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 = 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
SEQUENCE: 21
nnnnnnnnnnnnnnnnn gttttgtac tctcaagatt tagaaataaa tcttgcagaa  60
gctacaaaaga taaggcattca tgccgaaatc aacacctgt catttatgg cagggtttt 120
tcgttattta attttt                                         137

SEQ ID NO: 22      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: 22
nnnnnnnnnnnnnnnnn gttttgtac tctcagaat gcagaagcta caaagataag  60
gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttgtt tatttaattt 120
ttt                                         123

SEQ ID NO: 23      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: 23
nnnnnnnnnnnnnnnnn gttttgtac tctcagaat gcagaagcta caaagataag  60
gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgtttttt                                         110

```

-continued

```

SEQ ID NO: 24      moltype = DNA length = 102
FEATURE
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: 24
nnnnnnnnnn nnnnnnnnnn gtttagago tagaaatagc aagttaaat aaggctagtc 60
cgtttatcaac ttgaaaaagt ggcaccgagt cggtgcttt tt 102

SEQ ID NO: 25      moltype = DNA length = 88
FEATURE
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: 25
nnnnnnnnnn nnnnnnnnnn gtttagago tagaaatagc aagttaaat aaggctagtc 60
cgtttatcaac ttgaaaaagt gttttttt 88

SEQ ID NO: 26      moltype = DNA length = 76
FEATURE
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: 26
nnnnnnnnnn nnnnnnnnnn gtttagago tagaaatagc aagttaaat aaggctagtc 60
cgtttatcatt tttttt 76

SEQ ID NO: 27      moltype = DNA length = 12
FEATURE
misc_feature
1..12
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source
1..12
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 27
gttttagagc ta 12

SEQ ID NO: 28      moltype = DNA length = 31
FEATURE
misc_feature
1..31
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source
1..31
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 28
tagcaaggta aaataaggct agtccgaaaa t 31

SEQ ID NO: 29      moltype = length =
SEQUENCE: 29
000

SEQ ID NO: 30      moltype = RNA length = 12
FEATURE
misc_feature
1..12
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source
1..12
mol_type = other RNA
organism = synthetic construct

```

-continued

```

SEQUENCE: 30
gttttagagc ta                                12

SEQ ID NO: 31      moltype = DNA  length = 33
FEATURE          Location/Qualifiers
source           1..33
                 mol_type = unassigned DNA
                 organism = Homo sapiens

SEQUENCE: 31
ggacatcgat gtcacctcca atgactaggg tgg        33

SEQ ID NO: 32      moltype = DNA  length = 33
FEATURE          Location/Qualifiers
source           1..33
                 mol_type = unassigned DNA
                 organism = Homo sapiens

SEQUENCE: 32
cattggaggt gacatcgatg tcctccccat tgg        33

SEQ ID NO: 33      moltype = DNA  length = 33
FEATURE          Location/Qualifiers
source           1..33
                 mol_type = unassigned DNA
                 organism = Homo sapiens

SEQUENCE: 33
ggaagggcct gagtccgagc agaagaagaa ggg        33

SEQ ID NO: 34      moltype = DNA  length = 33
FEATURE          Location/Qualifiers
source           1..33
                 mol_type = unassigned DNA
                 organism = Homo sapiens

SEQUENCE: 34
ggtgtggcaga gggggccgaga ttgggtgttc agg        33

SEQ ID NO: 35      moltype = DNA  length = 33
FEATURE          Location/Qualifiers
source           1..33
                 mol_type = unassigned DNA
                 organism = Homo sapiens

SEQUENCE: 35
atgcaggagg gtggcgagag gggccgagat tgg        33

SEQ ID NO: 36      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: 36
aaaaccaccc ttctctctgg c                           21

SEQ ID NO: 37      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: 37
ggagattgga gacacggaga g                           21

SEQ ID NO: 38      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: 38
ctggaaagcc aatgcctgac                            20

SEQ ID NO: 39      moltype = DNA  length = 20
FEATURE          Location/Qualifiers
misc_feature    1..20

```

-continued

```

source          note = Description of Artificial Sequence: Synthetic primer
                1..20
                mol_type = other DNA
                organism = synthetic construct

SEQUENCE: 39
ggcagcaaac tccttgtcct                                20

SEQ ID NO: 40      moltype = DNA  length = 335
FEATURE          Location/Qualifiers
misc_feature     1..335
note = Description of Artificial Sequence: Synthetic
                polynucleotide
source          1..335
                mol_type = other DNA
                organism = synthetic construct

SEQUENCE: 40
gaggggcctat ttccccatgtat tccttcatac gatacaaggc tgtagagag  60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
aagtaataat ttcttggta gttgcagtt ttaaaattat gttttaaaat ggactatcat 180
atgcttaccg taacctgaaa gtatttcgtat ttcttggctt tatatatctt gtggaaagga 240
cgaaacaccg gaaccattca aaacagcata gcaagttaaa ataaggctag tccgttatca 300
acttgaaaaa gtggcaccga gtcggtgctt tttt                                335

SEQ ID NO: 41      moltype = DNA  length = 423
FEATURE          Location/Qualifiers
misc_feature     1..423
note = Description of Artificial Sequence: Synthetic
                polynucleotide
source          1..423
                mol_type = other DNA
                organism = synthetic construct

SEQUENCE: 41
gaggggcctat ttccccatgtat tccttcatac gatacaaggc tgtagagag  60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
aagtaataat ttcttggta gttgcagtt ttaaaattat gttttaaaat ggactatcat 180
atgcttaccg taacctgaaa gtatttcgtat ttcttggctt tatatatctt gtggaaagga 240
cgaaacaccg gtatgtatggat gttttttttt atggctgata aatttctt aatttctct 300
tgatttattt tataaaaatgtataaaataa ttcttggta accattcaa acagcatagc 360
aagttaaat aaggctagtc cggtatcaac ttgaaaaatggcaccggat cggtgctttt 420
ttt                                423

SEQ ID NO: 42      moltype = DNA  length = 339
FEATURE          Location/Qualifiers
misc_feature     1..339
note = Description of Artificial Sequence: Synthetic
                polynucleotide
source          1..339
                mol_type = other DNA
                organism = synthetic construct

SEQUENCE: 42
gaggggcctat ttccccatgtat tccttcatac gatacaaggc tgtagagag  60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
aagtaataat ttcttggta gttgcagtt ttaaaattat gttttaaaat ggactatcat 180
atgcttaccg taacctgaaa gtatttcgtat ttcttggctt tatatatctt gtggaaagga 240
cgaaacaccg gtttttagat ctatgctgtt ttgaatggc caaaaacggg tcttcgagaa 300
gacgttttag agctatgtt ttttgaatgg tcccaaaac                                339

SEQ ID NO: 43      moltype = DNA  length = 309
FEATURE          Location/Qualifiers
misc_feature     1..309
note = Description of Artificial Sequence: Synthetic
                polynucleotide
source          1..309
                mol_type = other DNA
                organism = synthetic construct

SEQUENCE: 43
gaggggcctat ttccccatgtat tccttcatac gatacaaggc tgtagagag  60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
aagtaataat ttcttggta gttgcagtt ttaaaattat gttttaaaat ggactatcat 180
atgcttaccg taacctgaaa gtatttcgtat ttcttggctt tatatatctt gtggaaagga 240
cgaaacaccg ggtcttcgag aagacctgtt tttagagctag aaatagcaag taaaataag 300
gcttagtcgg                                309

SEQ ID NO: 44      moltype = AA  length = 1648
FEATURE          Location/Qualifiers
REGION          1..1648

```

-continued

```

note = Description of Artificial Sequence: Synthetic
      polypeptide
source          1..1648
mol_type = protein
organism = synthetic construct

SEQUENCE: 44
MDYKDHGDY KDHIDYKDD DDKMAPKKR KVGIHGVPAK DKKYSIGLDI GTNSVGWAVI 60
TDEYKVPSK KFKVLGNTDRH SIKKNLIGAL LFDSGETAAE TRLKRTARR YTRRKNRIC 120
LQEIFSNEMA KVDDSSFFHRL EESFLVEEDK KHERHPIFGN IVDEVAYHEK YPTIYHLRK 180
LVSTDADL RLIYLALAHM IKFRGHFLIE GDLNPDNSDV DKLFIQLVQT YNQLFEEENPI 240
NASGVDAKAI LENLIAQLPQ EKKNGLFGNL IALSLGLTPN FKSNFDLAED 300
AKLQLSKDTY DDDLDNLLAQ IGDQYADLFL AAKNLSDAIL LSDILRVNTL ITKAPLSASM 360
IKRYDEHHQD LTLLKALVRQ QLPEKYKEIF FDQSKNGYAG YIDGGASQEE FYKFKIPILE 420
KMDGTEELLV KLNREDLLRQ QRTFNGSIP HQIHLGELHA ILTRRQEDFYP FLKDNRREKIE 480
KILTFRIPSEE VGPPLARGNSR FAWMTRKSEE TITPWNFEEV VDKGASAQSIE RERMTNFDKN 540
LPNEKVLPLKH SLLYEFVTY NELTKVKYV EGMRKPAFLS GEQKKAIVDL LFKTNRKVT 600
KQLKEDYFKK IECFDSVETIS GVEDRFNDSL GTYHDLLKII KDKDFLDNEE NEDILEDIVL 660
TTLTFEDREM IEERLKTYAH LFDDKVMKQL KRRRYTGWR LSRKLINGIR DKQSGKTIID 720
FLKSDGFAKR NFMQQLHDDS LTFKEDIQKA QVSGQGDLSL EHIANLAGSP AIKGILQTV 780
KVVDDELVKVM GRHKPENIVI EMARENQTTQ KGQKNSRERM KRIEEGIKEL GSQILKEHPV 840
ENTQLQNEKL YLYYLQNGRD MYVDQELDIN RLSDYDWDHI VPQSFLLKDS IDNVKLTSD 900
KNRGKSDNVP SEEVVKKMKV YWRQLLNALKI ITQRKFNDLNT KAERGGLSEL DKAGPIKRQL 960
VETRQITKHN AQILDLSRMNT KYDENDKLIR EVKVTILKSK LVSDFRKDFQ FYKVREINNY 1020
HHAHDAYLNA VVGTALIKKY PKELESEFVY DYKVYDVRK IAKSEQEIGK ATAKYFFYSN 1080
IMNFFKTEIT LANGEIRKRP LIETNGETGE IVDWKGRDFA TVRKVLMSMP VNIVVKTEVQ 1140
TGGFSKESIL PKRNSDKLIA RKKDWDPKYY GGFDSPTVAY SVLVAKVEK GKSKKLKSVK 1200
ELLGITIMER SSFEKNPIDE LEAKGYKEVK KDLIILKPKY SLFELENGRK RMLASAGELQ 1260
KGNGELALPSK YVNFLYFLASH YEKLKGSPED NEQKQLFVEQ HKHYLDEIE QISEFSKRV 1320
LADANLDKVL SAYNKHRDKP IREQAENIIH LFTLTLNLGAP AAFKYFDTI DRKRYTSTKE 1380
VLDATLHQIS ITGLYETRID LSQLGDDAAA VSKGEELFTV GPVPLVVELDG DVNGHFKFSVS 1440
GEGEGLDATYG KLTLPKFCITT GKLPVPWPWT VTTLTGYVQF FSRYPDHMKQ HDFFKKSAMPE 1500
GYVQERTIFF KDDGNYKTRA EVKFGDFTLV NRIELKGIDF KEDGNILGHK LEYNYNSHNV 1560
YIMADKQKNG IKVNFKIRHN IEDGSVQLAD HYQQNTPIGD GPVLLPDNHY LSTQSALSKD 1620
PNEKRDHMVL LEFVTAAGIT LGMDELYK 1648

SEQ ID NO: 45          moltype = AA length = 1625
FEATURE                  Location/Qualifiers
REGION          1..1625
note = Description of Artificial Sequence: Synthetic
      polypeptide
source          1..1625
mol_type = protein
organism = synthetic construct

SEQUENCE: 45
MDKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE 60
ATRLKRTARR YRTRRKNRIC YLQEIFSNEM AKVDDSSFFHRL EESFLVEED KKHERHPIFG 120
NIVDEVAYHE KPYTIYHLRK KLVSTDADL RLIYLALAHM MIKFRGHFLIE EGDLNPNSD 180
VDKLFQIQLVQ TYNQLFEENP INASGVDAKAI ILSARLSKSRL RLENLIAQLP GEKKNGLFGN 240
LIALSLGLTP NFKNSFDLAE DAKLQLSKDTY YDSDLNLLA QIGDQYADLFL LAAKNLSDAI 300
LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLKALVRQ QLPEKYKEI FFDQSKNGYA 360
GYIDGGASQEE FFYKFKIPIL EKMDGTEELL VKLNREDLLRQ KORTFDNGSI PHQIHLGELH 420
AILRRQEDFY PFLKDNREKIE KILTFRIPYY YVGPLARGNSR RFAWMTRKSE ETITPWNFEE 480
VVDKGASAQS PIERMTNFDL NLPNEKVLPLK HSLLYEFVTY YNELTKVKYV TEGMRKPAFL 540
SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK IECFDSVETIS SGVEDRFNDSL LGTYHDLLKI 600
IKDKDFLDNE ENEDILEDIVL LTLLTFEDREM MIEERLKTYAH LFKDNDKVLKQ LKRRRYTGWG 660
RLSRKLINGI RDKQSGKTIID DFLKSGDFAN RNFMQQLHDDS SLTFKEDIQKA AQVSGQGDLSL 720
HEHIANLAGS PAIKKGILQI KVVDDELVKM GRHKPENIVI EMARENQTTQ KGQKNSRERM 780
MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGRD MYVDQELDIN RLSDYDWDHI 840
IVPQSFLLKDD SIDNKVLTRS DKNRQKSDNVP PSSEVVKKMKV NYWRQLLNALKI LITQRKFNDL 900
TKAERGGLSE LDKAGFIKRQ LVETRQITKHN VAQILDLSRMNT KYDENDKLIR REVKVTILK 960
KLVSDFRKDF QFYKVREINNY YHHADAYLNA VVGTALIKKY PKELESEFVY GDYKVYDVRK 1020
MIAKSEQEIG KATAKYFFYNS NIMNFFKTEI TLANGEIRKRP LIETNGETGE EIVWDKGRDF 1080
ATVRKVLMSMP QVNIVVKTEV QTGGFSKESI LPKRNNSDKLIA RKKDWDPKYY GGFDSPTV 1140
YSVLUVVAKVE KGKSKKLKSVC KELLGITIME RSSFEKNPIDE LEAKGYKEVK KDLIILKPK 1200
YSLFELENGR KRMLASAGEL QKNGELALPSK YVNVFLYLAS HYEKLKGSPD DNEQKQLFVE 1260
QHKHYLDEII EQISEFSKRV ILADANLDKVL LSAYNKHRDKP IREQAENIIH LFTLTLNLGA 1320
PAAFKYFDTT IDRKRYTSTK EVLDAVLHQ SITGLYETRI DLSQLGDDAA AVSKGEELFT 1380
GVVPILVELD GDVNGHFKFSV SGEGEGLDATYG KLTLPKFCITT TGKLPVPWPWT VTTLTGYVQ 1440
CSFRYPDHMKQ QHDFFKSAMP EGYVQERTIF FKDDGNYKTRAEVKGEGDFTL VNRIELKGID 1500
FKEDGNILGH KLEYNNSHNV YVIMADKQKNG IGIKVNFKIRHN NIEDGSVQLA DHYQQNTPIG 1560
DGPVLLPDNHY YLSTQSALSK DPNEKRDHMVL LEFVTAAGIT TLGMDELYKK RPAATKKAGQ 1620
AKKKK

SEQ ID NO: 46          moltype = AA length = 1664
FEATURE                  Location/Qualifiers
REGION          1..1664

```

-continued

```

note = Description of Artificial Sequence: Synthetic
      polypeptide
source 1..1664
mol_type = protein
organism = synthetic construct

SEQUENCE: 46
MDYKDHG DY KDH D ID YKDD DDKMAPKKR K VGIHGVPA DKKYSIGLDI GTNSVGWAVI 60
TDEYKVPS KK FKVLGNTDRH SIKKNLIGAL LFD SGETAEE TRLKRTARR YTRRKNRIC Y 120
LQEIFSNE MA KVDDSFHRL EESFLVEEDK KHERHPIFGN IVDEVAYHEK YPTIYHLRK K 180
LV DSTDKADL RLIYLALAHM IKFRGHFELIE GDLNPDSNDV DKLF IQLVQT YNQLFEENPI 240
NASGVDAKAI LENLIAQLSR EKKNGLFGNL IALSLGLTPN FKS NFDLAED 300
AKLQLSKDTY DDDLDNLLAQ IGDQYADLFL AAKNLS DAIL LSDILRVNTE ITKAPLSASM 360
IKRYDEHHQD LTLLKALVRQ QLPEKYKEIF FDQSKNGYAG YIDGGASQEE FYKFIKP ILE 420
KMDGTEELV KLNREDLRLR QRTFDNSIP HQIHLGELHA ILR RQEDFYP FLKD NREKIE 480
KILTFRIPYY VGPLARGNSR FAWMTRKSEE TIT PWNFEEV VDKGASAQS F IERMTNFDKN 540
LPNEKVLPKH SLLYEYFTVY NELTKVKVYT EGMRKP AFLS GEQKKAIVDL LFKTNRKTV 600
KQLKEDYFKK IECFD SVEIS GV EDRFN ASL GTYHDLLKII KDKDFLDNEE NEDILEDIVL 660
TTLTFEDREM IEE RLKTYAH LFDDKVMK QL KR RYTG WGR LSRKLINGIR DKQSGKTILD 720
FLKSDGFANR NFMQLIHDDS LTFKEDIQKA QVSGQGDLSH EHIANLAGSP AIKKGILQTV 780
KV VDELVKV M GRHKPENIVI EMARENQTTQ KGQKNSRERM KRIEEGIKEL GSQILKEHPV 840
ENTQLQNEKL YLYYLQNGRD MYVDQELDIN RLSDYDV DHI VPQSFLKDD S IDNKVLTRSD 900
KNRGKSDNVP SEE VVKMKM K YWRQLLNALK I TQRKF DNLT KAERGGLSEL DKAGFIKRQL 960
VETRQITKHN AQILD SRMNT KYDEN DKL IR EVKVI TLKSK LVSDFRKDFQ FYKVREINNNY 1020
HHAHDAYLNA VVGTALIKKY PKLESEFVYG DYKVYDVRKM IAKSE QEI GK ATAKYFFY S 1080
IMNFFKTEIT LANGEIRKRP LIETNGETGE IWDKG RDFA TVRKVL SMPQ VNIVVKTEVQ 1140
TGGFSKESIL P KRN S DKLIA RKKDWDPK KY GGF D SPTVAY SVL VVAKVEK GSKKLKSVK 1200
ELL GITIMER SSFEKNP IDE LEAKY KEV K KDLI I KLPK Y SLFELENGRK RMLASAGELQ 1260
KG NELALPSK YVN FLYLASH YE KLKG SPED NEQKQLF VEQ HKHYLDEIIE QISEFSKRV 1320
LADANLDKVL SAYNKH RDKP IREQAENIIH LFTL TNL GAP AAFK YFDTI DRKRYTSTK 1380
VLDATLHQ S ITGLYETRID LS QLGGDAAA VSKGEELFTG VVPILVELDG DVNGH KFSVS 1440
GE GEGDATY G KLT KFIC TT GKL PVPW PAA T VTL TYGVQC FS RY PDH MKQ HDFF K SAMPE 1500
GYVQERTIFF KDDGNYKTRA EVKFEGDTLV NRIELKGIDF KEDGNILGHK LEYN YNSHNV 1560
YIMADKQKNG I KVNF KIRHN IEDGSVQLAD HYQONTPIGD GPVLLPDNH Y LSTQSALS K 1620
PNEKRDHMVL LEF VTAAGIT LGMD ELYKKR PAATKKAGQA KKK 1664

SEQ ID NO: 47
FEATURE moltype = AA length = 1423
REGION Location/Qualifiers
source 1..1423
note = Description of Artificial Sequence: Synthetic
      polypeptide
mol_type = protein
organism = synthetic construct

SEQUENCE: 47
MDYKDHG DY KDH D ID YKDD DDKMAPKKR K VGIHGVPA DKKYSIGLDI GTNSVGWAVI 60
TDEYKVPS KK FKVLGNTDRH SIKKNLIGAL LFD SGETAEE TRLKRTARR YTRRKNRIC Y 120
LQEIFSNE MA KVDDSFHRL EESFLVEEDK KHERHPIFGN IVDEVAYHEK YPTIYHLRK K 180
LV DSTDKADL RLIYLALAHM IKFRGHFELIE GDLNPDSNDV DKLF IQLVQT YNQLFEENPI 240
NASGVDAKAI LENLIAQLSR EKKNGLFGNL IALSLGLTPN FKS NFDLAED 300
AKLQLSKDTY DDDLDNLLAQ IGDQYADLFL AAKNLS DAIL LSDILRVNTE ITKAPLSASM 360
IKRYDEHHQD LTLLKALVRQ QLPEKYKEIF FDQSKNGYAG YIDGGASQEE FYKFIKP ILE 420
KMDGTEELV KLNREDLRLR QRTFDNSIP HQIHLGELHA ILR RQEDFYP FLKD NREKIE 480
KILTFRIPYY VGPLARGNSR FAWMTRKSEE TIT PWNFEEV VDKGASAQS F IERMTNFDKN 540
LPNEKVLPKH SLLYEYFTVY NELTKVKVYT EGMRKP AFLS GEQKKAIVDL LFKTNRKTV 600
KQLKEDYFKK IECFD SVEIS GV EDRFN ASL GTYHDLLKII KDKDFLDNEE NEDILEDIVL 660
TTLTFEDREM IEE RLKTYAH LFDDKVMK QL KR RYTG WGR LSRKLINGIR DKQSGKTILD 720
FLKSDGFANR NFMQLIHDDS LTFKEDIQKA QVSGQGDLSH EHIANLAGSP AIKKGILQTV 780
KV VDELVKV M GRHKPENIVI EMARENQTTQ KGQKNSRERM KRIEEGIKEL GSQILKEHPV 840
ENTQLQNEKL YLYYLQNGRD MYVDQELDIN RLSDYDV DHI VPQSFLKDD S IDNKVLTRSD 900
KNRGKSDNVP SEE VVKMKM K YWRQLLNALK I TQRKF DNLT KAERGGLSEL DKAGFIKRQL 960
VETRQITKHN AQILD SRMNT KYDEN DKL IR EVKVI TLKSK LVSDFRKDFQ FYKVREINNNY 1020
HHAHDAYLNA VVGTALIKKY PKLESEFVYG DYKVYDVRKM IAKSE QEI GK ATAKYFFY S 1080
IMNFFKTEIT LANGEIRKRP LIETNGETGE IWDKG RDFA TVRKVL SMPQ VNIVVKTEVQ 1140
TGGFSKESIL P KRN S DKLIA RKKDWDPK KY GGF D SPTVAY SVL VVAKVEK GSKKLKSVK 1200
ELL GITIMER SSFEKNP IDE LEAKY KEV K KDLI I KLPK Y SLFELENGRK RMLASAGELQ 1260
KG NELALPSK YVN FLYLASH YE KLKG SPED NEQKQLF VEQ HKHYLDEIIE QISEFSKRV 1320
LADANLDKVL SAYNKH RDKP IREQAENIIH LFTL TNL GAP AAFK YFDTI DRKRYTSTK 1380
VLDATLHQ S ITGLYETRID LS QLGGDAAA VSKGEELFTG VVPILVELDG DVNGH KFSVS 1440
GE GEGDATY G KLT KFIC TT GKL PVPW PAA T VTL TYGVQC FS RY PDH MKQ HDFF K SAMPE 1500
GYVQERTIFF KDDGNYKTRA EVKFEGDTLV NRIELKGIDF KEDGNILGHK LEYN YNSHNV 1560
YIMADKQKNG I KVNF KIRHN IEDGSVQLAD HYQONTPIGD GPVLLPDNH Y LSTQSALS K 1620
PNEKRDHMVL LEF VTAAGIT LGMD ELYKKR PAATKKAGQA KKK 1423

SEQ ID NO: 48
FEATURE moltype = AA length = 483
REGION Location/Qualifiers
source 1..483
note = Description of Artificial Sequence: Synthetic
      polypeptide
mol_type = protein

```

-continued

organism = synthetic construct

SEQUENCE: 48

MFLFLSLTSF LSSSRTLVSK GEEDNMAIIK EFMFRKVHME GSVNGHEFEI EGEGERPYE 60
 GTQATAKLKV TKGGLPLPFAWD ILSPLQFMYGS KAYVKHPADI PDYLKLSPPE GFKWERVMNF 120
 EDGGVVVTQ DSSLQDGEFI YKVKLRLGTFN PSDGPVIMQKK TMGWEASSER MYPEDGALKG 180
 EIKQLRKLKD GGHYDAEVKT TYKAKKPVQL PGAYNVNIKL DITSHNEDYT IVEQYERAEG 240
 RHSTGGMDEL YKGSKQLEL LSTSFDIQPN DLTLLTEAFT HTSYANEHRL LNVSHNERLE 300
 FLGDAVLQLI ISEYLFAKYK PTKEGDMSLK RSMVIREESL AGFSRCSFD AYIKLKGEE 360
 KSGGRRRTDI LGDLFEEAFLG ALLLDKGIDA VVRFLKQVMI PQVEKGNNFER VKDYKTCLQE 420
 SEV FLQTKGDVAI DYQVISEKGP AHAKQFEVSI VVNGAVLSKG LGKSKKLAEQ DAAKNALAQL 480
SEQ ID NO: 49 moltype = AA length = 483
FEATURE Location/Qualifiers
REGION 1..483
note = Description of Artificial Sequence: Synthetic
source polypeptide
 1..483
mol_type = protein
organism = synthetic construct

SEQUENCE: 49

MKQLEELLST SFIDIQFNDSL LLETAFTHTS YANEHRLLNV SHNERLEFLG DAVLQLIISE 60
 YLFAKYPKKT EGDMMSKLRSM IVREESLAGF SRFCSFDAYI KLKGEEKSG GRRRTILGD 120
 LFEAFLLGALL LDKGIDAVR FLKQVMPIQV EKGNFVERKD YTCTCLOEFLQ TKGDAVIDYQ 180
 VISEKGPAAA QKFQEVSVVN GAVLSKGLG SKKLAQDAA KNALAQSEV GSVSKEEDN 240
 MAIIKEFMRM KVHMEGSVNG HEFIEEGE GRPYEGTQTA KLKVTKGGLP FPAWADILSPQ 300
 FMYGSKAYVK HPADIPDYLK LSFPEGFKWE RVVMNFEDGGV VTVTQDSSLQ DGEFIYKVKL 360
 RGTNFPSDGP VMQKKTGMWE ASSERMYPED GALKGEIKQR LKLKDGGHYD AEVKTTTYKAK 420
 KPVLQPGAYN VNIKLDITSH NEDYATIVEQ ERAEGRHSTG GMDELYKKRP AATKKAGQAK 480
 KKK 483

SEQ ID NO: 50 moltype = AA length = 1423
FEATURE Location/Qualifiers
REGION 1..1423
note = Description of Artificial Sequence: Synthetic
source polypeptide
 1..1423
mol_type = protein
organism = synthetic construct

SEQUENCE: 50

MDYKDHGDY KDHDIDYKDD DDKMAPKKR KVGIHGVPAA DKKYSIGLAI GTNSVGAVI 60
 TDEYKVPSK FKVLGNTRH SIKKNLIGAL LFDSGETAEA TRLKRTARRR YTRRKNRICY 120
 LQEIFSNEMA KVDDSSFHRL EESFLVVEDK KHERHPIFGNV 1VDEVAYHEK YPTIYHLRK 180
 LVDSTDKAFL RLIYLALAHM IKFRGHFLIE GDLNPNDNSV DKLFIQLVQT YNQLFEENPI 240
 NASGVDAKAI LSARLSKSRA LENLIAQLPG EKKNGFLGNL IALSLGLTPN FKSNSFLDAED 300
 AKLQLSKDTY DDDDLNLLAQ IGQYADLFL AAKNLSDAIL LSDILRVNT EITKAPLSASM 360
 IKRYDEHQHD LTLLKALVRQ OLPEKYKEF FDQSKNGYAG YDGGASQEE FYFKFPILE 420
 KMDGTEELLV KLNREDLLRK QRTFDNGSIP HQIHLGELHA ILRROQDFYP FLKDNREKIE 480
 KILTFRIPYYV VGPLARGNSK FAWMTRKSEE TITPWNVFEV VDKGASAQS 1IERMTNFDN 540
 LPNEKVLPKH SLLYEYTFVY NELTKVKVYT EGMRKPFA 600
 QLKEDYFKK IECFDSVEIS GVEDRFNALS GTYHDLLKII KDKDFLDNEE NDEDILEDIVL 660
 TLTLFEDREM IEERLKTYAH LFDDKVMQL KRRRTGWWGR LSRKLINGIR DKQSGKTIID 720
 FLKSDGFANR NFMQLIHDDS LTFKEDIQKA QVSGQGDSLH EHIANLAGSP AIKKGLQTV 780
 KVVDDELVKVN GRHKPENIV EMARENTTQ KGKQNSRERM KRIEKGKEL GSQILKEHPV 840
 ENTQLQNEKL YLYLYQNQGRD MYVQELDN RLSLDYDVDPH VIQPSFLKDD5 IDNKVLTRSD 900
 KNRGKSDNVP SEEVVKMKM YWRQLLNAKL ITQRKFDDNL KAERGGLSEL DKAGFIKRQL 960
 VETRQITKHV AQILDSDRMT KYDENDKLIR EVKVITLKSX LVSDFRKDFQ FYKVRBENN 1020
 HHAHDAYLNA VVGTALIKK PKLESEFYVG DYKVDVDRKM IAKSEQEIGK ATAKYFFYSN 1080
 IMNFFKTEIT Langeirkrp LIETNGETGE IIVWDKGRDFV TVRKVLMSMQ VNIVKHTEVQ 1140
 TGGFSKESIL PKRNDSKLIA RKKWDWPKKY GGDPSVTWVY SVLUVVAKVEK GKSKKLKSVK 1200
 ELLGITIMER SSFEKNPIDE LEAKGYKEVK KDLIILPKY SLFELENGRK RMLASAGELQ 1260
 KGNELALPSK YVNFLYFLASH YEKLKGSPED NEQKQLFVEQ HKHYLDBIIIE QISEFSKRV 1320
 LADANLDKVL SAYNKHRDKP IREQAENI 1H LFTLTNLGAP AAFKYFDTI DRKRYTSTKE 1380
 VLDATLHQ5 ITGLYETRID LSQLGDKRP AATKKAGQAK KKK 1423

SEQ ID NO: 51 moltype = DNA length = 2012
FEATURE Location/Qualifiers
misc_feature 1..2012
note = Description of Artificial Sequence: Synthetic
source polynucleotide
 1..2012
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 51

gaatgtgcgc ctcagacccg ctteccctt gtccattgtc gtccaaaggag aatgggtct 60
 cactgggtga tttcgacta ccctgaggag ctggcacctg agggacaagg cccccccac 120

-continued

```
SEQ ID NO: 52          moltype = AA  length = 1153
FEATURE                Location/Qualifiers
REGION                 1..1153
note = Description of Artificial Sequence: Synthetic
                  polypeptide
source                1..1153
                      mol_type = protein
                      organism = synthetic construct
```

SEQUENCE :	52	01941813M - S74191010 CCR191128				
MKRPAATKKA	GQAKKKKSDL	VLGLDIGIGS	VGVGILINKVT	GEIIHKNSRI	FPAQAENNLL	60
VRRTNRRQQRV	LARRKKHRRV	RLRNRLFEESG	LITDFTPKISI	NLNPYQLRVK	GLTDELSNEE	120
FLIAFLNMVK	HGRISYLDDA	SDDGGNSVSD	YAQIVKENSEK	QLETFKTPGQI	QLEYQTYGQ	180
LRGDFTVEKD	GKKHRLINVF	PTSAVRSEAL	RILOTQQEFPN	PQITDEFINR	YLEILTGKRK	240
YYHGPNGNEKS	RTDYGRYRTS	GETLDNIFGI	LIGKCTFYPD	EFRAAKASYT	AQEFLNLLNDL	300
NNLTVPTEFT	KLSKEQKNQI	INYVKEKAM	GPAKLFKYIA	KLLSDFCVDADI	KGYRIDKSGK	360
AEIHTFEAYR	KMKTLETLDI	EQMDRETLTDI	LAYVLTNTE	REGIQALEAH	EFADGSFSQK	420
QVDELVQFRK	ANSSIFGKGW	HNFSVKLMM	LIPELYETSE	EQMTILTRLG	KQKTTSSSNK	480
TKYIDEKLLT	EBIYNPVVAK	SVRQAIKVNA	AAIKEYGFDF	NIVIEMARET	NEDDEKKAIQ	540
KIQKANKADE	DAAMLLKAANQ	YNGKAEPLHS	VFHGHQQLAT	KIRLWHHQGEG	RCLTYTGKTS	600
IHDLLINNSNQ	FEVDHILPLS	ITFDSSLANK	LVYATANQE	KQGRTPYQAL	DSMDDAWSFR	660
ELKAFVRESK	TLSNKKKEYL	LTTEEDISKFD	VRKKPIERNL	UDTRYTASRVV	LNALQEHHFR	720
HKIDTKVSVV	RGQFTSQLRR	HGWIEKTRDT	YHHHAVDALI	IAASSQLNLW	KKQKNTLVSY	780
SEQLDLIEDT	GELISDEYK	ESVFKPYQH	FVDTLKSKF	EDSLFLFSYQV	DSKEFNRKISD	840
ATIYATRQAK	VGKDKADETY	WLGKJIDT	QDGYDAFMKI	YKKDKSKFLM	YRHDPQTFEK	900
VIEPILENNP	NKQINKEGKE	VPCPNFLKYK	EEHGYIRKYS	KGKNGPEIKS	LKYDVSLSKLN	960
HIDITPKDSN	NKVVVLQSVSP	WRADVYFNKT	TGKYEILGLK	YADLQFEKGT	GTYKISQEKY	1020
NDIKKKEGV	SDSEFKFTLY	KNDLLLWVKT	ETKEQQLFRF	LSRTMPKQKH	YVELKPYDKQ	1080
KPEGGEALIK	VLGNVANSQG	CKKGLGKNSI	SIYKVRTDVL	GNQHIIKNEG	DPKPLDFKRP	1140
AATKKQAOAK	KKK					1153

```
SEQ ID NO: 53          moltype = DNA    length = 340
FEATURE                  Location/Qualifiers
misc_feature            1..340
note = Description of Artificial Sequence: Synthetic
                      polynucleotide
source                 1..340
mol_type = other DNA
organism = synthetic construct
```

```

SEQUENCE: 53
          Organism = synthetic construct

gggggcctat ttcccatgat tccttcata tgcatacata gatacaaggc tgtagagag 60
ataatggaa ttaatttgac tgtaaacaca aagatattac tacaaaatac gtgcgtgata 120
aagaatataat ttcttgggtt gtttgcgtt tttttttttt ggactatcat 180
atgccttaccg taacttgtaaa gttttcgat ttcttgggtt tatatatctt tgaaaagg 240

```

-continued

```

cgaaacaccg ttacttaat cttgcagaag ctacaaagat aaggcttcat gccgaaatca 300
acaccctgtc atttatggc agggtttt cgttattaa 340

SEQ ID NO: 54 moltype = DNA length = 360
FEATURE Location/Qualifiers
misc_feature 1..360
note = Description of Artificial Sequence: Synthetic
polynucleotide
misc_difference 288..317
note = a, c, t, g, unknown or other
source 1..360
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 54
gagggccatat ttcccatat tccttcatac ttgcatatac gatacaaggc tgtagagag 60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
aagtaataat ttcttggta gtttcgtt taaaattat gttttaaaat ggactatcat 180
atgcttaccg taacttgaaa gtatttcgtat ttcttggctt tatatatctt gtggaaagga 240
cgaaacaccg gtttttagag ctatgtgtt ttgaatggtc caaaaacnnn nnnnnnnnnn 300
nnnnnnnnnnnnnnnnn nnnnnnnngt tttagagctat gctgttttga atggcccaa aactttttt 360

SEQ ID NO: 55 moltype = DNA length = 318
FEATURE Location/Qualifiers
misc_feature 1..318
note = Description of Artificial Sequence: Synthetic
polynucleotide
misc_difference 250..269
note = a, c, t, g, unknown or other
source 1..318
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 55
gagggccatat ttcccatat tccttcatac ttgcatatac gatacaaggc tgtagagag 60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
aagtaataat ttcttggta gtttcgtt taaaattat gttttaaaat ggactatcat 180
atgcttaccg taacttgaaa gtatttcgtat ttcttggctt tatatatctt gtggaaagga 240
cgaaacaccn nnnnnnnnnnnnnnnn nnnnnnnnnng ttttagagct agaaatagca agttaaaata 300
aggctagtcc gttttttt 318

SEQ ID NO: 56 moltype = DNA length = 325
FEATURE Location/Qualifiers
misc_feature 1..325
note = Description of Artificial Sequence: Synthetic
polynucleotide
misc_difference 250..269
note = a, c, t, g, unknown or other
source 1..325
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 56
gagggccatat ttcccatat tccttcatac ttgcatatac gatacaaggc tgtagagag 60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
aagtaataat ttcttggta gtttcgtt taaaattat gttttaaaat ggactatcat 180
atgcttaccg taacttgaaa gtatttcgtat ttcttggctt tatatatctt gtggaaagga 240
cgaaacaccn nnnnnnnnnnnnnnnn nnnnnnnnnng ttttagagct agaaatagca agttaaaata 300
aggctagtcc gttatcattt tttttt 325

SEQ ID NO: 57 moltype = DNA length = 337
FEATURE Location/Qualifiers
misc_feature 1..337
note = Description of Artificial Sequence: Synthetic
polynucleotide
misc_difference 250..269
note = a, c, t, g, unknown or other
source 1..337
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 57
gagggccatat ttcccatat tccttcatac ttgcatatac gatacaaggc tgtagagag 60
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
aagtaataat ttcttggta gtttcgtt taaaattat gttttaaaat ggactatcat 180
atgcttaccg taacttgaaa gtatttcgtat ttcttggctt tatatatctt gtggaaagga 240
cgaaacaccn nnnnnnnnnnnnnnnn nnnnnnnnnng ttttagagct agaaatagca agttaaaata 300
aggctagtcc gttatcaact tgaaaaagtg tttttt 337

SEQ ID NO: 58 moltype = DNA length = 352

```

-continued

FEATURE	Location/Qualifiers
misc_feature	1..352 note = Description of Artificial Sequence: Synthetic polynucleotide
misc_difference	250..269 note = a, c, t, g, unknown or other
source	1..352 mol_type = other DNA organism = synthetic construct
SEQUENCE: 58	
gaggggcatat ttcccgtat tccttcatat ttgcataatac gataacaaggc tgtagagag 60	
ataaattggaa ttaatttgac tggaaacaca aagatattag tacaaaatac gtgacgtaga 120	
aagtaataat ttcttgggtt gtttcagtt taaaattat gttttaaaat ggactatcat 180	
atgcttaccg taacttggaaat gtatccgtat ttcttgggtt tatataatcc gtggaaagga 240	
cgaaacaccc nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ttttagagct agaaatagca agttaaaaata 300	
aggcttagtcc gttatcaact tgaaaaatgt gcaccgagtc ggtgtttt tt 352	
SEQ_ID NO: 59	moltype = DNA length = 5101
FEATURE	Location/Qualifiers
misc_feature	1..5101 note = Description of Artificial Sequence: Synthetic polynucleotide
source	1..5101 mol_type = other DNA organism = synthetic construct
SEQUENCE: 59	
cgttacataat ttacggtaa atggccgcgc tggctgaccg cccaaacgacc cccgccccatt 60	
gacgtcaata atgacgtat ttcccatagt aacgccaataa gggactttcc attgacgtca 120	
atggggtagg tatttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc 180	
aagtacgccc ctatggatcg tcaatgcggg taaatggccc gcctggatt atgcccagta 240	
catgacccta tgggacttcc ctatggatcg tcaatgcggg taaatggccc gcctggatt atgcccagta 300	
catggcggag gtgagcccca cgtttgtttt cacttccccc atctcccccc cctccccacc 360	
cccaattttt tattttatata ttttttattt attttgcgc gcgatgggggg cggggggggg 420	
ggggggggcgc ggcgcaggcg gggcgggggcg gggcgagggg cggggggggg cgaggcggag 480	
agggtgcggcg gcaagccatac agagccgcgc gtcctggaaat ttcccttttgc gggcgaggcg 540	
gccccgggggg cggccctataa aaaaaggcgaaat cgcgcggggcg gggggaggcg ctgcgcacgt 600	
gccttcgcgc cgtgccccgc tccgcgcgcg cttcgccgcg cccgcggcccg ctctgactga 660	
ccgcgttact cccacagggt acggggcggg acggcccttc tcctccggc tgtaattagc 720	
ttagcgaagag gtaagggtttt aagggtatgtt tggttgttgc ggttattaaat tttaatttacc 780	
tggagcacctt gcttgcataatc acttttttgc aggttggacc ggtgcacca tggactataa 840	
ggaccacgac ggagactaca aggttgcataa tatttgcattt aagagactgt acgataagat 900	
ggccccaagaa aagaagcgga aggtcggtat ccacggagtc ccacgcgcg acaagaagta 960	
caacatcgccg ctggacatcg gcacaaactc tggtggctgc ggcgtgatca ccgcacgacta 1020	
caagggtgcgc acaagaaatc tcaatggatgtt gggcaacaccc gacccgcaca gcatcaagaa 1080	
gaacctgttc gggcccttc tggtgcacag gggccaaacca ggcgcggccca cccggctgaa 1140	
gagaaccgcg aagaagaatg acaccagacg gaagaacccgg atctgtatc tgcaagagat 1200	
cttcagcaac gagatggccca aggtggacca cagottcttc cacagactgg aagagtctt 1260	
cttgggtggaa gaggttgcggaa agacacggcg gaccccccattt ttcggcaaca tcgtggacga 1320	
gggtggcttac caccggaaatg accccccattt ctaccacccgtt agaaagaaatc tggtggacag 1380	
caccgacaag gccgacacttc ggctgtatcta tctggccctt gcccacatga tcaagttccg 1440	
ggggccacttc tgatcgagg ggcgacccatc ccccgacaaac acgcacgtgg acaagctgtt 1500	
catccacgtt gtcgcacccatc aacaccatgtt gttcgaggaa aaccctatca acgcacgcgg 1560	
ctgtggacccg aaggccatcc ttttttttttgc actggaccaatc acgcacggc tggaaaatct 1620	
gatgcggccatc ctggccggcg agaagaagaa tggccctgttgc ggcacacttgc ttgccttgag 1680	
cttggggcttc accccccattt tcaatggatgtt gggccacccatc ggcgaggatg ccaaactgtca 1740	
gctggacccatc gacacccatc acgcacccatc ggcgacccatc ggcgaggatg ccaaactgtca 1800	
gtacggccac ctgtttctgg cccggccaaatc cttgtccgc acgcacccatc ggcgacccatc 1860	
cctggagatgt aacaccggaa tcaaccatccgcg ccccttgcgcg gctcttatca tcaagagata 1920	
cgcacggacac caccggggcc tggccctgttgc gggccacccatc ggcgaggatg ccaaactgtca 1980	
gaaggttttttcc gatggatgttgc gggccacccatc ggcgaggatg ccaaactgtca 2040	
cgggccggcc gacccatc ggcgacccatc ggcgaggatg ccaaactgtca 2100	
caccggggaa ctgtccgtgttgc agcttgcacccatc gggccacccatc ggcgacccatc 2160	
cgacaaacggc accatccatc accggatccatc cttgtccgc acgcacccatc ggcgacccatc 2220	
gcggggatgt ttttttttttcc tcccttgcggaa aacccggggaa agatccggaa agatccgttgc 2280	
cttccgcacatc cccttacttc tggcccttc ggcgaggatg ccaaactgtca 2340	
gaccggggaaatc gggccacccatc ggcgacccatc ggcgaggatg ccaaactgtca 2400	
cggttccgcgc cagatccatc tggcccttc ggcgacccatc ggcgacccatc ggcgacccatc 2460	
gaaggtgttgc cccaaatccatc ggcgacccatc ggcgacccatc ggcgacccatc ggcgacccatc 2520	
caaaatccatc tggcccttc ggcgacccatc ggcgacccatc ggcgacccatc ggcgacccatc 2580	
aaaggccatc gtttgcacccatc ggcgacccatc ggcgacccatc ggcgacccatc ggcgacccatc 2640	
agaggactac ttcaatggatgttgc gggccacccatc ggcgacccatc ggcgacccatc ggcgacccatc 2700	
tccgttacatc gggccacccatc ggcgacccatc ggcgacccatc ggcgacccatc ggcgacccatc 2760	
cttcctggatc aatggggatgttgc gggccacccatc ggcgacccatc ggcgacccatc ggcgacccatc 2820	
gtttggggatgttgc gggccacccatc ggcgacccatc ggcgacccatc ggcgacccatc ggcgacccatc 2880	
caaaatccatc tggcccttc ggcgacccatc ggcgacccatc ggcgacccatc ggcgacccatc 2940	
gtgtatccatc gggccacccatc ggcgacccatc ggcgacccatc ggcgacccatc ggcgacccatc 3000	

-continued

cgacggcttc	gccaacagaa	acttcatgca	gctgatccac	gacgacagcc	tgacctttaa	3060
agaggacatc	cagaagaagccc	aggtgtccgg	ccaggccgat	agcctgcacg	agcacattgc	3120
caatctggcc	ggcagccccg	ccattaagaa	gggcatectg	cagacagtga	aggtgggtgga	3180
cgagctcggt	aaagtgtatgg	gcccgcacaa	gccccgagaac	atcgatcg	aatggcgcag	3240
agagaaccag	accaccaga	agggacagaa	gaacagccgc	gagagaatga	agcggatcga	3300
agaggccatc	aaagagctgg	gcagocagat	cctgaaaagaa	cacccctggg	aaaacaccca	3360
gctgcagaac	gagaagctgt	acctgtacta	cctgcagaat	gggcgggata	tgtacgttgg	3420
ccaggaaactg	gacatcaacc	ggctgtccga	ctacgatgtg	gaccatatcg	tgccctcagag	3480
ctttctgaag	gacgactcca	tcgacaacaa	ggtgtctgacc	agaagcgaca	agaaccgggg	3540
caagagcgac	aacgtgccc	ccgaaagggt	cgtgaaqaag	atgaaact	actggcgca	3600
gctgctgaac	gccaagctgt	ttaccacag	aaagtgcac	aatctgacca	aggccgagag	3660
aggccgcttc	agcgaactgg	ataaggccg	cttcatcaag	agacatctgg	ttggaaacccg	3720
gcagatcaca	aagcacgtgg	cacagatct	ggactccccc	atgaacacta	agtacgcacga	3780
gaatgacaag	ctgatccggg	aagtggaaatg	gatcacccct	aagtccaa	ttgtgtccga	3840
tttccggaaag	gattttcagt	tttacaaag	gcccggatgc	aacaactacc	accacgccca	3900
cgacgcctac	ctgaaacggc	tcgtggaa	cgcctgtatc	aaaaatgtacc	ctaagcttga	3960
aagcgagttc	gtgtacggcg	actacaaggt	gtacgacgtg	cggaagatga	tcgccaagag	4020
cgagcaggaa	atccgcacgg	ctaccccaaa	gtacttcttc	tcagcaaca	tcatgaaact	4080
tttcaagacc	gagatttaccc	ttggcaacgg	cgagatccgg	aaggccctc	tgatcgagac	4140
aaacggcgaa	accggggaga	tcgtgtggaa	taaggccgg	gattttccca	ccgtgcggaa	4200
agtgctgagc	atgccccaa	tgaatatcg	aaaaaaagacc	gagggtcaga	caggcggtt	4260
cagccaaag	tctatctgc	ccaaagaggaa	cagcgataa	ctgatcgcca	gaaagaagga	4320
cttgggaccc	agaagaatcg	ggcggttcc	cagccccacc	ttggctatt	ctgtgttgg	4380
ggtggccaaa	gtggaaaagg	gcaatccaa	gaaaacttgc	atgtgaaag	agctgttgg	4440
gatcaccatc	atggaaagaa	gcagottcga	gaagaatccc	atcgactt	ttggaccaa	4500
gggctacaaa	gaagtggaaa	aggacatgt	catcaagctg	cctaactact	ccctgttgc	4560
gtctggaaaac	ggccggaa	gaatgttgc	cttgcggc	gaactgcaga	aggggaaacga	4620
acttggccctg	ccctccaaat	atgtgaaactt	cctgtaccc	ccggccact	atgagaagat	4680
gaagggttcc	cccgaggata	atgagcaga	acagctgtt	gtggaaacgc	acaagacta	4740
cctggacgag	atcatcgac	agatcagcga	gttctccaa	agagtgtatcc	ttggccgacgc	4800
taatctggac	aaagtgtatcc	ccgcctacaa	caagcacccg	gataagccca	tcagagagca	4860
ggccggagaat	atcatccacc	tgttttaccc	gaccaatctg	ggagccccctg	ccgccttcaa	4920
gtacttttgc	accaccatcg	accggaaagag	gtacaccacg	acccaaagg	tgctgacgc	4980
caccctgtatc	caccagagca	tcaccggct	gtacgagaca	cgatcgacc	tgtctca	5040
ggggggcgc	tttttttttc	tttagttgc	cagcttctt	atgtacgac	ggacgcttta	5100
a						5101

SEQ_ID NO: 60	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	
SEQUENCE: 60		
nnnnnnnnnn nnnnnnnnnn	gttattgtac tctcaagatt tagaaataaa tcttgcagaa	60
gctacaaaga taaggcttca	tgccgaaatc aacaccctgt cattttatgg cagggttttt	120
tcgttattta	atttttt	137
SEQ_ID NO: 61	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: 61		
nnnnnnnnnn nnnnnnnnnn	gttattgtac tctcagaaat gcagaagcta caaagataag	60
gttcatgcc	gaaatcaaca ccctgtcatt ttatggcagg gtgttttgc	120
ttt	tatttttttttt	123
SEQ_ID NO: 62	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	

-continued

```

SEQUENCE: 62          organism = synthetic construct
nnnnnnnnnnn nnnnnnnnnn gttattgtac tctcagaaaat gcagaagcta caaagataag  60
gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgtttttt             110

SEQ ID NO: 63          moltype = DNA  length = 137
FEATURE
misc_feature
1..137
note = Description of Artificial Sequence: Synthetic
polynucleotide
1..20
note = a, c, t, g, unknown or other
1..137
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 63          organism = synthetic construct
nnnnnnnnnnn nnnnnnnnnn gttattgtac tctcaagatt tagaaaataaa tcttgcaagaa  60
gttacaatga taaggcttca tgcccaaattt aacaccctgt cattttatgg cagggtgtt  120
tcgttattta atttttt             137

SEQ ID NO: 64          moltype = DNA  length = 123
FEATURE
misc_feature
1..123
note = Description of Artificial Sequence: Synthetic
polynucleotide
1..20
note = a, c, t, g, unknown or other
1..123
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 64          organism = synthetic construct
nnnnnnnnnnn nnnnnnnnnn gttattgtac tctcagaaaat gcagaagcta caatgataag  60
gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgtttcggtt             120
ttt             123

SEQ ID NO: 65          moltype = DNA  length = 110
FEATURE
misc_feature
1..110
note = Description of Artificial Sequence: Synthetic
polynucleotide
1..20
note = a, c, t, g, unknown or other
1..110
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 65          organism = synthetic construct
nnnnnnnnnnn nnnnnnnnnn gttattgtac tctcagaaaat gcagaagcta caatgataag  60
gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgtttttt             110

SEQ ID NO: 66          moltype = DNA  length = 107
FEATURE
misc_feature
1..107
note = Description of Artificial Sequence: Synthetic
polynucleotide
1..20
note = a, c, t, g, unknown or other
1..107
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 66          organism = synthetic construct
nnnnnnnnnnn nnnnnnnnnn gtttagago tggaaaca cagcgagttt aaataaggct  60
tagtccgtac tcaactgaa aagggtggcac cgattcggtt tttttttt             107

SEQ ID NO: 67          moltype = DNA  length = 4263
FEATURE
misc_feature
1..4263
note = Description of Artificial Sequence: Synthetic
polynucleotide
1..4263
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 67          organism = synthetic construct
atgaaaaggc cggcgccac gaaaaaggcc ggccaggcaa aaaagaaaaa gaccaagccc  60
tacagcatcg gcctggacat cggcaccaat agcgtggctt gggccgtgac caccgacaac 120
tacaagggtgc ccagcaagaa aatgaagggtt ctggcaaca cctccaagaa gtacatcaag  180
aaaaaacctgc tggcgtgct gctgttcgac agcggcattt cagccgaggc cagacggctg  240

```

-continued

aagagaaccg	ccagacggcg	gtacacccgg	cggagaaaaca	gaatcctgta	tctgaagag	300	
atcttcagca	cccgatggc	tacccggac	gacgccttct	tccagcggct	ggacgcacgc	360	
ttccctggtc	ccgacgacaa	gccccggac	aagtacccca	tcttcggcaa	cctggtgaa	420	
gagaaggcct	accacgacg	gttccccacc	atcttaccacc	tgagaaaaat	cctggccgac	480	
agcacaaga	aggccgacct	gagactggtg	tatctggccc	tggcccacat	gatcaagtag	540	
cggggccact	tcctgtatcg	ggggcgggtc	aacagcaaga	acaacgcacat	ccagaagaac	600	
ttccaggact	tcctggacac	ctacaacgc	atcttcgaga	gacgcctgtc	cctggaaaac	660	
agcaagcgc	ttggaaagat	cgtaaaggac	aagatcagca	agctggaaaa	gaaggacccg	720	
atccctgaagc	tggtccccgg	cgagaagaac	agcggaatct	tcagcggat	tctgaagctg	780	
atcgtgggca	accaggcc	cttcggaa	tgttcaacc	ttggacgagaa	agccagctg	840	
cacttcagca	aaagagacta	cgacggac	ctggaaaccc	tgtctggata	tatcggcgac	900	
gactacagcg	acgtgttct	gaaggccaa	aagctgtac	acgtctatct	gctgagccgc	960	
ttccctgacc	tgaccgacaa	cgagacagag	gccccactga	geagcgcct	gattaagccg	1020	
tacaacgacg	acaaagagga	tctggctct	ctgaaagat	acatccggaa	catcagctg	1080	
aaaacactca	atggatgtt	caaggacac	accaagaacg	getacgcgg	ctacatcgac	1140	
ggcaagacca	accaggaa	tttctatgt	tacctaa	gaatgttgc	cgagttcgag	1200	
ggggccact	actttctgga	aaaaatcgac	cgcgaggatt	tcctgeggaa	gcagcggacc	1260	
ttcgacaa	cgagcatccc	cttccat	cattctcg	aaatgcggc	catecttgac	1320	
aaggccca	atgttctacc	attctggcc	aagaacaa	agcggatcga	gaagatctg	1380	
acottccgca	ttccctacta	ctggggcccc	ctggccagag	gcaacagcga	ttttgcctgg	1440	
tccatccgga	agcgcaatga	gaagatcacc	ccctggaa	actggaggat	gategcacaaa	1500	
gagtccacg	ccggggccct	catcaacccg	atgaccacgt	tcgacccgt	cctgcccggag	1560	
gaaaaggatgc	tgccccaagca	cagectgtc	tcggagacat	tcaatgtt	taacgagctg	1620	
accaaaatgc	tggtttatcgc	cggatctatg	cgggactacc	atcttcgta	ctccaaacgag	1680	
aaaaaggaca	tcgtggggct	gtacttcaag	gacaagcgg	aagtggacca	taaggacatc	1740	
atcgatgtacc	tgcaacggccat	ctacggctca	gatggcatcg	agctgaagg	catcgagaag	1800	
cagtcaact	ccggccgt	cacataccac	gactgtgtc	acattatcaa	cgacaaagaa	1860	
tttctggacg	atcccgacaa	cgaggccatc	atcgaa	gaga	tcatccacac	1920	
tttgaggacc	ccggatgtat	caagcggcg	ctgagcaat	tcgagaa	cttcgacaa	1980	
agcgtgtcga	aaaagctgag	cagacggca	tacaccgg	ggggcaat	gagcgc	2040	
ctgatcaacg	gcacccgg	cgagaa	ggc	acaccaaa	tcctggacta	2100	
gacggcatca	gcaacccgaa	cttcatcg	ctgatccac	acgacgcct	gagttcaag	2160	
aaaagatcc	agaaggccca	gatcatcg	gacgaggaca	agggcaacat	caaagaatc	2220	
gtgaagtccc	tgcccccgc	ccccgc	cato	aagaaggaa	tcctgcag	2280	
gttgacgacg	tcgtgaaat	gatggggccg	aaaaagcc	agagcat	gttggaaat	2340	
gctagagaga	accatcac	caatcagg	aaagaca	ca	gccagcag	2400	
ctgaaaatg	ccctgaa	gctggggc	aa	atttctg	aaagaaat	2460	
ctgtccaa	tcgacaacaa	cgccctgc	gac	gacccgc	tctac	2520	
aatggca	acatgtat	aggcgcac	ctgat	atc	ccgcctg	2580	
atcgaccat	ttatccccca	ggcccttct	aa	agaca	ac	caaagtgt	2640
gtgtccctcg	ccagcaac	cgca	at	gtat	ccagcctg	2700	
aagagaaaga	ccttctggta	tcagtc	aaa	agca	tgat	gcca	2760
gacaacccg	ccaa	ggagacgg	ctgagcc	aa	gataa	ggc	2820
cagagac	tggtgg	ccggc	aca	aggc	tgcc	caat	2880
aaatggaa	aca	ccgg	aa	ccgg	ccgt	gatc	2940
ctgaatgtca	ccctgg	cc	agg	actt	ccat	2900	
atcaatgat	tcaccac	ccac	actt	gtt	ccat	3060	
ctgaaatgt	acc	ccat	ccat	gtt	ccat	3120	
tccttc	ccat	ccat	ccat	gtt	ccat	3180	
atctttaa	aga	ccat	ccat	gtt	ccat	3240	
gtgaatgt	ccat	ccat	ccat	gtt	ccat	3300	
cgggtgtc	tttat	ccat	ccat	gtt	ccat	3360	
ctggatcg	gg	ccat	ccat	gtt	ccat	3420	
aacttcaac	ccat	ccat	ccat	gtt	ccat	3480	
ta	ccat	ccat	ccat	gtt	ccat	3540	
aa	ccat	ccat	ccat	gtt	ccat	3600	
tacccgg	aaat	ccat	ccat	gtt	ccat	3660	
atcgatgtc	cta	ccat	ccat	gtt	ccat	3720	
tc	ccat	ccat	ccat	gtt	ccat	3780	
agcc	ttt	ccat	ccat	gtt	ccat	3840	
aacc	ccat	ccat	ccat	gtt	ccat	3900	
ctggatgt	ccat	ccat	ccat	gtt	ccat	3960	
ttcc	ccat	ccat	ccat	gtt	ccat	4020	
ggc	ccat	ccat	ccat	gtt	ccat	4080	
ttcc	ccat	ccat	ccat	gtt	ccat	4140	
gcc	ccat	ccat	ccat	gtt	ccat	4200	
ctgg	ccat	ccat	ccat	gtt	ccat	4260	
taa	ccat	ccat	ccat	gtt	ccat	4263	

```

SEQ ID NO: 68          moltype = DNA  length = 53
FEATURE
misc_feature          Location/Qualifiers
1..53
note = Description of Artificial Sequence: Synthetic primer
source               1..53
mol_type = other DNA
organism = synthetic construct

```

-continued

```

SEQUENCE: 68
tccttagcagg atttctgata ttactgtcac gtttttagagc tatgctgttt tga      53

SEQ ID NO: 69      moltype = DNA  length = 53
FEATURE          Location/Qualifiers
misc_feature     1..53
note = Description of Artificial Sequence: Synthetic primer
source          1..53
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 69
gtgacagtaa tattcagaat cctgcttagga gttttgggac cattcaaaac agc      53

SEQ ID NO: 70      moltype = DNA  length = 25
FEATURE          Location/Qualifiers
misc_feature     1..25
note = Description of Artificial Sequence: Synthetic primer
source          1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 70
gggtttcaag tctttgttagc aagag                          25

SEQ ID NO: 71      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: 71
gccaatgaac gggaaaccctt ggtc                           24

SEQ ID NO: 72      moltype = DNA  length = 25
FEATURE          Location/Qualifiers
misc_feature     1..25
note = Description of Artificial Sequence: Synthetic primer
misc_difference  1..4
note = a, c, t, g, unknown or other
source          1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 72
nnnngacgag gcaatggctg aaatc                          25

SEQ ID NO: 73      moltype = DNA  length = 25
FEATURE          Location/Qualifiers
misc_feature     1..25
note = Description of Artificial Sequence: Synthetic primer
misc_difference  1..4
note = a, c, t, g, unknown or other
source          1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 73
nnnnnttattt ggctcatatt tgctg                          25

SEQ ID NO: 74      moltype = DNA  length = 25
FEATURE          Location/Qualifiers
misc_feature     1..25
note = Description of Artificial Sequence: Synthetic primer
source          1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 74
cttacacca atcgctgcaa cagac                          25

SEQ ID NO: 75      moltype = DNA  length = 45
FEATURE          Location/Qualifiers
misc_feature     1..45
note = Description of Artificial Sequence: Synthetic primer
source          1..45
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 75
caaaaatttct agtttttttt gccttcccc ataaaaaccctt cctta      45

```

-continued

```

SEQ ID NO: 76      moltype = DNA  length = 45
FEATURE
misc_feature
1..45
note = Description of Artificial Sequence: Synthetic primer
source
1..45
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 76
agggttttat ggggaaaggc aaagaagact agaaatttg atacc          45

SEQ ID NO: 77      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = Description of Artificial Sequence: Synthetic primer
source
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 77
cttacgggtgc ataaagtcaa tttcc                           25

SEQ ID NO: 78      moltype = DNA  length = 21
FEATURE
misc_feature
1..21
note = Description of Artificial Sequence: Synthetic primer
source
1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 78
tggctcgatt tcagccattc c                               21

SEQ ID NO: 79      moltype = DNA  length = 43
FEATURE
misc_feature
1..43
note = Description of Artificial Sequence: Synthetic primer
misc_difference
33
note = a, c, t, g, unknown or other
source
1..43
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 79
cttgacgag gcaatggctg aaatcgagcc aaaaaagcgc aag          43

SEQ ID NO: 80      moltype = DNA  length = 43
FEATURE
misc_feature
1..43
note = Description of Artificial Sequence: Synthetic primer
misc_difference
34
note = a, c, t, g, unknown or other
source
1..43
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 80
cttgacgag gcaatggctg aaatcgagcc aaanaagcgc aag          43

SEQ ID NO: 81      moltype = DNA  length = 43
FEATURE
misc_feature
1..43
note = Description of Artificial Sequence: Synthetic primer
misc_difference
35
note = a, c, t, g, unknown or other
source
1..43
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 81
cttgacgag gcaatggctg aaatcgagcc aaaanagcgc aag          43

SEQ ID NO: 82      moltype = DNA  length = 43
FEATURE
misc_feature
1..43
note = Description of Artificial Sequence: Synthetic primer
misc_difference
36
note = a, c, t, g, unknown or other
source
1..43
mol_type = other DNA
organism = synthetic construct

```

-continued

```

SEQUENCE: 82
cttgacgag gcaatggctg aaatcgagcc aaaaangcgc aag 43

SEQ ID NO: 83      moltype = DNA  length = 43
FEATURE          Location/Qualifiers
misc_feature     1..43
note = Description of Artificial Sequence: Synthetic primer
37
note = a, c, t, g, unknown or other
1..43
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 83
cttgacgag gcaatggctg aaatcgagcc aaaaaancgc aag 43

SEQ ID NO: 84      moltype = DNA  length = 43
FEATURE          Location/Qualifiers
misc_feature     1..43
note = Description of Artificial Sequence: Synthetic primer
38
note = a, c, t, g, unknown or other
1..43
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 84
cttgacgag gcaatggctg aaatcgagcc aaaaaagngc aag 43

SEQ ID NO: 85      moltype = DNA  length = 46
FEATURE          Location/Qualifiers
misc_feature     1..46
note = Description of Artificial Sequence: Synthetic primer
39
note = a, c, t, g, unknown or other
1..46
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 85
cttgacgag gcaatggctg aaatcgagcc aaaaaagcnc aagaag 46

SEQ ID NO: 86      moltype = DNA  length = 46
FEATURE          Location/Qualifiers
misc_feature     1..46
note = Description of Artificial Sequence: Synthetic primer
40
note = a, c, t, g, unknown or other
1..46
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 86
cttgacgag gcaatggctg aaatcgagcc aaaaaagcgn aagaag 46

SEQ ID NO: 87      moltype = DNA  length = 46
FEATURE          Location/Qualifiers
misc_feature     1..46
note = Description of Artificial Sequence: Synthetic primer
41
note = a, c, t, g, unknown or other
1..46
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 87
cttgacgag gcaatggctg aaatcgagcc aaaaaagcgc nagaag 46

SEQ ID NO: 88      moltype = DNA  length = 23
FEATURE          Location/Qualifiers
misc_feature     1..23
note = Description of Artificial Sequence: Synthetic primer
1..23
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 88
gcgctttttt ggctcgattt cag 23

SEQ ID NO: 89      moltype = DNA  length = 40
FEATURE          Location/Qualifiers
misc_feature     1..40

```

-continued

```

misc_difference      note = Description of Artificial Sequence: Synthetic primer
31
note = a, c, t, g, unknown or other
source              1..40
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 89
caatggctga aatcgagccaaaaaagcgca nagaagaaatc                      40

SEQ ID NO: 90          moltype = DNA length = 40
FEATURE               Location/Qualifiers
misc_feature          1..40
note = Description of Artificial Sequence: Synthetic primer
misc_difference        32
note = a, c, t, g, unknown or other
source                1..40
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 90
caatggctga aatcgagccaaaaaagcgca anaagaaatc                      40

SEQ ID NO: 91          moltype = DNA length = 40
FEATURE               Location/Qualifiers
misc_feature          1..40
note = Description of Artificial Sequence: Synthetic primer
misc_difference        33
note = a, c, t, g, unknown or other
source                1..40
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 91
caatggctga aatcgagccaaaaaagcgca agnagaaatc                      40

SEQ ID NO: 92          moltype = DNA length = 40
FEATURE               Location/Qualifiers
misc_feature          1..40
note = Description of Artificial Sequence: Synthetic primer
misc_difference        34
note = a, c, t, g, unknown or other
source                1..40
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 92
caatggctga aatcgagccaaaaaagcgca agangaaatc                      40

SEQ ID NO: 93          moltype = DNA length = 40
FEATURE               Location/Qualifiers
misc_feature          1..40
note = Description of Artificial Sequence: Synthetic primer
misc_difference        35
note = a, c, t, g, unknown or other
source                1..40
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 93
caatggctga aatcgagccaaaaaagcgca agaanaaaatc                      40

SEQ ID NO: 94          moltype = DNA length = 44
FEATURE               Location/Qualifiers
misc_feature          1..44
note = Description of Artificial Sequence: Synthetic primer
misc_difference        36
note = a, c, t, g, unknown or other
source                1..44
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 94
caatggctga aatcgagccaaaaaagcgca agaagnaatc aacc                      44

SEQ ID NO: 95          moltype = DNA length = 44
FEATURE               Location/Qualifiers
misc_feature          1..44
note = Description of Artificial Sequence: Synthetic primer
misc_difference        37
note = a, c, t, g, unknown or other
source                1..44

```

-continued

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 95
caatggctga aatcgagcca aaaaagcgc a agaaganatc aacc        44

SEQ ID NO: 96      moltype = DNA length = 44
FEATURE           Location/Qualifiers
misc_feature      1..44
note = Description of Artificial Sequence: Synthetic primer
misc_difference   38
note = a, c, t, g, unknown or other
source            1..44
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 96
caatggctga aatcgagcca aaaaagcgc a agaagaanc aacc        44

SEQ ID NO: 97      moltype = DNA length = 44
FEATURE           Location/Qualifiers
misc_feature      1..44
note = Description of Artificial Sequence: Synthetic primer
misc_difference   39
note = a, c, t, g, unknown or other
source            1..44
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 97
caatggctga aatcgagcca aaaaagcgc a agaagaaaanc aacc        44

SEQ ID NO: 98      moltype = DNA length = 47
FEATURE           Location/Qualifiers
misc_feature      1..47
note = Description of Artificial Sequence: Synthetic primer
misc_difference   40
note = a, c, t, g, unknown or other
source            1..47
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 98
caatggctga aatcgagcca aaaaagcgc a agaagaaaatn aaccaggc        47

SEQ ID NO: 99      moltype = DNA length = 47
FEATURE           Location/Qualifiers
misc_feature      1..47
note = Description of Artificial Sequence: Synthetic primer
misc_difference   41
note = a, c, t, g, unknown or other
source            1..47
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 99
caatggctga aatcgagcca aaaaagcgc a agaagaaaatc naccaggc        47

SEQ ID NO: 100     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: 100
gatcctccat ccgtacaaacc cacaacctg g                         31

SEQ ID NO: 101     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: 101
aattccaggg ttgtgggttg tacggatgga g                         31

SEQ ID NO: 102     moltype = DNA length = 34
FEATURE           Location/Qualifiers
misc_feature      1..34

```

-continued

```

source          note = Description of Artificial Sequence: Synthetic primer
                1..34
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 102
catggatcct atttcttaat aactaaaaat atgg                                34

SEQ ID NO: 103      moltype = DNA  length = 33
FEATURE           Location/Qualifiers
misc_feature      1..33
note = Description of Artificial Sequence: Synthetic primer
source          1..33
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 103
catgaattca actcaacaag tctcagtgtg ctg                                33

SEQ ID NO: 104      moltype = DNA  length = 35
FEATURE           Location/Qualifiers
misc_feature      1..35
note = Description of Artificial Sequence: Synthetic primer
source          1..35
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 104
aaacatttt tctccattta ggaaaaagga tgctg                                35

SEQ ID NO: 105      moltype = DNA  length = 35
FEATURE           Location/Qualifiers
misc_feature      1..35
note = Description of Artificial Sequence: Synthetic primer
source          1..35
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 105
aaaacagcat ccttttcct aaatggagaa aaaaat                                35

SEQ ID NO: 106      moltype = DNA  length = 35
FEATURE           Location/Qualifiers
misc_feature      1..35
note = Description of Artificial Sequence: Synthetic primer
source          1..35
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 106
aaaccttaaa tcagtcacaa atagcagcaa aattg                                35

SEQ ID NO: 107      moltype = DNA  length = 35
FEATURE           Location/Qualifiers
misc_feature      1..35
note = Description of Artificial Sequence: Synthetic primer
source          1..35
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 107
aaaacaattt tgctgctatt tgtgactgat ttaag                                35

SEQ ID NO: 108      moltype = DNA  length = 35
FEATURE           Location/Qualifiers
misc_feature      1..35
note = Description of Artificial Sequence: Synthetic primer
source          1..35
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 108
aaactttca tcatacgacc aatctgctt atttg                                35

SEQ ID NO: 109      moltype = DNA  length = 35
FEATURE           Location/Qualifiers
misc_feature      1..35
note = Description of Artificial Sequence: Synthetic primer
source          1..35
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 109
aaaacacaata aaggcagattg gtcgtatgt gaaaa                                35

```

-continued

SEQ ID NO: 110 moltype = DNA length = 35
FEATURE Location/Qualifiers
misc_feature 1..35
note = Description of Artificial Sequence: Synthetic primer
source 1..35
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 110
aaactcgtcc agaaggattac gtaaaaagaaa tcgag 35

SEQ ID NO: 111 moltype = DNA length = 35
FEATURE Location/Qualifiers
misc_feature 1..35
note = Description of Artificial Sequence: Synthetic primer
source 1..35
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 111
aaaactcgat ttcttttacg ataacttctg gacga 35

SEQ ID NO: 112 moltype = DNA length = 35
FEATURE Location/Qualifiers
misc_feature 1..35
note = Description of Artificial Sequence: Synthetic primer
source 1..35
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 112
aaacaatctc tccaaaggttt ccttaaaaaat ctctg 35

SEQ ID NO: 113 moltype = DNA length = 35
FEATURE Location/Qualifiers
misc_feature 1..35
note = Description of Artificial Sequence: Synthetic primer
source 1..35
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 113
aaaacagaga ttttaagga aaccttggag agatt 35

SEQ ID NO: 114 moltype = DNA length = 35
FEATURE Location/Qualifiers
misc_feature 1..35
note = Description of Artificial Sequence: Synthetic primer
source 1..35
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 114
aaacgccatc gtcaggaaga agctatgctt gagtg 35

SEQ ID NO: 115 moltype = DNA length = 35
FEATURE Location/Qualifiers
misc_feature 1..35
note = Description of Artificial Sequence: Synthetic primer
source 1..35
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 115
aaaacactca agcatagctt cttcctgacg atggc 35

SEQ ID NO: 116 moltype = DNA length = 35
FEATURE Location/Qualifiers
misc_feature 1..35
note = Description of Artificial Sequence: Synthetic primer
source 1..35
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 116
aaacatctct atacttattg aaatttcttt gtatg 35

SEQ ID NO: 117 moltype = DNA length = 35
FEATURE Location/Qualifiers
misc_feature 1..35
note = Description of Artificial Sequence: Synthetic primer
source 1..35

-continued

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 117
aaaacataca aagaaatttc aataagtata gagat 35
SEQ ID NO: 118      moltype = DNA  length = 35
FEATURE          Location/Qualifiers
misc_feature    1..35
                note = Description of Artificial Sequence: Synthetic primer
source          1..35
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 118
aaacttagctg tgatagtcgg caaaaccagg ctccg 35
SEQ ID NO: 119      moltype = DNA  length = 35
FEATURE          Location/Qualifiers
misc_feature    1..35
                note = Description of Artificial Sequence: Synthetic primer
source          1..35
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 119
aaaacgaagg ctgggtttgc ggactatcac agcta 35
SEQ ID NO: 120      moltype = DNA  length = 35
FEATURE          Location/Qualifiers
misc_feature    1..35
                note = Description of Artificial Sequence: Synthetic primer
source          1..35
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 120
aaacatcgga aggtcgagca agtaattatc ttttg 35
SEQ ID NO: 121      moltype = DNA  length = 35
FEATURE          Location/Qualifiers
misc_feature    1..35
                note = Description of Artificial Sequence: Synthetic primer
source          1..35
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 121
aaaacacaaaag ataattactt gctcgacctt ccgat 35
SEQ ID NO: 122      moltype = DNA  length = 35
FEATURE          Location/Qualifiers
misc_feature    1..35
                note = Description of Artificial Sequence: Synthetic primer
source          1..35
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 122
aaacaagatg gtatcgaaaa gtaagtgaca ataag 35
SEQ ID NO: 123      moltype = DNA  length = 35
FEATURE          Location/Qualifiers
misc_feature    1..35
                note = Description of Artificial Sequence: Synthetic primer
source          1..35
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 123
aaaacttatt gtcacttact ttgcgatacc atctt 35
SEQ ID NO: 124      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: 124
gagacccttg agcttccgag actggtctca gttttgggac cattcaaac ag 52
SEQ ID NO: 125      moltype = DNA  length = 52

```

-continued

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: 125	
ttagagccat ctcggaaagct caaaggctcgttttttagagc tatgctgttt tg	52
SEQ ID NO: 126	moltype = DNA length = 35
FEATURE	Location/Qualifiers
misc_feature	1..35
	note = Description of Artificial Sequence: Synthetic primer
source	1..35
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 126	
aaactactttt acgcagcgcg gagttcgggtt ttttg	35
SEQ ID NO: 127	moltype = DNA length = 35
FEATURE	Location/Qualifiers
misc_feature	1..35
	note = Description of Artificial Sequence: Synthetic primer
source	1..35
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 127	
aaaaacaaaaaa accgaactcc gcgctgcgtta aagta	35
SEQ ID NO: 128	moltype = DNA length = 45
FEATURE	Location/Qualifiers
misc_feature	1..45
	note = Description of Artificial Sequence: Synthetic primer
source	1..45
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 128	
atgccggtagc tgccggggcct cttgcggat tacgaaatca tcctg	45
SEQ ID NO: 129	moltype = DNA length = 45
FEATURE	Location/Qualifiers
misc_feature	1..45
	note = Description of Artificial Sequence: Synthetic primer
source	1..45
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 129	
gtgactggcg atgctgtcgg aatggacgat cacactactc ttctt	45
SEQ ID NO: 130	moltype = DNA length = 50
FEATURE	Location/Qualifiers
misc_feature	1..50
	note = Description of Artificial Sequence: Synthetic primer
source	1..50
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 130	
ttaagaaaata atcttcatct aaaatataact tcagtcacct cctagctgac	50
SEQ ID NO: 131	moltype = DNA length = 48
FEATURE	Location/Qualifiers
misc_feature	1..48
	note = Description of Artificial Sequence: Synthetic primer
source	1..48
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 131	
attgatttga gtcagctagg aggtgactga agtatattttt agatgaag	48
SEQ ID NO: 132	moltype = DNA length = 85
FEATURE	Location/Qualifiers
misc_feature	1..85
	note = Description of Artificial Sequence: Synthetic primer
source	1..85
	mol_type = other DNA
	organism = synthetic construct

-continued

```

SEQUENCE: 132
gagacccttg agcttccgag actggctca gttttggac cattcaaac agcatagtc 60
taaaacctcg tagactattt ttgtc 85

SEQ ID NO: 133      moltype = DNA  length = 84
FEATURE
misc_feature        Location/Qualifiers
1..84
note = Description of Artificial Sequence: Synthetic primer
source
1..84
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 133
gagaccagtc tcggaagctc aaaggctcg ttttagatc atgctgttt gaatggccc 60
aaaacttcag cacactgaga ctg 84

SEQ ID NO: 134      moltype = DNA  length = 21
FEATURE
misc_feature        Location/Qualifiers
1..21
note = Description of Artificial Sequence: Synthetic primer
source
1..21
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 134
agtcatcca gcaacaaatg g 21

SEQ ID NO: 135      moltype = DNA  length = 31
FEATURE
misc_feature        Location/Qualifiers
1..31
note = Description of Artificial Sequence: Synthetic primer
source
1..31
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 135
cgtggtaaat cggtataacgt tccaaatgaa g 31

SEQ ID NO: 136      moltype = DNA  length = 22
FEATURE
misc_feature        Location/Qualifiers
1..22
note = Description of Artificial Sequence: Synthetic primer
source
1..22
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 136
tgctttttt cacaacaaag gg 22

SEQ ID NO: 137      moltype = DNA  length = 21
FEATURE
misc_feature        Location/Qualifiers
1..21
note = Description of Artificial Sequence: Synthetic primer
source
1..21
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 137
aagccaaatg ttggcaccac c 21

SEQ ID NO: 138      moltype = DNA  length = 22
FEATURE
misc_feature        Location/Qualifiers
1..22
note = Description of Artificial Sequence: Synthetic primer
source
1..22
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 138
gtatctttt cagtccatgt gg 22

SEQ ID NO: 139      moltype = DNA  length = 45
FEATURE
misc_feature        Location/Qualifiers
1..45
note = Description of Artificial Sequence: Synthetic primer
source
1..45
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 139
cgtttgtga actaatgggt gcaaattacg aatcttctcc tgacg 45

SEQ ID NO: 140      moltype = DNA  length = 45

```

-continued

FEATURE	Location/Qualifiers
misc_feature	1..45
	note = Description of Artificial Sequence: Synthetic primer
source	1..45
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 140	
cgtcaggaga agattcgtaa tttgcaccca ttagttcaac aaacg	45
SEQ ID NO: 141	moltype = DNA length = 48
FEATURE	Location/Qualifiers
misc_feature	1..48
	note = Description of Artificial Sequence: Synthetic primer
source	1..48
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 141	
gatattatgg agcctatttt tgtgggttt taggcataaa actatatg	48
SEQ ID NO: 142	moltype = DNA length = 48
FEATURE	Location/Qualifiers
misc_feature	1..48
	note = Description of Artificial Sequence: Synthetic primer
source	1..48
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 142	
catatagttt tatgcctaaa aacccacaaa aataggctcc ataataatc	48
SEQ ID NO: 143	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: 143	
attatttcctt aataactaaa aatatgg	27
SEQ ID NO: 144	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: 144	
cgtgtacaat tgctagcgta cggc	24
SEQ ID NO: 145	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: 145	
gcacccgtga tcactagtcc tagg	24
SEQ ID NO: 146	moltype = DNA length = 47
FEATURE	Location/Qualifiers
misc_feature	1..47
	note = Description of Artificial Sequence: Synthetic primer
source	1..47
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 146	
cctaggacta gtgatcaccg gtgcaaataat gagccaaata aatataat	47
SEQ ID NO: 147	moltype = DNA length = 44
FEATURE	Location/Qualifiers
misc_feature	1..44
	note = Description of Artificial Sequence: Synthetic primer
source	1..44
	mol_type = other DNA
	organism = synthetic construct

-continued

```

SEQUENCE: 147
gccgtacgct agcaatttgta cacgtttgtt gaactaatgg gtgc          44

SEQ ID NO: 148      moltype = DNA  length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = Description of Artificial Sequence: Synthetic primer
source            1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 148
ttcaatattt cccatatttgat tctcc                           25

SEQ ID NO: 149      moltype = DNA  length = 47
FEATURE           Location/Qualifiers
misc_feature      1..47
note = Description of Artificial Sequence: Synthetic primer
source            1..47
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 149
ccatattttt agtttataaag aaataataacc agccatcagt caccctcc        47

SEQ ID NO: 150      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: 150
agacgattca atagacaataa agg                               23

SEQ ID NO: 151      moltype = DNA  length = 45
FEATURE           Location/Qualifiers
misc_feature      1..45
note = Description of Artificial Sequence: Synthetic primer
source            1..45
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 151
gttttgggac cattcaaaac agcatagctc taaaacctcg tagac          45

SEQ ID NO: 152      moltype = DNA  length = 50
FEATURE           Location/Qualifiers
misc_feature      1..50
note = Description of Artificial Sequence: Synthetic primer
source            1..50
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 152
gctatgctgt ttgaatggt cccaaaacca ttatTTAAC acacgagggtg        50

SEQ ID NO: 153      moltype = DNA  length = 50
FEATURE           Location/Qualifiers
misc_feature      1..50
note = Description of Artificial Sequence: Synthetic primer
source            1..50
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 153
gctatgctgt ttgaatggt cccaaaacgc acccatttagt tcaacaaacg          50

SEQ ID NO: 154      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: 154
aattctttc ttcatcatcg gtc                                23

SEQ ID NO: 155      moltype = DNA  length = 24
FEATURE           Location/Qualifiers
misc_feature      1..24

```

-continued

```

source          note = Description of Artificial Sequence: Synthetic primer
1..24
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 155
aagaaagaat gaagattgtt catg                                24

SEQ ID NO: 156      moltype = DNA  length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = Description of Artificial Sequence: Synthetic primer
source          1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 156
ggtaactaatac aaaatagtga ggagg                                25

SEQ ID NO: 157      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: 157
gtttttcaaa atctgcgggtt gcg                                23

SEQ ID NO: 158      moltype = DNA  length = 26
FEATURE           Location/Qualifiers
misc_feature      1..26
note = Description of Artificial Sequence: Synthetic primer
source          1..26
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 158
aaaaattgaa aaaatggtgg aaacac                                26

SEQ ID NO: 159      moltype = DNA  length = 53
FEATURE           Location/Qualifiers
misc_feature      1..53
note = Description of Artificial Sequence: Synthetic primer
source          1..53
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 159
atttcgtaaa cggtatcggt ttcttttaaa gtttgggac cattcaaaac agc      53

SEQ ID NO: 160      moltype = DNA  length = 53
FEATURE           Location/Qualifiers
misc_feature      1..53
note = Description of Artificial Sequence: Synthetic primer
source          1..53
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 160
tttaaaagaa accgataccg tttacgaaat gtttagagc tatgctgtt tga      53

SEQ ID NO: 161      moltype = DNA  length = 53
FEATURE           Location/Qualifiers
misc_feature      1..53
note = Description of Artificial Sequence: Synthetic primer
source          1..53
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 161
aacacggatc ggtttctttt aaattcaatt gtttgggac cattcaaaac agc      53

SEQ ID NO: 162      moltype = DNA  length = 53
FEATURE           Location/Qualifiers
misc_feature      1..53
note = Description of Artificial Sequence: Synthetic primer
source          1..53
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 162
aattgaattt aaaagaaacc gataccgttt gtttagagc tatgctgtt tga      53

```

-continued

```

SEQ ID NO: 163      moltype = DNA  length = 39
FEATURE
misc_feature
1..39
note = Description of Artificial Sequence: Synthetic primer
source
1..39
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 163
gttccttaaa caaaaacggt atcggtttct tttaaattc          39

SEQ ID NO: 164      moltype = DNA  length = 47
FEATURE
misc_feature
1..47
note = Description of Artificial Sequence: Synthetic primer
source
1..47
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 164
gaaaaccgata ccgttttggt ttaaggaaca ggttaaagggc atttaac          47

SEQ ID NO: 165      moltype = DNA  length = 22
FEATURE
misc_feature
1..22
note = Description of Artificial Sequence: Synthetic primer
source
1..22
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 165
cgatttcagc cattgcctcg tc          22

SEQ ID NO: 166      moltype = DNA  length = 56
FEATURE
misc_feature
1..56
note = Description of Artificial Sequence: Synthetic primer
misc_difference
29..33
note = a, c, t, g, unknown or other
source
1..56
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 166
gcctttgacg aggcaatggc tgaaaatcggn nnnaaaaagc gcaagaagaa atcaac          56

SEQ ID NO: 167      moltype = DNA  length = 53
FEATURE
misc_feature
1..53
note = Description of Artificial Sequence: Synthetic primer
source
1..53
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 167
tccgtacaac ccacaaccct gctagtgagc gtttgggac cattcaaac agc          53

SEQ ID NO: 168      moltype = DNA  length = 53
FEATURE
misc_feature
1..53
note = Description of Artificial Sequence: Synthetic primer
source
1..53
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 168
gctcaactgc agggttgtgg gttgtacgga gtttagagc tatgctgttt tga          53

SEQ ID NO: 169      moltype = DNA  length = 23
FEATURE
misc_feature
1..23
note = Description of Artificial Sequence: Synthetic primer
source
1..23
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 169
ttgttgcac ttttccttct ttc          23

SEQ ID NO: 170      moltype = DNA  length = 41
FEATURE
misc_feature
1..41

```

-continued

```

source          note = Description of Artificial Sequence: Synthetic primer
                1..41
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 170  cagggttgtg ggttggcg atggagttaa ctcccatctc c           41
SEQ ID NO: 171      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: 171  gggaggttaac tccatcgcaa caaccacaa ccctgctagt g           41
SEQ ID NO: 172      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: 172  gtggtatcta tcgtgatgtg ac                         22
SEQ ID NO: 173      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: 173  ttaccgaaac ggaatttatac tgc                         23
SEQ ID NO: 174      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: 174  aaagctagag ttccgcaatt gg                         22
SEQ ID NO: 175      moltype = DNA  length = 37
FEATURE          Location/Qualifiers
misc_feature     1..37
note = Description of Artificial Sequence: Synthetic primer
source          1..37
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 175  gtgggttgta cggattgagt taactccat ctccctc           37
SEQ ID NO: 176      moltype = DNA  length = 38
FEATURE          Location/Qualifiers
misc_feature     1..38
note = Description of Artificial Sequence: Synthetic primer
source          1..38
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 176  gatgggagtt aactcaatcc gtacaaccca caaccctg           38
SEQ ID NO: 177      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: 177  gcttcaccta ttgcagcacc aattgaccac atgaagatag           40

```

-continued

```

SEQ ID NO: 178      moltype = DNA  length = 41
FEATURE
misc_feature
1..41
note = Description of Artificial Sequence: Synthetic primer
source
1..41
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 178
gtggtaatt ggtgctcaa tagtgaaac taatggat g          41

SEQ ID NO: 179      moltype = DNA  length = 39
FEATURE
misc_feature
1..39
note = Description of Artificial Sequence: Synthetic primer
source
1..39
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 179
ctgatttata ttaattttaa gacattatgc ttcacccctc          39

SEQ ID NO: 180      moltype = DNA  length = 40
FEATURE
misc_feature
1..40
note = Description of Artificial Sequence: Synthetic primer
source
1..40
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 180
gcataatgtc tcaaaaattaa tacaaatcg tgaatcatg          40

SEQ ID NO: 181      moltype = DNA  length = 52
FEATURE
misc_feature
1..52
note = Description of Artificial Sequence: Synthetic primer
source
1..52
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 181
gttttggac cattcaaaac agcatagctc taaaacgtga cagtaatatc ag          52

SEQ ID NO: 182      moltype = DNA  length = 53
FEATURE
misc_feature
1..53
note = Description of Artificial Sequence: Synthetic primer
source
1..53
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 182
gttttagagc tatgctgttt tgaatggtcc caaaacgctc actaggaggg ttg          53

SEQ ID NO: 183      moltype = DNA  length = 59
FEATURE
misc_feature
1..59
note = Description of Artificial Sequence: Synthetic primer
source
1..59
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 183
atactttacg cagcgccggag ttccgggttt taggagtggg agtatataca cgagtacat  59

SEQ ID NO: 184      moltype = DNA  length = 33
FEATURE
misc_feature
1..33
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source
1..33
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 184
gctcaatgc agggttgtgg gttgtacgga tgg          33

SEQ ID NO: 185      moltype = DNA  length = 33
FEATURE
misc_feature
1..33
note = Description of Artificial Sequence: Synthetic

```

-continued

```

source          oligonucleotide
               1..33
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 185
tcttagcagg atttcgtata ttactgtcac tgg                         33

SEQ ID NO: 186      moltype = DNA  length = 33
FEATURE           Location/Qualifiers
misc_feature      1..33
note = Description of Artificial Sequence: Synthetic
               oligonucleotide
source          1..33
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 186
tttaaaagaa accgataccg tttacgaaat tgg                         33

SEQ ID NO: 187      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: 187
ggaaccattc ataacagcat agcaagttat aataaggcta gtccgttatac aacttggaaa 60
agtggcaccg agtcgggtct tttt                           84

SEQ ID NO: 188      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
SEQUENCE: 188
gtttagagac tatgctgtta tgaatggtcc caaac                         36

SEQ ID NO: 189      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: 189
ggaaccattc aatacagcat agcaagttaa tataaggcta gtccgttatac aacttggaaa 60
agtggcaccg agtcgggtct tttt                           84

SEQ ID NO: 190      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
SEQUENCE: 190
gtttagagac tatgctgtat tgaatggtcc caaac                         36

SEQ ID NO: 191      moltype = DNA  length = 103
FEATURE           Location/Qualifiers
misc_feature      1..103
note = Description of Artificial Sequence: Synthetic
               polynucleotide
misc_difference   1..20
note = a, c, t, g, unknown or other
source          1..103
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 191
nnnnnnnnnn nnnnnnnnnn gtttagagc tagaaatagc aagttaaaat aaggctagtc 60

```

-continued

```

cgtttatcaac ttgaaaaagt ggcaccgagt cggtgcttt ttt          103
SEQ ID NO: 192      moltype = DNA  length = 103
FEATURE
misc_feature        Location/Qualifiers
1..103
note = Description of Artificial Sequence: Synthetic
polynucleotide
misc_difference     1..20
note = a, c, t, g, unknown or other
source              1..103
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 192
nnnnnnnnnn nnnnnnnnnn gtattagago tagaaatagc aagttaatat aaggctagtc 60
cgtttatcaac ttgaaaaagt ggcaccgagt cggtgcttt ttt          103

SEQ ID NO: 193      moltype = DNA  length = 123
FEATURE
misc_feature        Location/Qualifiers
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: 193
nnnnnnnnnn nnnnnnnnnn gtttagago tatgctgtt tggaaacaaa acagcatagc 60
aagttaaat aaggctagtc cgtttatcaac ttgaaaaagt ggcaccgagt cggtgcttt 120
ttt                  123

SEQ ID NO: 194      moltype = DNA  length = 123
FEATURE
misc_feature        Location/Qualifiers
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: 194
nnnnnnnnnn nnnnnnnnnn gtattagago tatgctgtat tggaaacaat acagcatagc 60
aagttaatat aaggctagtc cgtttatcaac ttgaaaaagt ggcaccgagt cggtgcttt 120
ttt                  123

SEQ ID NO: 195      moltype = DNA  length = 20
FEATURE
source             Location/Qualifiers
1..20
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 195
gtcacacctca atgactaggg                                     20

SEQ ID NO: 196      moltype = DNA  length = 23
FEATURE
source             Location/Qualifiers
1..23
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 196
gacatcgatg tcctccccat tgg                                     23

SEQ ID NO: 197      moltype = DNA  length = 23
FEATURE
source             Location/Qualifiers
1..23
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 197
gagtccgagc agaagaagaa ggg                                     23

SEQ ID NO: 198      moltype = DNA  length = 23
FEATURE
source             Location/Qualifiers
1..23
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 198

```

-continued

gccccaccgg ttgatgtgat ggg	23
SEQ ID NO: 199 moltype = DNA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = unassigned DNA organism = Homo sapiens	
SEQUENCE: 199 ggggcacaga tgagaaaactc agg	23
SEQ ID NO: 200 moltype = DNA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = unassigned DNA organism = Homo sapiens	
SEQUENCE: 200 gtacaaacgg cagaagctgg agg	23
SEQ ID NO: 201 moltype = DNA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = unassigned DNA organism = Homo sapiens	
SEQUENCE: 201 ggcagaagct ggaggaggaa ggg	23
SEQ ID NO: 202 moltype = DNA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = unassigned DNA organism = Homo sapiens	
SEQUENCE: 202 ggagcccttc ttcttctgct cgg	23
SEQ ID NO: 203 moltype = DNA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = unassigned DNA organism = Homo sapiens	
SEQUENCE: 203 ggcaaccac aaacccacga ggg	23
SEQ ID NO: 204 moltype = DNA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = unassigned DNA organism = Homo sapiens	
SEQUENCE: 204 gctcccatca catcaaccgg tgg	23
SEQ ID NO: 205 moltype = DNA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = unassigned DNA organism = Homo sapiens	
SEQUENCE: 205 gtggcgatt gccacgaaagc agg	23
SEQ ID NO: 206 moltype = DNA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = unassigned DNA organism = Homo sapiens	
SEQUENCE: 206 ggcagagtgc tgcttgctgc tgg	23
SEQ ID NO: 207 moltype = DNA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = unassigned DNA organism = Homo sapiens	
SEQUENCE: 207 gccctctgcgt gggcccaagc tgg	23
SEQ ID NO: 208 moltype = DNA length = 23 FEATURE Location/Qualifiers	

-continued

```

source          1..23
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 208
gagtggccag agtccagctt ggg                                23

SEQ ID NO: 209      moltype = DNA  length = 23
FEATURE
source          1..23
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 209
ggcctccccca aagcctggcc agg                                23

SEQ ID NO: 210      moltype = DNA  length = 23
FEATURE
source          1..23
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 210
ggggccgaga ttgggtgttc agg                                23

SEQ ID NO: 211      moltype = DNA  length = 23
FEATURE
source          1..23
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 211
gtggcgagag gggccgagat tgg                                23

SEQ ID NO: 212      moltype = DNA  length = 23
FEATURE
source          1..23
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 212
gagtgcggcc gagggcgggggc ggg                                23

SEQ ID NO: 213      moltype = DNA  length = 23
FEATURE
source          1..23
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 213
ggagtgcgcg cggaggcgaaaaa cggg                                23

SEQ ID NO: 214      moltype = DNA  length = 23
FEATURE
source          1..23
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 214
ggagaggaggat gcccggagg cggt                                23

SEQ ID NO: 215      moltype = DNA  length = 20
FEATURE
misc_feature    1..20
               note = Description of Artificial Sequence: Synthetic primer
source          1..20
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 215
ccatcccctt ctgtgaatgt                                20

SEQ ID NO: 216      moltype = DNA  length = 20
FEATURE
misc_feature    1..20
               note = Description of Artificial Sequence: Synthetic primer
source          1..20
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 216
ggagattgga gacacggaga                                20

SEQ ID NO: 217      moltype = DNA  length = 20
FEATURE

```

-continued

misc_feature	1..20	
source	note = Description of Artificial Sequence: Synthetic primer	
	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 217		
aagcaccgac tcggtgccac		20
SEQ ID NO: 218	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
source	note = Description of Artificial Sequence: Synthetic primer	
	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 218		
tcacacctcaa tgacttagggg		20
SEQ ID NO: 219	moltype = DNA length = 22	
FEATURE	Location/Qualifiers	
misc_feature	1..22	
source	note = Description of Artificial Sequence: Synthetic primer	
	1..22	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 219		
caagttgata acggactagc ct		22
SEQ ID NO: 220	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23	
source	note = Description of Artificial Sequence: Synthetic primer	
	1..23	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 220		
agtccgagca gaagaagaag ttt		23
SEQ ID NO: 221	moltype = DNA length = 25	
FEATURE	Location/Qualifiers	
misc_feature	1..25	
source	note = Description of Artificial Sequence: Synthetic primer	
	1..25	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 221		
tttcaagttg ataacggact agcct		25
SEQ ID NO: 222	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
source	note = Description of Artificial Sequence: Synthetic primer	
	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 222		
aaacagcaga ttgcgttga		20
SEQ ID NO: 223	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
source	note = Description of Artificial Sequence: Synthetic primer	
	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 223		
tcatccgctc gatgaagctc		20
SEQ ID NO: 224	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
source	note = Description of Artificial Sequence: Synthetic primer	
	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 224		

-continued

tccaaaatca agtggggcga	20
SEQ ID NO: 225	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: 225	
tgatgaccct tttggctccc	20
SEQ ID NO: 226	moltype = DNA length = 45
FEATURE	Location/Qualifiers
misc_feature	1..45
	note = Description of Artificial Sequence: Synthetic primer
source	1..45
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 226	
gaggaattct tttttgtty gaatatgttg gaggtttttt ggaag	45
SEQ ID NO: 227	moltype = DNA length = 42
FEATURE	Location/Qualifiers
misc_feature	1..42
	note = Description of Artificial Sequence: Synthetic primer
source	1..42
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 227	
gagaagctta aataaaaaac racaatactc aacccaacaa cc	42
SEQ ID NO: 228	moltype = DNA length = 17
FEATURE	Location/Qualifiers
misc_feature	1..17
	note = Description of Artificial Sequence: Synthetic primer
source	1..17
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 228	
caggaaacag ctatgac	17
SEQ ID NO: 229	moltype = DNA length = 39
FEATURE	Location/Qualifiers
misc_feature	1..39
	note = Description of Artificial Sequence: Synthetic primer
source	1..39
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 229	
gcctctagag gtacctgagg gcctatttc catgattcc	39
SEQ ID NO: 230	moltype = DNA length = 133
FEATURE	Location/Qualifiers
misc_feature	1..133
	note = Description of Artificial Sequence: Synthetic primer
misc_difference	92..111
	note = a, c, t, g, unknown or other
source	1..133
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 230	
acctctagaa aaaaaggcacc gactcggtgc cacttttca agttgataac ggacttagcct	60
tattnaact tgcttatttct agctctaaaa cnnnnnnnnn nnnnnnnnn nggtgttcg	120
tcctttccac aag	133
SEQ ID NO: 231	moltype = DNA length = 133
FEATURE	Location/Qualifiers
misc_feature	1..133
	note = Description of Artificial Sequence: Synthetic primer
misc_difference	92..111
	note = a, c, t, g, unknown or other
source	1..133
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 231	

-continued

```

acctctagaa aaaaaggcacc gactcgggtgc cacttttca agttgataac ggactagcct 60
tatattaact tgctatttct agctctaata cnnnnnnnnn nnnnnnnnnn nggtgttcg 120
tccttccac aag 133

SEQ ID NO: 232 moltype = DNA length = 153
FEATURE Location/Qualifiers
misc_feature 1..153
note = Description of Artificial Sequence: Synthetic primer
misc_difference 112..131
note = a, c, t, g, unknown or other
source 1..153
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 232
acctctagaa aaaaaggcacc gactcgggtgc cacttttca agttgataac ggactagcct 60
tatattaact tgctatgctg ttttgttcc aaacacgcat agctctaaa cnnnnnnnnn 120
nnnnnnnnnn nggtgttcg tccttccac aag 153

SEQ ID NO: 233 moltype = DNA length = 153
FEATURE Location/Qualifiers
misc_feature 1..153
note = Description of Artificial Sequence: Synthetic primer
misc_difference 112..131
note = a, c, t, g, unknown or other
source 1..153
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 233
acctctagaa aaaaaggcacc gactcgggtgc cacttttca agttgataac ggactagcct 60
tatattaact tgctatgctg tatttgttcc aatacagcat agctctaata cnnnnnnnnn 120
nnnnnnnnnn nggtgttcg tccttccac aag 153

SEQ ID NO: 234 moltype = DNA length = 22
FEATURE Location/Qualifiers
misc_difference 20
note = a, c, t, g, unknown or other
source 1..22
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 234
aggccccagt ggctgctcn aa 22

SEQ ID NO: 235 moltype = DNA length = 22
FEATURE Location/Qualifiers
misc_difference 20
note = a, c, t, g, unknown or other
source 1..22
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 235
acatcaaccg gtggcgcatn at 22

SEQ ID NO: 236 moltype = DNA length = 22
FEATURE Location/Qualifiers
misc_difference 20
note = a, c, t, g, unknown or other
source 1..22
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 236
aagggtgtggt tccagaaccn ac 22

SEQ ID NO: 237 moltype = DNA length = 22
FEATURE Location/Qualifiers
misc_difference 20
note = a, c, t, g, unknown or other
source 1..22
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 237
ccatcacatc aaccgggtgn ag 22

SEQ ID NO: 238 moltype = DNA length = 22
FEATURE Location/Qualifiers
misc_difference 20
note = a, c, t, g, unknown or other

```

-continued

```

source          1..22
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 238
aacacggcaga agctggagggn ta                                22

SEQ ID NO: 239      moltype = DNA  length = 22
FEATURE
misc_difference    20
note = a, c, t, g, unknown or other
source            1..22
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 239
ggcagaagct ggaggaggan tt                                22

SEQ ID NO: 240      moltype = DNA  length = 22
FEATURE
misc_difference    20
note = a, c, t, g, unknown or other
source            1..22
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 240
ggtgtggttc cagaaccgggn tc                                22

SEQ ID NO: 241      moltype = DNA  length = 22
FEATURE
misc_difference    20
note = a, c, t, g, unknown or other
source            1..22
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 241
aaccggagga caaagtacan tg                                22

SEQ ID NO: 242      moltype = DNA  length = 22
FEATURE
misc_difference    20
note = a, c, t, g, unknown or other
source            1..22
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 242
ttccagaacc ggaggacaan ca                                22

SEQ ID NO: 243      moltype = DNA  length = 22
FEATURE
misc_difference    20
note = a, c, t, g, unknown or other
source            1..22
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 243
gtgtgggtcc agaaccggan ct                                22

SEQ ID NO: 244      moltype = DNA  length = 22
FEATURE
misc_difference    20
note = a, c, t, g, unknown or other
source            1..22
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 244
tccagaacccg gaggacaan cc                                22

SEQ ID NO: 245      moltype = DNA  length = 22
FEATURE
misc_difference    20
note = a, c, t, g, unknown or other
source            1..22
               mol_type = unassigned DNA
               organism = Homo sapiens
SEQUENCE: 245
cagaagctgg aggaggaagn cg                                22

```

-continued

```

SEQ ID NO: 246      moltype = DNA length = 22
FEATURE
misc_difference 20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 246
catcaaccgg tggcgcattn ga                                22

SEQ ID NO: 247      moltype = DNA length = 22
FEATURE
misc_difference 20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 247
gcagaagctg gaggaggaan gt                                22

SEQ ID NO: 248      moltype = DNA length = 22
FEATURE
misc_difference 20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 248
cctccctccc tggcccaggn gc                                22

SEQ ID NO: 249      moltype = DNA length = 22
FEATURE
misc_difference 20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 249
tcatctgtgc ccctccctcn aa                                22

SEQ ID NO: 250      moltype = DNA length = 22
FEATURE
misc_difference 20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 250
gggaggacat cgatgtcacn at                                22

SEQ ID NO: 251      moltype = DNA length = 22
FEATURE
misc_difference 20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 251
caaacggcag aagctggagn ac                                22

SEQ ID NO: 252      moltype = DNA length = 22
FEATURE
misc_difference 20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 252
gggtgggcaa ccacaaaccn ag                                22

SEQ ID NO: 253      moltype = DNA length = 22
FEATURE
misc_difference 20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA

```

-continued

```

SEQUENCE: 253          organism = Homo sapiens
ggtgggcaac cacaaaccn ta                                22

SEQ ID NO: 254          moltype = DNA  length = 22
FEATURE
misc_difference          Location/Qualifiers
20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 254          organism = Homo sapiens
ggctcccatc acatcaaccn tt                                22

SEQ ID NO: 255          moltype = DNA  length = 22
FEATURE
misc_difference          Location/Qualifiers
20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 255          organism = Homo sapiens
gaagggcctg agtccgagcn tc                                22

SEQ ID NO: 256          moltype = DNA  length = 22
FEATURE
misc_difference          Location/Qualifiers
20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 256          organism = Homo sapiens
caaccgggtgg cgcattgccn tg                                22

SEQ ID NO: 257          moltype = DNA  length = 22
FEATURE
misc_difference          Location/Qualifiers
20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 257          organism = Homo sapiens
aggaggaagg gcctgagtcn ca                                22

SEQ ID NO: 258          moltype = DNA  length = 22
FEATURE
misc_difference          Location/Qualifiers
20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 258          organism = Homo sapiens
agtggagga ggaagggccn ct                                22

SEQ ID NO: 259          moltype = DNA  length = 22
FEATURE
misc_difference          Location/Qualifiers
20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 259          organism = Homo sapiens
gcattgccac gaagcaggcn cc                                22

SEQ ID NO: 260          moltype = DNA  length = 22
FEATURE
misc_difference          Location/Qualifiers
20
note = a, c, t, g, unknown or other
source
1..22
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 260          organism = Homo sapiens
attgccacga agcaggccan cg                                22

SEQ ID NO: 261          moltype = DNA  length = 22
FEATURE
Location/Qualifiers

```

-continued

```

misc_difference      20
                      note = a, c, t, g, unknown or other
source               1..22
                      mol_type = unassigned DNA
                      organism = Homo sapiens
SEQUENCE: 261
agaaccggag gacaaagtan ga                                22

SEQ ID NO: 262      moltype = DNA length = 22
FEATURE             Location/Qualifiers
misc_difference     20
                      note = a, c, t, g, unknown or other
source               1..22
                      mol_type = unassigned DNA
                      organism = Homo sapiens
SEQUENCE: 262
tcaaccggtg gcgcattgcn gt                                22

SEQ ID NO: 263      moltype = DNA length = 22
FEATURE             Location/Qualifiers
misc_difference     20
                      note = a, c, t, g, unknown or other
source               1..22
                      mol_type = unassigned DNA
                      organism = Homo sapiens
SEQUENCE: 263
gaagctggag gaggaagggn gc                                22

SEQ ID NO: 264      moltype = DNA length = 123
FEATURE             Location/Qualifiers
misc_feature        1..123
                      note = Description of Artificial Sequence: Synthetic
                      polynucleotide
source              1..123
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 264
ccaatgggga ggacatcgat gtcacctcaa atgacttaggg tgggcaacca caaaccacg  60
agggcagagt gtcgttgcg tctggccagg cccctgcgtg ggcccaagct ggactctggc  120
cac                                         123

SEQ ID NO: 265      moltype = DNA length = 121
FEATURE             Location/Qualifiers
misc_feature        1..121
                      note = Description of Artificial Sequence: Synthetic
                      polynucleotide
source              1..121
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 265
cgagcagaag aagaagggtc cccatcacat caaccggtg cgcatggcca cgaaggcaggc  60
caatggggag gacatcgatg tcacctccaa tgacttagggt gggcaaccac aaacccacga 120
g                                         121

SEQ ID NO: 266      moltype = DNA length = 128
FEATURE             Location/Qualifiers
misc_feature        1..128
                      note = Description of Artificial Sequence: Synthetic
                      polynucleotide
source              1..128
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 266
ggagggacaaa gtacaaacgg cagaaggctgg aggaggaagg gcctgagttc gagcagaaga  60
agaagggttc cccatcacatc aaccgggtggc gcattgccac gaaggcaggcc aatggggagg 120
acatcgat                                         128

SEQ ID NO: 267      moltype = DNA length = 130
FEATURE             Location/Qualifiers
misc_feature        1..130
                      note = Description of Artificial Sequence: Synthetic
                      polynucleotide
source              1..130
                      mol_type = other DNA
                      organism = synthetic construct
SEQUENCE: 267

```

-continued

```

agaagctgga ggaggaaggg cctgagtccg agcagaagaa gaagggctcc catcacatca 60
acccgtggcg cattgccacg aagcaggcca atggggagga catcgatgtc acctccaatg 120
actagggtgg 130

SEQ ID NO: 268      moltype = DNA length = 125
FEATURE
misc_feature
1..125
note = Description of Artificial Sequence: Synthetic
polynucleotide
source
1..125
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 268
ccctcagtttt cccatcaggc tctcagctca gcctgagtg tgaggccccca gtggctgtc 60
tggggcetc ctgagtttct catctgtgcc cctccctccc tggcccaggtaa gaaggtgtgg 120
ttcca 125

SEQ ID NO: 269      moltype = DNA length = 129
FEATURE
misc_feature
1..129
note = Description of Artificial Sequence: Synthetic
polynucleotide
source
1..129
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 269
tcatctgtgc ccctccctcc ctggccagg tgaaggctgt gttccagaac cggaggacaa 60
agtacaaaacg gcagaagctg gaggaggaag ggcctgagtc cgagcagaag aagaagggtc 120
cccatcaca 129

SEQ ID NO: 270      moltype = DNA length = 129
FEATURE
misc_feature
1..129
note = Description of Artificial Sequence: Synthetic
polynucleotide
source
1..129
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 270
ctccaaatgac taggggtgggc aaccacaaac ccacgaggcc agagtgtc ttgctgtgg 60
ccaggccccc gcgtggggccc aagctggact ctggccactc cttggccagg ctttggggag 120
gcctggagt 129

SEQ ID NO: 271      moltype = DNA length = 127
FEATURE
misc_feature
1..127
note = Description of Artificial Sequence: Synthetic
polynucleotide
source
1..127
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 271
ctgcttgcg ctggccaggc ccctgcgtgg qcccaagctg gactctggcc actccctggc 60
caggcttgg ggaggctgg agtcatggcc ccacaggctg tgaagccgg ggccgcatt 120
gacagag 127

SEQ ID NO: 272      moltype = DNA length = 25
FEATURE
misc_feature
1..25
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source
1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 272
gaaaattaata cgactcacta taggg 25

SEQ ID NO: 273      moltype = DNA length = 126
FEATURE
misc_feature
1..126
note = Description of Artificial Sequence: Synthetic
polynucleotide
source
1..126
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 273

```

-continued

```

aaaaaaagcac cgactcggtg ccacttttc aagttgataa cggactagcc ttatTTAAC 60
ttgtctatttc tagctctaaa acaacgcga gcgtgacacc accctatagt gagtcgtatt 120
aatttc 126

SEQ ID NO: 274 moltype = DNA length = 126
FEATURE Location/Qualifiers
misc_feature 1..126
note = Description of Artificial Sequence: Synthetic
polynucleotide
source 1..126
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 274
aaaaaaagcac cgactcggtg ccacttttc aagttgataa cggactagcc ttatTTAAC 60
ttgtctatttc tagctctaaa acgcaacaat taatagactg gacctatagt gagtcgtatt 120
aatttc 126

SEQ ID NO: 275 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: 275
tctttcttcg gctatgacac ttccagcaaa aggttagggcg ggctgcccaga cggcttcccg 60
gcgcgtcgcgat cacacccgat gatgttgcg ccccccgaag ctcccttcggg gctgcattgg 120
cgctcccgat cccgctccagg gcgagcgcgt tttaaatagc caggcccccg attgcaaaaga 180
catttatagcg actatccaaa gccatattca aacacctaga ttaactaccac ttctacacag 240
gcacactcgag cttgtgatcg cactcgctca agggggggcc tttctctttt cgtttcagtc 300
acaacccggca aacatgttcc catatgttcc tccagattac gtttcgcgcg agaaaaaagcg 360
caaggcggcaa gggtccgaca agaagtacag catccggctt gacatccggca ccaactctgt 420
gggctggggc gtgatcaccg acgagttacaa ggtgcccgcg aagaaattca aggtgcttgg 480
caaacccggc cccgacacgac tcaaaagaaatccgatccggca gcccctgttgt tcgcacaggcg 540
cgaacacggcc gggccaccc ggctgaaagag aaccggccgcg agaagataca ccagacggaa 600
gaaccggatc tgctatcttcg aagagatttt cagcaacggag atggccaagg tggacacag 660
cttcttccac agacttggaaag agtcttccctt ggttggaaagag gataagaagc acgagcggca 720
ccccatcttc ggcaacatcg tggacggatgg ggcttaccac gagaagtacc ccacccatcta 780
ccaccttgaga aagaaacttgg tggacggatcg cggacaaaggc gacccctggc tgatctatct 840
ggccctggcc ccatgtatca agttccgggg ccacttcttcg atcgaggggcg accttgaaccc 900
cgacaacacgc gacgttgcgca agtctgttcat ccagctggcg cagacatcaca accagctgtt 960
cgaggaaaaac cccatcaacgc ccacggccatgg gacccatctgt ctgcacatcc 1020
gagcaagacg acacggctgg aaaaatctgtt cggccatctgtt cccggcgaga agaagaatgg 1080
cctgttccggc aacctgttcc ccctggatcc gggcttgcacc cccaaacttca agagcaactt 1140
cgacctggcc gaggatgttcc aactgttgcgtt gggatggccg accttgcgtt acgacacttgg 1200
caacccgttcc gcccacatcg tggacggatcg cggccatctgtt ctgttgcgcg ccaagaaccc 1260
gtccggacggcc atccctgttca gggacatctt cggatgttcc accggatccca ccaaggcccc 1320
cctggatggcc tctatgtatca agatatacgatcc cggccatccatcc cccctgttca 1380
agctctctgtt cggccatccatcc tggatgttcc gtttccatcc accggatccca 1440
gaacccgttcc gcccacatcg tggatgttcc gggatggccg agccagccgcg gaagatgttcc acaagtttcat 1500
caagccatccatcg tggggaaatggatccatcc cggatgttcc tggatgttcc tgaacagaga 1560
ggacccgttcc cggccatccatcc cggatgttcc cccatccatcc agatccacatcc 1620
ggggatggatccatcc cggccatccatcc tggatgttcc gggatgttcc tgaaggacaa 1680
cccgccggaaatggatccatcc tggatgttccatcc cggccatccatcc tggatgttcc 1740
caggggaaatggatccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 1800
gaacttccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 1860
caacttccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 1920
gtacttccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 1980
ggccggatccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2040
cccgaaatggatccatcc cggccatccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc 2100
ctccgttccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2160
tctgttccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2220
ggaaatggatccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2280
gaaaacatccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2340
caccggatccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2400
caagacaatccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2460
gatccacacgc gacacgttccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc 2520
ggggatccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2580
catccgttccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2640
cgagaacatccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2700
cagccggatccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2760
gaaaagaacacccgttccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2820
gcagaatggatccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2880
cgatgttccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 2940
gctggaccatcc tggatgttccatcc cggccatccatcc tggatgttccatcc cggccatccatcc 3000

```

-continued

gaagaagatg	aagaactact	ggcggcagct	gctgaacgcc	aagctgatttta	cccagagaaaa	3060
gttcgacaat	ctgaccaagg	ccgagagagg	cggcctgac	gaactggata	aggccggctt	3120
catcaagaga	cagctgggg	aaaccggca	gatcacaaag	cacgtggcac	agatcttgaa	3180
ctccccatg	aacactaagt	acgacgagaa	tgacaagctg	atccggaaag	tgaaaagtat	3240
caccctgaag	tccaagctgg	tgtccgattt	ccggaaggat	ttccagttt	acaaaatgcg	3300
cgagatcaac	aactaccac	acgeccacga	cgectactt	aacgcggctg	tggaacccgc	3360
cctgtatcaa	aagtacccta	agctggaaag	cgagttcg	tacgggact	acaagggtta	3420
cgacgtcggt	aagatgtatc	ccaaagcgca	tgagaaatc	ggcaaggctt	ccgcaactgt	3480
cttcttctac	agcaacatca	tgaactttt	caagaccgag	attaccctgg	ccaacggcga	3540
gatccggaa	cggcctctg	tcgagacaaa	cgccggaaac	ggggagatcg	tgtgggatata	3600
gggcgggat	tttgcacccg	tcgagacaaa	cgccggaaac	ggggagatcg	tatctgtgaa	3660
aaagacccgg	gtgcacacag	cgccgtttag	caaaaggtct	atccctccca	agaggaaacag	3720
cgataagctg	atcgccagaa	agaaggactg	ggaccctta	aaagtacggcg	gttcgcacag	3780
ccccacccgt	gcttattctg	tgctgtgtgt	ggccaaatgt	gaaaaggcga	agtccaaagaa	3840
actgaagatg	tgaaagagc	tgctgtgtgt	caccatcg	gaaagaagca	gttcgagaa	3900
gaatccatc	gactttctgg	aagccaaagg	ctacaaatgt	gtggaaaaaa	acctgtatcat	3960
caagctgcct	aagtactccc	tgttcgatgt	ggaaaacggc	cggaagagaa	tgctggctc	4020
tgccggcgaa	ctgcacaaagg	gaaacgact	ggccctgccc	tccaaatatg	tgaacttcc	4080
gtacccctgg	ggccctggcc	ggccctggcc	gggtcccccc	gaggaaatg	agcagaacaa	4140
gctgttgtgt	gaacacgcac	gacacttct	ggacgagatc	atccggcaga	tcagcgatgt	4200
ctccaaagaga	gtgatctgg	ccgacgctaa	tctggacaaa	gtgtgtccg	cctacaacaa	4260
gcacccggat	aagccatca	gagacggcgg	cgagaatata	atccacccgt	ttacccctgac	4320
caatctggaa	ggccctggcc	ccttcgatgt	ctttgacacc	accatcgacc	ggaagaggta	4380
caccacccac	aaaggatgtc	tggacggcc	ccttcgatcc	cgacgatca	ccggccctgt	4440
cgagacacgg	atcgacctgt	ctcagctggg	aggcgacacg	cccaagaaga	agagaaaggt	4500
ggaggccacg	taaggatccg	gcaagactgg	cccccgttgg	caacgcaaca	gtgagccct	4560
ccctagtgt	tttggggatg	tgactatgt	ttcgtgtgt	ggccaaacggg	tcaacccgaa	4620
catggata	cccgcccttg	catttcgt	cagaatgtaa	cgtcagttga	tggact	4677

SEQ ID NO: 276	moltype = DNA	length = 3150
FEATURE	Location/Qualifiers	
misc_feature	1..3150	
	note = Description of Artificial Sequence: Synthetic	
	polynucleotide	
source	1..3150	
	mol_type = other DNA	
	organism = synthetic construct	

SEQUENCE: 276						
tctttcttgc	gtatgacac	ttccagcaaa	aggtagggcg	ggctgegaga	cggttcccg	60
gcgcgtcgat	caacacggat	gatgttcga	ccccccga	ctccttcggg	gctgcatgg	120
cggtcccgat	ccgctccagg	cgcaacgtgt	tttaaatagc	caggcccccg	attgcaaaa	180
cattatagcg	agcttacaaa	gccatattca	aacacctaga	tcactaccac	ttctacacag	240
gcaactcgag	tttgcgtatc	cacttcgt	agggggccgc	tcttcctt	cgtttca	300
acaacccgc	aaatcgatc	agaagaagag	gaagggttaac	acgatataca	tcgctaagaa	360
cgacttcct	gacatcgac	tggctgtat	cccggttac	actctggctg	accattacgg	420
tgagcgttta	gctcgacac	agttggccct	tgagcatgag	tcttacgaga	tgggtgaagc	480
acgcttcgc	aatgtgtttg	agcgtcaatc	taaagctgtt	gagggttgcg	ataacgttc	540
cgccaaacgc	ctcatcaacta	ccctactccc	taaagatgtat	gacacgtatca	acgactgtt	600
tgaggaaatg	aaagctaa	gcggcaagcg	cccgacacgg	tttcagttcc	tgcaagaaat	660
caagccggaa	ggcgtagcgt	acatcacca	taagaccact	ctggcttgc	taaccagtgc	720
tgacaataac	accgttcagg	ctgttagca	cgcaatecggt	ccggccat	aggacgggc	780
tcgttcgtgt	cgatccgt	accttgcgt	taagcactt	aagaaaaacg	ttgagaaaca	840
acttcaacaag	ccgcgttggc	acgttttacaa	qaaagcattt	atgcgttgc	tcgaggtct	900
catgtctct	aaaggcttac	tcgggtggca	ggcgtgttct	tcgtggcata	aggaagactc	960
tatttcgtt	ggagatcgct	gcatcgatgt	gtcattgtt	tcaacccgaa	tggttatgt	1020
acaccgc	aatgtgtggc	taggtgtca	agactctgtt	actatcgaa	tcgcacccgt	1080
atacgcgtat	gtatgttca	cccggttgcgg	tgccgttgc	ggcatcttc	cgatgttca	1140
accttgcgtt	gttcccttca	agccgttggac	tggcattact	ggtgtgttgc	attgggttca	1200
cggtcgtgt	cttcgttgc	tgggtgcgt	tcacgtatc	aaagactgt	tcgcgttac	1260
agacgttac	atgtgttgc	tgatgttgc	gattaatcg	gcccataatcg	ccgcgttgc	1320
aatcaacaat	aaatgttgc	ccgttgc	cgatcatcc	aaatgttgc	attgtccgt	1380
cgaggacatc	ccgtcgat	agcgtgttgc	actcccgat	aaaccggaa	acatcgacat	1440
gaatccctgt	gttccatcc	cggttgcgt	tgctgttgc	gtgtgttgc	gcaaggacaa	1500
gggtcgttgc	tccgttgcgt	tcgttgcgt	gatgttgc	gatgttgc	tttgcgttgc	1560
taaccataag	ggccatcgat	tcccttacaa	catggactgg	cgccgttgc	tttgcgttgc	1620
gtcaatgttc	aacccgc	gttacatcc	gatgttgc	gatgttgc	tttgcgttgc	1680
taaaatccat	ggtaaggaa	tttactactg	gttacatcc	gatgttgc	tttgcgttgc	1740
tgtcgacaa	gttccgttcc	ctgacgttgc	caagtttac	gatgttgc	tttgcgttgc	1800
catggcttgc	gttacatcc	caactggat	gatgttgc	gatgttgc	tttgcgttgc	1860
ctgttcgtt	gggttgcgt	tttgcgttgc	gatgttgc	gatgttgc	tttgcgttgc	1920
ctgttcgtt	ccgttgcgt	tttgcgttgc	atccgttgc	tctccgt	tttgcgttgc	1980
gttccgttgc	gatgttgcgt	tttgcgttgc	atccgttgc	tctccgt	tttgcgttgc	2040
catctacgg	attgttgc	agaaatgttgc	cgatgttgc	gatgttgc	tttgcgttgc	2100
gaccgataac	gaatgttgc	ccgttgcgt	tttgcgttgc	atccgttgc	tctccgt	2160
caagcttgc	actaaggc	tttgcgttgc	atccgttgc	gatgttgc	tttgcgttgc	2220
gactaaggcg	tcgttgcgt	ccgttgcgt	tttgcgttgc	atccgttgc	tctccgt	2280

-continued

agtgtctggaa	gataccatc	agccagctat	tgattccggc	aagggtctga	tgttcaactca	2340
gccgaatcaq	gctgctggat	acatggctaa	gctgatttgg	gaatctgtga	gcgtgacgggt	2400
ggtagctcg	gttgaagcaa	tgaactggct	taagtctgt	gctaagaatgc	ttggctgtga	2460
ggtaaagat	aagaagact	gagagatct	tcgcaagcgt	tcgcaatgtgc	attgggtta	2520
tcttgtatgtt	ttccctgtgt	ggcagaatgtt	caagaagcct	atccagcgc	gttggaaacct	2580
gatgtttctc	ggtcagttcc	gcttacagcc	taccatataa	accacaaaag	atagccgat	2640
tgatgcacac	aaacaggagt	ctggtatcgc	tcctaaacttt	gtacacagcc	aaagaeggtag	2700
ccacccctcg	aagactgttag	tgtggggcaca	cgagaatgc	ggaatcgaa	ctttgcact	2760
gattcagcac	ttcttgcgtt	cgatggccgc	tgacgtcg	aacctgttca	aaggactgtcg	2820
cgaaatctatg	gttgacacat	atagtcctt	tgatgtactg	tgatgtttct	acgaccaggta	2880
cgctgaccag	ttgcacgagt	ctcaatttgg	caaaatgc	gcacttccgg	ctaaaggtaa	2940
cttgaaccc	cgtgacatct	tagactcg	cttcgcgtt	cgcttaaggat	ccggcaagac	3000
tggcccgct	tggcaacca	acatgtggcc	cctccctatgt	gtgtttgggg	atgtgactat	3060
gttattctgt	gttggccac	gggtcaaccc	gaacagatgtt	ataccgcct	tggcatattcc	3120
tgtcaaatq	taacqtcgt	tgatgtqat				3150

```

SEQ ID NO: 277          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: 277
gaaattaata cgactcacta tannnnnnnn nnnnnnnnnn nngttttaga gctagaaaata 60
gcaaggtaaa ataaggctag tcggtatca acttgaaaaa gtggcacgga gtcgggtctt 120
ttttt

```

SEQUENCE: 278

tgcggattt	cacaccgcat	caggtggcac	tttcgggga	aatgtgcgcg	gaacccttat	60
tttgttattt	ttctaaatac	atccaat	gtatcgctc	atgaggatt	caaaaaggat	120
cttcacttag	atccctttaa	ataaaaat	aagtttaaa	tcatactaaa	gtatatatga	180
gtaaacctgg	tctgacagg	accatgttt	aatcgttag	gcacccatct	cagcgtatcg	240
tctatccgt	tcatccatag	ttgcgtact	ccccgtcg	tagataacta	cgataacggg	300
gggccttacca	tctggccca	gtgtcgat	gataccgca	gaccacgt	caccggct	360
agatttatca	gcaataaaacc	aggcagcgg	aaggccgg	cgcagaagtg	gtctgtcaac	420
tttacccgc	tccatccat	ctttaattt	ttgcgggaa	ftgatggta	tgatgtcgcc	480
agttaatagt	ttgcgcac	ttgttgc	tgctacaggc	atcgttgt	cacgcgtc	540
gtttgttgc	gttttgcattt	gtccgggtt	ccaaacgtca	aggcgaggta	catgatcccc	600
catgttgtgc	aaaaaaaggcg	ttatgtctt	ccgttcccg	atcggtgtca	gaagtagta	660
ggccgcgtgt	ttatctacta	ttgttatgg	agcaactgt	atttcttta	ctgtcgatgcc	720
atccgtaaaga	tgcttttctg	tgactgggt	gtactcaacc	aagtcaattt	gagaatagt	780
tatgcggcga	ccgagggtgt	cttgcggcgc	gtcaatacgg	gataataccg	cgccacatag	840
cagaacttta	aaatgtgtca	tcattggaaa	acgttttcgg	ggggaaaaac	tctcaatggat	900
tttccatccgt	tttagatccca	gttcgtatgt	accacccat	gcacccaaat	gatccatcg	960
atcttttact	ttcaccaggc	tttctgggt	agcaaaaaca	ggaaggccaa	atgcgc当地	1020
aaagggaata	agggcgacac	ggaaaatgtt	aatactcata	cttccctt	ttcaatatta	1080
tttagatcc	tatcagggtt	attgttctat	gaccaaaaatc	ccttaacgt	agtttgcgt	1140
ccactgtacgg	tcagacccc	tagaaaaat	caaaaggat	tcttgatgc	ctttttttt	1200
gcgcgtaaatc	tgctgttgc	aaacaaaaaa	accaccccta	ccacgggtgg	tttttttgc	1260
ggatcaagag	ctaccaactc	tttttccaa	ggtaactggc	ttcagcagag	cgcagatacc	1320
aaatactgtt	tttcttaggt	agccgtatgt	aggccaccac	ttcaaaact	ctgtacgacc	1380
gcctactatcc	ctcgctctgc	taatccgtt	accagggtgt	tttgcgtgt	gcgataatgc	1440
tggtcttacc	gggtgtgtact	caagaatgtt	gttacccggat	aaaggccgc	gggttgggtgt	1500
aacgggggggt	tcgtgcacac	agccaggctt	ggagcgaaac	acccatccgg	aacttggatata	1560
cctacagcgt	gagctatgag	aaagcgccac	gttcccgaa	gggagaaagg	cgacaggtt	1620
tccggtaaagg	ggcagggtcg	gaacaggaga	gcgcacagg	gagcttccgg	ggggaaaacgc	1680
ctgggtatctt	tatgttcc	ttcggttcc	ccacccctgt	cttgcgtgt	gattttgtgt	1740
atgtcgatca	ggggggccga	gcctatggaa	aaacggccgc	aaacggccgt	ttttatcggtt	1800
cctggccctt	tgctggcc	ttgtccat	gttcttctt	gcgttatccc	ctgatgttgt	1860
ggataaccgt	attaccgc	ttgagtgt	tgataccgt	ccgcgcagcc	gaacgaccga	1920
gcccgcgtgg	tcagtgtacgg	aggaacgggt	cgctgtgggt	tgatcgat	gggtcgatgt	1980
tttgcgtatgt	gctacaggac	tgatgttgc	ggctatgt	ggggggaaag	cttgcgttgc	2040
gcgcgtatgg	ggggccggcc	gttcggaa	ccgcataccg	ccccgtggc	gcacccatcc	2100
gggtataaaaq	ccccccaccc	cqaacqgtq	cctccactt	caqcaaaaa	cqacqactt	2160

-continued

tacatacgcg	actattctgc	cgctatacat	aaccactca	ctagcttaag	atcccatcaa	2220
gttgcatgc	cgggcgcgcc	agaaggagcg	cagccaaac	aggatgtatgt	tttatgggggt	2280
atttgacac	ttgcac	tatccggaa	ccccctggcc	cacaaggct	aggcgccaaat	2340
gcaagcagg	cgcatgcgc	ccctggagcg	gtgccctcc	gataaaccgg	ccagggggcc	2400
tatgttctt	actttttac	aagagaagtc	actcaacatc	ttaaaaatggc	caggtgagtc	2460
gacgagcag	cccgccggat	caggoagegt	gcttgcagat	ttgacttgc	acgcccgc	2520
tgtgtcgac	aaaggctttt	getctctgt	cgctgtctc	aecgcacat	aaccctgcgt	2580
cgcgcgtttc	atttgcagg	gatccggat	accatgtacc	catacgatgt	tccqattac	2640
gcttcgecga	agaaaaaagcg	caaggctgaa	gegtccgaca	agaagtacag	catggcctg	2700
gacatcgcca	ccaaactctg	gggttggcc	gtgatcaccc	acgagatcaa	ggtgcggcag	2760
aagaaattca	aaagggtgtgg	caacacgc	cggcacagca	tcaagaagaa	cctgtatcg	2820
gcctgtgt	ttgcacagcg	cgaaacagcc	gaggccaccc	ggctgaagag	aaccgcacaga	2880
agaagatata	ccagacggaa	gaacccgatc	tgctatctgc	aagagatctt	cagcaacagag	2940
atggccaaag	ttggacacag	cttctccca	agactggaa	agtccctct	ggtggaaagag	3000
gataagaagc	acgagcggca	ccccatcttc	ggaacacatc	tggacgagg	gcccattacc	3060
gagaagtttac	ccacatcta	ccacccatc	aqaaacttgc	ttgacacgac	cgacaaggcc	3120
gacctgeggc	tgatctatct	ggccctggcc	cacatgtatc	agttccgggg	ccacttctg	3180
atcgaggccg	acctgaaaccc	cgacaaac	gacgttggaca	agctgttcat	ccagctgtgt	3240
cagacatcaca	accagcttgtt	cgaggaaaac	cccatcaac	ccagcggcgt	ggacgcacaa	3300
gcacatctgt	ctggccgact	gagaagag	agacggcttgc	aaaatctgt	cgcccaactg	3360
cccgccgaga	agaagaatgg	ccttgcgg	aacctgattt	ccctgagcct	gggcctgacc	3420
ccaaacttca	agagaaactt	cgacccctggc	gaggatgcca	aactgcagct	gagaacggac	3480
acatcagac	acgacatgg	caacccatgc	ccggcagatc	cgccgcactg	3540	
tttctggccg	ccaaagaaactt	gtccgaccc	atctctgt	gacgatctt	gagagtgaac	3600
accgagatca	ccaaggcccc	cctgagcgc	tctatgtatc	agagatacga	cgagcacac	3660
caggacccgt	ccctgtgt	agcttctgc	ccggcagcgc	ttgcctgagaa	gtacaaagag	3720
attttcttcg	accagggc	gaacggctac	ggccgttaca	tttgcggccgg	agccagccag	3780
gaagagtttac	aaaggatctat	caaggccatc	ttggaaaaaa	ttgacggccac	cgaggaaact	3840
ctctgtgaagc	tgaacagaga	ggacccatgc	cggaaggac	ggacccatc	caacggcgc	3900
atccccccacc	agatccacat	ggggagag	ctccgcattt	tgccggccgc	ggaagat	3960
tacccatcc	tgaaggacaa	ccggggaaa	atcgaga	ttctgtac	ccgcacatccc	4020
tactacgtgg	ggccctgtgc	cgaggaaaac	agcagatc	cttggatgac	cagaagag	4080
gaggaaacc	tcacccctg	gaacccatc	gaagtgg	acaaggccgc	ttccgccc	4140
agtttcatcg	agccgtatgc	caacccatc	gttgcac	ccaaac	gggtgtgccc	4200
aagcaca	tgctgtac	gttgcac	gtgtataa	agctgac	aaatgt	4260
gtgacccgg	aatgatgaa	ggccgcctt	ctgacggcgc	agcagaaaaa	ggccatctgt	4320
gacactgtgt	tcaagacaa	cgccggaaat	accgtgaa	agctgaaaaa	ggactactt	4380
aagaaaatcg	agtgttgc	ctccgtgg	atctccgc	ttgaagatc	gttcaacgc	4440
tcctggcc	cataccac	tctgtgt	ggaaatttca	acaaggactt	ccttggacat	4500
gaggaaaac	aggacatctt	ggaagat	gtgtac	tgacatgtt	tgaggac	4560
gagatgtatc	acggacgg	aaaaacat	gcccac	tgcacgacaa	agtgtatg	4620
cagctgaa	ggcggagata	cacccgttgc	ggcaggatc	gcccgaat	gatcaacgg	4680
atccggac	acgacttcc	caagacata	cttgatttcc	tgaatcc	ccggcc	4740
aacagaaact	tatgtatc	gatccac	gagaccc	cctttaa	ggacatcc	4800
aaagccagg	tgccggcc	ggggatago	ctgcac	gacat	tctggcc	4860
agcccccc	ttaagaaggg	catctgc	acgtgt	ttgtgg	gatgt	4920
gtatggggc	ggcacaagg	cgagacat	gtgtac	ttggcc	gacccgac	4980
acccagaagg	gacagaa	cagccgg	gagaat	ttggcc	gacccgac	5040
gagctggcc	gcacatcc	ggaaaac	ccctgg	ttggcc	ggatc	5100
aagctgtacc	tgtactac	gcacat	ggggat	ttgtgg	ttggcc	5160
atcaacccgc	tgccgacta	cgatgt	ccat	ttgtgg	ttggcc	5220
gactccatc	acaacaagg	gtgt	ttgtgg	ttgtgg	ttggcc	5280
tgcccttc	ccggatgt	ttggcc	ttgtgg	ttgtgg	ttggcc	5340
aagctgtat	cccgag	ttggcc	ttgtgg	ttgtgg	ttggcc	5400
gaactgttgc	ggccggctt	ttggcc	ttgtgg	ttgtgg	ttggcc	5460
cacgtggc	agatcttgc	ttggcc	ttgtgg	ttgtgg	ttggcc	5520
atccggcc	tggatgt	ccat	ttgtgg	ttgtgg	ttggcc	5580
ttccagttt	acaaatgt	ttggcc	ttgtgg	ttgtgg	ttggcc	5640
aaacccgtc	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	5700
acccgtc	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	5760
ggcaagg	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	5820
attaccctgg	ccaaacgg	ttggcc	ttgtgg	ttgtgg	ttggcc	5880
ggggagatcg	tgtggata	ttggcc	ttgtgg	ttgtgg	ttggcc	5940
ccccaatgt	atatgt	ttggcc	ttgtgg	ttgtgg	ttggcc	6000
atccctgc	ccggat	ttggcc	ttgtgg	ttgtgg	ttggcc	6060
aagtacccgc	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	6120
gaaaagg	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	6180
gaaaagg	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	6240
gtgaaaa	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	6300
cgaaagg	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	6360
tccaaat	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	6420
gaggataat	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	6480
atcgac	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	6540
atcgac	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	6600
atccac	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	6660
accatcgac	ttggcc	ttgtgg	ttgtgg	ttgtgg	ttggcc	6720

-continued

cagagcatca	ccggcctgta	cgagacacgg	atcgaccctgt	ctcagctggg	aggcgacacg	6780
cccaagaaga	agagaaaaggt	ggggcccago	taacatatga	ttcgaatgtc	tttcttgcgc	6840
tatgacactt	ccagcaaaag	gtagggcggg	ctgcgagacg	gettcccgcc	gtgcgtatgca	6900
acaccgtatg	tgcttcgacc	ccccaaagct	cttccggggc	tgcataggcg	ctccatgtcc	6960
gctccagggc	gagcgctgtt	taaatagcca	ggcccccgat	tgcaaaagaca	ttatagcgag	7020
ctacccaaggc	catattcaaa	cacctagatc	actaccactt	ctacacaggc	cactcgagct	7080
tgtgtatcgca	ctccgctaaag	ggggcgctc	ttccttctcg	tttcgtatcac	aaccgcggaaa	7140
catgacacaa	gaatccctgt	tacttctcg	ccgtattgtat	tcggatgtt	cctacgcgg	7200
cctgcggaaac	gaccaggaaat	tctggggagt	gagtcgacga	gcaageccgg	cgatcgaggc	7260
agcgtgttgtt	cgatgttgac	ttgcacacgc	cgcatgtgt	cgacgaaggc	ttttggctcc	7320
tctgtcgctgt	tctcaaggcg	catetaacc	tgcgtcgcc	tttcatttg	cagccgcgtgg	7380
ccccccggcgc	ctggaggagag	ctcggtctgc	cggtgcggcc	gggtgtcgcc	gtgcggcgcc	7440
agagcaccaa	ccccgtactg	gtcggcgacg	cgggcccggt	gatcaagctg	ttcggcgagc	7500
actgtgtcgcc	tccggagacg	ctegcgctgg	agtccggaggc	gtacgcggtc	ctggcgacgc	7560
ccccgggtcc	ggtggccggc	ctccctcgcc	gacccggatc	gcccggcgcc	acggggaccc	7620
ggccgtggcc	ctacccgtgt	atgacccgga	tgacccggac	cacccggccg	tccgcgtatgg	7680
acggcacgac	cgacccggaaac	gctgtgtcg	ccctggccgg	cgaaactcgcc	cggtgtcg	7740
ggccggctca	cagggtggcc	ctgacccggg	acaccgtgtc	caccccccatt	tccggaggct	7800
tcccggaact	gtcggggaa	cgccggccgg	cgacccgtcg	ggaccacccg	gggtggggct	7860
acccctcgcc	ccgggtgtcg	gaccggctgg	aggactggct	gcccggacgt	gacacgtcg	7920
tggccggccg	cgaaaccggg	tgcgtccacg	gacccgtcg	cgggacccaa	atccgtgg	7980
acccctggccg	gaccgggttc	accggggatcg	tcgacttccac	cgacgttat	gccccggagact	8040
cccgctacag	ctcggtgca	ctgcattctca	acgccttcgg	ggccggacccg	gagatccctgg	8100
cccgctgtcg	cgacccggccg	cagtggaa	ggaccggagg	cttcggccgc	gaactgtcg	8160
cettcacctt	ctgcacacgc	ttcgagggt	tcgaggagac	cccgctggat	cttcggcgct	8220
taaccgatcc	ggaggaaactg	gcccggatcc	tctggggggcc	gcccggacacc	gccccccggc	8280
cctgataagg	atccggaaag	actggggccg	cttggcaacg	caacagttag	ccctccctca	8340
gtgtgtttgg	ggatgtgtact	atgtattgt	gtgttggcca	acgggtcaac	ccgaacagat	8400
tgataccccc	cttggccattt	cctgtcagaa	tgtaacgtca	gttgcgttgc	tttttttttt	8452

SEQ ID NO: 279	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: 279						
gttttagagc	tatgtctgttt	tgaatggtcc	aaaaacggaa	gggcctgagt	ccgagcagaa	60
gaagaagttt	tagagctatg	ctgtttgaa	tggtcccaa	ac		102
SEQ ID NO: 280	moltype = DNA length = 100					
FEATURE	Location/Qualifiers					
source	1..100					
	mol_type = unassigned DNA					
	organism = Homo sapiens					
SEQUENCE: 280						
cgaggaggacaa	agtacaaacg	gcagaagctg	gaggaggaa	ggcctgagtc	cgagcagaag	60
aagaagggt	cccatcacat	caaccgggtgg	cgcatggcca			100
SEQ ID NO: 281	moltype = DNA length = 50					
FEATURE	Location/Qualifiers					
source	1..50					
	mol_type = unassigned DNA					
	organism = Homo sapiens					
SEQUENCE: 281						
agctggagga	ggaaggggct	gagtcggcgc	agaagaagaa	gggctccac		50
SEQ ID NO: 282	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					
SEQUENCE: 282						
gagtccgagc	agaagaagaa	gttttagagc				30
SEQ ID NO: 283	moltype = DNA length = 49					
FEATURE	Location/Qualifiers					
misc_feature	1..49					
	note = Description of Artificial Sequence: Synthetic oligonucleotide					
source	1..49					

-continued

SEQUENCE: 283	mol_type = other DNA organism = synthetic construct	
agctggagga ggaaggggcct gagtccgago agaagagaag ggctcccat		49
SEQ ID NO: 284	moltype = DNA length = 53	
FEATURE	Location/Qualifiers	
source	1..53	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 284		
ctggaggagg aagggcctga gtccgagcag aagaagaagg gtcacatca cat		53
SEQ ID NO: 285	moltype = DNA length = 52	
FEATURE	Location/Qualifiers	
source	1..52	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 285		
ctggaggagg aagggcctga gtccgagcag aagaagaagg ctccatcac at		52
SEQ ID NO: 286	moltype = DNA length = 54	
FEATURE	Location/Qualifiers	
source	1..54	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 286		
ctggaggagg aagggcctga gtccgagcag aagaagaagg ggctccatc acat		54
SEQ ID NO: 287	moltype = DNA length = 50	
FEATURE	Location/Qualifiers	
source	1..50	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 287		
ctggaggagg aagggcctga gtccgagcag aagaaggcgt cccatcacat		50
SEQ ID NO: 288	moltype = DNA length = 47	
FEATURE	Location/Qualifiers	
source	1..47	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 288		
ctggaggagg aagggcctga gcccgagcag aagggtccc atcacat		47
SEQ ID NO: 289	moltype = RNA length = 66	
FEATURE	Location/Qualifiers	
misc_feature	1..66	
	note = Description of Artificial Sequence: Synthetic	
	oligonucleotide	
misc_difference	1..20	
	note = a, c, t, g, unknown or other	
source	1..66	
	mol_type = other RNA	
	organism = synthetic construct	
modified_base	56	
	mod_base = OTHER	
	note = Thymine	
modified_base	59	
	mod_base = OTHER	
	note = Thymine	
SEQUENCE: 289		
nnnnnnnnnn nnnnnnnnnn gtttagago tagaaatagc aagttaaat aaggctagtc		60
cgtttt		66
SEQ ID NO: 290	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: 290		
gagtcgagc agaagaagaa		20

-continued

SEQ ID NO: 291	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: 291	
gacatcgatg tcctccccat	20
SEQ ID NO: 292	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: 292	
gtcacacctcca atgactaggg	20
SEQ ID NO: 293	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: 293	
attgggtgtt cagggcagag	20
SEQ ID NO: 294	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: 294	
gtggcgcagag gggcccgagat	20
SEQ ID NO: 295	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: 295	
ggggcccgaga ttgggtgttc	20
SEQ ID NO: 296	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: 296	
gtgccattag ctaaatgcata	20
SEQ ID NO: 297	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: 297	
gttaccaccca cagggtgccag	20

-continued

```

SEQ ID NO: 298      moltype = RNA  length = 20
FEATURE
misc_feature
1..20
note = Description of Artificial Sequence: Synthetic
      oligonucleotide
source
1..20
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 298
gaaagctct gggccaggaa                                         20

SEQ ID NO: 299      moltype = DNA   length = 48
FEATURE
misc_feature
1..48
note = Description of Artificial Sequence: Synthetic
      oligonucleotide
source
1..48
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 299
ctggaggagg aagggcctga gtccgagc aagaagaagg gctccccat          48

SEQ ID NO: 300      moltype = RNA  length = 20
FEATURE
misc_feature
1..20
note = Description of Artificial Sequence: Synthetic
      oligonucleotide
source
1..20
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 300
gagtccgagc agaagaagat                                         20

SEQ ID NO: 301      moltype = RNA  length = 20
FEATURE
misc_feature
1..20
note = Description of Artificial Sequence: Synthetic
      oligonucleotide
source
1..20
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 301
gagtccgagc agaagaagata                                         20

SEQ ID NO: 302      moltype = RNA  length = 20
FEATURE
misc_feature
1..20
note = Description of Artificial Sequence: Synthetic
      oligonucleotide
source
1..20
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 302
gagtccgagc agaagaacaa                                         20

SEQ ID NO: 303      moltype = RNA  length = 20
FEATURE
misc_feature
1..20
note = Description of Artificial Sequence: Synthetic
      oligonucleotide
source
1..20
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 303
gagtccgagc agaagatgaa                                         20

SEQ ID NO: 304      moltype = RNA  length = 20
FEATURE
misc_feature
1..20
note = Description of Artificial Sequence: Synthetic
      oligonucleotide
source
1..20
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 304

```

-continued

gagtccgagc agaagttagaa	20
SEQ ID NO: 305	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: 305	
gagtccgagc agatgaagaa	20
SEQ ID NO: 306	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: 306	
gagtccgagc acaagaagaa	20
SEQ ID NO: 307	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: 307	
gagtccgagg agaagaagaa	20
SEQ ID NO: 308	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: 308	
gagtccgtgc agaagaagaa	20
SEQ ID NO: 309	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: 309	
gagtccggagc agaagaagaa	20
SEQ ID NO: 310	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: 310	
gagaccggagc agaagaagaa	20
SEQ ID NO: 311	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

-continued

```

SEQUENCE: 311
aatgacaagg ttgcttagcggttggg 24

SEQ ID NO: 312      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: 312
aaaacggaaag ggcctgagtc cgagcagaag aagaagttt 39

SEQ ID NO: 313      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: 313
aacacaggggc cgagattggg tgttcaggggc agagggtttt 39

SEQ ID NO: 314      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: 314
aaaacggaaag ggcctgagtc cgagcagaag aagaagttt 38

SEQ ID NO: 315      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: 315
aacggggggg ggggcacaga tgagaaaactc agggttttag 40

SEQ ID NO: 316      moltype = DNA  length = 38
FEATURE          Location/Qualifiers
source           1..38
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 316
agcccttctt cttctgctcg gactcaggcc cttcctcc 38

SEQ ID NO: 317      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: 317
cagggggggg ggggcacaga tgagaaaactc aggaggcccc 40

SEQ ID NO: 318      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: 318
ggcaatgcgc caccgggttga tggatgggaccccttcttag gaggccccca gagcagccac 60

```

-continued

tggggcctca acactcaggc	80
SEQ ID NO: 319 moltype = DNA length = 23	
FEATURE Location/Qualifiers	
source 1..23	
mol_type = unassigned DNA	
organism = Homo sapiens	
SEQUENCE: 319 gtcacacctcca atgactaggg tgg	23
SEQ ID NO: 320 moltype = length =	
SEQUENCE: 320 000	
SEQ ID NO: 321 moltype = length =	
SEQUENCE: 321 000	
SEQ ID NO: 322 moltype = DNA length = 33	
FEATURE Location/Qualifiers	
source 1..33	
mol_type = unassigned DNA	
organism = Homo sapiens	
SEQUENCE: 322 catcgatgtc ctccccattg gcctgcttcg tgg	33
SEQ ID NO: 323 moltype = DNA length = 33	
FEATURE Location/Qualifiers	
source 1..33	
mol_type = unassigned DNA	
organism = Homo sapiens	
SEQUENCE: 323 ttctgtggcaa tgcgccaccg gttgtatgtga tgg	33
SEQ ID NO: 324 moltype = DNA length = 33	
FEATURE Location/Qualifiers	
source 1..33	
mol_type = unassigned DNA	
organism = Homo sapiens	
SEQUENCE: 324 tcgtggcaat ggcgccaccgg ttgtatgtat ggg	33
SEQ ID NO: 325 moltype = DNA length = 33	
FEATURE Location/Qualifiers	
source 1..33	
mol_type = unassigned DNA	
organism = Homo sapiens	
SEQUENCE: 325 tccagcttct gccgtttgtta ctttgtccctc cgg	33
SEQ ID NO: 326 moltype = DNA length = 33	
FEATURE Location/Qualifiers	
source 1..33	
mol_type = unassigned DNA	
organism = Homo sapiens	
SEQUENCE: 326 ggaggggagggg gcacagatga gaaactcagg agg	33
SEQ ID NO: 327 moltype = DNA length = 33	
FEATURE Location/Qualifiers	
source 1..33	
mol_type = unassigned DNA	
organism = Homo sapiens	
SEQUENCE: 327 agggggccgag attgggtgtt cagggcagag agg	33
SEQ ID NO: 328 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: 328 aacacccgggt cttcgagaag acctgtttta gagctagaaaa tagcaagtta aaat	54

-continued

```

SEQ ID NO: 329      moltype = DNA  length = 54
FEATURE
misc_feature
1..54
note = Description of Artificial Sequence: Synthetic
      oligonucleotide
source
1..54
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 329
caaaacgggt cttcgagaag acgtttaga gctatgctgt tttgaatggt cccaa          54

SEQ ID NO: 330      moltype = DNA  length = 33
FEATURE
source
1..33
mol_type = unassigned DNA
organism = Mus musculus
SEQUENCE: 330
caaggcactga gtgccattag ctaaatgcattt agg                         33

SEQ ID NO: 331      moltype = DNA  length = 33
FEATURE
source
1..33
mol_type = unassigned DNA
organism = Mus musculus
SEQUENCE: 331
aatgcattttt gtaccacccca caggtgccag ggg                         33

SEQ ID NO: 332      moltype = DNA  length = 33
FEATURE
source
1..33
mol_type = unassigned DNA
organism = Mus musculus
SEQUENCE: 332
acacacatgg gaaaggctctt gggccaggaa agg                         33

SEQ ID NO: 333      moltype = DNA  length = 37
FEATURE
source
1..37
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 333
ggaggaggta gtatacagaa acacagagaa gtagaat                         37

SEQ ID NO: 334      moltype = DNA  length = 37
FEATURE
source
1..37
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 334
agaatgtaga ggagtcacag aaactcagca cttagaaa                         37

SEQ ID NO: 335      moltype = DNA  length = 98
FEATURE
misc_feature
1..98
note = Description of Artificial Sequence: Synthetic
      oligonucleotide
source
1..98
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 335
ggacgaaaca ccggAACCAT tcAAAACAGC atAGCAAGGT AAAATAAGGC tagTCGGTTA  60
tcaacttgaa aaAGTGGCAC CGAGTCGGTG CTTTTTTT                         98

SEQ ID NO: 336      moltype = DNA  length = 186
FEATURE
misc_feature
1..186
note = Description of Artificial Sequence: Synthetic
      polynucleotide
source
1..186
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 336
ggacgaaaca ccggTAGAT taagtattgt ttatGGCTG atAAATTCT ttGAATTCT  60
ccttgattat ttgttataaa agttataaaa taatCTTGTG ggaACCATC aaaACAGCAT 120
agcaagttaa aataaggcta gtccgttac aacttggaaaa agtggcacccg agtCGGTGCT 180

```

-continued

ttttt	186
SEQ ID NO: 337	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: 337	
gggttttaga gctatgctgt tttgaatggt cccaaaacgg gtcttcgaga agacgttta	60
gagctatgct gtttgaatg gtcggaaaac tttt	95
SEQ ID NO: 338	moltype = length =
SEQUENCE: 338	
000	
SEQ ID NO: 339	moltype = length =
SEQUENCE: 339	
000	
SEQ ID NO: 340	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: 340	
gtggaaaggc cgaaacaccg ggtcttcgag aagacctgtt ttagagctag aaatagcaag	60
ttaaaaataag gctagtccgt tttt	84
SEQ ID NO: 341	moltype = length =
SEQUENCE: 341	
000	
SEQ ID NO: 342	moltype = length =
SEQUENCE: 342	
000	
SEQ ID NO: 343	moltype = DNA length = 88
FEATURE	Location/Qualifiers
misc_feature	1..88
	note = Description of Artificial Sequence: Synthetic
	oligonucleotide
source	1..88
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 343	
gttttagagc tatgtgtttt tgaatggtcc caaaactgag accaaaggtc tcgttttaga	60
gctatgctgt tttgaatggt cccaaaac	88
SEQ ID NO: 344	moltype = DNA length = 35
FEATURE	Location/Qualifiers
misc_feature	1..35
	note = Description of Artificial Sequence: Synthetic
	oligonucleotide
source	1..35
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 344	
aaacacggagg gcctgagtcc gagcagaaga agaag	35
SEQ ID NO: 345	moltype = DNA length = 35
FEATURE	Location/Qualifiers
misc_feature	1..35
	note = Description of Artificial Sequence: Synthetic
	oligonucleotide
source	1..35
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 345	
aaaacttctt ctctgtctcg gactcaggcc cttcc	35

-continued

```

SEQ ID NO: 346      moltype = RNA  length = 46
FEATURE
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: 346
nnnnnnnnnn nnnnnnnnng ttattgtact ctcaagattt attttt          46

SEQ ID NO: 347      moltype = RNA  length = 91
FEATURE
misc_feature
1..91
note = Description of Artificial Sequence: Synthetic
      oligonucleotide
source
1..91
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 347
gttacttaaa tcttgcagaa gctacaaaaga taaggcttca tgccgaaatc aacaccctgt 60
cattttatgg cagggtgttt tcgttattta a                                91

SEQ ID NO: 348      moltype = DNA   length = 70
FEATURE
source
1..70
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 348
tttcttagt ctgagttct gtgacttcctc tacattctac ttctctgtgt ttctgtatac 60
tacccctccc          70

SEQ ID NO: 349      moltype = DNA   length = 122
FEATURE
source
1..122
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 349
ggagggaaggg cctgagtcgg agcagaagaa gaagggtctc catcacatca accggtgccg 60
cattgccacg aagcaggcca atggggagga catcgatgtc acctccaatg actagggtgg 120
gc                                         122

SEQ ID NO: 350      moltype = RNA   length = 48
FEATURE
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: 350
acnnnnnnnnn nnnnnnnnnn nngttttaga gctatgtc          48

SEQ ID NO: 351      moltype = RNA   length = 67
FEATURE
misc_feature
1..67
note = Description of Artificial Sequence: Synthetic
      oligonucleotide
source
1..67
mol_type = other RNA
organism = synthetic construct
modified_base
24
mod_base = OTHER
note = Thymine
SEQUENCE: 351
agcatagcaa gttaaaataa ggctagtcgg ttatcaactt gaaaaagtgg caccgagtcg 60
gtgctt          67

SEQ ID NO: 352      moltype = RNA   length = 62
FEATURE
misc_feature
1..62
note = Description of Artificial Sequence: Synthetic

```

-continued

```

oligonucleotide
misc_difference      1..20
note = a, c, u, g, unknown or other
source               1..62
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 352
nnnnnnnnnnn nnnnnnnnnn gtttagago tagaaatagc aagttaaaat aaggctagtc 60
cg                               62

SEQ ID NO: 353      moltype = DNA length = 73
FEATURE
misc_feature        1..73
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source              1..73
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 353
tgaatggtcc caaaacggaa gggcctgagt ccgagcagaa gaagaagtt tagagctatg 60
ctgttttcaa tgg                               73

SEQ ID NO: 354      moltype = DNA length = 69
FEATURE
source              1..69
Location/Qualifiers
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 354
ctgggtttcc acctctctgc cctgaacacc caatctcgcc ccctctcgcc accctctgc 60
atttctgtt                               69

SEQ ID NO: 355      moltype = DNA length = 138
FEATURE
source              1..138
Location/Qualifiers
mol_type = unassigned DNA
organism = Mus musculus
SEQUENCE: 355
acccaaggcac tgagtgccat tagctaaatg catagggtac cacccacagg tgccagggc 60
ctttcccaa gtccccagcc ccttctccaa ccttctctgg cccagaggt ttcccatgtg 120
tgtggcttga ccctttga                               138

SEQ ID NO: 356      moltype = DNA length = 20
FEATURE
misc_feature        1..20
note = Description of Artificial Sequence: Synthetic primer
source              1..20
Location/Qualifiers
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 356
gtgctttgca gagggcctacc                               20

SEQ ID NO: 357      moltype = DNA length = 20
FEATURE
misc_feature        1..20
note = Description of Artificial Sequence: Synthetic primer
source              1..20
Location/Qualifiers
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 357
cctggagcgc atgcagtagt                               20

SEQ ID NO: 358      moltype = DNA length = 22
FEATURE
misc_feature        1..22
note = Description of Artificial Sequence: Synthetic primer
source              1..22
Location/Qualifiers
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 358
accttctgtg tttccaccat tc                               22

SEQ ID NO: 359      moltype = DNA length = 20
FEATURE
misc_feature        1..20
note = Description of Artificial Sequence: Synthetic primer
source              1..20
Location/Qualifiers

```

-continued

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 359
ttggggagtg cacagacttc                                         20

SEQ ID NO: 360      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: 360
ggctccctgg gttcaaagta                                         20

SEQ ID NO: 361      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: 361
agaggggtct ggatgtcgta a                                         21

SEQ ID NO: 362      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: 362
tagctctaaa acttcttctt ctgctcggac                                         30

SEQ ID NO: 363      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: 363
ctagccattat tttaacttgc tatgtgttt                                         30

SEQ ID NO: 364      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: 364
nnnnnnnnnn nnnnnnnnnn gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgtttatcaac ttgaaaaagt ggcaccgagt cggtgctt                           99

SEQ ID NO: 365      moltype = DNA  length = 12
FEATURE           Location/Qualifiers
source            1..12
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 365
tagcgggtaa gc                                         12

SEQ ID NO: 366      moltype = DNA  length = 12
FEATURE           Location/Qualifiers
source            1..12
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 366
tcggtgacat gt                                         12

SEQ ID NO: 367      moltype = DNA  length = 12

```

-continued

FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 367	
actccccgtaa	12
SEQ ID NO: 368	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 368	
actgcgtgtt aa	12
SEQ ID NO: 369	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 369	
acgtcgcctg at	12
SEQ ID NO: 370	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 370	
taggtcgacc ag	12
SEQ ID NO: 371	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 371	
ggcggttaatg at	12
SEQ ID NO: 372	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 372	
tgtcgcatagt ta	12
SEQ ID NO: 373	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 373	
atggaaaacgc at	12
SEQ ID NO: 374	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 374	
gcccgaaattcc tc	12
SEQ ID NO: 375	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 375	
gcatggtacg ga	12
SEQ ID NO: 376	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens

-continued

SEQUENCE: 376		
cggtactt ac		12
SEQ ID NO: 377	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 377		
gctctgtgcg ta		12
SEQ ID NO: 378	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 378		
tacggtaagt cg		12
SEQ ID NO: 379	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 379		
cacgaaatta cc		12
SEQ ID NO: 380	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 380		
aaccaagata cg		12
SEQ ID NO: 381	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 381		
gagtcgatac gc		12
SEQ ID NO: 382	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 382		
gtctcacgat cg		12
SEQ ID NO: 383	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 383		
tcgtcggttg ca		12
SEQ ID NO: 384	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 384		
actccgtagt ga		12
SEQ ID NO: 385	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 385		
caggacgtcc gt		12
SEQ ID NO: 386	moltype = DNA length = 12	

-continued

FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 386	
tcgtatccct ac	12
SEQ ID NO: 387	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 387	
tttcaaggcc gg	12
SEQ ID NO: 388	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 388	
cgcgcgggtgga at	12
SEQ ID NO: 389	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 389	
gaaccccggtcc ta	12
SEQ ID NO: 390	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 390	
gattcatcag cg	12
SEQ ID NO: 391	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 391	
acacccgggtct tc	12
SEQ ID NO: 392	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 392	
atcgtgcct aa	12
SEQ ID NO: 393	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 393	
gcgtcaatgt tc	12
SEQ ID NO: 394	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 394	
ctccgttatct cg	12
SEQ ID NO: 395	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens

-continued

SEQUENCE: 395 ccgattcctt cg	12
SEQ ID NO: 396 FEATURE source	moltype = DNA length = 12 Location/Qualifiers 1..12 mol_type = unassigned DNA organism = Homo sapiens
SEQUENCE: 396 tgcgcccca gt	12
SEQ ID NO: 397 FEATURE source	moltype = DNA length = 12 Location/Qualifiers 1..12 mol_type = unassigned DNA organism = Homo sapiens
SEQUENCE: 397 taacgtcggaa gc	12
SEQ ID NO: 398 FEATURE source	moltype = DNA length = 12 Location/Qualifiers 1..12 mol_type = unassigned DNA organism = Homo sapiens
SEQUENCE: 398 aagggtcgccc at	12
SEQ ID NO: 399 FEATURE source	moltype = DNA length = 12 Location/Qualifiers 1..12 mol_type = unassigned DNA organism = Homo sapiens
SEQUENCE: 399 gtcggggact at	12
SEQ ID NO: 400 FEATURE source	moltype = DNA length = 12 Location/Qualifiers 1..12 mol_type = unassigned DNA organism = Homo sapiens
SEQUENCE: 400 ttcgagcgtt tt	12
SEQ ID NO: 401 FEATURE source	moltype = DNA length = 12 Location/Qualifiers 1..12 mol_type = unassigned DNA organism = Homo sapiens
SEQUENCE: 401 tgagtctcg ag	12
SEQ ID NO: 402 FEATURE source	moltype = DNA length = 12 Location/Qualifiers 1..12 mol_type = unassigned DNA organism = Homo sapiens
SEQUENCE: 402 tttacgcaga gg	12
SEQ ID NO: 403 FEATURE source	moltype = DNA length = 12 Location/Qualifiers 1..12 mol_type = unassigned DNA organism = Homo sapiens
SEQUENCE: 403 aggaagtatc gc	12
SEQ ID NO: 404 FEATURE source	moltype = DNA length = 12 Location/Qualifiers 1..12 mol_type = unassigned DNA organism = Homo sapiens
SEQUENCE: 404 actcgatacc at	12
SEQ ID NO: 405	moltype = DNA length = 12

-continued

FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 405	
cgctacatag ca	12
SEQ ID NO: 406	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 406	
ttcataaaccg gc	12
SEQ ID NO: 407	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 407	
ccaaacgggtt aa	12
SEQ ID NO: 408	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 408	
cgattcccttc gt	12
SEQ ID NO: 409	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 409	
cgtcatgaat aa	12
SEQ ID NO: 410	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 410	
agtggcgatg ac	12
SEQ ID NO: 411	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 411	
cccctacggc ac	12
SEQ ID NO: 412	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 412	
gccaaacccgc ac	12
SEQ ID NO: 413	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 413	
tgggacacccg gt	12
SEQ ID NO: 414	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens

-continued

SEQUENCE: 414		
ttgactgcgg cg		12
SEQ ID NO: 415	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 415		
actatgcgt a gg		12
SEQ ID NO: 416	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 416		
tcacccaaag cg		12
SEQ ID NO: 417	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 417		
gcaggacgtc cg		12
SEQ ID NO: 418	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 418		
acaccgaaaa cg		12
SEQ ID NO: 419	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 419		
cggtgtattg ag		12
SEQ ID NO: 420	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 420		
cacgaggtat gc		12
SEQ ID NO: 421	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 421		
taaaagcgacc cg		12
SEQ ID NO: 422	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 422		
cttagtcggc ca		12
SEQ ID NO: 423	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 423		
cggaaaacgtg gc		12
SEQ ID NO: 424	moltype = DNA length = 12	

-continued

FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 424	
cgtgccctga ac	12
SEQ ID NO: 425	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 425	
tttaccatcg aa	12
SEQ ID NO: 426	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 426	
cgttagccatg tt	12
SEQ ID NO: 427	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 427	
cccaaaacggt ta	12
SEQ ID NO: 428	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 428	
gcgttattcag aa	12
SEQ ID NO: 429	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 429	
tcgatggtaa ac	12
SEQ ID NO: 430	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 430	
cgacttttg ca	12
SEQ ID NO: 431	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 431	
tcgacgactc ac	12
SEQ ID NO: 432	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens
SEQUENCE: 432	
acgcgtcaga ta	12
SEQ ID NO: 433	moltype = DNA length = 12
FEATURE	Location/Qualifiers
source	1..12
	mol_type = unassigned DNA
	organism = Homo sapiens

-continued

SEQUENCE: 433		
cgtacggcac ag		12
SEQ ID NO: 434	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 434		
ctatgccgtg ca		12
SEQ ID NO: 435	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 435		
cgcgtcagat at		12
SEQ ID NO: 436	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 436		
aagatcggtt gc		12
SEQ ID NO: 437	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 437		
tttcgcaagg ag		12
SEQ ID NO: 438	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 438		
gtcgtggact ac		12
SEQ ID NO: 439	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 439		
ggtcgtcatc aa		12
SEQ ID NO: 440	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 440		
gttaaacagcg tg		12
SEQ ID NO: 441	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 441		
tagctaaccg tt		12
SEQ ID NO: 442	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 442		
agtaaaaggcg ct		12
SEQ ID NO: 443	moltype = DNA length = 12	

-continued

```

FEATURE          Location/Qualifiers
source           1..12
                  mol_type = unassigned DNA
                  organism = Homo sapiens
SEQUENCE: 443   ggtaatttcg tg                               12
SEQ ID NO: 444   moltype = RNA  length = 69
FEATURE          Location/Qualifiers
misc_feature     1..69
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..69
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 444   gtcacacctcca atgactaggg gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
                  cgttttttt                                         69
SEQ ID NO: 445   moltype = RNA  length = 69
FEATURE          Location/Qualifiers
misc_feature     1..69
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..69
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 445   gacatcgatg tcctccccat gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
                  cgttttttt                                         69
SEQ ID NO: 446   moltype = RNA  length = 69
FEATURE          Location/Qualifiers
misc_feature     1..69
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..69
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 446   gagtcggagc agaagaagaa gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
                  cgttttttt                                         69
SEQ ID NO: 447   moltype = RNA  length = 69
FEATURE          Location/Qualifiers
misc_feature     1..69
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..69
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 447   gggcccgaga ttgggtgttc gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
                  cgttttttt                                         69
SEQ ID NO: 448   moltype = RNA  length = 69
FEATURE          Location/Qualifiers
misc_feature     1..69
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..69
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 448   gtggcgagag gggccgagat gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
                  cgttttttt                                         69
SEQ ID NO: 449   moltype = RNA  length = 76
FEATURE          Location/Qualifiers
misc_feature     1..76
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..76
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 449

```

-continued

```

gtcacctcca atgactaggg gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcatt tttttt          76

SEQ ID NO: 450      moltype = RNA  length = 76
FEATURE           Location/Qualifiers
misc_feature      1..76
note = Description of Artificial Sequence: Synthetic
            oligonucleotide
source            1..76
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 450
gacatcgatg tcctcccat gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcatt tttttt          76

SEQ ID NO: 451      moltype = RNA  length = 76
FEATURE           Location/Qualifiers
misc_feature      1..76
note = Description of Artificial Sequence: Synthetic
            oligonucleotide
source            1..76
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 451
gagtccgagc agaagaagaa gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcatt tttttt          76

SEQ ID NO: 452      moltype = RNA  length = 76
FEATURE           Location/Qualifiers
misc_feature      1..76
note = Description of Artificial Sequence: Synthetic
            oligonucleotide
source            1..76
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 452
ggggccgaga ttgggtgttc gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcatt tttttt          76

SEQ ID NO: 453      moltype = RNA  length = 76
FEATURE           Location/Qualifiers
misc_feature      1..76
note = Description of Artificial Sequence: Synthetic
            oligonucleotide
source            1..76
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 453
gtggcgagag gggccgagat gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcatt tttttt          76

SEQ ID NO: 454      moltype = RNA  length = 88
FEATURE           Location/Qualifiers
misc_feature      1..88
note = Description of Artificial Sequence: Synthetic
            oligonucleotide
source            1..88
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 454
gtcacctcca atgactaggg gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt gttttttt          88

SEQ ID NO: 455      moltype = RNA  length = 88
FEATURE           Location/Qualifiers
misc_feature      1..88
note = Description of Artificial Sequence: Synthetic
            oligonucleotide
source            1..88
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 455
gacatcgatg tcctcccat gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt gttttttt          88

SEQ ID NO: 456      moltype = RNA  length = 88

```

-continued

```

FEATURE          Location/Qualifiers
misc_feature    1..88
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source          1..88
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 456
gagtccgagc agaagaagaa gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt gttttttt                         88

SEQ ID NO: 457      moltype = RNA  length = 88
FEATURE          Location/Qualifiers
misc_feature    1..88
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source          1..88
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 457
ggggccgaga ttgggtgttc gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt gttttttt                         88

SEQ ID NO: 458      moltype = RNA  length = 88
FEATURE          Location/Qualifiers
misc_feature    1..88
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source          1..88
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 458
gtgcgcgagag gggccgagat gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt gttttttt                         88

SEQ ID NO: 459      moltype = RNA  length = 103
FEATURE          Location/Qualifiers
misc_feature   1..103
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source          1..103
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 459
gtcacacctca atgacttaggg gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgcttt ttt           103

SEQ ID NO: 460      moltype = RNA  length = 103
FEATURE          Location/Qualifiers
misc_feature   1..103
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source          1..103
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 460
gacatcgatc tcctccccat gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgcttt ttt           103

SEQ ID NO: 461      moltype = RNA  length = 103
FEATURE          Location/Qualifiers
misc_feature   1..103
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source          1..103
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 461
gagtccgagc agaagaagaa gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcaac ttgaaaaagt ggcaccgagt cggtgcttt ttt           103

SEQ ID NO: 462      moltype = RNA  length = 103
FEATURE          Location/Qualifiers
misc_feature   1..103
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide

```

-continued

```

source          1..103
               mol_type = other RNA
               organism = synthetic construct
SEQUENCE: 462
ggggccgaga ttgggtgttc gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcaac ttgaaaagt ggcaccgagt cggtgcttt ttt                      103

SEQ ID NO: 463      moltype = RNA  length = 103
FEATURE           Location/Qualifiers
misc_feature      1..103
note = Description of Artificial Sequence: Synthetic
               oligonucleotide
source          1..103
               mol_type = other RNA
               organism = synthetic construct
SEQUENCE: 463
gtggcgagag gggccgagat gtttagago tagaaatagc aagtaaaaat aaggctagtc 60
cgttatcaac ttgaaaagt ggcaccgagt cggtgcttt ttt                      103

SEQ ID NO: 464      moltype = DNA  length = 120
FEATURE           Location/Qualifiers
misc_feature      1..120
note = Description of Artificial Sequence: Synthetic
               polynucleotide
source          1..120
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 464
gtgaaaagga cgaaacacccg ggttccgag aagacctgtt ttagagctag aaatagcaag 60
ttaaaaataag gctagtcgt tatcaacttg aaaaagtggc accgagtcgg tgctttttt 120

SEQ ID NO: 465      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: 465
tcggtgcgct gtttgatttc ttcttgcgt ttttggctt                                40

SEQ ID NO: 466      moltype = RNA  length = 26
FEATURE           Location/Qualifiers
misc_feature      1..26
note = Description of Artificial Sequence: Synthetic
               oligonucleotide
source          1..26
               mol_type = other RNA
               organism = synthetic construct
SEQUENCE: 466
gatttcttct tgcgctttt gttta                                         26

SEQ ID NO: 467      moltype = DNA  length = 26
FEATURE           Location/Qualifiers
misc_feature      1..26
note = Description of Artificial Sequence: Synthetic
               oligonucleotide
misc_difference   22..26
note = a, c, t, g, unknown or other
source          1..26
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 467
tgatttcttc ttgcgtttt tnnnnn                                         26

SEQ ID NO: 468      moltype = DNA  length = 26
FEATURE           Location/Qualifiers
misc_feature      1..26
note = Description of Artificial Sequence: Synthetic
               oligonucleotide
misc_difference   21
note = a, c, t, g, unknown or other
source          1..26
               mol_type = other DNA
               organism = synthetic construct

```

-continued

```

SEQUENCE: 468
tgatttcttc ttgcgcgttt ntggct                                26

SEQ ID NO: 469      moltype = DNA  length = 26
FEATURE          Location/Qualifiers
misc_feature     1..26
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_difference   2
                  note = a, c, t, g, unknown or other
source           1..26
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 469
tnatttcttc ttgcgcgttt ttggct                                26

SEQ ID NO: 470      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: 470
gatttcttct tgcgcttttt tgg                                23

SEQ ID NO: 471      moltype = DNA  length = 34
FEATURE          Location/Qualifiers
misc_feature     1..34
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source            1..34
                  mol_type = other DNA
                  organism = synthetic construct
CDS               1..33

SEQUENCE: 471
tccatccgta caaccccacaa ccctgctagt gagc                                34

SEQ ID NO: 472      moltype = AA   length = 11
FEATURE          Location/Qualifiers
REGION           1..11
                  note = Description of Artificial Sequence: Synthetic peptide
source            1..11
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 472
SIRTTHNPAS E                                              11

SEQ ID NO: 473      moltype = DNA  length = 34
FEATURE          Location/Qualifiers
misc_feature     1..34
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source            1..34
                  mol_type = other DNA
                  organism = synthetic construct
CDS               1..33

SEQUENCE: 473
tccatcgcaa caaccccacaa ccctgctagt gagc                                34

SEQ ID NO: 474      moltype = AA   length = 11
FEATURE          Location/Qualifiers
REGION           1..11
                  note = Description of Artificial Sequence: Synthetic peptide
source            1..11
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 474
SIATTHNPAS E                                              11

SEQ ID NO: 475      moltype = DNA  length = 34
FEATURE          Location/Qualifiers
misc_feature     1..34
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide

```

-continued

```

source          1..34
               mol_type = other DNA
               organism = synthetic construct
CDS            1..33
SEQUENCE: 475 tcaatccgta caacccacaa ccctgctagt gago           34
SEQ ID NO: 476          moltype = DNA  length = 42
FEATURE          Location/Qualifiers
source           1..42
               mol_type = unassigned DNA
               organism = Homo sapiens
CDS            1..36
SEQUENCE: 476 caattgaatt taaaagaaac cgataccgtt ttggtttaag ga           42
SEQ ID NO: 477          moltype = AA   length = 12
FEATURE          Location/Qualifiers
source           1..12
               mol_type = protein
               organism = Homo sapiens
SEQUENCE: 477 QLNLKETDTV LV                           12
SEQ ID NO: 478          moltype = DNA  length = 42
FEATURE          Location/Qualifiers
source           1..42
               mol_type = unassigned DNA
               organism = Homo sapiens
CDS            1..42
SEQUENCE: 478 caattgaatt taaaagaaac cgataccgtt tacgaaattt ga           42
SEQ ID NO: 479          moltype = AA   length = 14
FEATURE          Location/Qualifiers
source           1..14
               mol_type = protein
               organism = Homo sapiens
SEQUENCE: 479 QLNLKETDTV YEIG                           14
SEQ ID NO: 480          moltype = DNA  length = 34
FEATURE          Location/Qualifiers
source           1..34
               mol_type = unassigned DNA
               organism = Homo sapiens
CDS            2..34
SEQUENCE: 480 tcctaaaaaa ccgaactccg cgctgcgtaa agta           34
SEQ ID NO: 481          moltype = AA   length = 11
FEATURE          Location/Qualifiers
source           1..11
               mol_type = protein
               organism = Homo sapiens
SEQUENCE: 481 PKKPNSALRK V                           11
SEQ ID NO: 482          moltype = DNA  length = 34
FEATURE          Location/Qualifiers
source           1..34
               mol_type = unassigned DNA
               organism = Homo sapiens
CDS            2..34
SEQUENCE: 482 tcctacaaaa ccgaactccg cgctgcgtaa agta           34
SEQ ID NO: 483          moltype = AA   length = 11
FEATURE          Location/Qualifiers
source           1..11
               mol_type = protein
               organism = Homo sapiens
SEQUENCE: 483 PTKPNSALRK V                           11
SEQ ID NO: 484          moltype = DNA  length = 33

```

-continued

FEATURE	Location/Qualifiers	
source	1..33	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 484	tgcgctggtt gatttcttct tgcgcttttt tgg	33
SEQ ID NO: 485	moltype = DNA length = 33	
FEATURE	Location/Qualifiers	
source	1..33	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 485	tacgctggtt gatttcttct tgcgcttttt ttg	33
SEQ ID NO: 486	moltype = DNA length = 27	
FEATURE	Location/Qualifiers	
source	1..27	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 486	ggaggggtttt atggggaaag gccattg	27
SEQ ID NO: 487	moltype = DNA length = 29	
FEATURE	Location/Qualifiers	
source	1..29	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 487	gtaaaaaaaaga agactagaaaaa ttttgatac	29
SEQ ID NO: 488	moltype = DNA length = 46	
FEATURE	Location/Qualifiers	
source	1..46	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 488	ggaggggtttt atggggaaag gcaaagaaga ctagaaattt tgatac	46
SEQ ID NO: 489	moltype = DNA length = 27	
FEATURE	Location/Qualifiers	
source	1..27	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 489	aggtgaagca taatgtctca aaaaaaa	27
SEQ ID NO: 490	moltype = DNA length = 29	
FEATURE	Location/Qualifiers	
source	1..29	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 490	attttattaa tacaaatcag tgaaaatcat	29
SEQ ID NO: 491	moltype = DNA length = 46	
FEATURE	Location/Qualifiers	
source	1..46	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 491	aggtgaagca taatgtctca aaattaatac aaatcagtga aatcat	46
SEQ ID NO: 492	moltype = DNA length = 36	
FEATURE	Location/Qualifiers	
source	1..36	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
CDS	1..36	
SEQUENCE: 492	aatttaaaag aaaccgatac cgtttacgaa attgga	36
SEQ ID NO: 493	moltype = AA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
	mol_type = protein	

-continued

```

SEQUENCE: 493          organism = Homo sapiens
NLKETDTVYE IG                      12

SEQ ID NO: 494          moltype = DNA length = 36
FEATURE          Location/Qualifiers
source           1..36
mol_type = unassigned DNA
organism = Homo sapiens
CDS              1..30
SEQUENCE: 494          aatttaaaag aaaccgatac cgtttgggtt taagga            36

SEQ ID NO: 495          moltype = AA length = 10
FEATURE          Location/Qualifiers
source           1..10
mol_type = protein
organism = Homo sapiens
SEQUENCE: 495          NLKETDTVLV                         10

SEQ ID NO: 496          moltype = DNA length = 36
FEATURE          Location/Qualifiers
source           1..36
mol_type = unassigned DNA
organism = Homo sapiens
CDS              1..36
SEQUENCE: 496          tgggatccaa aaaaatatgg tggttttgat agtcca            36

SEQ ID NO: 497          moltype = AA length = 12
FEATURE          Location/Qualifiers
source           1..12
mol_type = protein
organism = Homo sapiens
SEQUENCE: 497          WDPKKYGGFD SP                         12

SEQ ID NO: 498          moltype = DNA length = 36
FEATURE          Location/Qualifiers
source           1..36
note = Description of Artificial Sequence: Synthetic
oligonucleotide
1..36
mol_type = other DNA
organism = synthetic construct
CDS              1..36
SEQUENCE: 498          tgggatccaa aaaaatatgg tggttttgat agtcca            36

SEQ ID NO: 499          moltype = AA length = 12
FEATURE          Location/Qualifiers
REGION           1..12
note = Description of Artificial Sequence: Synthetic peptide
1..12
mol_type = protein
organism = synthetic construct
SEQUENCE: 499          WDPKKYCGFD SP                         12

SEQ ID NO: 500          moltype = DNA length = 35
FEATURE          Location/Qualifiers
source           1..35
note = Description of Artificial Sequence: Synthetic
oligonucleotide
1..35
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 500          aaactacttt acgcagcgcg gagttcggtt ttttg            35

SEQ ID NO: 501          moltype = DNA length = 4104
FEATURE          Location/Qualifiers
source           1..4104
mol_type = unassigned DNA
organism = Homo sapiens

```

-continued

```
SEQ ID NO: 502          moltype = AA  length = 1368
FEATURE                  Location/Qualifiers
source                   1..1368
                        mol type = protein
```

-continued

organism = Homo sapiens

SEQUENCE: 502
MDKKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIIKKNLIGA LLFDGSGETAE 60
ATRLKRTRARR YRTRRKNCRIC YLQEIFSNEM AKVDDSSFHR LEESFLVEED KKHERHPIFG 120
NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLNPDNSD 180
VDKLFITQLVQ TYNQLFEEENP INASGVDAKA ILSARLSKS RLENLIAQLP GEKKNGLFGN 240
LIALSISGLNL NFKSNFDLAE DAKLQLSKDT YDDDLNDLNA QIGDQYADLPL LAAKNLSDAI 300
LLSDILRVNT EITKAPLSSA MIKRYDEHHQ DLTLKLVALR QOLPEKYKEI FFDQSKNGYA 360
GYIDGGASQE EFYKFKIPIL EKMDGTEELL VKLNRDPLL RQRTFDNGSI PHQIHLGELH 420
AILRRQEDFY PFLKDNRREKI EKILTFPRIPY YVGLPLARGNS RFAWMTRKSE ETITTPWNFEE 480
VVDKGASAQS FIERMTNFID NLPNEKVLPK HSLLYEFYTV YNELTKVKVV TEGMRKPAPL 540
SEQEKKAIVD LLFKTNRKVVT VKQLKEDYFK KIECPDSVEI SGVEDRFNAS LGTYHDLLKI 600
IKDKDFLDNE ENEDILEDIV LTTLTLEFDRE MIEERLRLKTYA HLFDDKVMQK LKRRRYTGWG 660
RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSQGDSL 720
HEHIANLAGS PAIKKGILQT VVVKVDELVKV MGRHKPENIV IAMARENQTT QKGQKNSRBR 780
MKRIEEGIKE LGSQILKEPH VENTVQNEK LYLYLQLQNGR DMVYDQELDI NRLSDYDVA 840
IPVQSFLLKD SIDAKVLTRS DKARGKSDNV PSEEVVKKMK NYWRQOLLNAK LITQRKFDSL 900
TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDLSRMN TKYDENDKLI REVKVLTLKS 960
KLVSDFRKDF QFYKVREINN YHHAAHAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK 1020
MIAKSEQEIG KATAKYFFYS NIMNNFFKTEI TLANGEIRK PLIETNGETG EIVWDKGDRK 1080
ATVRKVLMSM OVNIVKVTEV OTGGFSKEST LPKRNSDLKL ARKKWDWPKK YYGGFDSPTVA 1140
YSVLVVAKVE KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDLI1KLPK 1200
YSLFELENGR KRMLASAGEL QKGNEELALPS KYVNPNFLYAS HYEKLKGSP E DNEQKQLFVE 1260
QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRDK PIREQAENII HLFTLTNLGA 1320
PAAFKYPFDTT IDRKRYTSTK EVLDATLHQ SITGLYETRI DLSQLGGD 1368

SEQ ID NO: 503
FEATURE moltype = DNA length = 15
source Location/Qualifiers
.1..15
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 503
cagaagaaga agggc 15

SEQ ID NO: 504
FEATURE moltype = DNA length = 51
source Location/Qualifiers
.1..51
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 504
ccaatgggaa ggacatcgat gtcacctcca atgactaggg tgggtggcaa c 51

SEQ ID NO: 505
FEATURE moltype = DNA length = 15
source Location/Qualifiers
.1..15
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 505
ctctggccac tccct 15

SEQ ID NO: 506
FEATURE moltype = DNA length = 52
source Location/Qualifiers
.1..52
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 506
acatcgatgt caccctcaat gacaaggctt ctgcgggtgg gcaaccacaa ac 52

SEQ ID NO: 507
FEATURE moltype = DNA length = 1733
misc_feature Location/Qualifiers
.1..1733
note = Description of Artificial Sequence: Synth
polynucleotide
source .1..1733
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 507
ccgtttaaac aattctgcag gaattagtt attaatagta atcaattacg gggcattag 60
ttcatatggcc atatatggag ttccgcgtt cataacttac ggttaatggc ccgcctggc 120
gaccgccccaa cgaccccccgc ccattgcacgt caataatgac gtatgttccc atatgtaaacgc 180
caataggagc ttccatttgta cgtcaatggg tggagtattt acggtaaact gcccacttgg 240
cagtcacatca agtgtatcat atgccaagta cggcccttat tgacgtcaat gacggttaaat 300
ggccgcgcgtt gcattatgcc cagtagatgtt cctttagggg ctttctact tggcgttaca 360
tctacgtt atgtcatcgctt attacatggc tcgagggttag ccccaacgttc tgcttcactc 420
tccccatctc cccccccctcc ccaccccaaa ttgttattt atttatttt taatttttt 480
gtgcagcgtt ggggggggggg ggggggggggg ggcgcgcgc ggggggggggg ggggggggggg 540

-continued

gggggcgggg	cggggcgagg	cggagagggtg	cgccggcagc	caatcagagc	ggcgcgctcc	600
gaaaatttcc	ttttatggcg	agggcgccgc	ggcgccggcc	ctataaaaag	cgaagcgccg	660
ggcgggggega	agtgcgtcg	cgctgcettc	gccccgtgcc	ccgctccggc	ggcgctcgc	720
ggccgcggcc	cggctctga	ctgaccgcgt	tactcccaca	ggtgagcggg	cgggacggcc	780
cttctctcc	gggtgtaat	tagegcttgg	ttaatgacg	gcttgtttct	tttctgtggc	840
tgcgtaaag	ccttgggggg	ctccgggagg	gccttgggt	egggggggagc	ggctcggggg	900
gtgcgtgcgt	gtgtgtgtgc	gtgggggagg	cegcgtcggt	ctcgegtgc	cccgccggct	960
gtgagcgtg	cgggcgccgc	gccccgttt	gtcgctccg	cagtgtcgcc	gaggggagcg	1020
cggccggggg	cggtgccccg	cggtgcgggg	ggggctgcga	ggggaaacaaa	ggctgcgtgc	1080
gggggtgtgt	cggtgggggg	tgagcagggg	gtgtggggcc	gtcggtcggt	ctgaaccccc	1140
cctctgaccc	ccttcccca	gttgcgtgcg	aegggccggc	ttcgggtgcg	gggttcgtta	1200
cgggggcggtt	cgcggggctc	gcgcgtccgg	gccccgggtt	gcccgcggg	gggggtgcgg	1260
gccccggggg	gcccgcctgg	gcccgggggg	gctcgggggg	ggggcgcggc	ggccccccgg	1320
gccccggggg	ctgtcgaggg	cgccggcggcc	gcgcatttc	ccttttatgg	taatcggtcg	1380
agaggggcgt	gggacttc	ttgtccccaa	tctgtcggg	gcccgaatct	ggggaggcc	1440
gccccacccc	ctcttgcggg	cgccccggca	agcggtcgcc	cgccggcagg	aaggaaatgg	1500
gccccggggg	ccttcgtgcg	tcgcgcgcgc	gcccgtccct	tctccctctc	cagcctcggt	1560
gtctgtcccg	gggggacggc	tgccctcggt	ggggacgggg	caggccgggg	ttcggcttct	1620
ggcggtgtgc	cgccggctct	agacgcctcg	ctaaccatgt	tcatgccttc	ttcttttctc	1680
taacagtcct	ggggacacgt	ctgggttattg	tgctgtctca	tatctttggc	aaa	1733

```

SEQ ID NO: 508      moltype = DNA    length = 4269
FEATURE          Location/Qualifiers
misc_feature      1..4269
                   note = Description of Artificial Sequence: Synthetic
                   polynucleotide
source            1..4269
                   mol_type = other DNA
                   organism = synthetic construct

```

SEQUENCE: 508							
atggactata	aggaccacga	cggagactac	aaggatcatg	atattgatta	caaagacgat	60	
gacgataaga	tggccccc	aaa	gaagaagcg	aaaggtcgtt	tccacggagt	cccagcagcc	120
gacaagaaga	atagcatcg	cctggacatc	ggcaccaact	ctgtgggtcg	ggccgtgtatc	180	
accgacgagt	acaagggtgc	cagcaagaaa	ttcaagggtc	ttggcaacac	cgaccggcac	240	
agcatcaaga	ttttttatggat	cgaggccctg	ctgttcgacaa	gccccggaaac	agcccgaggcc	300	
accgggtga	agagaacccg	cagaagaga	tacaccacag	ggaagaacccg	gatctgttat	360	
ctgcaagaga	tcttcagcaa	cgagatggcc	aagggtggacg	acagcttctt	ccacagactg	420	
gaagagtcct	tccttgggtt	gaaggataag	aagcacgcgc	ggcaccccat	tttcggcaac	480	
atcggtggac	agggtggctt	ccacggaga	tacccca	tcttaccact	gagaagaagaaa	540	
ctgggtggaca	gcacccactt	ggccgaccc	cggtgtatct	atctggccct	ggccccatcg	600	
atcaagttcc	ggggccactt	cctgtatcg	ggccgaccc	accccgacaa	cagcgacgtg	660	
gacaaggtgt	tcatccagct	ggtgcagacc	tacaaccacg	tgttcgagga	aaacccatc	720	
aaacggccacg	cgctggacgc	caaggccatc	ctgtctgcgc	gactgaccaa	gaggcagacgg	780	
ctggaaaatc	tgatcgccca	getggccgg	gagaagaaga	atggccctgt	cgggaaactcg	840	
attggccctga	gctgtggccct	gaccccaac	ttcaagacca	acttcgact	ggcccgaggat	900	
gccaaactgc	agctgagcaa	ggacacccatc	gacgacgacc	tggacaacct	gtctggcccg	960	
atcgccgacc	agtacccgc	cctgtttcg	ggccccaaggaa	acctgtccga	cgccatctcg	1020	
ctgaggcaca	tcttggatgt	gaacacccgg	atccacaa	ccccctgg	cgccctctat	1080	
atcaagagat	acgacgagca	ccacccaggac	ctgacccctgc	tgaaagctc	cggtggccag	1140	
cagctgcctg	agaagtacaa	agagattttc	ttcgaccaga	gcaagaacccg	ctacggccgc	1200	
taatttgcac	ggggggccac	ccacccggaa	ttcttacaa	gttcatcaaccc	cattctggaa	1260	
aaagatggac	gcacccggac	actgtcgat	ttcatcaaccc	gagggact	gtctggggaa	1320	
cagccggaccc	tccacaacccg	cacatcccc	cacccatcc	acccggggaa	gtctgcaccc	1380	
attctgcggc	ggcaggaa	tttttacca	ttcttgcagg	acaacccgg	aaagatcgag	1440	
aaagatctga	ctttccgc	cccccactac	gtggcccttc	ttggccagggg	aaacacccg	1500	
tttcgccttgc	tgaccggaaa	gagccgggg	accatccac	ccttggaa	acttggaaat	1560	
gttgacacgg	gggttccgc	ccacggatcc	atcgacggaa	ttccacaa	cgataaaaa	1620	
ctggcccaacg	agaagggtgt	gccccggac	agectgtgt	acgagttact	caccgtgtat	1680	
aaacggatgt	ccaaatgtaa	atacgtgacc	gagggatgt	gaaacccgc	tttgcgttgc	1740	
ggcgacacg	aaaaggccat	cgtggaccc	ctgttc	ttggccaggaa	actgtggcc	1800	
aacggatgt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1860	
ggcggtggaa	atcggttcaa	cgccccc	tttttttttt	tttttttttt	tttttttttt	1920	
aaaggacaagg	acttccgg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1980	
accctgcac	tgtttggat	cacccatcc	tttttttttt	tttttttttt	tttttttttt	2040	
ctgttcgac	aaacatgtat	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2100	
ctgagccgg	actgtatcaa	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2160	
tttctgtat	ccggccgg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2220	
ctgacccctt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2280	
gacccat	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2340	
aagggtgg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2400	
aaaaatggc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2460	
aaacggatgt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2520	
aaaaacaccc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2580	
atgtacgtgg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2640	
gtgcctcaga	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2700	
aaacggccgg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2760	

-continued

tactggcgc	agctgctgaa	cgttcaagctg	attacccaga	gaaagtgcg	caatctgacc	2820
aaggccgaga	gaggcggcct	gagcgaaactg	gataaggccc	gcttcatcaa	gagacagctg	2880
gtggaaaccc	ggcagatcac	aaagcacgtg	gcacagatcc	tggactcccg	gatgaacact	2940
aaagtacgac	agaatgacaa	gctgatccgg	gaagtggaaa	tgatcacct	gaagtccaa	3000
ctgggtgtccg	atttccggaa	ggatttccag	tttacaaag	tgcgcgagat	caacaactac	3060
caccacgccc	acgacgccta	cctgaaacgc	gtcggtggaa	ccgcctcgat	caaaaagtag	3120
cctaagctgg	aaagcgagtt	cggtgtacgg	gactacaagg	tgtacgacgt	gcggaagatg	3180
atcgccaaga	gogagcagga	aatcgcaag	gctaccgc	agtacttctt	ctacagcaac	3240
atcatgaact	tttcaagac	cgagattacc	ctggccaac	gcgagatccg	gaagcggcct	3300
ctgatcgaga	caaaccggc	aaccggggag	atcgtgtgg	ataaggggccg	ggattttgcc	3360
accgtgtgg	aaagtgttag	catggggccaa	gtgaaatatcg	tgaaaaagac	cgagggtcg	3420
acaggcgcct	ttagaaaga	gtctatcc	cccaagagga	acagcgat	gctgatcgcc	3480
agaaagaagg	actgggacc	taagaagtac	ggcggttc	acagccccac	cgtggcttat	3540
tctgtgtgg	tttgtggccaa	agtggaaaag	ggcaagtcc	agaaactgaa	gagtgtgaaa	3600
gagctgtgg	ggatcaccat	catggaaaga	agcagcttcg	agaagaatcc	catcgactt	3660
ctgaaagc	agggtacaa	aaagtggaaa	aaggacctg	tcatcaact	gcctaagtac	3720
tccctgtcg	actggaaaa	cggcggaaag	agaatgtgg	cctctgc	cgaactgcag	3780
aaggaaacg	aactggccct	gcctccaaa	tatgtgaact	tcctgtact	ggccagccac	3840
tatgagaage	tgaaggcgtc	ccccggagg	aatgagcaga	aacagctgtt	tgtggaaacag	3900
cacaaggact	acctggacg	gatcatcg	catagcgg	agttctccaa	gagagtgtac	3960
ctggccgacg	ctaatactg	caaagtgt	tccgc	taca	acaaggaccc	4020
atcagagagc	aggccgagaa	tatccat	ctgtttacc	tgaccaat	gggagccct	4080
gcccgcctca	actacttgc	caccatc	gaccggaa	ggtagccac	caccaaaagag	4140
gtgctggacg	ccaccctgt	ccacca	ggacggag	atcaccggc	tgtacgagac	4200
ctgtctc	ctggaggcga	caaaggcc	gcccgcac	aaaaggccg	ccaggcaaaa	4260
aagaaaaag						4269

SEQ ID NO: 509	moltype = DNA	length = 780
FEATURE	Location/Qualifiers	
misc_feature	1..780	
	note = Description of Artificial Sequence: Synthetic	
	polynucleotide	
source	1..780	
	mol_type = other DNA	
	organism = synthetic construct	

SEQUENCE: 509							
ggaa	gcgg	ccactaactt	ctcccgttg	aaacaagcg	gggatgtcg	agagaatccc	60
ggggc	cagt	gcaaggcga	ggagctgtt	accgggg	tgc	ccatct	120
gacgg	gcg	taaacggc	caatcg	gtgtccgg	aggggc	aggatgc	180
tacgg	caac	tgacc	ttgc	accacggc	agtc	ccctgg	240
accct	ctgt	ccacc	ctac	cgatgt	cc	ctccat	300
aagc	cacg	actt	ccat	gtcc	ccat	cgaccat	360
ttctt	caag	ccgg	ctca	ggc	cc	gcac	420
ctgg	gtt	accc	ctaa	ggc	cc	gacacc	480
caca	agg	gtat	cgac	gact	cc	atctgg	540
aacgg	cat	actt	caac	tcata	tc	atggc	600
ccgg	acc	agg	tgat	caat	ca	acggc	660
cact	act	acc	ccat	ccat	cc	gtgtct	720
gtc	ctgt	ggc	gtac	ggc	cc	ggccaca	780

SEQ ID NO: 510	moltype = DNA	length = 597
FEATURE	Location/Qualifiers	
misc_feature	1..597	
	note = Description of Artificial Sequence: Synthetic	
	polynucleotide	
source	1..597	
	mol_type = other DNA	
	organism = synthetic construct	

SEQUENCE: 510						
cgataatcaa	cctctggatt	acaaaatttg	tgaagattt	actggat	ttaactatgt	60
tgtcc	ttt	acgtatgt	gtatcg	ttaatgc	ttgtatcat	120
ccgtatgg	ttc	ttt	cctctt	taa	atctgg	180
gttgg	ggc	ttt	tttgc	ttgt	gtct	240
actcttgg	ggc	ttt	tttgc	ttgt	gtct	300
ccctatttgc	acggc	ttt	tttgc	ttgt	gtct	360
gtctgttggc	act	ttt	tttgc	ttgt	gtct	420
gtctgcctgt	gtt	ttt	tttgc	ttgt	gtct	480
cctcaatcca	gggg	ttt	tttgc	ttgt	gtct	540
tcttcg	ccct	ttt	tttgc	ttgt	gtct	597

SEQ ID NO: 511	moltype = DNA	length = 210
FEATURE	Location/Qualifiers	
misc_feature	1..210	
	note = Description of Artificial Sequence: Synthetic	
	polynucleotide	
source	1..210	

-continued

-continued

cgaggggaaat tgaacagggtg taaaatttgg a gggacaagac ttcccacaga ttttcggtt 120
tgtcggaaat ttttttaata a gggcaaaaata aggaaaatgg gaggataggt agtcatctgg 180
ggttttatgc agccaaaactt cagggttta ttgttgcgtg tccgcctcg agtatcc 240
atcgaggtag attaaagaca tgctcaccc agttttatac ttcctcgat gagatccta 300
ctacatgtt aaaaaaaaaatcgt gtcgcgatgt agactatgt a gcaaaaaattt taatctttt 360
taaagagccc agtacttcat atccatttt cccgctcctt ctgcagcctt atcaaaaagg 420
attttagaac actcattttt gccccatttt cattttatc actgggtat ccaaccctta 480
gacagagcat tggcattttt cttttctcgat tctttagaaat ctgtatgact atgaaaccag 540
acagatgtt tacatcaccc acaaatcgag gctgtatgtt ggccctcaac atgcgttc 600
ttttataact ctttagtaca ctttttgtt atcccttgcc ttgatccta atttctgt 660
tctatcacct ctcccgatcg gtgttgcgtt acatttggc ctatcttcag tccaggagt 720
tttacaaacaa tagatgtt gagaatccaa cctaaatggc aacttccac tcccatgat 780
gcctctcc tttttttccat tttttttttt gaggatattt ccattatgg tttccatgg 840
gatgttccat ccccccaatatt tacctgtatgt atcttacata ttgcaggct gatattttt 900
gacattttaa ggttattttt attttatggc cacatgtatgtt ttgatctgc ttactaaat 960
ttttgttcat tttttttttt tttttttttt aaaaagggttgg tttttttttt aatggaaatg tcaagggtt 1020
gttgttcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1080
ctcttgact ggcttgactt atggcatctt acacgttattt gctggctaa atgttattttt 1140
gcacaacttcc ttcaggactt ataaatttttgc ttgactgttgc gccaacacca agttaaatgt 1200
ttaaagcaaca aatgttatttttgc tttttttttt tttttttttt tttttttttt tttttttttt 1260
gtcttataat aataatccatc aagggttgcgg gagggttgcgg gaggatcaagc acacagact 1320
getcttccatc aagggttgcgg gagggttgcgg gaggatcaagc acacagact 1380
taatgggatc tgatggccctc ttctgggtgt tctgaagacc acaaggttat tcacattttt 1440
taataatataat cttcccttccat ttctttttttt tttttttttt tttttttttt tttttttttt 1500
aattttatgtt aatgtatggaa tactttgttgc ttgttccat tttttttttt tttttttttt 1560
tactctttttaa gcactggaaa ttgttaccaag gaactaattt tttttttttt tttttttttt 1620
gacagaggag ccataacttcg agacttgcgg gatacagaag accaaatgcgg actttatgt 1680
ctttttcttc acataagca ataaagaaat aaaaatttttgc ttctttagat tttttttttt 1740
tttaactgtca gctttttactt acttttttgc ttctatcttca acaactgttgc agttaatgtt 1800
geagccattttaa acataatgttgc ttgttgcgg gttttttttt tttttttttt tttttttttt 1860
cagagttagt ccagccatgtt gttttttttt tttttttttt tttttttttt tttttttttt 1920
ggcagagacaa gggcagatctt tgatgttgcgg cccaggcttgc ttctatccatc aatgttccatc 1980
tagtagatcc acgaatataaccc acagaaaaaccc ttgttgggggg gggggctctg agatccatca 2040
aaatttataat tgaagatcc tcttgcgg tttttttttt tttttttttt tttttttttt tttttttttt 2100
ttctgtatggat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2160
tgttggggat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2220
ctatggctttaa atggcatgttcc tttttttttt tttttttttt tttttttttt tttttttttt 2280
atttccagggg tgggggttggg aaatctttca aacactaaaaa ttgttcccttta atttttttttt 2340
taaaaaatgg gtttataataat aaacccatca aatatttttgc tttttttttt tttttttttt 2400
atttttatgtt gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2460
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2520
ttgcatcgat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2580
aaagtcaaat taagaatttgc tttttttttt tttttttttt tttttttttt tttttttttt 2640
gatagttttt cctggggatcc ggttctcaactt tttttttttt tttttttttt tttttttttt 2700
gttagataaa gtttgcgttgc tttttttttt tttttttttt tttttttttt tttttttttt 2760
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2820
gttggtaaat gtgtttaatgtt gatggaaact tttttttttt tttttttttt tttttttttt 2880
aaaaaaaaccc ccagggttgcgtt tagagcatgc tttttttttt tttttttttt tttttttttt 2940
caagagtatg tactttggcag gtttctcgat tttttttttt tttttttttt tttttttttt 3000
ccaaagaaactt acaggatggaa ttacatgttgc tttttttttt tttttttttt tttttttttt 3060
aaaatattcc aaataatatt tttttttttt tttttttttt tttttttttt tttttttttt 3120
tccgggtatag aggccatgtt gggaaacttcaatggatgg tttttttttt tttttttttt 3180
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3240
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3300
cctacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3360
gattttttat atcaaggccaa acatgttgc tttttttttt tttttttttt tttttttttt 3420
gttaggtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3480
acatgttcaat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3540
agtaaccctt cttatagagc acctttgttgc gtttccatc tttttttttt tttttttttt 3600
tttggatc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3660
tttggatc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3720
agaagatattt cttatgttgc tttttttttt tttttttttt tttttttttt tttttttttt 3780
gttaggtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3840
ggccctttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3900
aaataacttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3960
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 4020
cttggacttcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 4080
tttgcgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 4140
atataactat aaccatcaactt aactccactt ggttggatttt tttttttttt tttttttttt 4200
agtggatc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 4260
ctatgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 4320
actagttagt cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 4386

SEQ ID NO: 515
FEATURE
misc_feature

```
moltyp = DNA  length = 1846  
Location/Qualifiers  
1..1846
```

-continued

-continued

```

misc_feature      1..23
                  note = Description of Artificial Sequence: Synthetic primer
source           1..23
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 517
gagggcctat ttcccatgat tcc                                23

SEQ ID NO: 518      moltype = DNA  length = 22
FEATURE          Location/Qualifiers
misc_feature     1..22
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..22
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 518
cttgtggaaa ggacgaaaca cc                                22

SEQ ID NO: 519      moltype = DNA  length = 45
FEATURE          Location/Qualifiers
misc_feature     1..45
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_difference   4..23
                  note = a, c, t, g, unknown or other
source           1..45
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 519
aacnnnnnnn nnnnnnnnnn nnnggtgtt cgtccttcc acaag        45

SEQ ID NO: 520      moltype = DNA  length = 28
FEATURE          Location/Qualifiers
misc_feature     1..28
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..28
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 520
cctgagtgtt gagggccccag tggctgct                                28

SEQ ID NO: 521      moltype = DNA  length = 37
FEATURE          Location/Qualifiers
misc_feature     1..37
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..37
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 521
acgagggcag agtgctgctt gctgctggcc agggccc                 37

SEQ ID NO: 522      moltype = DNA  length = 68
FEATURE          Location/Qualifiers
misc_feature     1..68
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..68
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 522
catcaggctc tcagctcagc ctgagtgtt aggccctgct ggccaggccc ctgcgtggc  60
ccaagctg                                         68

SEQ ID NO: 523      moltype = DNA  length = 27
FEATURE          Location/Qualifiers
misc_feature     1..27
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..27
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 523
gagggcctat ttcccatgat tccttca                                27

```

-continued

```

SEQ ID NO: 524      moltype = DNA length = 125
FEATURE
misc_feature
1..125
note = Description of Artificial Sequence: Synthetic
polynucleotide
84..103
note = a, c, t, g, unknown or other
1..125
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 524
aaaaaaagca ccgactcggt gccactttt caagttgata acggacttagc cttattttaa 60
cttgctattt cttagctcaa aacnnnnnnn nnnnnnnnn nnnggtt cgtccttcc 120
acaag                                         125

SEQ ID NO: 525      moltype = DNA length = 111
FEATURE
misc_feature
1..111
note = Description of Artificial Sequence: Synthetic
polynucleotide
9..28
note = a, c, t, g, unknown or other
1..111
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 525
gaaacaccnn nnnnnnnnnn nnnnnnnnnt ttagagcta gaaatagcaa gttaaaaataa 60
ggctagtcg ttatcaactt gaaaaagtgg caccgagtcg gtgttttt t           111

SEQ ID NO: 526      moltype = DNA length = 20
FEATURE
misc_feature
1..20
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 526
agctgttta ctggtcggct                                         20

SEQ ID NO: 527      moltype = DNA length = 20
FEATURE
misc_feature
1..20
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 527
aatggataca cctggtcgaa                                         20

SEQ ID NO: 528      moltype = DNA length = 20
FEATURE
misc_feature
1..20
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 528
caatggatac acctggtcga                                         20

SEQ ID NO: 529      moltype = DNA length = 68
FEATURE
source
1..68
mol_type = unassigned DNA
organism = Homo sapiens
SEQUENCE: 529
accatgtata ccacttgggc tttggcagta gctaactgca ctaaatataa tataaggagg 60
gttttatg                                         68

SEQ ID NO: 530      moltype =
SEQUENCE: 530
000

```

-continued

SEQ ID NO: 531	moltype =	length =
SEQUENCE: 531		
000		
SEQ ID NO: 532	moltype =	length =
SEQUENCE: 532		
000		
SEQ ID NO: 533	moltype =	length =
SEQUENCE: 533		
000		
SEQ ID NO: 534	moltype =	length =
SEQUENCE: 534		
000		
SEQ ID NO: 535	moltype =	length =
SEQUENCE: 535		
000		
SEQ ID NO: 536	moltype =	length =
SEQUENCE: 536		
000		
SEQ ID NO: 537	moltype =	length =
SEQUENCE: 537		
000		

1. (canceled)
2. An engineered CRISPR-Cas system comprising:
 - (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 chimeric RNA comprises a guide sequence capable of hybridizing to a target sequence adjacent to a protospacer adjacent motif (PAM) in a genomic locus of interest of a eukaryotic cell, a tracr-mate sequence capable of hybridizing to a tracr sequence, and a tracr sequence comprising 40 or more nucleotides in length; wherein the chimeric RNA is capable of forming a CRISPR complex with the Cas9 protein in the eukaryotic cell, and wherein the guide sequence is capable of directing sequence-specific binding of the CRISPR complex to the target sequence adjacent to the PAM in the genomic locus of interest of the eukaryotic cell.
3. The engineered CRISPR-Cas system of claim 2, wherein the PAM is NGG.
4. The engineered CRISPR-Cas system of claim 2, wherein the tracr sequence comprises 50 or more nucleotides in length.
5. The engineered CRISPR-Cas system of claim 2, wherein the chimeric RNA comprises

NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAUGCAAGUUAAAAAU
AAGGCUAGUCGGUUAUCA.
6. The engineered CRISPR-Cas system of claim 2, wherein the chimeric RNA further comprises a poly-U sequence.
7. The engineered CRISPR-Cas system of claim 2, wherein the chimeric RNA is encoded by SEQ ID NO:25.
8. The engineered CRISPR-Cas system of claim 2, wherein the chimeric RNA comprises one or more modified nucleotides.
9. The engineered CRISPR-Cas system of claim 2, wherein the chimeric RNA comprises one or more methylated nucleotides or nucleotide analogs.
10. The engineered CRISPR-Cas system of claim 2, wherein the two or more NLSs are independently selected from the group consisting of PKKKRKV, KRPAATKK-AGQAKKKK, PAAKRVKLD, QRQRNELKRSP, NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAK-PRNQGGY, RMRIZFKNKGKDTELRRRRVEVS-VELRKAKKDEQILKRRNV, VSRKRPPR, PPKKARED, PQPKKKPL, SALIKKKKKMAP, DRLRR, PKQKKRK, RKLKKKIKKL, REKKKFLKRR, KRKGDEVVDGVDEVAKKKSKK, and RKCLQAGMNLEARKTKK
11. The engineered CRISPR-Cas system of claim 10, wherein at least one of the NLSs comprises PKKKRKV.
12. The engineered CRISPR-Cas system of claim 2, wherein the Cas9 protein comprises a mutation of D10A, H840A, N854A, or N863A.
13. The engineered CRISPR-Cas system of claim 12, wherein the Cas9 protein is fused with at least one heterologous protein domain.
14. The engineered CRISPR-Cas system 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 engineered CRISPR-Cas system of claim 2, wherein the polynucleotide encoding the Cas9 protein is codon-optimized for expression in the eukaryotic cell.

16. The engineered CRISPR-Cas system of claim 2, wherein the polynucleotide encoding the Cas9 protein comprises a polyadenylation signal.

17. The engineered CRISPR-Cas system of claim 2, wherein the CRISPR-Cas system is comprised in a liposome for delivery.

18. The engineered CRISPR-Cas system of claim 2, further comprising an exogenous polynucleotide for targeted integration into a DNA break introduced by the CRISPR complex.

19. The engineered CRISPR-Cas system of claim 2, wherein the CRISPR complex being directed by the guide sequence is capable of cleaving the genomic locus of interest, thereby modifying the eukaryotic cell with an insertion, deletion, or substitution of one or more nucleotides in the genomic locus of interest.

20. An engineered CRISPR-Cas system comprising:

(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, RQRRLNELKRSP, NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAK-PRNQGGY, RMRIZFKNKGKDTELRRRVEVS-VELRKAKKDEQILKRRNV, VSRKRPRP, PPK-KARED, PQPKKKPL, SALIKKKKKMAP, DRLRR, PKQKKRK, RKLKKKIKKL, REKKKFLKRR, KRKGDEVGVDEVAKKSKK, and RKCLQAGMNLEARKTKK;

(b) a CRISPR-Cas system chimeric RNA or a polynucleotide encoding the chimeric RNA, wherein the chimeric RNA comprises NNNNNNNNNNNNNNNNNNNNNNGUUUJAGAGC-UAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCA, wherein NNNNNNNNNNNNNNNNNNNNN is a guide sequence capable of hybridizing to a target sequence adjacent to a protospacer adjacent motif (PAM) in a genomic locus of interest of a eukaryotic cell;

wherein the chimeric RNA and the Cas9 protein are capable of forming a CRISPR complex in the eukaryotic cell, and wherein the guide sequence is capable of directing sequence-specific binding of the CRISPR complex to the target sequence adjacent to the PAM in the genomic locus of interest of the eukaryotic cell.

21. An engineered CRISPR-Cas system comprising:

(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, RQRRLNELKRSP, NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAK-PRNQGGY, RMRIZFKNKGKDTELRRRVEVS-VELRKAKKDEQILKRRNV, VSRKRPRP, PPK-KARED, PQPKKKPL, SALIKKKKKMAP, DRLRR, PKQKKRK, RKLKKKIKKL, REKKKFLKRR, KRKGDEVGVDEVAKKSKK, and RKCLQAGMNLEARKTKK, wherein the polynucleotide is codon-optimized for expression in a eukaryotic cell and comprises a polyadenylation signal;

(b) a CRISPR-Cas system chimeric RNA comprising NNNNNNNNNNNNNNNNNNNNNGUUUUAGAGC-UAGAAAUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCA, wherein NNNNNNNNNNNNNNNNNNN is a guide sequence capable of hybridizing to a target sequence adjacent to a protospacer adjacent motif (PAM) in a genomic locus of interest of a eukaryotic cell;

wherein the chimeric RNA and the Cas9 protein are capable of forming a CRISPR complex in the eukaryotic cell, and wherein the guide sequence is capable of directing sequence-specific binding of the CRISPR complex to the target sequence adjacent to the PAM in the genomic locus of interest, wherein the CRISPR complex being directed by the guide sequence is capable of cleaving the genomic locus of interest in the eukaryotic cell.

* * * * *