(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 May 2016 (12.05.2016)

(10) International Publication Number
WO 2016/073433 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 15/62* (2006.01)

(21) International Application Number:
PCT/US2015/058760

(22) International Filing Date:
3 November 2015 (03.11.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/075,999 6 November 2014 (06.11.2014) US

(71) Applicant: **E. I. DU PONT DE NEMOURS AND COMPANY** [US/US]; Chestnut Run Plaza, 974 Centre Road, P.O. Box 2915, Wilmington, Delaware 19805 (US).

(72) Inventors: **FRISCH, Ryan L.**; 27 Berkley Drive, Newark, Delaware 19702 (US). **FAN, Xiaochun**; 175 Pratt Lane, West Chester, Pennsylvania 19382 (US). **HONG, Seung-Pyo**; 9 Pierson's Ridge, Hockessin, Delaware 19707 (US). **JACKSON, Ethel Noland**; 49 Guyencourt Road, Greenville, Delaware 19807 (US).

(74) Agent: **STOOP, Johan M.**; E. I. du Pont de Nemours and Company, Legal Patent Records Center, Chestnut Run Plaza 721/2340, 974 Centre Road, PO Box 2915 Wilmington, Delaware 19805 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

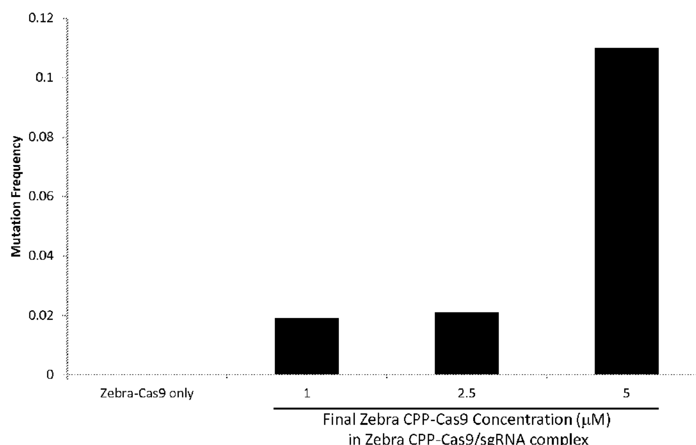
— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2016/073433 A1

(54) Title: PEPTIDE-MEDIATED DELIVERY OF RNA-GUIDED ENDONUCLEASE INTO CELLS

FIGURE 8



(57) Abstract: A composition is disclosed that comprises at least one protein component of an RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP), wherein the RGEN protein component and CPP are covalently or non-covalently linked to each other in an RGEN protein-CPP complex. The RGEN protein-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a cell. The RGEN protein component of an RGEN protein-CPP complex in certain embodiments can be associated with a suitable RNA component to provide an RGEN capable of specific DNA targeting. Further disclosed are compositions comprising at least one protein component of a guide polynucleotide/Cas endonuclease complex and at least one CPP, as well as methods of delivering RGEN proteins into microbial cells, as well as methods of targeting DNA with RGENs.

TITLE

PEPTIDE-MEDIATED DELIVERY OF RNA-GUIDED ENDONUCLEASE INTO CELLS

This application claims the benefit of U.S. Provisional Application No.
5 62/075999 filed November 6, 2014, incorporated herein in its entirety by reference.

FIELD OF INVENTION

The invention is in the field of molecular biology. Specifically, this invention pertains to delivery of protein components of RNA-guided endonucleases into cells using cell-penetrating peptides.

10 ## REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

The official copy of the sequence listing is submitted electronically via EFS-Web as an ASCII formatted sequence listing with a file named 20151013_CL6273PCT_SequenceListing_ST25.txt created October 13, 2015, and having a size of 384 kilobytes and is filed concurrently with the specification. The
15 sequence listing contained in this ASCII-formatted document is part of the specification and is herein incorporated by reference in its entirety.

BACKGROUND

A way to understand the function of a gene within an organism is to inhibit its expression. Inhibition of gene expression can be accomplished, for example, by
20 interrupting or deleting the DNA sequence of the gene, resulting in “knock-out” of the gene (Austin et al., *Nat. Genetics* 36:921-924). Gene knock-outs mostly have been carried out through homologous recombination (HR), a technique applicable across a wide array of organisms from bacteria to mammals. Another way for studying gene function can be through genetic “knock-in”, which is also usually
25 performed by HR. HR for purposes of gene targeting (knock-out or knock-in) can employ the presence of an exogenously supplied DNA having homology with the target site (“donor DNA”).

HR for gene targeting has been shown to be enhanced when the targeted DNA site contains a double-strand break (Rudin et al., *Genetics* 122:519-534; Smih
30 et al., *Nucl. Acids Res.* 23:5012-5019). Strategies for introducing double-strand

breaks to facilitate HR-mediated DNA targeting have therefore been developed. For example, zinc finger nucleases have been engineered to cleave specific DNA sites leading to enhanced levels of HR at the site when a donor DNA was present (Bibikova et al., *Science* 300:764; Bibikova et al., *Mol. Cell. Biol.* 21:289-297).
5 Similarly, artificial meganucleases (homing endonucleases) and transcription activator-like effector (TALE) nucleases have also been developed for use in HR-mediated DNA targeting (Epinat et al., *Nucleic Acids Res.* 31: 2952-2962; Miller et al., *Nat. Biotech.* 29:143-148).

Loci encoding CRISPR (clustered regularly interspaced short palindromic
10 repeats) DNA cleavage systems have been found exclusively in about 40% of bacterial genomes and most archaeal genomes (Horvath and Barrangou, *Science* 327:167-170; Karginov and Hannon, *Mol. Cell* 37:7-19). In particular, the CRISPR-associated (Cas) RNA-guided endonuclease (RGEN), Cas9, of the type II CRIPSR system has been developed as a means for introducing site-specific DNA strand
15 breaks that stimulate HR with donor DNA (U.S. Provisional Appl. No. 61/868,706, filed August 22, 2013). The sequence of the RNA component of Cas9 can be designed such that Cas9 recognizes and cleaves DNA containing (i) sequence complementary to a portion of the RNA component and (ii) a protospacer adjacent motif (PAM) sequence.

20 Native Cas9/RNA complexes comprise two RNA sequences, a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA). A crRNA contains, in the 5'-to-3' direction, a unique sequence complementary to a target DNA site and a portion of a sequence encoded by a repeat region of the CRISPR locus from which the crRNA was derived. A tracrRNA contains, in the 5'-to-3' direction, a sequence
25 that anneals with the repeat region of crRNA and a stem loop-containing portion. Recent work has led to the development of guide RNAs (gRNA), which are chimeric sequences containing, in the 5'-to-3' direction, a crRNA linked to a tracrRNA (U.S. Provisional Appl. No. 61/868,706, filed August 22, 2013).

Protein and RNA components for performing Cas9-mediated DNA targeting
30 in a cell have been provided in some studies through recombinant DNA expression strategies. For example, Cas9 protein has been expressed in cells using nucleic acid-based expression systems. Methods of expressing RNA components such as

gRNA in certain cell types have included using RNA polymerase III (Pol III) promoters, which allow for transcription of RNA with precisely defined, unmodified, 5'- and 3'-ends (DiCarlo et al., *Nucleic Acids Res.* 41: 4336-4343; Ma et al., *Mol. Ther. Nucleic Acids* 3:e161). These protein and RNA expression techniques have been applied in cells of several different species including maize and soybean (U.S. Provisional Appl. No. 61/868,706, filed August 22, 2013), as well as humans, mouse, zebrafish, *Trichoderma* and *Sacchromyces cerevisiae.*

Despite these advances, other means of providing protein and RNA components in a cell, such as a microbial cell, to mediate Cas9-mediated DNA targeting are of interest.

SUMMARY OF INVENTION

In one embodiment, the invention concerns a composition comprising at least one protein component of an RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP), wherein the protein component and CPP are covalently, or non-covalently, linked to each other in an RGEN protein-CPP complex, and wherein the RGEN protein-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a microbial cell.

In a second embodiment, the protein component of the RGEN is associated with at least one RNA component that comprises a sequence complementary to a target site sequence on a chromosome or episome in the microbial cell, wherein the RGEN can bind to the target site sequence, and optionally cleave one or both DNA strands at the target site sequence. In a third embodiment, the RNA component comprises a guide RNA (gRNA) comprising a CRISPR RNA (crRNA) operably linked to a trans-activating CRISPR RNA (tracrRNA). In a fourth embodiment, the RGEN can cleave one or both DNA strands at the target site sequence.

In a fifth embodiment, the RGEN comprises a CRISPR-associated (Cas) protein-9 (Cas9) amino acid sequence.

In a sixth embodiment, the RGEN protein component and CPP are covalently linked.

In a seventh embodiment, the RGEN protein component and CPP are non-covalently linked.

In an eighth embodiment, the CPP is cationic or amphipathic.

In a ninth embodiment, the CPP comprises (i) a CPP from an Epstein-Barr virus Zebra trans-activator protein, (ii) a CPP having 6 or more contiguous arginine residues, (iii) a transportan-10 (TP10) CPP, or (iv)a CPP from a vascular endothelium cadherin protein.

5 In a tenth embodiment, the RGEN protein-CPP complex can traverse a cell wall and cell membrane of a microbial cell.

An eleventh embodiment concerns a microbial cell comprising a composition disclosed herein.

A twelfth embodiment concerns a method of delivering a protein component 10 of an RNA-guided endonuclease (RGEN) into a microbial cell. This method comprises contacting a microbial cell with a composition comprising the RGEN protein component and at least one cell-penetrating peptide (CPP), wherein the RGEN protein component and CPP are covalently, or non-covalently, linked to each other in an RGEN protein-CPP complex. As a result of this contacting step, the 15 RGEN protein-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of the microbial cell, and thereby gain entry to the microbial cell.

In a thirteenth embodiment, with respect to the method, (i) the composition further comprises at least one RNA component that is associated with the protein component of the RGEN, or (ii) the microbial cell comprises the RNA component, 20 wherein the RNA component associates with the protein component of the RGEN after the RGEN protein-CPP complex enters the microbial cell; wherein the RNA component in (i) or (ii) comprises a sequence complementary to a target site sequence on a chromosome or episome in the microbial cell, and wherein the RGEN can bind to the target site sequence, and optionally cleave one or both DNA 25 strands at the target site sequence. In a fourteenth embodiment, the RGEN can cleave one or both DNA strands at the target site sequence. In a fifteenth embodiment, the microbial cell further comprises a donor polynucleotide comprising at least one sequence homologous to a sequence at or near the target site sequence, wherein the donor polynucleotide integrates at or near the target site 30 sequence by homologous recombination.

A sixteenth embodiment concerns a polynucleotide sequence comprising a nucleotide sequence encoding an RGEN protein-CPP fusion protein that comprises a protein component of an RNA-guided endonuclease (RGEN) and at least one cell-

penetrating peptide (CPP), wherein optionally, the nucleotide sequence is operably linked to a promoter sequence.

A seventeenth embodiment concerns a method of producing an RGEN protein-CPP fusion protein. This method comprises: (a) providing a polynucleotide sequence comprising a nucleotide sequence encoding an RGEN protein-CPP fusion protein that comprises a protein component of an RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP), wherein optionally, the nucleotide sequence is operably linked to a promoter sequence; (b) expressing the RGEN protein-CPP fusion protein from the polynucleotide sequence, thereby producing the RGEN protein-CPP fusion protein, wherein the expressing is optionally performed in a microbial cell; and (c) optionally, isolating the RGEN protein-CPP fusion protein produced in step (b).

An eighteenth embodiment concerns a composition comprising at least one protein component of a guide polynucleotide/Cas endonuclease complex and at least one cell-penetrating peptide (CPP), wherein the protein component and CPP are covalently, or non-covalently, linked to each other in a guide polynucleotide/Cas endonuclease-CPP complex, wherein the guide polynucleotide/Cas endonuclease-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a microbial cell.

A nineteenth embodiment concerns a method for modifying a target site in the genome of a microbial cell. This method comprises providing a guide polynucleotide, a cell-penetrating peptide (CPP) and a Cas endonuclease to a microbial cell, wherein the guide polynucleotide, Cas endonuclease and CPP are covalently, or non-covalently, linked to each other in a guide polynucleotide/Cas endonuclease-CPP complex, and wherein the guide polynucleotide/Cas endonuclease-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of the microbial cell

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCES

Figure 1: pZUFCas9 plasmid (SEQ ID NO:6) contains the *Yarrowia* codon-optimized Cas9 expression cassette set forth in SEQ ID NO:5. Origins of replication (ARS 18, f1 ori, ColE1) are in cross-hatch, and selectable markers (Ura3, Amp) are in grey. Refer to Example 1.

Figure 2A: pBAD/HisB plasmid (SEQ ID NO:10) for expressing heterologous proteins in *E. coli*. pBAD promoter is in white. Origin of replication is in cross-hatch. Refer to Example 1.

Figure 2B: pRF48 plasmid (SEQ ID NO:11) for expressing Cas9-NLS ("Cas9" in figure) in *E. coli*. Origin of replication is in cross-hatch. Refer to Example 1.

Figure 3A: pRF144 plasmid (SEQ ID NO:20) for expressing 6xHis-Zebra CPP-Cas9-NLS fusion in *E. coli*. Origin of replication is in cross-hatch. Refer to Example 1.

Figure 3B: pRF145 plasmid (SEQ ID NO:21) for expressing 6xHis-PolyR CPP-Cas9-NLS fusion in E. coli. Origin of replication is in cross-hatch. Refer to Example 1.

Figure 3C: pRF146 plasmid (SEQ ID NO:22) for expressing 6xHis-TP10 CPP-Cas9-NLS fusion in *E. coli*. Origin of replication is in cross-hatch. Refer to Example 1.

Figure 3D: pRF162 plasmid (SEQ ID NO:23) for expressing 6xHis-pVEC CPP-Cas9-NLS fusion in *E. coli*. Origin of replication is in cross-hatch. Refer to Example 1.

Figure 4: SDS-PAGE separation of purification fractions of 6xHis-Zebra-Cas9-NLS. Lysates, washes, elution fractions, and molecular weight standards are indicated. Refer to Example 1.

Figure 5: A structural model of a single guide polynucleotide such as a single guide RNA (sgRNA). A variable targeting (VT) domain is shown in gray. A Cas9 endonuclease recognition (CER) domain is shown in black.

Figure 6: *In vitro* transcription of RGR sgRNA (targeting Can1-1 locus) off of template derived from plasmid pRF46 (SEQ ID NO:30). *In vitro* transcription reactions incubated for 2, 4, 6 and 18 hours produced similar levels of sgRNA. Ribozyme autocatalytic cleavage products were also produced. Refer to Example 2.

Figure 7: *In vitro* cleavage assay using Zebra CPP-Cas9 complexed with sgRNA specific for Can1-1 target site. A DNA polynucleotide (982 bp) containing the Can1-1 target site was included in each reaction. Each reaction was electrophoretically resolved on a 1.2% gel. "Target only", "sgRNA only", "Zebra-Cas9 only", and "Zebra-Cas9 only (2xFT)" (FT, freeze-thaw) reactions did not cleave the target polynucleotide. "Zebra-Cas9/sgRNA", "Zebra-Cas9/sgRNA (2xFT)", and "Cas9/sgRNA" (wild type Cas9) reactions cleaved the target polynucleotide in a specific manner as indicated by the resulting cleavage products. Refer to Example 3.

Figure 8: Measuring the genome-targeting efficiency of Zebra CPP-Cas9 (not associated with sgRNA) and Zebra CPP-Cas9/gRNA complexes after contact

thereof with *Yarrowia lipolytica* cells. The final concentration of Zebra-Cas9 used alone was 5 µM, while different final concentrations (1-5 µM) of Zebra CPP-Cas9 were used in the sgRNA complexes. Mutation frequency is reported as the proportion of yeast colonies (grown on non-selective medium after contacting cells
5 with either Zebra CPP-Cas9 or Zebra CPP-Cas9/gRNA) that scored as resistant to canavanine upon transfer to canavanine-containing medium. Refer to Example 4.

Figure 9: Example of PAGE gel analysis of CPP-dsRED purification. 12.5% PAGE gel stained with Simply blue stain. Lane 1: Molecular weight standard, Lane 2: clarified cell extract tp10-dsREDexpress, Lane 3: clarified-cell extract post bead
10 treatment tp10-dsREDexpress, lane 4: final protein solution tp10-dsREDexpress, Lane 5 clarified cell extract MPG-dsREDexpress, Lane 3: clarified-cell extract post bead treatment MPG-dsREDexpress, lane 4: final protein solution MPG-dsREDexpress.

Table 1

15 Summary of Nucleic Acid and Protein SEQ ID Numbers

Description	Nucleic acid SEQ ID NO.	Protein SEQ ID NO.
Streptococcus pyogenes Cas9 open reading frame codon-optimized for expression in *Y. lipolytica.*	1 (4107 bases)	
Streptococcus pyogenes Cas9 including C-terminal linker and SV40 NLS ("Cas9-NLS"); open reading frame codon-optimized for expression in *Y. lipolytica*.	2 (4140 bases)	3 (1379 aa)
Y. lipolytica FBA1 promoter.	4 (543 bases)	
Cas9-NLS expression cassette (FBA1 promoter and Cas9-NLS open reading frame).	5 (4683 bases)	
pZUFCas9 plasmid.	6 (10706 bases)	
Cas9-NLS forward PCR primer.	7 (35 bases)	
Cas9-NLS reverse PCR primer.	8 (31 bases)	
EcoRI-Cas9-NLS-HinDIII PCR product	9 (4166 bases)	
pBAD/HisB plasmid	10 (4092 bases)	
pRF48 plasmid	11 (8237 bases)	
Zebra cell-penetrating peptide (CPP), from Epstein-Barr virus Zebra trans-activator protein		12 (54 aa)

pVEC CPP, from murine endothelial cadherin protein		13 (18 aa)
TP10 CPP, from neuropeptide galanin protein		14 (21 aa)
Poly-arginine (PolyR) CPP		15 (17 aa)
NcoI-6xHis-Zebra CPP-EcoRI	16 (194 bases)	
NcoI-6xHis-pVEC CPP-EcoRI	17 (86 bases)	
NcoI-6xHis-TP10 CPP-EcoRI	18 (95 bases)	
NcoI-6xHis-PolyR CPP-EcoRI	19 (83 bases)	
pRF144 plasmid, encoding Zebra CPP-Cas9 fusion protein	20 (8294 bases)	
pRF145 plasmid, encoding PolyR CPP-Cas9 fusion protein	21 (8183 bases)	
pRF146 plasmid, encoding TP10 CPP-Cas9 fusion protein	22 (8195 bases)	
pRF162 plasmid, encoding pVEC CPP-Cas9 fusion protein	23 (8186 bases)	
Cas9 endonuclease recognition (CER) domain of a gRNA.	24 (80 bases)	
Y. lipolytica Can1-1 target site, or alternatively, DNA encoding Can1-1 variable target domain of a gRNA.	25 (20 bases)	
Hammerhead (HH) ribozyme.	26 (43 bases)	
HDV ribozyme.	27 (68 bases)	
HH-sgRNA-HDV (RGR) pre-sgRNA expression cassette, or alternatively, “RGR” expression cassette (for targeting Can1-1 locus)	28 (211 bases)	
T7 RNA polymerase promoter	29 (20 bases)	
pRF46 plasmid	30 (2875 bases)	
T7 forward primer	31 (20 bases)	
gRNArev1 reverse primer	32 (20 bases)	
IV-up primer	33 (21 bases)	
IV-down primer	34 (20 bases)	
Can1 cleavage assay DNA sequence	35 (982 bases)	

RNA loop-forming sequence (GAAA).	36 (4 bases)	
RNA loop-forming sequence (CAAA).	37 (4 bases)	
RNA loop-forming sequence (AAAG).	38 (4 bases)	
Zebra CPP-Cas9-NLS fusion protein		39 (1434 aa)
PolyR CPP-Cas9-NLS fusion protein		40 (1397 aa)
TP10 CPP-Cas9-NLS fusion protein		41 (1401 aa)
pVEC CPP-Cas9-NLS fusion protein		42 (1398 aa)
Example of a Cas9 target site:PAM sequence.	43 (23 bases)	
PAM sequence NGG.	44 (3 bases)	
PAM sequence NNAGAA.	45 (6 bases)	
PAM sequence NNAGAAW.	46 (7 bases)	
PAM sequence NGGNG.	47 (5 bases)	
PAM sequence NNNNGATT.	48 (8 bases)	
PAM sequence NAAAAC.	49 (6 bases)	
PAM sequence NG.	50 (2 bases)	
TracrRNA mate sequence example 1.	51 (22 bases)	
TracrRNA mate sequence example 2.	52 (15 bases)	
TracrRNA mate sequence example 3.	53 (12 bases)	
TracrRNA mate sequence example 4.	54 (13 bases)	
TracrRNA example 1.	55 (60 bases)	
TracrRNA example 2.	56 (45 bases)	
TracrRNA example 3.	57 (32 bases)	
TracrRNA example 4.	58 (85 bases)	
TracrRNA example 5.	59 (77 bases)	
TracrRNA example 6.	60 (65 bases)	
gRNA example 1.	61 (131 bases)	

gRNA example 2.	62 (117 bases)	
gRNA example 3.	63 (104 bases)	
gRNA example 4.	64 (99 bases)	
gRNA example 5.	65 (81 bases)	
gRNA example 6.	66 (68 bases)	
gRNA example 7.	67 (100 bases)	
Tat-derived CPP (GRKKRRQRRR)		68 (10 aa)
Tat-derived CPP (RKKRRQRRR)		69 (9 aa)
Tat-derived CPP (RKKRRQRR)		70 (8 aa)
Penetratin CPP (RQIKIWFQNRRMKWKK)		71 (16 aa)
Polyarginine CPP (THRLPRRRRRR)		72 (11 aa)
Polyarginine CPP (GGRRARRRRRR)		73 (11 aa)
pVEC CPP (shorter version), from murine endothelial cadherin protein		74 (17 aa)
CPP comprising $(KFF)_3K$		75 (10 aa)
MAP peptide CPP		76 (18 aa)
CPP (RRQRRTSKLMKR)		77 (12 aa)
CPP (KALAWEAKLAKALAKALAKHLAKALAKALKCEA)		78 (33 aa)
Proline-rich CPP repeat VHLPPP		79 (6 aa)
Proline-rich CPP repeat VHRPPP		80 (6 aa)
MPG peptide CPP		81 (27 aa)
Pep-1 peptide CPP		82 (21 aa)
hCT CPP example 1		83 (24 aa)
hCT CPP example 2		84 (18 aa)
his tagged dsRED		85
E. coli codon optimized dsRED	86	
pBAD/HisB	87	
pRF161	88	

TAT		89
TLM		90
MPG1		91
pep1		92
CFFKDEL		93
his-TAT E.coli optimized	94	
his-TLM E.coli optimized	95	
his-MPG1 E. coli optimized	96	
his-pep1 E. coli optimized	97	
his-CFFKDEL E. coli optimized	98	
pRF224	99	
pRF214	100	
pRF213	101	
pRF217	102	
pRF216	103	
oligo 36	104	
His-Zebra PCR	105	
His-tp10 PCR	106	
His-pVEC PCR	107	
pRF144	108	
pRF162	109	
pRF146	110	
oligo 153	111	
pRF186	112	
pRF192	113	
pRF190	114	
his-CFFKDEL-Cas9		115
his-MPG1-Cas9		116
pRF48	117	
pRF243	118	
pRF238	119	
galK gene	120	
galE gene		121
galT gene		122
CER domain I	123	

CER encoding DNA PCR	124	
pRF291	125	
CER forward	126	
universal reverse	127	
universal forward T7 primer	128	
galK2-1 forward primer	129	
galK2-1 reverse primer	130	
galK2-1 sgRNA in vitro transcription template	131	
T7 promoter	132	
DNA encoding galK2-1 variable targeting domain	133	
galK2-1 target site	134	
galK2-1 sgRNA	135	
his-MPG1-dsREDexpress;		136
pVEC-dsREDexpress		137
CFFKDEL-dsREDexpress		138
TLM-dsREDexpress		139
Zebra-dsREDexpress		140
pep1-dsREDexpress		141
tp10-dsREDexpress		142
Zebra-Cas9		143
pVEC-Cas9		144

DETAILED DESCRIPTION OF THE INVENTION

The disclosures of all cited patent and non-patent literature are incorporated herein by reference in their entirety.

5 As used herein, the term "invention" or "disclosed invention" is not meant to be limiting, but applies generally to any of the inventions defined in the claims or described herein. These terms are used interchangeably herein.

The term "cell" herein refers to any type of cell such as a prokaryotic or eukaryotic cell. A eukaryotic cell has a nucleus and other membrane-enclosed 10 structures (organelles), whereas a prokaryotic cell lacks a nucleus. A cell in certain embodiments can be a mammalian cell or non-mammalian cell. Non-mammalian cells can be eukaryotic or prokaryotic. For example, a non-mammalian cell herein can refer to a microbial cell or cell of a non-mammalian multicellular organism such as a plant, insect, nematode, avian species, amphibian, reptile, or fish.

A microbial cell herein can refer to a fungal cell (e.g., yeast cell), prokaryotic cell, protist cell (e.g., algal cell), euglenoid cell, stramenopile cell, or oomycete cell, for example. A prokaryotic cell herein can refer to a bacterial cell or archaeal cell, for example. Fungal cells (e.g., yeast cells), protist cells (e.g., algal cells), euglenoid cells, stramenopile cells, and oomycete cells represent examples of eukaryotic microbial cells. A eukaryotic microbial cell has a nucleus and other membrane-enclosed structures (organelles), whereas a prokaryotic cell lacks a nucleus.

The term "yeast" herein refers to fungal species that predominantly exist in unicellular form. Yeast can alternatively be referred to as "yeast cells". A yeast herein can be characterized as either a conventional yeast or non-conventional yeast, for example.

The term "conventional yeast" ("model yeast") herein generally refers to *Saccharomyces* or *Schizosaccharomyces* yeast species. Conventional yeast in certain embodiments are yeast that favor homologous recombination (HR) DNA repair processes over repair processes mediated by non-homologous end-joining (NHEJ).

The term "non-conventional yeast" herein refers to any yeast that is not a *Saccharomyces* or *Schizosaccharomyces* yeast species. Non-conventional yeast are described in Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology: Practical Protocols (K. Wolf, K.D. Breunig, G. Barth, Eds., Springer-Verlag, Berlin, Germany, 2003) and Spencer et al. (Appl. Microbiol. Biotechnol. 58:147-156), which are incorporated herein by reference. Non-conventional yeast in certain embodiments may additionally (or alternatively) be yeast that favor NHEJ DNA repair processes over repair processes mediated by HR. Definition of a non-conventional yeast along these lines – preference of NHEJ over HR – is further disclosed by Chen et al. (*PLoS ONE* 8:e57952), which is incorporated herein by reference. Preferred non-conventional yeast herein are those of the genus *Yarrowia* (e.g., *Yarrowia lipolytica*).

The term "plant" herein refers to whole plants, plant organs, plant tissues, plant cells, seeds and progeny of the same. Plant cells include, without limitation, cells from seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen and microspores. Plant parts include differentiated and undifferentiated tissues including, but not limited to

roots, stems, shoots, leaves, pollens, seeds, tumor tissue and various forms of cells and culture (e.g., single cells, protoplasts, embryos, and callus tissue). The plant tissue may be in plant or in a plant organ, tissue or cell culture. The term "plant organ" refers to plant tissue or a group of tissues that constitute a morphologically and functionally distinct part of a plant. The term "genome" refers to the entire complement of genetic material (genes and non-coding sequences) that is present in each cell of an organism, or virus or organelle; and/or a complete set of chromosomes inherited as a (haploid) unit from one parent. "Progeny" comprises any subsequent generation of a plant.

A transgenic plant includes, for example, a plant which comprises within its genome a heterologous polynucleotide introduced by a transformation step. The heterologous polynucleotide can be stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant DNA construct. A transgenic plant can also comprise more than one heterologous polynucleotide within its genome. Each heterologous polynucleotide may confer a different trait to the transgenic plant. A heterologous polynucleotide can include a sequence that originates from a foreign species, or, if from the same species, can be substantially modified from its native form. Transgenic plant material can include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The alterations of a plant genome (chromosomal or extra-chromosomal) by conventional plant breeding methods, by a genome editing procedure described herein that does not result in an insertion of a foreign polynucleotide, or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation are not intended to be regarded as transgenic.

A fertile plant is a plant that produces viable male and female gametes and is self-fertile. Such a self-fertile plant can produce a progeny plant without the contribution from any other plant of a gamete and the genetic material contained therein. Male-sterile plants include plants that do not produce male gametes that

are viable or otherwise capable of fertilization. Female-sterile plants include plants that do not produce female gametes that are viable or otherwise capable of fertilization. It is recognized that male-sterile and female-sterile plants can be female-fertile and male-fertile, respectively. It is further recognized that a male-fertile (but female-sterile) plant can produce viable progeny when crossed with a female-fertile plant and that a female-fertile (but male-sterile) plant can produce viable progeny when crossed with a male-fertile plant.

The term "RNA-guided endonuclease" (RGEN) herein refers to a complex comprising at least one CRISPR (clustered regularly interspaced short palindromic repeats)-associated (Cas) protein and at least one RNA component. The terms "protein component of an RGEN" and "RGEN protein component" are used interchangeably herein and refer to a Cas protein, which is, or forms part of, the endonuclease component of an RGEN. A protein component in certain embodiments can be a complete endonuclease (e.g., Cas9); such a protein component can alternatively be referred to as "the endonuclease component" of an RGEN. An RGEN herein typically has specific DNA targeting activity, given its association with at least one RNA component.

Briefly, an RNA component of an RGEN contains sequence that is complementary to a DNA sequence in a target site sequence. Based on this complementarity, an RGEN can specifically recognize and cleave a particular DNA target site sequence. An RGEN herein can comprise Cas protein(s) and suitable RNA component(s) of any of the four known CRISPR systems (Horvath and Barrangou, *Science* 327:167-170) such as a type I, II, or III CRISPR system. An RGEN in preferred embodiments comprises a Cas9 endonuclease (CRISPR II system) and at least one RNA component (e.g., a crRNA and tracrRNA, or a gRNA).

The term "CRISPR" (clustered regularly interspaced short palindromic repeats) refers to certain genetic loci encoding factors of class I, II, or III DNA cleavage systems, for example, used by bacterial and archaeal cells to destroy foreign DNA (Horvath and Barrangou, *Science* 327:167-170). Components of CRISPR systems are taken advantage of herein in a heterologous manner for DNA targeting in cells.

The terms "type II CRISPR system" and "type II CRISPR-Cas system" are used interchangeably herein and refer to a DNA cleavage system utilizing a Cas9

endonuclease in complex with at least one RNA component. For example, a Cas9 can be in complex with a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA). In another example, a Cas9 can be in complex with a guide RNA. Thus, crRNA, tracrRNA, and guide RNA are non-limiting examples of RNA
5 components herein.

The term CRISPR-associated ("Cas") endonuclease herein refers to a Cas protein encoded by a Cas gene. A Cas endonuclease, when in complex with a suitable RNA component, is capable of cleaving all or part of a specific DNA target sequence in certain embodiments. For example, it is can be capable of introducing
10 a single- or double-strand break in a specific DNA target sequence; it can alternatively be characterized as being able to cleave one or both strands of a specific DNA target sequence. A Cas endonuclease can unwind the DNA duplex at the target sequence and cleaves at least one DNA strand, as mediated by recognition of the target sequence by a crRNA or guide RNA that is in complex with
15 the Cas. Such recognition and cutting of a target sequence by a Cas endonuclease typically occurs if the correct protospacer-adjacent motif (PAM) is located at or adjacent to the 3' end of the DNA target sequence. Alternatively, a Cas protein herein may lack DNA cleavage or nicking activity, but can still specifically bind to a DNA target sequence when complexed with a suitable RNA component. A
20 preferred Cas protein herein is Cas9.

Any guided endonuclease can be used in the methods disclosed herein. Such endonucleases include, but are not limited to Cas9 and Cpf1 endonucleases. Many endonucleases have been described to date that can recognize specific PAM sequences (see for example – US patent applications 62/162377 filed May 15,
25 2015 and 62/162353 filed May 15, 2015 and Zetsche B et al. 2015. Cell 163, 1013) and cleave the target DNA at a specific positions. It is understood that based on the methods and embodiments described herein utilizing a guided Cas system, one can now tailor these methods such that they can utilize any guided endonuclease system.

30 "Cas9" (formerly referred to as Cas5, Csn1, or Csx12) herein refers to a Cas endonuclease of a type II CRISPR system that forms a complex with crRNA and tracrRNA, or with a guide RNA, for specifically recognizing and cleaving all or part of a DNA target sequence. Cas9 protein comprises an RuvC nuclease domain and an

HNH (H-N-H) nuclease domain, each of which cleaves a single DNA strand at a target sequence (the concerted action of both domains leads to DNA double-strand cleavage, whereas activity of one domain leads to a nick). In general, the RuvC domain comprises subdomains I, II and III, where domain I is located near the N-terminus of Cas9 and subdomains II and III are located in the middle of the protein, flanking the HNH domain (Hsu et al, *Cell* 157:1262-1278). "Apo-Cas9" refers to Cas9 that is not complexed with an RNA component. Apo-Cas9 can bind DNA, but does so in a non-specific manner, and cannot cleave DNA (Sternberg et al., *Nature* 507:62-67).

The term "RNA component" herein refers to an RNA component of an RGEN containing a ribonucleic acid sequence that is complementary to a strand of a DNA target sequence. This complementary sequence is referred to herein as a "guide sequence" or "variable targeting domain" sequence (Figure 5). Examples of suitable RNA components herein include crRNA and guide RNA. RNA components in certain embodiments (e.g., guide RNA alone, crRNA + tracrRNA) can render an RGEN competent for specific DNA targeting.

The term "CRISPR RNA" (crRNA) herein refers to an RNA sequence that can form a complex with one or more Cas proteins (e.g., Cas9) and provides DNA binding specificity to the complex. A crRNA provides DNA binding specificity since it contains "guide sequence" ("variable targeting domain" [VT]) that is complementary to a strand of a DNA target sequence. A crRNA further comprises a "repeat sequence" ("tracr RNA mate sequence") encoded by a repeat region of the CRISPR locus from which the crRNA was derived. A repeat sequence of a crRNA can anneal to sequence at the 5'-end of a tracrRNA. crRNA in native CRISPR systems is derived from a "pre-crRNA" transcribed from a CRISPR locus. A pre-crRNA comprises spacer regions and repeat regions; spacer regions contain unique sequence complementary to a DNA target site sequence. Pre-crRNA in native systems is processed to multiple different crRNAs, each with a guide sequence along with a portion of repeat sequence. CRISPR systems utilize crRNA, for example, for DNA targeting specificity.

The term "trans-activating CRISPR RNA" (tracrRNA) herein refers to a non-coding RNA used in type II CRISPR systems, and contains, in the 5'-to-3' direction,

(i) a sequence that anneals with the repeat region of CRISPR type II crRNA and (ii) a stem loop-containing portion (Deltcheva et al., *Nature* 471:602-607).

The terms "guide RNA" (gRNA) and "single guide RNA" (sgRNA) are used interchangeably herein. A gRNA herein can refer to a chimeric sequence containing a crRNA operably linked to a tracrRNA. Alternatively, a gRNA can refer to a synthetic fusion of a crRNA and a tracrRNA, for example. A gRNA can also be characterized in terms of having a guide sequence (variable targeting domain) followed by a Cas endonuclease recognition (CER) domain. A CER domain can comprise a tracrRNA mate sequence followed by a tracrRNA sequence.

A "CRISPR DNA" (crDNA) can optionally be used instead of an RNA component. A crDNA has a DNA sequence corresponding to the sequence of a crRNA as disclosed herein. A crDNA can be used with a tracrRNA in a crDNA/tracrRNA complex, which in turn can be associated with an RGEN protein component. U.S. Appl. No. 61/953,090 discloses crDNA and the methods of its use in RGEN-mediated DNA targeting. It is contemplated that any disclosure herein regarding a crRNA can similarly apply to using a crDNA, accordingly. Thus, in embodiments herein incorporating a crDNA, an "RNA-guided endonuclease" (RGEN) could instead be referred to as a complex comprising at least one Cas protein and at least one crDNA.

As used herein, the term "guide polynucleotide", relates to a polynucleotide sequence that can form a complex with a Cas endonuclease and enables the Cas endonuclease to recognize and optionally cleave a DNA target site. The guide polynucleotide can be a single molecule or a double molecule. The guide polynucleotide sequence can be a RNA sequence, a DNA sequence, or a combination thereof (an RNA-DNA combination sequence). Optionally, the guide polynucleotide can comprise at least one nucleotide, phosphodiester bond or linkage modification such as, but not limited, to Locked Nucleic Acid (LNA), 5-methyl dC, 2,6-Diaminopurine, 2'-Fluoro A, 2'-Fluoro U, 2'-O-Methyl RNA, Phosphorothioate bond, linkage to a cholesterol molecule, linkage to a polyethylene glycol molecule, linkage to a spacer 18 (hexaethylene glycol chain) molecule, or 5' to 3' covalent linkage resulting in circularization. A guide polynucleotide that solely comprises ribonucleic acids is also referred to as a "guide RNA".

The guide polynucleotide can be a double molecule (also referred to as duplex guide polynucleotide) comprising a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that is complementary to a nucleotide sequence in a target DNA and a second nucleotide sequence domain (referred to as Cas endonuclease recognition domain or CER domain) that interacts with a Cas endonuclease polypeptide. The CER domain of the double molecule guide polynucleotide comprises two separate molecules that are hybridized along a region of complementarity. The two separate molecules can be RNA, DNA, and/or RNA-DNA- combination sequences. In some embodiments, the first molecule of the duplex guide polynucleotide comprising a VT domain linked to a CER domain ("crNucleotide") is referred to as "crDNA" (when composed of a contiguous stretch of DNA nucleotides) or "crRNA" (when composed of a contiguous stretch of RNA nucleotides), or "crDNA-RNA" (when composed of a combination of DNA and RNA nucleotides). In some embodiments the second molecule of the duplex guide polynucleotide comprising a CER domain is referred to as "tracrRNA" (when composed of a contiguous stretch of RNA nucleotides) or "tracrDNA" (when composed of a contiguous stretch of DNA nucleotides) or "tracrDNA-RNA" (when composed of a combination of DNA and RNA nucleotides).

The guide polynucleotide can also be a single molecule comprising a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that is complementary to a nucleotide sequence in a target DNA and a second nucleotide domain (referred to as Cas endonuclease recognition domain or CER domain) that interacts with a Cas endonuclease polypeptide. By "domain" it is meant a contiguous stretch of nucleotides that can be RNA, DNA, and/or RNA-DNA-combination sequence. The VT domain and/or the CER domain of a single guide polynucleotide can comprise an RNA sequence, a DNA sequence, or a, RNA-DNA-combination sequence. In some embodiments the single guide polynucleotide comprises a crNucleotide (comprising a VT domain linked to a CER domain) linked to a tracrNucleotide (comprising a CER domain), wherein the linkage is a nucleotide sequence comprising an RNA sequence, a DNA sequence, or an RNA-DNA combination sequence. The single guide polynucleotide being comprised of sequences from the crNucleotide and tracrNucleotide may be referred to as "single guide RNA" (when composed of a contiguous stretch of RNA nucleotides) or "single

guide DNA" (when composed of a contiguous stretch of DNA nucleotides) or "single guide RNA-DNA" (when composed of a combination of RNA and DNA nucleotides).

Thus, a guide polynucleotide and a type II Cas endonuclease in certain embodiments can form a complex with each other (referred to as a "guide polynucleotide/Cas endonuclease complex" or also referred to as "guide polynucleotide/Cas endonuclease system"), wherein the guide polynucleotide/Cas endonuclease complex can direct the Cas endonuclease to target a genomic target site in a cell (e.g., plant cell), optionally enabling the Cas endonuclease to introduce a single- or double-strand break into the genomic target site. A guide polynucleotide/Cas endonuclease complex can be linked to at least one CPP, wherein such complex is capable of binding to, and optionally creating a single- or double-strand break to, a target site of a cell (e.g., a plant cell).

The term "variable targeting domain" or "VT domain" is used interchangeably herein and refers to a nucleotide sequence that is complementary to one strand (nucleotide sequence) of a double strand DNA target site. The percent complementation between the first nucleotide sequence domain (VT domain) and the target sequence can be at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 63%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%. The variable target domain can be at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 nucleotides in length. In some embodiments, the variable targeting domain comprises a contiguous stretch of 12 to 30 nucleotides. The variable targeting domain can be composed of a DNA sequence, an RNA sequence, a modified DNA sequence, a modified RNA sequence (see, e.g., modifications described herein), or any combination thereof.

The term "Cas endonuclease recognition domain" or "CER domain" of a guide polynucleotide is used interchangeably herein and relates to a nucleotide sequence (such as a second nucleotide sequence domain of a guide polynucleotide), that interacts with a Cas endonuclease polypeptide. A CER domain can be composed of a DNA sequence, an RNA sequence, a modified DNA sequence, a modified RNA sequence (see, e.g., modifications described herein), or any combination thereof.

The terms "target site", "target sequence", "target DNA", "DNA target sequence", "target locus", "protospacer" and the like are used interchangeably herein. A target site sequence refers to a polynucleotide sequence on a chromosome, episome, or any other DNA molecule in the genome of a cell to which
5 an RGEN herein can recognize, bind to, and optionally nick or cleave. A target site can be (i) an endogenous/native site in the cell, (ii) heterologous to the cell and therefore not be naturally occurring in the genome, or (iii) found in a heterologous genomic location compared to where it natively occurs.

A target site sequence herein is at least 13 nucleotides in length and has a
10 strand with sufficient complementarity to a guide sequence (of a crRNA or gRNA) to be capable of hybridizing with the guide sequence and direct sequence-specific binding of a Cas protein or Cas protein complex to the target sequence (if a suitable PAM is adjacent to the target sequence in certain embodiments). A cleavage/nick site (applicable with a endonucleolytic or nicking Cas) can be within the target
15 sequence (e.g., using a Cas9) or a cleavage/nick site could be outside of the target sequence (e.g., using a Cas9 fused to a heterologous endonuclease domain such as one derived from a FokI enzyme). It is also possible for a target site sequence to be bound by an RGEN lacking cleavage or nicking activity.

An "artificial target site" or "artificial target sequence" herein refers to a target
20 sequence that has been introduced into the genome of a cell. An artificial target sequence in some embodiments can be identical in sequence to a native target sequence in the genome of the cell, but be located at a different position (a heterologous position) in the genome, or it can different from the native target sequence if located at the same position in the genome of the cell.

25 An "episome" herein refers to a DNA molecule that can exist in a cell autonomously (can replicate and pass on to daughter cells) apart from the chromosomes of the cell. Episomal DNA can be either native or heterologous to a cell. Examples of native episomes herein include mitochondrial DNA (mtDNA) and chloroplast DNA. Examples of heterologous episomes herein include plasmids and
30 yeast artificial chromosomes (YACs).

A "protospacer adjacent motif" (PAM) herein refers to a short sequence that is recognized by an RGEN herein. The sequence and length of a PAM herein can

differ depending on the Cas protein or Cas protein complex used, but are typically 2, 3, 4, 5, 6, 7, or 8 nucleotides long, for example.

The terms "5'-cap" and "7-methylguanylate (m^7G) cap" are used interchangeably herein. A 7-methylguanylate residue is located on the 5' terminus of RNA transcribed by RNA polymerase II (Pol II) in eukaryotes. A capped RNA herein has a 5'-cap, whereas an uncapped RNA does not have such a cap.

The terminology "uncapped", "not having a 5'-cap", and the like are used interchangeably herein to refer to RNA lacking a 5'-cap and optionally having, for example, a 5'-hydroxyl group instead of a 5'-cap. Uncapped RNA can better accumulate in the nucleus following transcription, since 5'-capped RNA is subject to nuclear export.

The terms "ribozyme", "ribonucleic acid enzyme" and "self-cleaving ribozyme" are used interchangeably herein. A ribozyme refers to one or more RNA sequences that form secondary, tertiary, and/or quaternary structure(s) that can cleave RNA at a specific site, particularly at a *cis*-site relative to the ribozyme sequence (i.e., auto-catalytic, or self-cleaving). The general nature of ribozyme nucleolytic activity has been described (e.g., Lilley, *Biochem. Soc. Trans.* 39:641-646). A "hammerhead ribozyme" (HHR) herein may comprise a small catalytic RNA motif made up of three base-paired stems and a core of highly conserved, non-complementary nucleotides that are involved in catalysis. Pley et al. (*Nature* 372:68-74) and Hammann et al. (*RNA* 18:871-885), which are incorporated herein by reference, disclose hammerhead ribozyme structure and activity. A hammerhead ribozyme herein may comprise a "minimal hammerhead" sequence as disclosed by Scott et al. (*Cell* 81:991-1002, incorporated herein by reference), for example.

The terms "targeting", "gene targeting", "DNA targeting", "editing", "gene editing" and "DNA editing" are used interchangeably herein. DNA targeting herein may be the specific introduction of an indel, knock-out, or knock-in at a particular DNA sequence, such as in a chromosome or episome of a cell. In general, DNA targeting can be performed herein by cleaving one or both strands at a specific DNA sequence in a cell with a Cas protein associated with a suitable RNA component. Such DNA cleavage, if a double-strand break (DSB), can prompt NHEJ processes which can lead to indel formation at the target site. Also, regardless of whether the cleavage is a single-strand break (SSB) or DSB, HR processes can be prompted if a

suitable donor DNA polynucleotide is provided at the DNA nick or cleavage site. Such an HR process can be used to introduce a knock-out or knock-in at the target site, depending on the sequence of the donor DNA polynucleotide. Alternatively, DNA targeting herein can refer to specific association of a Cas/RNA component complex herein to a target DNA sequence, where the Cas protein does or does not cut a DNA strand (depending on the status of the Cas protein's endonucleolytic domains).

The term "indel" herein refers to an insertion or deletion of a nucleotide base or bases in a target DNA sequence in a chromosome or episome. Such an insertion or deletion may be of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more bases, for example. An indel in certain embodiments can be even larger, at least about 20, 30, 40, 50, 60, 70, 80, 90, or 100 bases. If an indel is introduced within an open reading frame (ORF) of a gene, oftentimes the indel disrupts wild type expression of protein encoded by the ORF by creating a frameshift mutation.

The terms "knock-out", "gene knock-out" and "genetic knock-out" are used interchangeably herein. A knock-out represents a DNA sequence of a cell herein that has been rendered partially or completely inoperative by targeting with a Cas protein; such a DNA sequence prior to knock-out could have encoded an amino acid sequence, or could have had a regulatory function (e.g., promoter), for example. A knock-out may be produced by an indel (by NHEJ, prompted by Cas-mediated cleavage), or by specific removal of sequence (by HR, prompted by Cas-mediated cleavage or nicking, when a suitable donor DNA polynucleotide is also used), that reduces or completely destroys the function of sequence at, adjoining, or near the targeting site. A knocked out DNA polynucleotide sequence herein can alternatively be characterized as being partially or totally disrupted or downregulated, for example.

The terms "knock-in", "gene knock-in" and "genetic knock-in" are used interchangeably herein. A knock-in represents the replacement or insertion of a DNA sequence at a specific DNA sequence in a cell by targeting with a Cas protein (by HR, prompted by Cas-mediated cleavage or nicking, when a suitable donor DNA polynucleotide is also used). Examples of knock-ins are a specific insertion of a heterologous amino acid coding sequence in a coding region of a gene, or a specific insertion of a transcriptional regulatory element in a genetic locus.

The terms "donor polynucleotide", "donor DNA", "targeting polynucleotide" and "targeting DNA" are used interchangeably herein. A donor polynucleotide refers to a DNA sequence that comprises at least one sequence that is homologous to a sequence at or near a DNA target site (e.g., a sequence specifically targeted by a Cas protein herein). A suitable donor polynucleotide is able to undergo HR with a DNA target site if the target site contains a SSB or DSB (such as can be introduced using certain Cas proteins herein associated with an appropriate RNA component). A "homologous sequence" within a donor polynucleotide herein can, for example, comprise or consist of a sequence of at least about 25 nucleotides, for example, having 100% identity with a sequence at or near a target site, or at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity with a sequence at or near a target site.

In certain embodiments, a donor DNA polynucleotide can have two homologous sequences separated by a sequence (or base pair) that is heterologous to sequence at a target site. These two homologous sequences of such a donor polynucleotide can be referred to as "homology arms", which flank the heterologous sequence. HR between a target site and a donor polynucleotide with two homology arms typically results in the replacement of a sequence at the target site with the heterologous sequence of the donor polynucleotide (target site sequence located between DNA sequences homologous to the homology arms of the donor polynucleotide is replaced by the heterologous sequence of the donor polynucleotide). In a donor polynucleotide with two homology arms, the arms can be separated by 1 or more nucleotides (i.e., the heterologous sequence in the donor polynucleotide can be at least 1 nucleotide in length). Various HR procedures that can be performed in a cell herein are disclosed, for example, in DNA Recombination: Methods and Protocols: 1st Edition (H. Tsubouchi, Ed., Springer-Verlag, New York, 2011), which is incorporated herein by reference.

The terms "cell-penetrating peptide" (CPP) and "protein transduction domain" (PTD) are used interchangeably herein. A CPP refers to a peptide , typically of about 5-60 amino acid residues in length, that can facilitate cellular uptake of molecular cargo, particularly one or more RGEN protein components herein (e.g., Cas9 protein). Such protein cargo can be associated with one or more CPPs through covalent or non-covalent linkage. A CPP can also be characterized in

certain embodiments as being able to facilitate the movement or traversal of molecular cargo across/through one or more of a lipid bilayer, micelle, cell membrane, organelle membrane, vesicle membrane, or cell wall. A CPP herein can be cationic, amphipathic, or hydrophobic in certain embodiments. Examples of CPPs useful herein, and further description of CPPs in general, are disclosed in Schmidt et al. (*FEBS Lett.* 584:1806-1813), Holm et al. (*Nature Protocols* 1:1001-1005), Yandek et al. (*Biophys. J.* 92:2434-2444), Morris et al. (*Nat. Biotechnol.* 19:1173-1176), and U.S. Patent Appl. Publ. No. 2014/0068797, which are all incorporated herein by reference.

A "cationic", or "polycationic", CPP herein refers to a CPP having a high relative abundance (at least 60%) of positively charged amino acids such as lysine (K), arginine (R), and/or histidine (H).

An "amphipathic", or "amphiphilic", CPP herein refers to a CPP with an amino acid sequence containing an alternating pattern of polar/charged residues and non-polar, hydrophobic residues. An amphipathic CPP can alternatively be characterized as possessing both hydrophilic and lipophilic properties.

A "hydrophobic", or "lipophilic", CPP herein contains mostly, or only, non-polar residues with low net charge and/or hydrophobic amino acid groups.

The terms "covalently linked", "covalently attached", "covalently associated", "covalent linkage", "covalent interaction" and the like are used interchangeably herein. A covalent linkage herein can be via a peptide bond(s) or chemical crosslink(s), for example. A covalent linkage can be direct, for example, where there is a covalent link directly between (directly linking) an RGEN protein component and a CPP (e.g., there is a chemical bond [sharing of electrons] between an atom of an RGEN protein component and an atom of a CPP). A covalent linkage can alternatively be indirect, for example, where an RGEN protein component and a CPP are linked to each other through at least one intermediary factor. Such an intermediary factor, or group of intermediary factors that are themselves covalently linked together, is covalently linked to the RGEN protein component and CPP. Thus, an intermediary factor or group thereof can be characterized as being a bridge between an RGEN protein component and a CPP.

The terms "fusion protein", "protein fusion", "chimeric protein" and the like are used interchangeably herein. A fusion protein herein contains at least two different

(heterologous) amino acid sequences linked together within a single polypeptide. Fusion proteins are typically produced by genetic engineering processes in which DNA sequences encoding different amino acid sequences are joined together to encode a single protein containing the different amino acid sequences. Examples of fusion proteins herein include RGEN protein-CPP fusions (RGEN protein amino acid sequence fused to one or more CPP amino acid sequences).

The terms "non-covalently linked", "non-covalently attached", "non-covalently associated", "non-covalent linkage", "non-covalent interaction" and the like are used interchangeably herein. A non-covalent linkage herein refers to an interaction between atoms in which electrons are not shared. This type of interaction is weaker than a covalent linkage. Hydrophobic interactions represent an example of a non-covalent linkage that may occur between an RGEN protein component and one or more CPPs. Other examples of non-covalent linkages that may apply herein include electrostatic forces (e.g., ionic, hydrogen bonding) and Van der Waals forces (London Dispersion forces).

An "RGEN protein-CPP complex" as used herein refers to a complex between a protein component of an RGEN and at least one CPP, where the RGEN and CPP interact via covalent or non-covalent linkage. Both RGEN and CPP components in this complex typically retain all of, or some of (e.g., at least 50%), their respective activity/function as disclosed herein. For example, in embodiment in which the RGEN protein component is Cas9, the Cas9 in a Cas9-CPP complex is capable of associating with a suitable RNA component (e.g., gRNA) and targeting the Cas9-CPP complex to a DNA target site in a cell.

The terms "traverse", "travel through", "cross through", "goes across" and the like are used interchangeably herein.

The terms "cell membrane", "plasma membrane", and "cytoplasmic membrane" are used interchangeably herein and refer to a biological membrane that separates the interior of a cell from its exterior. A cell membrane typically comprises a phospholipid bilayer with proteins embedded therein. Among several other functions, a cell membrane can serve as an attachment surface for extracellular structures such as cell wall or glycocalyx structures. Detailed information regarding cell membrane lipid bilayers is provided in Molecular Biology

of the Cell. 4th Edition (B. Alberts et al., Eds., Garland Science, New York, 2002), which is incorporated herein by reference.

The term "cell wall" herein refers to a tough, flexible (but sometimes fairly rigid) layer that surrounds some types of non-mammalian cells (e.g., bacteria, 5 plants, algae, fungi such as yeast). It is located outside the cell membrane and provides structural support and protection to cells. A major function of a cell wall in certain embodiments is to help maintain cell osmotic pressure. Fungal cell (e.g., yeast cell) walls generally comprise chitin, and algal cells walls generally comprise glycoproteins and polysaccharides. Plant cell walls generally comprise mostly 10 polysaccharides with lesser amounts of other components (e.g., phenolic esters, structural proteins). "Primary cell wall" and/or "secondary cell wall" may be used to characterize a plant cell wall, where the secondary wall is located inside the primary wall. Lignin is a major component of secondary walls. Bacterial cell walls generally comprise peptidoglycan as the main constituent. In certain aspects, such as in 15 bacteria, a cell wall can further comprise at its outer layer a glycocalyx, which is generally a coat of polysaccharides.

The term "leucine zipper domain" herein refers to a dimerization domain characterized by the presence of a leucine residue every seventh residue in a stretch of approximately 35 residues. Leucine zipper domains form dimers held 20 together by an alpha-helical coiled coil. A coiled coil has 3.5 residues per turn, which means that every seventh residue occupies an equivalent position with respect to the helix axis. The regular array of leucines inside the coiled coil stabilizes the structure by hydrophobic and Van der Waals interactions.

The terms "percent by volume", "volume percent", "vol %" and "v/v %" are 25 used interchangeably herein. The percent by volume of a solute in a solution can be determined using the formula: [(volume of solute)/(volume of solution)] x 100%.

The terms "percent by weight", "weight percentage (wt %)" and "weight-weight percentage (% w/w)" are used interchangeably herein. Percent by weight refers to the percentage of a material on a mass basis as it is comprised in a 30 composition, mixture, or solution.

The terms "polynucleotide", "polynucleotide sequence", and "nucleic acid sequence" are used interchangeably herein. These terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of DNA or RNA that is

single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof. Nucleotides (ribonucleotides or deoxyribonucleotides) can be referred to by a single letter
5 designation as follows: "A" for adenylate or deoxyadenylate (for RNA or DNA, respectively), "C" for cytidylate or deoxycytidylate (for RNA or DNA, respectively), "G" for guanylate or deoxyguanylate (for RNA or DNA, respectively), "U" for uridylate (for RNA), "T" for deoxythymidylate (for DNA), "R" for purines (A or G), "Y" for pyrimidines (C or T), "K" for G or T, "H" for A or C or T, "I" for inosine, "W" for A or T,
10 and "N" for any nucleotide (e.g., N can be A, C, T, or G, if referring to a DNA sequence; N can be A, C, U, or G, if referring to an RNA sequence). Any RNA sequence (e.g., crRNA, tracrRNA, gRNA) disclosed herein may be encoded by a suitable DNA sequence.

The term "isolated" as used herein refers to a polynucleotide or polypeptide
15 molecule that has been completely or partially purified from its native source. In some instances, the isolated polynucleotide or polypeptide molecule is part of a greater composition, buffer system or reagent mix. For example, the isolated polynucleotide or polypeptide molecule can be comprised within a cell or organism in a heterologous manner. Compositions herein comprising a protein component of
20 an RGEN and a cell-penetrating peptide can be considered isolated compositions. These compositions contain heterologous components and do not occur in nature.

The term "gene" as used herein refers to a DNA polynucleotide sequence that expresses an RNA (RNA is transcribed from the DNA polynucleotide sequence) from a coding region, which RNA can be a messenger RNA (encoding a protein) or
25 a non-protein-coding RNA (e.g., a crRNA, tracrRNA, or gRNA herein). A gene may refer to the coding region alone, or may include regulatory sequences upstream and/or downstream to the coding region (e.g., promoters, 5'-untranslated regions, 3'-transcription terminator regions). A coding region encoding a protein can alternatively be referred to herein as an "open reading frame" (ORF). A gene that is
30 "native" or "endogenous" refers to a gene as found in nature with its own regulatory sequences; such a gene is located in its natural location in the genome of a host cell. A "chimeric" gene refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature (i.e., the

regulatory and coding regions are heterologous with each other). Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. A "foreign" or "heterologous" gene refers to a gene that is introduced into the host organism by gene transfer. Foreign/heterologous genes can comprise native genes inserted into a non-native organism, native genes introduced into a new location within the native host, or chimeric genes. The polynucleotide sequences in certain embodiments disclosed herein are heterologous. A "transgene" is a gene that has been introduced into the genome by a gene delivery procedure (e.g., transformation). A "codon-optimized" open reading frame has its frequency of codon usage designed to mimic the frequency of preferred codon usage of the host cell.

A "mutated gene" is a gene that has been altered through human intervention. Such a "mutated gene" has a sequence that differs from the sequence of the corresponding non-mutated gene by at least one nucleotide addition, deletion, or substitution. In certain embodiments of the disclosure, the mutated gene comprises an alteration that is made by using a guide polynucleotide/Cas endonuclease system as disclosed herein. A mutated plant is a plant comprising at least one mutated gene.

A "non-native" amino acid sequence or polynucleotide sequence comprised in a cell or organism herein does not occur in a native (natural) counterpart of such cell or organism.

"Regulatory sequences" as used herein refer to nucleotide sequences located upstream of a gene's transcription start site (e.g., promoter), 5' untranslated regions, and 3' non-coding regions, and which may influence the transcription, processing or stability, or translation of an RNA transcribed from the gene. Regulatory sequences herein may include promoters, enhancers, silencers, 5' untranslated leader sequences, introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites, stem-loop structures, and other elements involved in regulation of gene expression. One or more regulatory elements herein may be heterologous to a coding region herein.

A "promoter" as used herein refers to a DNA sequence capable of controlling the transcription of RNA from a gene. In general, a promoter sequence is upstream of the transcription start site of a gene. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different
5 promoters found in nature, or even comprise synthetic DNA segments. Promoters that cause a gene to be expressed in a cell at most times under all circumstances are commonly referred to as "constitutive promoters". One or more promoters herein may be heterologous to a coding region herein.

A "strong promoter" as used herein refers to a promoter that can direct a
10 relatively large number of productive initiations per unit time, and/or is a promoter driving a higher level of gene transcription than the average transcription level of the genes in a cell.

A plant promoter is a promoter capable of initiating transcription in a plant cell; for a review of plant promoters, see Potenza *et al.*, (2004) *In Vitro Cell Dev Biol*
15 40:1-22. Constitutive promoters include, for example, the core promoter of the Rsyn7 promoter and other constitutive promoters disclosed in WO99/43838 and U.S. Patent No. 6072050; the core CaMV 35S promoter (Odell *et al.*, (1985) *Nature* 313:810-2); rice actin (McElroy *et al.*, (1990) *Plant Cell* 2:163-71); ubiquitin (Christensen *et al.*, (1989) *Plant Mol Biol* 12:619-32; Christensen *et al.*, (1992) *Plant*
20 *Mol Biol* 18:675-89); pEMU (Last *et al.*, (1991) *Theor Appl Genet* 81:581-8); MAS (Velten *et al.*, (1984) *EMBO J* 3:2723-30); ALS promoter (U.S. Patent No. 5659026), and the like. Other constitutive promoters are described in, for example, U.S. Patent Nos. 5608149; 5608144; 5604121; 5569597; 5466785; 5399680; 5268463; 5608142 and 6177611. In some examples, an inducible promoter may be used.
25 Pathogen-inducible promoters induced following infection by a pathogen include, but are not limited to those regulating expression of PR proteins, SAR proteins, beta-1,3-glucanase, chitinase, *etc.*

Chemical-regulated promoters can be used to modulate the expression of a gene in a plant through the application of an exogenous chemical regulator. The
30 promoter may be a chemical-inducible promoter, where application of the chemical induces gene expression, or a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters include, but are not limited to, the maize In2-2 promoter, activated by benzene sulfonamide

herbicide safeners (De Veylder *et al.*, (1997) *Plant Cell Physiol* 38:568-77), the maize GST promoter (GST-II-27, WO93/01294), activated by hydrophobic electrophilic compounds used as pre-emergent herbicides, and the tobacco PR-1a promoter (Ono *et al.*, (2004) *Biosci Biotechnol Biochem* 68:803-7) activated by
5 salicylic acid. Other chemical-regulated promoters include steroid-responsive promoters (see, for example, the glucocorticoid-inducible promoter (Schena *et al.*, (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-5; McNellis *et al.*, (1998) *Plant J* 14:247-257); tetracycline-inducible and tetracycline-repressible promoters (Gatz *et al.*, (1991) *Mol Gen Genet* 227:229-37; U.S. Patent Nos. 5814618 and 5789156).

10 Tissue-preferred promoters can be utilized to target enhanced expression within a particular plant tissue. Tissue-preferred promoters include, for example, Kawamata *et al.*, (1997) *Plant Cell Physiol* 38:792-803; Hansen *et al.*, (1997) *Mol Gen Genet* 254:337-43; Russell *et al.*, (1997) *Transgenic Res* 6:157-68; Rinehart *et al.*, (1996) *Plant Physiol* 112:1331-41; Van Camp *et al.*, (1996) *Plant Physiol*
15 112:525-35; Canevascini *et al.*, (1996) *Plant Physiol* 112:513-524; Lam, (1994) *Results Probl Cell Differ* 20:181-96; and Guevara-Garcia *et al.*, (1993) *Plant J* 4:495-505. Leaf-preferred promoters include, for example, Yamamoto *et al.*, (1997) *Plant J* 12:255-65; Kwon *et al.*, (1994) *Plant Physiol* 105:357-67; Yamamoto *et al.*, (1994) *Plant Cell Physiol* 35:773-8; Gotor *et al.*, (1993) *Plant J* 3:509-18; Orozco *et*
20 *al.*, (1993) *Plant Mol Biol* 23:1129-38; Matsuoka *et al.*, (1993) *Proc. Natl. Acad. Sci. USA* 90:9586-90; Simpson *et al.*, (1958) *EMBO J* 4:2723-9; Timko *et al.*, (1988) *Nature* 318:57-8. Root-preferred promoters include, for example, Hire *et al.*, (1992) *Plant Mol Biol* 20:207-18 (soybean root-specific glutamine synthase gene); Miao *et al.*, (1991) *Plant Cell* 3:11-22 (cytosolic glutamine synthase (GS)); Keller and
25 Baumgartner, (1991) *Plant Cell* 3:1051-61 (root-specific control element in the GRP 1.8 gene of French bean); Sanger *et al.*, (1990) *Plant Mol Biol* 14:433-43 (root-specific promoter of *A. tumefaciens* mannopine synthase (MAS)); Bogusz *et al.*, (1990) *Plant Cell* 2:633-41 (root-specific promoters isolated from *Parasponia andersonii* and *Trema tomentosa*); Leach and Aoyagi, (1991) *Plant Sci* 79:69-76 (*A.*
30 *rhizogenes* rolC and rolD root-inducing genes); Teeri *et al.*, (1989) *EMBO J* 8:343-50 (*Agrobacterium* wound-induced TR1' and TR2' genes); VfENOD-GRP3 gene promoter (Kuster *et al.*, (1995) *Plant Mol Biol* 29:759-72); and rolB promoter (Capana *et al.*, (1994) *Plant Mol Biol* 25:681-91; phaseolin gene (Murai *et al.*, (1983)

Science 23:476-82; Sengopta-Gopalen *et al.*, (1988) *Proc. Natl. Acad. Sci. USA* 82:3320-4). See also, U.S. Patent Nos. 5837876; 5750386; 5633363; 5459252; 5401836; 5110732 and 5023179.

Seed-preferred promoters include both seed-specific promoters active during seed development, as well as seed-germinating promoters active during seed germination. See, Thompson *et al.*, (1989) *BioEssays* 10:108. Seed-preferred promoters include, but are not limited to, Cim1 (cytokinin-induced message); cZ19B1 (maize 19 kDa zein); and milps (myo-inositol-1-phosphate synthase); (WO00/11177; and U.S. Patent 6225529). For dicots, seed-preferred promoters include, but are not limited to, bean beta-phaseolin, napin, beta-conglycinin, soybean lectin, cruciferin, and the like. For monocots, seed-preferred promoters include, but are not limited to, maize 15 kDa zein, 22 kDa zein, 27 kDa gamma zein, waxy, shrunken 1, shrunken 2, globulin 1, oleosin, and nuc1. See also, WO00/12733, where seed-preferred promoters from *END1* and *END2* genes are disclosed.

The terms "3' non-coding sequence", "transcription terminator" and "terminator" as used herein refer to DNA sequences located downstream of a coding sequence. This includes polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression.

The term "cassette" as used herein refers to a promoter operably linked to a DNA sequence encoding a protein-coding RNA or non-protein-coding RNA. A cassette may optionally be operably linked to a 3' non-coding sequence.

The terms "upstream" and "downstream" as used herein with respect to polynucleotides refer to "5' of" and "3' of", respectively.

The term "expression" as used herein refers to (i) transcription of RNA (e.g., mRNA or a non-protein coding RNA such as crRNA, tracrRNA, or gRNA) from a coding region, or (ii) translation of a polypeptide from mRNA.

When used to describe the expression of a gene or polynucleotide sequence, the terms "down-regulation", "disruption", "inhibition", "inactivation", and "silencing" are used interchangeably herein to refer to instances when the transcription of the polynucleotide sequence is reduced or eliminated. This results in the reduction or elimination of RNA transcripts from the polynucleotide sequence, which results in a

reduction or elimination of protein expression derived from the polynucleotide sequence (if the gene comprised an ORF). Alternatively, down-regulation can refer to instances where protein translation from transcripts produced by the polynucleotide sequence is reduced or eliminated. Alternatively still, down-regulation can refer to instances where a protein expressed by the polynucleotide sequence has reduced activity. The reduction in any of the above processes (transcription, translation, protein activity) in a cell can be by about 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% relative to the transcription, translation, or protein activity of a suitable control cell. Down-regulation can be the result of a targeting event as disclosed herein (e.g., indel, knock-out), for example.

The terms "control cell" and "suitable control cell" are used interchangeably herein and may be referenced with respect to a cell in which a particular modification (e.g., over-expression of a polynucleotide, down-regulation of a polynucleotide) has been made (i.e., an "experimental cell"). A control cell may be any cell that does not have or does not express the particular modification of the experimental cell. Thus, a control cell may be an untransformed wild type cell or may be genetically transformed but does not express the genetic transformation. For example, a control cell may be a direct parent of the experimental cell, which direct parent cell does not have the particular modification that is in the experimental cell. Alternatively, a control cell may be a parent of the experimental cell that is removed by one or more generations. Alternatively still, a control cell may be a sibling of the experimental cell, which sibling does not comprise the particular modification that is present in the experimental cell.

The term "increased" as used herein may refer to a quantity or activity that is at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 50%, 100%, or 200% more than the quantity or activity for which the increased quantity or activity is being compared. The terms "increased", "elevated", "enhanced", "greater than", and "improved" are used interchangeably herein. The term "increased" can be used to characterize the expression of a polynucleotide encoding a protein, for example, where "increased expression" can also mean "over-expression".

The term "operably linked" as used herein refers to the association of two or more nucleic acid sequences such that that the function of one is affected by the

other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence. That is, the coding sequence is under the transcriptional control of the promoter. Coding sequences can be operably linked to regulatory sequences, for example. Also, for example, a crRNA can be operably linked (fused to) a tracrRNA herein such that the tracrRNA mate sequence of the crRNA anneals with 5' sequence of the tracrRNA. Such operable linkage may comprise a suitable loop-forming sequence such as GAAA (SEQ ID NO:36), CAAA (SEQ ID NO:37), or AAAG (SEQ ID NO:38). Also, for example, an RGEN can be operably linked (fused to) one or more CPPs.

The term "recombinant" as used herein refers to an artificial combination of two otherwise separated segments of sequence, e.g., by chemical synthesis or by the manipulation of isolated segments of nucleic acids by genetic engineering techniques.

Methods for preparing recombinant constructs/vectors herein (e.g., a DNA polynucleotide encoding an RNA component cassette herein, or a DNA polynucleotide encoding a Cas protein or Cas-CPP fusion protein herein) can follow standard recombinant DNA and molecular cloning techniques as described by J. Sambrook and D. Russell (Molecular Cloning: A Laboratory Manual, 3rd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001); T.J. Silhavy et al. (Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1984); and F.M. Ausubel et al. (Short Protocols in Molecular Biology, 5th Ed. Current Protocols, John Wiley and Sons, Inc., NY, 2002), for example.

The term "transformation" as used herein refers to the transfer of a nucleic acid molecule into a host organism or host cell by any method. A nucleic acid molecule that has been transformed into an organism/cell may be one that replicates autonomously in the organism/cell, or that integrates into the genome of the organism/cell, or that exists transiently in the cell without replicating or integrating. Non-limiting examples of nucleic acid molecules suitable for transformation are disclosed herein, such as plasmids and linear DNA molecules.

A "transgenic plant" herein includes, for example, a plant which comprises within its genome a heterologous polynucleotide introduced by a transformation step. The heterologous polynucleotide can be stably integrated within the genome

such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant DNA construct. A transgenic plant can also comprise more than one heterologous polynucleotide within its genome. Each heterologous polynucleotide may confer a different trait to the transgenic plant. Transgenic plant material can include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The alterations of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods, by genome editing procedures that does not result in an insertion of a foreign polynucleotide, or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation are not intended to be regarded as transgenic.

A "phenotypic marker" is a screenable or selectable marker that includes visual markers and selectable markers whether it is a positive or negative selectable marker. Any phenotypic marker can be used. Specifically, a selectable or screenable marker comprises a DNA segment that allows one to identify, or select for or against a molecule or a cell that contains it, often under particular conditions. These markers can encode an activity, such as, but not limited to, production of RNA, peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds or compositions and the like.

Examples of selectable markers include, but are not limited to, DNA segments that comprise restriction enzyme sites; DNA segments that encode products which provide resistance against otherwise toxic compounds including antibiotics, such as, spectinomycin, ampicillin, kanamycin, tetracycline, Basta, neomycin phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT); DNA segments that encode products which are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); DNA segments that encode products which can be readily identified (e.g., phenotypic markers such as beta-galactosidase, GUS; fluorescent proteins such as green fluorescent protein (GFP), cyan (CFP), yellow (YFP), red (RFP), and cell surface proteins); the generation of

new primer sites for PCR (e.g., the juxtaposition of two DNA sequence not previously juxtaposed), the inclusion of DNA sequences not acted upon or acted upon by a restriction endonuclease or other DNA modifying enzyme, chemical, etc.; and, the inclusion of a DNA sequences required for a specific modification (e.g., methylation) that allows its identification.

Additional selectable markers include genes that confer resistance to herbicidal compounds, such as glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D). See for example, Yarranton, (1992) *Curr Opin Biotech* 3:506-11; Christopherson *et al.*, (1992) *Proc. Natl. Acad. Sci. USA* 89:6314-8; Yao *et al.*, (1992) *Cell* 71:63-72; Reznikoff, (1992) *Mol Microbiol* 6:2419-22; Hu *et al.*, (1987) *Cell* 48:555-66; Brown *et al.*, (1987) *Cell* 49:603-12; Figge *et al.*, (1988) *Cell* 52:713-22; Deuschle *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:5400-4; Fuerst *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:2549-53; Deuschle *et al.*, (1990) *Science* 248:480-3; Gossen, (1993) Ph.D. Thesis, University of Heidelberg; Reines *et al.*, (1993) *Proc. Natl. Acad. Sci. USA* 90:1917-21; Labow *et al.*, (1990) *Mol Cell Biol* 10:3343-56; Zambretti *et al.*, (1992) *Proc. Natl. Acad. Sci. USA* 89:3952-6; Baim *et al.*, (1991) *Proc. Natl. Acad. Sci. USA* 88:5072-6; Wyborski *et al.*, (1991) *Nucleic Acids Res* 19:4647-53; Hillen and Wissman, (1989) *Topics Mol Struc Biol* 10:143-62; Degenkolb *et al.*, (1991) *Antimicrob Agents Chemother* 35:1591-5; Kleinschnidt *et al.*, (1988) *Biochemistry* 27:1094-104; Bonin, (1993) Ph.D. Thesis, University of Heidelberg; Gossen *et al.*, (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-51; Oliva *et al.*, (1992) *Antimicrob Agents Chemother* 36:913-9; Hlavka *et al.*, (1985) *Handbook of Experimental Pharmacology*, Vol. 78 (Springer-Verlag, Berlin); Gill *et al.*, (1988) *Nature* 334:721-4.

The terms “sequence identity” or “identity” as used herein with respect to polynucleotide or polypeptide sequences refer to the nucleic acid residues or amino acid residues in two sequences that are the same when aligned for maximum correspondence over a specified comparison window. Thus, “percentage of sequence identity” or “percent identity” refers to the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two

sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the results by 100 to yield the percentage of sequence identity. It would be understood that, when calculating sequence identity between a DNA sequence and an RNA sequence, T residues of the DNA sequence align with, and can be considered "identical" with, U residues of the RNA sequence. For purposes of determining percent complementarity of first and second polynucleotides, one can obtain this by determining (i) the percent identity between the first polynucleotide and the complement sequence of the second polynucleotide (or vice versa), for example, and/or (ii) the percentage of bases between the first and second polynucleotides that would create canonical Watson and Crick base pairs.

The Basic Local Alignment Search Tool (BLAST) algorithm, which is available online at the National Center for Biotechnology Information (NCBI) website, may be used, for example, to measure percent identity between or among two or more of the polynucleotide sequences (BLASTN algorithm) or polypeptide sequences (BLASTP algorithm) disclosed herein. Alternatively, percent identity between sequences may be performed using a Clustal algorithm (e.g., ClustalW or ClustalV). For multiple alignments using a Clustal method of alignment, the default values may correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using a Clustal method may be KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids, these parameters may be KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. Alternatively still, percent identity between sequences may be performed using an EMBOSS algorithm (e.g., needle) with parameters such as GAP OPEN=10, GAP EXTEND=0.5, END GAP PENALTY=false, END GAP OPEN=10, END GAP EXTEND=0.5 using a BLOSUM matrix (e.g., BLOSUM62).

Herein, a first sequence that is "complementary" to a second sequence can alternatively be referred to as being in the "antisense" orientation with the second sequence.

Various polypeptide amino acid sequences and polynucleotide sequences are disclosed herein as features of certain embodiments of the disclosed invention. Variants of these sequences that are at least about 70-85%, 85-90%, or 90%-95% identical to the sequences disclosed herein can be used. Alternatively, a variant amino acid sequence or polynucleotide sequence can have at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity with a sequence disclosed herein. The variant amino acid sequence or polynucleotide sequence has the same function/activity of the disclosed sequence, or at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% of the function/activity of the disclosed sequence.

All the amino acid residues disclosed herein at each amino acid position of Cas9 proteins herein are examples. Given that certain amino acids share similar structural and/or charge features with each other (i.e., conserved), the amino acid at each position in a Cas9 can be as provided in the disclosed sequences or substituted with a conserved amino acid residue ("conservative amino acid substitution") as follows:

1. The following small aliphatic, nonpolar or slightly polar residues can substitute for each other: Ala (A), Ser (S), Thr (T), Pro (P), Gly (G);
2. The following polar, negatively charged residues and their amides can substitute for each other: Asp (D), Asn (N), Glu (E), Gln (Q);
3. The following polar, positively charged residues can substitute for each other: His (H), Arg (R), Lys (K);
4. The following aliphatic, nonpolar residues can substitute for each other: Ala (A), Leu (L), Ile (I), Val (V), Cys (C), Met (M); and
5. The following large aromatic residues can substitute for each other: Phe (F), Tyr (Y), Trp (W).

Advances have been made in expressing protein and RNA components in cells for performing RGEN-mediated DNA targeting therein (e.g., U.S. Provisional Appl. Nos. 61/868,706 and 62/036,652). Such strategies typically have entailed recombinant DNA expression in the target cells. Additional means of providing

protein and RNA components in a cell to mediate RGEN-mediated DNA targeting are of interest.

Embodiments of the disclosed invention concern a composition comprising at least one protein component of an RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP), wherein the RGEN protein component and CPP are covalently or non-covalently linked to each other in an RGEN protein-CPP complex. The RGEN protein-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a cell.

Significantly, certain embodiments of the disclosed invention can be used to deliver an RGEN already associated (pre-associated) with an RNA component into a cell. Such embodiments may avoid the need to deliver a DNA construct into cells for expressing an RGEN RNA component, thus averting any potentially unwanted effects of introducing exogenous DNA into cells. The disclosed invention is flexible, however, since in certain other embodiments an RNA component can be provided (e.g., expressed) in a cell into which an RGEN protein-CPP complex is being delivered. An RNA component provided in this manner can associate with an RGEN protein component after delivery/entry of the RGEN protein-CPP complex into the cell. Regardless of the mode of RNA component delivery, an RGEN protein-CPP complex herein is able to associate with an RNA component, forming an RGEN-CPP complex that can target a specific DNA sequence in the cell. Thus, the disclosed invention offers substantial flexibility for providing an RGEN in cells to perform RGEN-mediated DNA targeting.

Compositions disclosed in certain embodiments comprise at least one protein component of an RGEN. An RGEN herein refers to a complex comprising at least one Cas protein and at least one RNA component. Thus, an RGEN protein component can refer to a Cas protein such as Cas9. Examples of suitable Cas proteins include one or more Cas endonucleases of type I, II, or III CRISPR systems (Bhaya et al., *Annu. Rev. Genet.* 45:273-297, incorporated herein by reference). A type I CRISPR Cas protein can be a Cas3 or Cas4 protein, for example. A type II CRISPR Cas protein can be a Cas9 protein, for example. A type III CRISPR Cas protein can be a Cas10 protein, for example. A Cas9 protein is used in certain preferred embodiments. A Cas protein in certain embodiments may be a bacterial or archaeal protein. Type I-III CRISPR Cas proteins herein are typically prokaryotic

in origin; type I and III Cas proteins can be derived from bacterial or archaeal species, whereas type II Cas proteins (i.e., a Cas9) can be derived from bacterial species, for example. In other embodiments, suitable Cas proteins include one or more of Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10,
5 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.

In other aspects of the disclosed invention, a Cas protein herein can be from
10 any of the following genera: *Aeropyrum, Pyrobaculum, Sulfolobus, Archaeoglobus, Haloarcula, Methanobacteriumn, Methanococcus, Methanosarcina, Methanopyrus, Pyrococcus, Picrophilus, Thernioplasnia, Corynebacterium, Mycobacterium, Streptomyces, Aquifrx, Porphvromonas, Chlorobium, Thermus, Bacillus, Listeria, Staphylococcus, Clostridium, Thermoanaerobacter, Mycoplasma, Fusobacterium,*
15 *Azarcus, Chromobacterium, Neisseria, Nitrosomonas, Desulfovibrio, Geobacter, Myrococcus, Campylobacter, Wolinella, Acinetobacter, Erwinia, Escherichia, Legionella, Methylococcus, Pasteurella, Photobacterium, Salmonella, Xanthomonas, Yersinia, Streptococcus, Treponema, Francisella*, or *Thermotoga*. Alternatively, a Cas protein herein can be encoded, for example, by any of SEQ ID
20 NOs:462-465, 467-472, 474-477, 479-487, 489-492, 494-497, 499-503, 505-508, 510-516, or 517-521 as disclosed in U.S. Appl. Publ. No. 2010/0093617, which is incorporated herein by reference.

An RGEN protein component can comprise a Cas9 amino acid sequence, for example. An RGEN comprising this type of protein component typically can be
25 characterized as having Cas9 as the endonuclease component of the RGEN. The amino acid sequence of a Cas9 protein herein, as well as certain other Cas proteins herein, may be derived from a *Streptococcus* (e.g., *S. pyogenes, S. pneumoniae, S. thermophilus, S. agalactiae, S. parasanguinis, S. oralis, S. salivarius, S. macacae, S. dysgalactiae, S. anginosus, S. constellatus, S. pseudoporcinus, S. mutans*),
30 *Listeria* (e.g., *L. innocua*), *Spiroplasma* (e.g., *S. apis, S. syrphidicola*), *Peptostreptococcaceae, Atopobium, Porphyromonas* (e.g., *P. catoniae*), *Prevotella* (e.g., *P. intermedia*), *Veillonella, Treponema* (e.g., *T. socranskii, T. denticola*), *Capnocytophaga, Finegoldia* (e.g., *F. magna*), *Coriobacteriaceae* (e.g., *C.*

bacterium), *Olsenella* (e.g., *O. profusa*), *Haemophilus* (e.g., *H. sputorum*, *H. pittmaniae*), *Pasteurella* (e.g., *P. bettyae*), *Olivibacter* (e.g., *O. sitiensis*), *Epilithonimonas* (e.g., *E. tenax*), *Mesonia* (e.g., *M. mobilis*), *Lactobacillus*, *Bacillus* (e.g., *B. cereus*), *Aquimarina* (e.g., *A. muelleri*), *Chryseobacterium* (e.g., *C. palustre*), *Bacteroides* (e.g., *B. graminisolvens*), *Neisseria* (e.g., *N. meningitidis*), *Francisella* (e.g., *F. novicida*), or *Flavobacterium* (e.g., *F. frigidarium*, *F. soli*) species, for example. An *S. pyogenes* Cas9 is preferred in certain aspects herein. As another example, a Cas9 protein can be any of the Cas9 proteins disclosed in Chylinski et al. (*RNA Biology* 10:726-737), which is incorporated herein by reference.

Accordingly, the sequence of a Cas9 protein herein can comprise, for example, any of the Cas9 amino acid sequences disclosed in GenBank Accession Nos. G3ECR1 (*S. thermophilus*), WP_026709422, WP_027202655, WP_027318179, WP_027347504, WP_027376815, WP_027414302, WP_027821588, WP_027886314, WP_027963583, WP_028123848, WP_028298935, Q03JI6 (*S. thermophilus*), EGP66723, EGS38969, EGV05092, EHI65578 (*S. pseudoporcinus*), EIC75614 (*S. oralis*), EID22027 (*S. constellatus*), EIJ69711, EJP22331 (*S. oralis*), EJP26004 (*S. anginosus*), EJP30321, EPZ44001 (*S. pyogenes*), EPZ46028 (*S. pyogenes*), EQL78043 (*S. pyogenes*), EQL78548 (*S. pyogenes*), ERL10511, ERL12345, ERL19088 (*S. pyogenes*), ESA57807 (*S. pyogenes*), ESA59254 (*S. pyogenes*), ESU85303 (*S. pyogenes*), ETS96804, UC75522, EGR87316 (*S. dysgalactiae*), EGS33732, EGV01468 (*S. oralis*), EHJ52063 (*S. macacae*), EID26207 (*S. oralis*), EID33364, EIG27013 (*S. parasanguinis*), EJF37476, EJO19166 (*Streptococcus* sp. BS35b), EJU16049, EJU32481, YP_006298249, ERF61304, ERK04546, ETJ95568 (*S. agalactiae*), TS89875, ETS90967 (*Streptococcus* sp. SR4), ETS92439, EUB27844 (*Streptococcus* sp. BS21), AFJ08616, EUC82735 (*Streptococcus* sp. CM6), EWC92088, EWC94390, EJP25691, YP_008027038, YP_008868573, AGM26527, AHK22391, AHB36273, Q927P4, G3ECR1, or Q99ZW2 (*S. pyogenes*), which are incorporated by reference. A variant of any of these Cas9 protein sequences may be used, but should have specific binding activity, and optionally cleavage or nicking activity, toward DNA when associated with an RNA component herein. Such a variant may comprise an amino acid sequence that is at least about 80%, 81%,

82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the amino acid sequence of the reference Cas9.

Alternatively, a Cas9 protein herein can be encoded by any of SEQ ID NOs:462 (*S. thermophilus*), 474 (*S. thermophilus*), 489 (*S. agalactiae*), 494 (*S. agalactiae*), 499 (*S. mutans*), 505 (*S. pyogenes*), or 518 (*S. pyogenes*) as disclosed in U.S. Appl. Publ. No. 2010/0093617 (incorporated herein by reference), for example. Alternatively still, a Cas9 protein herein can comprise the amino acid sequence of SEQ ID NO:3, or residues 1-1368, 2-1368, or 2-1379, of SEQ ID NO:3, for example. Alternatively still, a Cas9 protein may comprise an amino acid sequence that is at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to any of the foregoing amino acid sequences, for example. Such a variant Cas9 protein should have specific binding activity, and optionally cleavage or nicking activity, toward DNA when associated with an RNA component herein.

The origin of a Cas protein used herein (e.g., Cas9) may be from the same species from which the RNA component(s) is derived, or it can be from a different species. For example, an RGEN comprising a Cas9 protein derived from a *Streptococcus* species (e.g., *S. pyogenes* or *S. thermophilus*) may be complexed with at least one RNA component having a sequence (e.g., crRNA repeat sequence, tracrRNA sequence) derived from the same *Streptococcus* species. Alternatively, the origin of a Cas protein used herein (e.g., Cas9) may be from a different species from which the RNA component(s) is derived (the Cas protein and RNA component(s) may be heterologous to each other); such heterologous Cas/RNA component RGENs should have DNA targeting activity.

Determining binding activity and/or endonucleolytic activity of a Cas protein herein toward a specific target DNA sequence may be assessed by any suitable assay known in the art, such as disclosed in U.S. Patent No. 8697359, which is disclosed herein by reference. A determination can be made, for example, by expressing a Cas protein and suitable RNA component in a cell, and then examining the predicted DNA target site for the presence of an indel (a Cas protein in this particular assay would typically have complete endonucleolytic activity [double-strand cleaving activity]). Examining for the presence of an alteration/modification

(e.g., indel) at the predicted target site could be done via a DNA sequencing method or by inferring alteration/modification formation by assaying for loss of function of the target sequence, for example. In another example, Cas protein activity can be determined by expressing a Cas protein and suitable RNA component in a cell that has been provided a donor DNA comprising a sequence homologous to a sequence in at or near the target site. The presence of donor DNA sequence at the target site (such as would be predicted by successful HR between the donor and target sequences) would indicate that targeting occurred. In still another example, Cas protein activity can be determined using an *in vitro* assay in which a Cas protein and suitable RNA component are mixed together along with a DNA polynucleotide containing a suitable target sequence. This assay can be used to detect binding (e.g., gel-shift) by Cas proteins lacking cleavage activity, or cleavage by Cas proteins that are endonucleolytically competent.

A Cas protein herein such as a Cas9 can further comprise a heterologous nuclear localization sequence (NLS) in certain aspects. A heterologous NLS amino acid sequence herein may be of sufficient strength to drive accumulation of a Cas protein, or Cas protein-CPP complex, in a detectable amount in the nucleus of a cell herein, for example. An NLS may comprise one (monopartite) or more (e.g., bipartite) short sequences (e.g., 2 to 20 residues) of basic, positively charged residues (e.g., lysine and/or arginine), and can be located anywhere in a Cas amino acid sequence but such that it is exposed on the protein surface. An NLS may be operably linked to the N-terminus or C-terminus of a Cas protein herein, for example. Two or more NLS sequences can be linked to a Cas protein, for example, such as on both the N- and C-termini of a Cas protein. Non-limiting examples of suitable NLS sequences herein include those disclosed in U.S. Patent Nos. 6660830 and 7309576 (e.g., Table 1 therein), which are both incorporated herein by reference. Another example of an NLS useful herein includes amino acid residues 1373-1379 of SEQ ID NO:3. A Cas protein as disclosed herein can be fused with a CPP (an example of a Cas protein covalently linked to a CPP), for example. It would be understood that such a Cas-CPP fusion protein can also comprise an NLS as described above. It would also be understood that, in embodiments in which a Cas protein is fused with an amino acid sequence targeting a different organelle (e.g., mitochondria), such a Cas protein typically would not contain an NLS.

In certain embodiments, a Cas protein and its respective RNA component (e.g., crRNA) that directs DNA-specific targeting by the Cas protein can be heterologous to a cell, in particular a non-prokaryotic cell. The heterologous nature of these RGEN components is due to that Cas proteins and their respective RNA components are only known to exist in prokaryotes (bacteria and archaea).

In some embodiments, a Cas protein is part of a fusion protein comprising one or more heterologous protein domains (e.g., 1, 2, 3, or more domains in addition to the Cas protein). These embodiments can encompass a Cas protein that is covalently linked to a CPP and one or more additional heterologous amino acid sequences, for example. Other embodiments can encompass a Cas protein that is covalently linked to one or more additional heterologous amino acid sequences not including a CPP, for example (a CPP would be non-covalently linked to a Cas fusion protein in such embodiments). A fusion protein comprising a Cas protein may comprise any additional protein sequence, and optionally a linker sequence between any two domains, such as between Cas and a first heterologous domain. Examples of protein domains that may be fused to a Cas protein herein include, without limitation, epitope tags (e.g., histidine [His, poly-histidine], V5, FLAG, influenza hemagglutinin [HA], myc, VSV-G, thioredoxin [Trx]), reporters (e.g., glutathione-5-transferase [GST], horseradish peroxidase [HRP], chloramphenicol acetyltransferase [CAT], beta-galactosidase, beta-glucuronidase [GUS], luciferase, green fluorescent protein [GFP], HcRed, DsRed, cyan fluorescent protein [CFP], yellow fluorescent protein [YFP], blue fluorescent protein [BFP]), and domains having one or more of the following activities: methylase activity, demethylase activity, transcription activation activity (e.g., VP16 or VP64), transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity and nucleic acid binding activity. A Cas protein in other embodiments may be in fusion with a protein that binds DNA molecules or other molecules, such as maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD), GAL4A DNA binding domain, and herpes simplex virus (HSV) VP16. Additional domains that may be part of a fusion protein comprising a Cas protein herein are disclosed in U.S. Patent Appl. Publ. No. 2011/0059502, which is incorporated herein by reference. In certain embodiments in which a Cas protein is fused to a heterologous protein (e.g., a transcription factor), the Cas protein has

DNA recognition and binding activity (when in complex with a suitable RNA component herein), but no DNA nicking or cleavage activity. A Cas protein as disclosed herein can be fused with a CPP (an example of a Cas protein covalently linked to a CPP), for example. It would be understood that such a Cas-CPP fusion protein can also be fused with one or more heterologous domains as described above, if desired.

Other examples of heterologous domains that can be linked to a Cas protein herein include amino acid sequences targeting the protein to a particular organelle (i.e., localization signal). Examples of organelles that can be targeted include mitochondria and chloroplasts. Typically, such targeting domains are used instead of an NLS when targeting extra-nuclear DNA sites. A mitochondrial targeting sequence (MTS) can be situated at or near the N-terminus of a Cas protein, for example. MTS examples are disclosed in U.S. Patent Appl. Publ. Nos. 2007/0011759 and 2014/0135275, which are incorporated herein by reference. A chloroplast targeting sequence can be as disclosed in U.S. Patent Appl. Publ. No. 2010/0192262 or 2012/0042412, for example, which are incorporated herein by reference.

The protein component of an RGEN can be associated with at least one RNA component (thereby constituting a complete RGEN) that comprises a sequence complementary to a target site sequence on a chromosome or episome in a cell, for example. The RGEN in such embodiments can bind to the target site sequence, and optionally cleave one or both DNA strands at the target site sequence. An RGEN can cleave one or both strands of a DNA target sequence, for example. An RGEN can cleave both strands of a DNA target sequence in another example. It would be understood that in all these embodiments, an RGEN protein component can be covalently or non-covalently linked to at least one CPP in an RGEN protein-CPP complex. The association of an RGEN protein-CPP complex with an RNA component herein can be characterized as forming an RGEN-CPP complex. Any disclosure herein regarding an RGEN can likewise apply to the RGEN component of an RGEN-CPP complex, unless otherwise noted.

An RGEN herein that can cleave both strands of a DNA target sequence typically comprises a Cas protein that has all of its endonuclease domains in a functional state (e.g., wild type endonuclease domains or variants thereof retaining

some or all activity in each endonuclease domain). Thus, a wild type Cas protein (e.g., a Cas9 protein disclosed herein), or a variant thereof retaining some or all activity in each endonuclease domain of the Cas protein, is a suitable example of an RGEN that can cleave both strands of a DNA target sequence. A Cas9 protein comprising functional RuvC and HNH nuclease domains is an example of a Cas protein that can cleave both strands of a DNA target sequence. An RGEN herein that can cleave both strands of a DNA target sequence typically cuts both strands at the same position such that blunt-ends (i.e., no nucleotide overhangs) are formed at the cut site.

An RGEN herein that can cleave one strand of a DNA target sequence can be characterized herein as having nickase activity (e.g., partial cleaving capability). A Cas nickase (e.g., Cas9 nickase) herein typically comprises one functional endonuclease domain that allows the Cas to cleave only one strand (i.e., make a nick) of a DNA target sequence. For example, a Cas9 nickase may comprise (i) a mutant, dysfunctional RuvC domain and (ii) a functional HNH domain (e.g., wild type HNH domain). As another example, a Cas9 nickase may comprise (i) a functional RuvC domain (e.g., wild type RuvC domain) and (ii) a mutant, dysfunctional HNH domain.

Non-limiting examples of Cas9 nickases suitable for use herein are disclosed by Gasiunas et al. (*Proc. Natl. Acad. Sci. U.S.A.* 109:E2579-E2586), Jinek et al. (*Science* 337:816-821), Sapranauskas et al. (*Nucleic Acids Res.* 39:9275-9282) and in U.S. Patent Appl. Publ. No. 2014/0189896, which are incorporated herein by reference. For example, a Cas9 nickase herein can comprise an *S. thermophilus* Cas9 having an Asp-31 substitution (e.g., Asp-31-Ala) (an example of a mutant RuvC domain), or a His-865 substitution (e.g., His-865-Ala), Asn-882 substitution (e.g., Asn-882-Ala), or Asn-891 substitution (e.g., Asn-891-Ala) (examples of mutant HNH domains). Also for example, a Cas9 nickase herein can comprise an *S. pyogenes* Cas9 having an Asp-10 substitution (e.g., Asp-10-Ala), Glu-762 substitution (e.g., Glu-762-Ala), or Asp-986 substitution (e.g., Asp-986-Ala) (examples of mutant RuvC domains), or a His-840 substitution (e.g., His-840-Ala), Asn-854 substitution (e.g., Asn-854-Ala), or Asn-863 substitution (e.g., Asn-863-Ala) (examples of mutant HNH domains). Regarding *S. pyogenes* Cas9, the three RuvC subdomains are generally located at amino acid residues 1-59, 718-769 and 909-

1098, respectively, and the HNH domain is located at amino acid residues 775-908 (Nishimasu et al., *Cell* 156:935-949).

A Cas9 nickase herein can be used for various purposes in cells, if desired. For example, a Cas9 nickase can be used to stimulate HR at or near a DNA target site sequence with a suitable donor polynucleotide. Since nicked DNA is not a substrate for NHEJ processes, but is recognized by HR processes, nicking DNA at a specific target site should render the site more receptive to HR with a suitable donor polynucleotide.

As another example, a pair of Cas9 nickases can be used to increase the specificity of DNA targeting. In general, this can be done by providing two Cas9 nickases that, by virtue of being associated with RNA components with different guide sequences, target and nick nearby DNA sequences on opposite strands in the region for desired targeting. Such nearby cleavage of each DNA strand creates a DSB (i.e., a DSB with single-stranded overhangs), which is then recognized as a substrate for NHEJ (leading to indel formation) or HR (leading to recombination with a suitable donor polynucleotide, if provided). Each nick in these embodiments can be at least about 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, or 100 (or any integer between 5 and 100) bases apart from each other, for example. One or two Cas9 nickase proteins herein can be used in a Cas9 nickase pair as described above. For example, a Cas9 nickase with a mutant RuvC domain, but functioning HNH domain (i.e., Cas9 $HNH^+/RuvC^-$), could be used (e.g., *S. pyogenes* Cas9 $HNH^+/RuvC^-$). Each Cas9 nickase (e.g., Cas9 $HNH^+/RuvC^-$) would be directed to specific DNA sites nearby each other (up to 100 base pairs apart) by using suitable RNA components herein with guide RNA sequences targeting each nickase to each specific DNA site.

An RGEN in certain embodiments can bind to a DNA target site sequence, but does not cleave any strand at the target site sequence. Such an RGEN may comprise a Cas protein in which all of its nuclease domains are mutant, dysfunctional. For example, a Cas9 protein herein that can bind to a DNA target site sequence, but does not cleave any strand at the target site sequence, may comprise both a mutant, dysfunctional RuvC domain and a mutant, dysfunctional HNH domain. Non-limiting examples of such a Cas9 protein comprise any of the RuvC and HNH nuclease domain mutations disclosed above (e.g., an *S. pyogenes*

Cas9 with an Asp-10 substitution such as Asp-10-Ala and a His-840 substitution such as His-840-Ala). A Cas protein herein that binds, but does not cleave, a target DNA sequence can be used to modulate gene expression, for example, in which case the Cas protein could be fused with a transcription factor (or portion thereof)
5 (e.g., a repressor or activator, such as any of those disclosed herein). For example, a Cas9 comprising an *S. pyogenes* Cas9 with an Asp-10 substitution (e.g., Asp-10-Ala) and a His-840 substitution (e.g., His-840-Ala) can be fused to a VP16 or VP64 transcriptional activator domain. The guide sequence used in the RNA component of such an RGEN would be complementary to a DNA sequence in a
10 gene promoter or other regulatory element (e.g., intron), for example.

An RGEN herein can bind to a target site sequence, and optionally cleave one or both strands of the target site sequence, in a chromosome, episome, or any other DNA molecule in the genome of a cell. This recognition and binding of a target sequence is specific, given that an RNA component of the RGEN comprises a
15 sequence (guide sequence) that is complementary to a strand of the target sequence. A target site in certain embodiments can be unique (i.e., there is a single occurrence of the target site sequence in the subject genome).

The length of a target sequence herein can be at least 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides; between 13-30
20 nucleotides; between 17-25 nucleotides; or between 17-20 nucleotides, for example. This length can include or exclude a PAM sequence. Also, a strand of a target sequence herein has sufficient complementarity with a guide sequence (of a crRNA or gRNA) to hybridize with the guide sequence and direct sequence-specific binding of a Cas protein or Cas protein complex to the target sequence (if a suitable PAM is
25 adjacent to the target sequence, see below). The degree of complementarity between a guide sequence and a strand of its corresponding DNA target sequence is at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%, for example. A target site herein may be located in a sequence encoding a gene product (e.g., a protein or an RNA) or a non-coding sequence (e.g., a regulatory
30 sequence or a "junk" sequence), for example.

A PAM (protospacer-adjacent motif) sequence may be adjacent to the target site sequence. A PAM sequence is a short DNA sequence recognized by an RGEN herein. The associated PAM and first 11 nucleotides of a DNA target sequence are

likely important to Cas9/gRNA targeting and cleavage (Jiang et al., Nat. Biotech. 31:233-239). The length of a PAM sequence herein can vary depending on the Cas protein or Cas protein complex used, but is typically 2, 3, 4, 5, 6, 7, or 8 nucleotides long, for example. A PAM sequence is immediately downstream from, or within 2, or 3 nucleotides downstream of, a target site sequence that is complementary to the strand in the target site that is in turn complementary to an RNA component guide sequence, for example. In embodiments herein in which an RGEN is an endonucleolytically active Cas9 protein complexed with an RNA component, Cas9 binds to the target sequence as directed by the RNA component and cleaves both strands immediately 5' of the third nucleotide position upstream of the PAM sequence. Consider the following example of a target site:PAM sequence:

5'-NNNNNNNNNNNNNNNNNNNNXGG-3' (SEQ ID NO:43).

N can be A, C, T, or G, and X can be A, C, T, or G in this example sequence (X can also be referred to as N_{PAM}). The PAM sequence in this example is XGG (underlined). A suitable Cas9/RNA component complex would cleave this target immediately 5' of the double-underlined N. The string of N's in SEQ ID NO:43) represents target sequence that is at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical, for example, with a guide sequence in an RNA component herein (where any T's of the DNA target sequence would align with any U's of the RNA guide sequence). A guide sequence of an RNA component of a Cas9 complex, in recognizing and binding at this target sequence (which is representive of target sites herein), would anneal with the complement sequence of the string of N's; the percent complementarity between a guide sequence and the target site complement is at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%, for example. If a Cas9 nickase is used to target SEQ ID NO:43) in a genome, the nickase would nick immediately 5' of the double-underlined N or at the same position of the complementary strand, depending on which endonuclease domain in the nickase is dysfunctional. If a Cas9 having no nucleolytic activity (both RuvC and HNH domains dysfuntional) is used to target SEQ ID NO:43 in a genome, it would recognize and bind the target sequence, but not make any cuts to the sequence.

A PAM herein is typically selected in view of the type of RGEN being employed. A PAM sequence herein may be one recognized by an RGEN

comprising a Cas, such as Cas9, derived from any of the species disclosed herein from which a Cas can be derived, for example. In certain embodiments, the PAM sequence may be one recognized by an RGEN comprising a Cas9 derived from *S. pyogenes*, *S. thermophilus, S. agalactiae, N. meningitidis, T. denticola,* or *F. novicida.* For example, a suitable Cas9 derived from *S. pyogenes* could be used to target genomic sequences having a PAM sequence of NGG (SEQ ID NO:44; N can be A, C, T, or G). As other examples, a suitable Cas9 could be derived from any of the following species when targeting DNA sequences having the following PAM sequences: *S. thermophilus* (NNAGAA [SEQ ID NO:45]), *S. agalactiae* (NGG [SEQ ID NO:44]), NNAGAAW [SEQ ID NO:46, W is A or T], NGGNG [SEQ ID NO:47]), *N. meningitidis* (NNNNGATT [SEQ ID NO:48]), *T. denticola* (NAAAAC [SEQ ID NO:49]), or *F. novicida* (NG [SEQ ID NO:50]) (where N's in all these particular PAM sequences are A, C, T, or G). Other examples of Cas9/PAMs useful herein include those disclosed in Shah et al. (*RNA Biology* 10:891-899) and Esvelt et al. (*Nature Methods* 10:1116-1121), which are incorporated herein by reference. Examples of target sequences herein follow SEQ ID NO:43, but with the 'XGG' PAM replaced by any one of the foregoing PAMs.

An RNA component herein can comprise a sequence complementary to a target site sequence in a chromosome or episome in a cell. An RGEN can specifically bind to a target site sequence, and optionally cleave one or both strands of the target site sequence, based on this sequence complementary. Thus, the complementary sequence of an RNA component in certain embodiments of the disclosed invention can also be referred to as a guide sequence or variable targeting domain.

The guide sequence of an RNA component (e.g., crRNA or gRNA) herein can be at least 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 ribonucleotides in length; between 13-30 ribonucleotides in length; between 17-25 ribonucleotides in length; or between 17-20 ribonucleotides in length, for example. In general, a guide sequence herein has sufficient complementarity with a strand of a target DNA sequence to hybridize with the target sequence and direct sequence-specific binding of a Cas protein or Cas protein complex to the target sequence (if a suitable PAM is adjacent to the target sequence). The degree of complementarity between a guide sequence and its corresponding DNA target

sequence is at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%, for example. The guide sequence can be engineered accordingly to target an RGEN to a DNA target sequence in a cell.

An RNA component herein can comprise a crRNA, for example, which comprises a guide sequence and a repeat (tracrRNA mate) sequence. The guide sequence is typically located at or near (within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more bases) the 5' end of the crRNA. Downstream the guide sequence of a crRNA is a "repeat" or "tracrRNA mate" sequence that is complementary to, and can hybridize with, sequence at the 5' end of a tracrRNA. Guide and tracrRNA mate sequences can be immediately adjacent, or separated by 1, 2, 3, 4 or more bases, for example. A tracrRNA mate sequence has, for example, at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence complementarity to the 5' end of a tracrRNA. In general, degree of complementarity can be with reference to the optimal alignment of the tracrRNA mate sequence and 5' end of the tracrRNA sequence, along the length of the shorter of the two sequences. The length of a tracrRNA mate sequence herein can be at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18 ribonucleotides in length, for example, and hybridizes with sequence of the same or similar length (e.g., plus or minus 1, 2, 3, 4, or 5 bases) at the 5' end of a tracrRNA. Suitable examples of tracrRNA mate sequences herein comprise SEQ ID NO:51 (guuuuuguacucucaagauuua), SEQ ID NO:52 (guuuuuguacucuca), SEQ ID NO:53 (guuuuagagcua), or SEQ ID NO:54 (guuuuagagcuag), or variants thereof that (i) have at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity and (ii) can anneal with the 5'-end sequence of a tracrRNA. The length of a crRNA herein can be at least about 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, or 48 ribonucleotides; or about 18-48 ribonucleotides; or about 25-50 ribonucleotides, for example.

A tracrRNA can be included along with a crRNA in embodiments in which a Cas9 protein of a type II CRISPR system is comprised in the RGEN. A tracrRNA herein comprises in 5'-to-3' direction (i) a sequence that anneals with the repeat region (tracrRNA mate sequence) of crRNA and (ii) a stem loop-containing portion. The length of a sequence of (i) can be the same as, or similar with (e.g., plus or minus 1, 2, 3, 4, or 5 bases), any of the tracrRNA mate sequence lengths disclosed above, for example. The total length of a tracrRNA herein (i.e., sequence

components [i] and [ii]) can be at least about 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, or 90 (or any integer between 30 and 90) ribonucleotides, for example. A tracrRNA may further include 1, 2, 3, 4, 5, or more uracil residues at the 3'-end, which may be present by virtue of expressing the tracrRNA with a transcription terminator sequence.

A tracrRNA herein can be derived from any of the bacterial species listed above from which a Cas9 sequence can be derived, for example. Examples of suitable tracrRNA sequences include those disclosed in U.S. Patent No. 8697359 and Chylinski et al. (*RNA Biology* 10:726-737), which are incorporated herein by reference. A preferred tracrRNA herein can be derived from a *Streptococcus* species tracrRNA (e.g., *S. pyogenes*, *S. thermophilus*). Other suitable examples of tracrRNAs herein may comprise:

SEQ ID NO:55:
uagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugc,
SEQ ID NO:56:
uagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaagug, or
SEQ ID NO:57:
uagcaaguuaaaauaaggcuaguccguuauca,
which are derived from *S. pyogenes* tracrRNA. Other suitable examples of tracrRNAs herein may comprise:
SEQ ID NO:58:
uaaaucuugcagaagcuacaaagauaaggcuucaugccgaaaucaacacccugucauuuuauggcagg guguuuucguuauuuaa,
SEQ ID NO:59:
ugcagaagcuacaaagauaaggcuucaugccgaaaucaacacccugucauuuuauggcaggguguuuu cguuauuua, or
SEQ ID NO:60:
ugcagaagcuacaaagauaaggcuucaugccgaaaucaacacccugucauuuuauggcagggugu,
which are derived from *S. thermophilus* tracrRNA.
Still other examples of tracrRNAs herein are variants of these tracrRNA SEQ ID NOs that (i) have at least about 80%, 85%, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% sequence identity therewith and (ii) can function as a tracrRNA (e.g., 5'-end sequence can anneal to tracrRNA mate sequence of a crRNA, sequence

downstream from the 5'-end sequence can form one or more hairpins, variant tracrRNA can form complex with a Cas9 protein).

An RNA component of an RGEN disclosed herein (or said another way, an RNA component that may be associated with an RGEN protein component) can
5 comprise, for example, a guide RNA (gRNA) comprising a crRNA operably linked to, or fused to, a tracrRNA. The crRNA component of a gRNA in certain preferred embodiments is upstream of the tracrRNA component (i.e., such a gRNA comprises, in 5'-to-3' direction, a crRNA operably linked to a tracrRNA). Any crRNA and/or tracrRNA (and/or portion thereof, such as a crRNA repeat sequence,
10 tracrRNA mate sequence, or tracrRNA 5'-end sequence) as disclosed herein (e.g., above embodiments) can be comprised in a gRNA, for example.

The tracrRNA mate sequence of the crRNA component of a gRNA herein should be able to anneal with the 5'-end of the tracrRNA component, thereby forming a hairpin structure. Any of the above disclosures regarding lengths of, and
15 percent complementarity between, tracrRNA mate sequences (of crRNA component) and 5'-end sequences (of tracrRNA component) can characterize the crRNA and tracrRNA components of a gRNA, for example. To facilitate this annealing, the operable linkage or fusion of the crRNA and tracrRNA components preferably comprises a suitable loop-forming ribonucleotide sequence (i.e., a loop-
20 forming sequence may link the crRNA and tracrRNA components together, forming the gRNA). Suitable examples of RNA loop-forming sequences include GAAA (SEQ ID NO:36), CAAA (SEQ ID NO:37) and AAAG (SEQ ID NO:38). However, longer or shorter loop sequences may be used, as may alternative loop sequences. A loop sequence preferably comprises a ribonucleotide triplet (e.g., AAA) and an
25 additional ribonucleotide (e.g., C or G) at either end of the triplet.

A gRNA herein forms a hairpin ("first hairpin") with annealing of its tracrRNA mate sequence (of the crRNA component) and tracrRNA 5'-end sequence portions. One or more (e.g., 1, 2, 3, or 4) additional hairpin structures can form downstream from this first hairpin, depending on the sequence of the tracrRNA component of the
30 gRNA. A gRNA may therefore have up to five hairpin structures, for example. A gRNA may further include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more residues following the end of the gRNA sequence, which may be present by virtue of expressing the gRNA with a

transcription terminator sequence, for example. These additional residues can be all U residues, or at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% U residues, for example, depending on the choice of terminator sequence.

Non-limiting examples of suitable gRNAs useful in the disclosed invention may comprise:

SEQ ID NO:61:
NNNNNNNNNNNNNNNNNNNNguuuuuguacucucaagauuuaGAAAuaaaucuugcagaagcuacaaagauaaggcuucaugccgaaaucaacacccugucauuuuauggcaggguguuuucguuauuuaa,

SEQ ID NO:62:
NNNNNNNNNNNNNNNNNNNNguuuuuguacucucaGAAAugcagaagcuacaaagauaaggcuucaugccgaaaucaacacccugucauuuuauggcaggguguuuucguuauuuaa,

SEQ ID NO:63:
NNNNNNNNNNNNNNNNNNNNguuuuuguacucucaGAAAugcagaagcuacaaagauaaggcuucaugccgaaaucaacacccugucauuuuauggcagggugu,

SEQ ID NO:64:
NNNNNNNNNNNNNNNNNNNNguuuuuguacucucaGAAAuagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugc,

SEQ ID NO:65:
NNNNNNNNNNNNNNNNNNNNguuuuagagcuaGAAAuagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaagug,

SEQ ID NO:66:
NNNNNNNNNNNNNNNNNNNNguuuuagagcuaGAAAuagcaaguuaaaauaaggcuaguccguuauca, or

SEQ ID NO:67:
NNNNNNNNNNNNNNNNNNNNguuuuagagcuaGAAAuagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuu.

In each of SEQ ID NOs:61-67, the single-underlined sequence represents a crRNA portion of the gRNA. Each "N" represents a ribonucleotide base (A, U, G, or C) of a suitable guide sequence. The first block of lower case letters represents tracrRNA mate sequence. The second block of lower case letters represents a tracrRNA portion of the gRNA. The double-underlined sequence approximates that portion of tracrRNA sequence that anneals with the tracrRNA mate sequence to form a first

hairpin. A loop sequence (GAAA, SEQ ID NO:36) is shown in capital letters, which operably links the crRNA and tracrRNA portions of each gRNA. Other examples of gRNAs herein include variants of the foregoing gRNAs that (i) have at least about 80%, 85%, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% sequence identity (excluding guide sequence in this calculation) with these sequences, and (ii) can function as a gRNA that specifically targets a Cas9 protein to bind with, and optionally nick or cleave, a target DNA sequence.

A gRNA herein can also be characterized in terms of having a guide sequence (VT domain) followed by a Cas endonuclease recognition (CER) domain. A CER domain comprises a tracrRNA mate sequence followed by a tracrRNA sequence. Examples of CER domains useful herein include those comprised in SEQ ID NOs:61-67 above (the CER domain in each is the sequence following the N's of the VT domain). Another suitable example of a CER domain is SEQ ID NO:24 (see Examples), which comprises in 5'-to-3' direction the tracrRNA mate sequence of SEQ ID NO:53, the loop-forming sequence of SEQ ID NO:36 (GAAA), and the tracrRNA sequence of SEQ ID NO:55.

An RNA component of an RGEN optionally does not have a 5'-cap (7-methylguanylate [m^7G] cap) (i.e., such an RNA component does not have an m^7G cap at its 5'-terminus). An RNA component herein can have, for example, a 5'-hydroxyl group instead of a 5'-cap. Alternatively, an RNA component herein can have, for example, a 5' phosphate instead of a 5'-cap. It is believed that an RNA component in these embodiments can better accumulate in the nucleus (such as after its transcription in the nucleus, or after its RGEN-mediated import into the nucleus, depending on how the RNA component is provided herein), since 5'-capped RNA (i.e., RNA having 5' m^7G cap) is subject to nuclear export. Preferred examples of uncapped RNA components herein include suitable gRNAs, crRNAs, and/or tracrRNAs. In certain embodiments, an RNA component herein lacks a 5'-cap, and optionally has a 5'-hydroxyl group instead, by virtue of RNA autoprocessing by a ribozyme sequence at the 5'-end of a precursor of the RNA component (i.e., a precursor RNA comprising a ribozyme sequence upstream of an RNA component such as a gRNA undergoes ribozyme-mediated autoprocessing to remove the ribozyme sequence, thereby leaving the downstream RNA component

without a 5'-cap). In certain other embodiments, an RNA component herein is not produced by transcription from an RNA polymerase III (Pol III) promoter.

A cell-penetrating peptide (CPP) herein can be about 5-30, 5-25, 5-20, 10-30, 10-25, or 10-20 amino acid residues in length, for example. As other examples, a CPP can be about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 amino acid residues in length. Yet in further aspects herein, a CPP can be up to about 35, 40, 45, 50, 55, or 60 amino acid residues in length.

A CPP disclosed herein can be cationic or amphipathic, for example. A cationic CPP herein typically comprises at least about 60% positively charged amino acids such as lysine (K), arginine (R), and/or histidine (H). Alternatively, a cationic CPP can comprise, for example, at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95% positively charged amino acids (e.g., R residues; K residues; K and R residues; K, R and H residues). A cationic CPP can be characterized as being arginine-rich (e.g., comprising at least 70% or 80% R residues) or lysine-rich (e.g., comprising at least 70% or 80% L residues) in certain embodiments. Examples of cationic CPPs useful herein are disclosed in Schmidt et al. (*FEBS Lett.* 584:1806-1813) and Wender et al. (polylysine; *Proc. Natl. Acad. Sci. USA* 97:13003-13008), which are incorporated herein by reference. Other examples of cationic CPPs comprise GRKKRRQRRR (SEQ ID NO:68), RKKRRQRRR (SEQ ID NO:69), or RKKRRQRR (SEQ ID NO:70), which were originally derived from HIV Tat protein, and penetratin (RQIKIWFQNRRMKWKK, SEQ ID NO:71), which was originally derived for the Antennapedia homeodomain protein of *Drosophila*.

Another example of a cationic CPP comprises a polyarginine sequence having a number of contiguous arginines sufficient to direct entry of the CPP and its cargo (e.g., RGEN protein component or RGEN) into a cell. The number of contiguous arginine residues in such a polyarginine sequence can be at least 3, 4, 5, 6, 7, 8, 9, 10, or 10-50 arginines, for instance. In certain aspects herein, a CPP can have 6 or more contiguous arginine residues (e.g., 6-7, 6-8, 6-9, or 6-10 arginine residues). "PolyR" (GGGGRRRRRRRRRLLLL, SEQ ID NO:15) can be comprised in a polyarginine CPP, if desired. Other polyarginine CPP examples comprise THRLPRRRRRR (SEQ ID NO:72) or GGRRARRRRRR (SEQ ID NO:73). In some embodiments, a CPP is an activatable CPP ("ACPP") (Aguilera et al., *Integr*

Biol. (Camb) 1:371-381; incorporated herein by reference). ACPPs typically comprise a polycationic CPP (e.g., nine contiguous arginines) connected via a cleavable linker to a matching polyanion (e.g., nine contiguous glutamates), which reduces the net charge to nearly zero and thereby inhibits CPP adhesion and uptake into cells. Upon cleavage of the linker, the polyanion is released, locally unmasking the polycation portion and its inherent adhesiveness, thereby allowing CPP cell entry. Another example herein is a polylysine CPP; any of the above embodiments of polyarginine, but in which R is replaced with K, are examples of polylysine CPPs herein.

An amphipathic CPP herein comprises an amino acid sequence containing an alternating pattern of polar/charged residues and non-polar, hydrophobic residues. The following CPPs are believed to be amphipathic, and are useful in certain aspects (regardless of whether amphipathic terminology perfectly applies): a CPP comprising transportan-10 (TP10) peptide (e.g., AGYLLGKINLKACAACAKKIL, SEQ ID NO:14); a CPP from a vascular endothelium cadherin protein, such as a CPP comprising a pVEC peptide (e.g., LIILRRRIRKQAHAHSK, SEQ ID NO:74; LLIILRRRIRKQAHAHSK, SEQ ID NO:13); a CPP from an Epstein-Barr virus Zebra trans-activator protein, such as a CPP comprising a Zebra peptide (e.g., ECDSELEIKRYKRVRVASRKCRAKFKQLLQHYREVAAAKSSENDRLRLLLKQMC, SEQ ID NO:12); a CPP comprising a $(KFF)_3K$ peptide (e.g., KFFKFFKFFK, SEQ ID NO:75); a CPP comprising a MAP peptide (KLALKLALKALKAALKLA, SEQ ID NO:76); a CPP comprising RRQRRTSKLMKR (SEQ ID NO:77); a CPP comprising KALAWEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO:78). Other amphipathic CPPs suitable herein include proline-rich CPPs, such as those comprising at least 3, 4, 5, 6, 7, or 8 repeats of VHLPPP (SEQ ID NO:79) or VRLPPP (SEQ ID NO:80).

As other examples, a CPP herein may comprise an MPG peptide (e.g., GALFLGFLGAAGSTMGAWSQPKSKRKV, SEQ ID NO:81); a Pep-1 peptide (e.g., KETWWETWWTEWSQPKKKRKV, SEQ ID NO:82); or a CPP from a human calcitonin protein, such as an hCT peptide (e.g., LGTYTQDFNKFHTFPQTAIGVGAP, SEQ ID NO:83; CGNLSTCMLGTYTQDFNK, SEQ ID NO:84). Still other examples of CPPs herein include those disclosed in Regberg et al. (*Int. J. Pharm.* 464:111-116), which is incorporated herein by reference.

A CPP suitable herein can alternatively comprise an amino acid sequence that is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to any of the CPP amino acid sequences disclosed herein, for example. Such a variant CPP protein should have CPP activity, such as the ability to mediate cellular uptake of molecular cargo (e.g., an amino acid sequence comprising one or more RGEN protein components [e.g., Cas9], or an amino acid sequence comprising one or more RGEN protein components [e.g., Cas9] associated with an RNA component). Testing the activity of a variant CPP can be done any number of ways, such as by covalently linking it with a fluorescent protein (e.g., GFP) and measuring the degree of fluorescence emitted from a cell contacted with a the CPP-fluorescent protein complex.

A CPP herein can be modified, if desired, to render it even more capable of carrying RGEN protein cargo from outside a cell to inside a cell. For example, a CPP can be modified to have a lipid group at either its N- or C-terminus. Suitable lipid groups herein include acyl groups such as stearyl and myristyl groups. Other examples of lipid groups are acyl groups with 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18 carbons. Conditions for modifying peptides with lipid groups useful herein are disclosed in Regberg et al. (*Int. J. Pharm.* 464:111-116) and Anko et al. (*Biochim. Biophys. Acta – Biomembranes* 1818:915-924) for example, which are incorporated herein by reference.

An RGEN protein component and at least one CPP herein can be covalently linked to each other in an RGEN protein-CPP complex in certain aspects herein. For example, an RGEN protein component and at least one CPP can be fused together in a single amino acid sequence (i.e., an RGEN protein component and at least one CPP can be comprised within a fusion protein). Thus, an example of covalent linkage herein can be via a peptide bond in which the amino acid sequence of an RGEN protein component is fused with the amino acid sequence of a CPP, such that both these amino acid sequences are contained in a single amino acid sequence. Such a fusion protein (or "chimeric protein"), can be characterized as an RGEN protein-CPP fusion herein. In those embodiments in which an RNA component is associated with an RGEN protein component, such a fusion protein can be characterized as an RGEN-CPP fusion.

One or more CPPs can be located at the N-terminus or C-terminus of an RGEN protein-CPP fusion, for example. Alternatively, one or more CPPs can be located at both the N- and C-termini of an RGEN protein-CPP fusion. Alternatively still, one or more CPPs can be located within the amino acid sequence of an RGEN protein-CPP fusion. Embodiments herein comprising more than one CPP can comprise at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 CPPs, or 5-10, 5-20, or 10-20 CPPs. The CPPs fused to the RGEN protein component can be the same or different (e.g., 2, 3, 4, or more different types of CPPs). One or more CPPs can be fused directly to the amino acid sequence of an RGEN protein, and/or can be fused to a heterologous domain(s) (e.g., NLS or other organelle-targeting sequence such as an MTS) that is fused with an RGEN protein.

A fusion between a CPP and an RGEN protein component herein can be direct (i.e., CPP amino acid sequence is directly linked to RGEN amino acid sequence by a peptide bond). Alternatively, a fusion between a CPP and an RGEN protein component can be via an intermediary amino acid sequence (this is an example of a CPP and RGEN protein component being indirectly linked). Examples of an intermediary amino acid sequence include suitable linker sequences comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, or 100 amino acid residues such as glycine, serine, alanine and/or proline. Suitable amino acid linkers are disclosed in U.S. Patent Nos. 8828690, 8580922 and 5990275, for example, which are incorporated herein by reference. Other examples of intermediary amino acid sequences can comprise one or more other types of proteins and/or domains. For example, a marker protein (e.g., a fluorescent protein such as any of those disclosed herein) can be comprised in an intermediary amino acid sequence.

A composition comprising a covalent complex of an RGEN protein component and at least one CPP, such as in a fusion protein, can be used with any cell type disclosed herein. Optionally, however, this composition can be used with non-mammalian cells such as yeast, fungi, and plants, but excludes use on mammalian cells.

Examples of RGEN protein-CPP fusion proteins herein can comprise SEQ ID NO:39 (Zebra CPP-Cas9-NLS fusion protein), 40 (PolyR CPP-Cas9-NLS fusion protein), 41 (TP10 CPP-Cas9-NLS fusion protein), or 42 (pVEC CPP-Cas9-NLS

fusion protein). SEQ ID NOs:39-42 are examples of Cas9-CPP fusion proteins. Other examples of RGEN protein-CPP fusion proteins comprise an amino acid sequence that is at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to any of SEQ ID NOs:39-42. Such a variant fusion protein should have (i) a CPP domain that can mediate cellular uptake of the fusion protein, and (ii) a Cas9 protein with specific binding activity, and optionally cleavage or nicking activity, toward DNA when associated with an RNA component. SEQ ID NO:39, 40, 41 and 42 comprise Zebra CPP (SEQ ID NO:12), PolyR CPP (SEQ ID NO:15), TP10 CPP (SEQ ID NO:14) and pVEC CPP (SEQ ID NO:13), respectively, operably linked to Cas9 (*S. pyogenes*)-NLS protein (residues 2-1379 of SEQ ID NO:3).

In certain embodiments, the protein component of a guide polynucleotide/Cas endonuclease system can be fused to a CPP, wherein the CPP comprises:

(i) a CPP from an Epstein-Barr virus Zebra trans-activator protein,

(ii) a CPP having 6 or more contiguous arginine residues,

(iii) a transportan-10 (TP10) CPP,

(iv) a CPP from a vascular endothelium cadherin protein, or

(vi) a CPP selected from the group consisting of a synthetic non-arginine CPP, a histidine-rich nona-arginine CPP and a Pas nona-arginine CPP.

Examples of synthetic nona-arginine, histidine-rich nona-arginine, and Pas nona-arginine CPPs are disclosed in, for example, Liu et al. (*Advanced Studies in Biology* 5(2):71-88, HIKARI Ltd), which is incorporated herein by reference.

Another example of how an RGEN protein component and at least one CPP can be covalently linked is via crosslinking (chemical crosslinking). Thus, an example of an RGEN protein-CPP complex herein can comprise an RGEN protein crosslinked to at least one CPP. Crosslinking herein refers to a process of chemically joining two or more molecules (an RGEN protein component and at least one CPP, in this case) by a covalent bond(s). Crosslinking can be performed using any number of processes known in the art, such as those disclosed in U.S. Patent Appl. Publ. No. 2011/0190813, U.S. Patent No. 8642744, and Bioconjugate Techniques, 2nd Edition (G.T. Hermanson, Academic Press, 2008), which are all incorporated herein by reference.

Typically, a CPP can be modified and/or synthesized to contain a suitable protein linking group at its N-terminus, C-terminus, and/or an amino acid side group, for the purpose of crosslinking the CPP to an RGEN protein component. A "protein linking group" refers to a group that is capable of reacting directly, either spontaneously or after activation (e.g., light), with an accessible side chain functional group of an RGEN protein component under suitable conditions (e.g., aqueous conditions) to produce a covalently link the CPP to the RGEN protein. A protein linking group may react with the side chain functional groups of a Lys, Cys, Ser, Thr, Tyr, His, or Arg amino acid residue in an RGEN protein, for example, to produce a covalent linkage to the protein. Either a homobifunctional (e.g., capable of linking amine to amine) or heterobifunctional (e.g., capable of linking amine to thiol) protein linking group can be used, for example. A protein linking group on a CPP can also react with a terminal group (e.g., N-terminus) of an RGEN protein in certain embodiments. Suitable protein linking groups herein include amino-reactive (e.g., NHS ester or imidoester), thiol (sulfhydryl)-reactive (e.g., a maleimide such as BMOE, BMB, or BMH), hydroxyl-reactive, imidazolyl-reactive, or guanidinyl-reactive groups. Exemplary protein linking groups include active esters (e.g., an amino-reactive NHS ester), and thiol-reactive maleimide or iodoacetamide groups. Further exemplary protein linking groups useful herein and methods of using them are described in Bioconjugate Techniques, 2nd Edition (G.T. Hermanson, Academic Press, 2008), for example.

A protein linking group herein typically can produce a link between a CPP and an RGEN protein with a backbone of 20 atoms or less in length. For example, such a link can be between 1 and 20 atoms in length, or about 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18 or 20 carbon atoms in length. A link may be linear, branched, cyclic or a single atom in certain embodiments. In certain cases, one, two, three, four or five or more carbon atoms of a linker backbone may be substituted with a sulfur, nitrogen or oxygen heteroatom. The bonds between backbone atoms may be saturated or unsaturated (usually not more than one, two, or three unsaturated bonds in the linker backbone). A linker may include, without limitation, an oligo(ethylene glycol); ether; thioether; tertiary amine; or alkyl group, which may be straight or branched (e.g., methyl, ethyl, n-propyl, iso-propyl, n-butyl, n-pentyl, t-butyl). As other examples, a linker backbone may include a cyclic group such as an

aryl, a heterocycle, or a cycloalkyl group, where 2 or more atoms (e.g., 2, 3 or 4 atoms) of the cyclic group are included in the backbone.

More than one type of CPP (e.g., 2, 3, 4, or more different types of CPPs) can be crosslinked to an RGEN protein component in certain embodiments. The ratio (molar ratio) of CPP(s) to RGEN protein that can be used when crosslinking can be at least about 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1 15:1, 20:1, 30:1, 40:1, or 50:1, for example. In other aspects, the average number of CPPs crosslinked to an RGEN protein may be at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25, or at least 5-10, 5-15, 5-20, or 5-25.

An RGEN protein component and at least one CPP can be crosslinked into a complex further comprising one or more other proteins/peptides/domains, if desired. Such other elements can optionally be used to bridge an RGEN protein component with a CPP, and may include any of the intermediary amino acid sequences described above.

An RGEN protein component and at least one CPP herein can be non-covalently linked to each other in an RGEN protein-CPP complex in certain aspects herein. Though not intending to be held to any particular theory or mechanism, it is contemplated that a non-covalent linkage between an RGEN protein component and at least one CPP can be due to electrostatic, Van der Waals, and/or hydrophobic forces. In those embodiments in which an RNA component is associated with an RGEN protein component, such embodiments can be characterized as comprising an RGEN that is non-covalently linked to at least one CPP in an RGEN-CPP complex. A composition comprising an RGEN protein component and CPP that are non-covalently linked can optionally be characterized as a mixture of these components.

In certain embodiments, an RGEN protein component is non-covalently linked to at least one CPP with an amino acid sequence consisting of the CPP amino acid sequence only. Such a CPP, while not having any “non-CPP” amino acid sequence, can optionally comprise a modification such as a lipid group as disclosed herein.

Alternatively, a CPP that is non-covalently linked to an RGEN protein component may be comprised in a fusion protein having both CPP amino acid

sequence and one or more heterologous amino acid sequences (non-RGEN protein sequences). A heterologous sequence in such embodiments can be that of a domain or a protein (e.g., a fluorescent protein such as any of those disclosed herein, or any domain/protein listed in the above disclosure regarding Cas fusions).
5 Another example is fusing a dimerization domain to a CPP, which dimerization domain is able to bind to a dimerization domain linked or fused to an RGEN protein component.

Leucine zipper domains are examples of dimerization domains herein. Leucine zipper domains can represent those from natural proteins known to contain
10 such domains (e.g., transcription factors), or can be synthetically designed. A leucine zipper domain linked to a CPP can associate ("zip together") with a leucine zipper domain of an RGEN protein component, thereby linking the CPP and RGEN protein component in a non-covalent complex. A pair of leucine zipper domains for non-covalently linking a CPP and an RGEN protein component can be the same
15 (such a domain pair forms a homodimeric leucine zipper) or different (such a domain pair forms a heterodimeric leucine zipper). Examples of leucine zipper domains include those disclosed in U.S. Patent Appl. Publ. Nos. 2003/0108869 and 2004/0147721. In certain aspects, a homodimeric leucine zipper can be formed using a leucine zipper domain from a GCN4 transcription factor, while in other
20 aspects a heterodimeric leucine zipper can be formed using leucine zipper domains from fos and jun transcription factors, respectively.

A non-covalent complex of an RGEN protein component and at least one CPP can further comprise one or more other proteins/peptides/domains, if desired. Such other elements can optionally be used to bridge an RGEN protein component
25 with a CPP, and may include any of the intermediary amino acid sequences described above.

More than one type of CPP (e.g., 2, 3, 4, or more different types of CPPs) can be non-covalently linked to an RGEN protein component in certain embodiments. The ratio (molar ratio) of CPP(s) to RGEN protein that can be used
30 to prepare such a complex can be at least about 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1 15:1, 20:1, 30:1, 40:1, or 50:1, for example. In other aspects, the average number of CPPs non-covalently linked to an RGEN protein may be at least 1, 2, 3,

4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25, or at least 5-10, 5-15, 5-20, or 5-25.

In certain embodiments, a non-covalent complex of an RGEN protein component and at least one CPP can be prepared by mixing an appropriate amount of each component (e.g., such as to obtain a ratio of CPP to RGEN protein disclosed above) in an aqueous medium. A suitable aqueous medium can comprise a buffer solution such as PBS or a serum-free medium such as DMEM, for example. The mixture can be incubated for about 30, 60, 90, or 120 minutes at a temperature of about 4 to 45 °C, for example, to allow formation of a non-covalent RGEN protein-CPP complex. A suitable volume (e.g., a minimum volume that adequately covers/immerses cells being treated) of this solution comprising the complex can be applied to a cell in a cell type-appropriate manner. In embodiments in which an RNA component is associated with an RGEN protein component, such formation of an RGEN can comprise adding an RNA component before, at the same time of, or after incubating a CPP with the RGEN protein component.

A composition comprising a non-covalent complex of an RGEN protein component and at least one CPP can be used with any cell type disclosed herein. Optionally, however, this composition can be used with non-mammalian cells such as yeast, fungi, and plants, but excludes use on mammalian cells.

An RGEN protein-CPP complex, as it may exist in a composition before application to cells can be at least about 30%, 40%, 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% pure, for example. Such purity can be on a protein basis in certain embodiments. As an example, if the purity of a complex is at least 80%, this would mean that at least 80% of all the protein in a composition is constituted by the complex. Complex purity alternatively can take into account not only purity on a protein basis, but also in account of other biomolecules (e.g., lipids, saccharides, and/or nucleic acids). As an example, if the purity of a complex is at least 80%, this could mean that at least 80% of all the biomolecules in the composition herein is constituted by the complex. In certain embodiments, compounds such as carbohydrates, salts, and/or lipids and the like do not affect the determination of percent purity herein. All these disclosures regarding purity can also apply to an RGEN-CPP complex (i.e., RGEN protein component of complex is associated with an RNA component).

A composition herein is preferably aqueous, wherein the solvent in which an RGEN protein-CPP complex or RGEN-CPP complex is dissolved is at least about 70, 75, 80, 85, 90, 95, 98, or 99 wt% water. The concentration of a complex in a composition can be at least about 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, 7.0, 8.0, 9.0 or 10.0 μM, or about 0.5 to 5.0 μM, 0.5 to 2.5 μM, 1.0 to 5.0 μM, 1.0 to 2.5 μM, or 2.5 to 5.0 μM, for example. It would be understood that such compositions can be in a liquid state.

The pH of a composition in certain embodiments can be between about 4.0 to about 10.0. Alternatively, the pH can be about 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5 or 10.0. pH can be adjusted or controlled by the addition or incorporation of a suitable buffer, including but not limited to: HEPES, phosphate (e.g., PBS), Tris, Tris-HCl, citrate, or a combination thereof. Buffer concentration in a composition herein can be from 0 mM to about 100 mM, or about 10, 20, or 50 mM, for example. A HEPES buffer (e.g., ~25 mM HEPES, such as 25 mM HEPES/KOH pH 7.5, 200 mM KCl, 20% glycerol, 1 mM DTT) can be used in certain aspects.

A composition herein can optionally comprise other components in addition to an RGEN protein-CPP complex or RGEN-CPP complex. For example, the composition can comprise one or more salts such as a sodium salt (e.g., NaCl, Na_2SO_4). Other non-limiting examples of salts include those having (i) an aluminum, ammonium, barium, calcium, chromium (II or III), copper (I or II), iron (II or III), hydrogen, lead (II), lithium, magnesium, manganese (II or III), mercury (I or II), potassium, silver, sodium strontium, tin (II or IV), or zinc cation, and (ii) an acetate, borate, bromate, bromide, carbonate, chlorate, chloride, chlorite, chromate, dichromate, dihydrogen phosphate, ferricyanide, ferrocyanide, fluoride, hydrogen carbonate, hydrogen phosphate, hydrogen sulfate, hydrogen sulfide, hydrogen sulfite, hydride, hydroxide, hypochlorite, iodate, iodide, nitrate, nitride, nitrite, oxalate, oxide, perchlorate, permanganate, peroxide, phosphate, phosphide, phosphite, silicate, stannate, stannite, sulfate, sulfide, sulfite, tartrate, or thiocyanate anion. Thus, any salt having a cation from (i) above and an anion from (ii) above can be in a composition herein, for example. A salt can be present at a wt% of about .01 to about 10.00 (or any hundredth increment between .01 and 10.00), for example.

An RGEN protein-CPP complex herein can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a cell. In those embodiments in which an RGEN protein component is associated with an RNA component (thereby constituting a complete RGEN), an RGEN-CPP complex similarly has this cell membrane/cell wall-traversing ability. Either an RGEN protein-CPP complex or an RGEN-CPP complex can traverse a cell wall and cell membrane in certain aspects herein.

An RGEN protein-CPP or RGEN-CPP complex herein can optionally traverse a cell wall that comprises a glycocalyx (capsule). These embodiments typically are with regard to prokaryotic cells (e.g., bacteria), some of which may have a glycocalyx depending on species type and growth conditions.

Though not intending to be held to any particular theory or mechanism, it is believed that a CPP herein may deliver an RGEN protein component into a cell via an endocytic process. Examples of such a process might include macropinocytosis, clathrin-mediated endocytosis, caveolae/lipid raft-mediated endocytosis, and/or receptor mediated endocytosis mechanisms (e.g., scavenger receptor-mediated uptake, proteoglycan-mediated uptake).

Once an RGEN protein-CPP or RGEN-CPP complex is inside a cell, it can traverse an organelle membrane such as a nuclear membrane or mitochondrial membrane, for example. This ability depends on, in certain embodiments, the presence of at least one organelle-targeting sequence (e.g., NLS, MTS) being included with the RGEN protein. Still, in other embodiments, the ability to traverse an organelle membrane such as a nuclear membrane or mitochondrial membrane does not depend on the presence of an organelle-targeting sequence (i.e., a CPP[s] in such embodiments may be responsible for allowing RGEN traversal into an organelle such as the nucleus or mitochondria).

A cell herein can be a mammalian cell or a non-mammalian cell, the latter of which is used in certain preferred embodiments. In certain other aspects, a cell herein can be as it exists (i) in an organism/tissue *in vivo*, (ii) in a tissue or group of cells *ex vivo*, or (iii) in an *in vitro* state.

A microbial cell herein can be as it exists in an isolated state (e.g., *in vitro* cells, cultured cells) or a non-isolated state.

A microbial cell in certain embodiments is a fungal cell such as a yeast cell. A yeast in certain aspects herein can be one that reproduces asexually (anamorphic) or sexually (teleomorphic). While yeast herein typically exist in unicellular form, certain types of these yeast may optionally be able to form pseudohyphae (strings of connected budding cells). In still further aspects, a yeast may be haploid or diploid, and/or may have the ability to exist in either of these ploidy forms.

Examples of yeast herein include conventional yeast and non-conventional yeast. Conventional yeast in certain embodiments are yeast that favor homologous recombination (HR) DNA repair processes over repair processes mediated by non-homologous end-joining (NHEJ). Examples of conventional yeast herein include species of the genera *Saccharomyces* (e.g., *S. cerevisiae*, which is also known as budding yeast, baker's yeast, and/or brewer's yeast; *S. bayanus; S. boulardii; S. bulderi; S. cariocanus; S. cariocus; S. chevalieri; S. dairenensis; S. ellipsoideus; S. eubayanus; S. exiguus; S. florentinus; S. kluyveri; S. martiniae; S. monacensis; S. norbensis; S. paradoxus; S. pastorianus; S. spencerorum; S. turicensis; S. unisporus; S. uvarum; S. zonatus*) and *Schizosaccharomyces* (e.g., *S. pombe*, which is also known as fission yeast; *S. cryophilus; S. japonicus; S. octosporus*).

A non-conventional yeast herein is not a conventional yeast such as a *Saccharomyces* (e.g., *S. cerevisiae*) or *Schizosaccharomyces* (e.g., *S. pombe*) species. Non-conventional yeast in certain embodiments can be yeast that favor NHEJ DNA repair processes over repair processes mediated by HR. Conventional yeasts such as *S. cerevisiae* and *S. pombe* typically exhibit specific integration of donor DNA with short flanking homology arms (30-50 bp) with efficiencies routinely over 70%, whereas non-conventional yeasts such as *Pichia pastoris*, *Pichia stipitis, Hansenula polymorpha*, *Yarrowia lipolytica* and *Kluyveromyces lactis* usually show specific integration with similarly structured donor DNA at efficiencies of less than 1% (Chen et al., *PLoS ONE* 8:e57952). Thus, a preference for HR processes can be gauged, for example, by transforming yeast with a suitable donor DNA and determining the degree to which it is specifically recombined with a genomic site predicted to be targeted by the donor DNA. A preference for NHEJ (or low preference for HR), for example, would be manifest if such an assay yielded a high degree of random integration of the donor DNA in the yeast genome. Assays for

determining the rate of specific (HR-mediated) and/or random (NHEJ-mediated) integration of DNA in yeast are known in the art (e.g., Ferreira and Cooper, *Genes Dev.* 18:2249-2254; Corrigan et al., *PLoS ONE* 8:e69628; Weaver et al., *Proc. Natl. Acad. Sci. U.S.A.* 78:6354-6358; Keeney and Boeke, *Genetics* 136:849-856).

5 Given their low level of HR activity, non-conventional yeast herein can (i) exhibit a rate of specific targeting by a suitable donor DNA having 30-50 bp flanking homology arms of less than about 1%, 2%, 3%, 4%, 5%,, 6%, 7%, or 8%, for example, and/or (ii) exhibit a rate of random integration of the foregoing donor DNA of more than about 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, or 10 75%, for example. These rates of (i) specific targeting and/or (ii) random integration of a suitable donor DNA can characterize a non-conventional yeast as it exists before being provided an RGEN as disclosed herein. An aim for providing an RGEN to a non-conventional yeast in certain embodiments is to create site-specific DNA single-strand breaks (SSB) or double-strand breaks (DSB) for biasing the 15 yeast toward HR at the specific site. Thus, providing a suitable RGEN in a non-conventional yeast typically should allow the yeast to exhibit an increased rate of HR with a particular donor DNA. Such an increased rate can be at least about 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, or 10-fold higher than the rate of HR in a suitable control (e.g., same non-conventional yeast transformed with the same donor DNA, but lacking a 20 suitable RGEN).

A non-conventional yeast herein can be cultivated following any means known in the art, such as described in Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology: Practical Protocols (K. Wolf, K.D. Breunig, G. Barth, Eds., Springer-Verlag, Berlin, Germany, 2003), Yeasts in Natural and Artificial 25 Habitats (J.F.T. Spencer, D.M. Spencer, Eds., Springer-Verlag, Berlin, Germany, 1997), and/or Yeast Biotechnology: Diversity and Applications (T. Satyanarayana, G. Kunze, Eds., Springer, 2009), all of which are incorporated herein by reference.

Non-limiting examples of non-conventional yeast herein include yeasts of the following genera: *Yarrowia*, *Pichia*, *Schwanniomyces, Kluyveromyces, Arxula,* 30 *Trichosporon, Candida, Ustilago, Torulopsis, Zygosaccharomyces, Trigonopsis, Cryptococcus, Rhodotorula, Phaffia, Sporobolomyces, Pachysolen,* and *Moniliella*. A suitable example of a *Yarrowia* species is *Y. lipolytica*. Suitable examples of *Pichia* species include *P. pastoris*, *P. methanolica*, *P. stipitis*, *P. anomala* and *P.*

angusta. Suitable examples of *Schwanniomyces* species include *S. castellii*, *S. alluvius*, *S. hominis*, *S. occidentalis*, S. capriottii, *S. etchellsii*, *S. polymorphus*, *S. pseudopolymorphus*, *S. vanrijiae* and *S. yamadae*. Suitable examples of *Kluyveromyces* species include *K. lactis*, *K. marxianus*, *K. fragilis*, *K. drosophilarum*,
5 *K. thermotolerans*, *K. phaseolosporus*, *K. vanudenii*, *K. waltii*, *K. africanus* and *K. polysporus*. Suitable examples of *Arxula* species include *A. adeninivorans* and *A. terrestre*. Suitable examples of *Trichosporon* species include *T. cutaneum*, *T. capitatum*, *T. inkin* and *T. beemeri*. Suitable examples of *Candida* species include *C. albicans, C. ascalaphidarum, C. amphixiae, C. antarctica, C. apicola, C.*
10 *argentea, C. atlantica, C. atmosphaerica, C. blattae, C. bromeliacearum, C. carpophila, C. carvajalis, C. cerambycidarum, C. chauliodes, C. corydali, C. dosseyi, C. dubliniensis, C. ergatensis, C. fructus, C. glabrata, C. fermentati, C. guilliermondii, C. haemulonii, C. insectamens, C. insectorum, C. intermedia, C. jeffresii, C. kefyr, C. keroseneae, C. krusei, C. lusitaniae, C. lyxosophila, C. maltosa,*
15 *C. marina, C. membranifaciens, C. milleri, C. mogii, C. oleophila, C. oregonensis, C. parapsilosis, C. quercitrusa, C. rugosa, C. sake, C. shehatea, C. temnochilae, C. tenuis, C. theae, C. tolerans, C. tropicalis, C. tsuchiyae, C. sinolaborantium, C. sojae, C. subhashii, C. viswanathii, C. utilis, C. ubatubensis* and *C. zemplinina*. Suitable examples of *Ustilago* species include *U. avenae*, *U. esculenta*, *U. hordei*,
20 *U. maydis*, *U. nuda* and *U. tritici*. Suitable examples of *Torulopsis* species include *T. geochares*, *T. azyma*, *T. glabrata* and *T. candida*. Suitable examples of *Zygosaccharomyces* species include *Z. bailii, Z. bisporus, Z. cidri, Z. fermentati, Z. florentinus, Z. kombuchaensis, Z. lentus, Z. mellis, Z. microellipsoides, Z. mrakii, Z. pseudorouxii* and *Z. rouxii*. Suitable examples of *Trigonopsis* species include *T.*
25 *variabilis*. Suitable examples of *Cryptococcus* species include *C. laurentii, C. albidus, C. neoformans, C. gattii, C. uniguttulatus, C. adeliensis, C. aerius, C. albidosimilis, C. antarcticus, C. aquaticus, C. ater, C. bhutanensis, C. consortionis, C. curvatus, C. phenolicus, C. skinneri, C. terreus* and *C. vishniacci*. Suitable examples of *Rhodotorula* species include *R. acheniorum, R. tula, R. acuta, R.*
30 *americana, R. araucariae, R. arctica, R. armeniaca, R. aurantiaca, R. auriculariae, R. bacarum, R. benthica, R. biourgei, R. bogoriensis, R. bronchialis, R. buffonii, R. calyptogenae, R. chungnamensis, R. cladiensis, R. corallina, R. cresolica, R. crocea, R. cycloclastica, R. dairenensis, R. diffluens, R. evergladiensis, R. ferulica,*

R. foliorum, R. fragaria, R. fujisanensis, R. futronensis, R. gelatinosa, R. glacialis, R. glutinis, R. gracilis, R. graminis, R. grinbergsii, R. himalayensis, R. hinnulea, R. histolytica, R. hylophila, R. incarnata, R. ingeniosa, R. javanica, R. koishikawensis, R. lactosa, R. lamellibrachiae, R. laryngis, R. lignophila, R. lini, R. longissima, R.
5 *ludwigii, R. lysinophila, R. marina, R. martyniae-fragantis, R. matritensis, R. meli, R. minuta, R. mucilaginosa, R. nitens, R. nothofagi, R. oryzae, R. pacifica, R. pallida, R. peneaus, R. philyla, R. phylloplana, R. pilatii, R. pilimanae, R. pinicola, R. plicata, R. polymorpha, R. psychrophenolica, R. psychrophila, R. pustula, R. retinophila, R. rosacea, R. rosulata, R. rubefaciens, R. rubella, R. rubescens, R. rubra, R.*
10 *rubrorugosa, R. rufula, R. rutila, R. sanguinea, R. sanniei, R. sartoryi, R. silvestris, R. simplex, R. sinensis, R. slooffiae, R. sonckii, R. straminea, R. subericola, R. suganii, R. taiwanensis, R. taiwaniana, R. terpenoidalis, R. terrea, R. texensis, R. tokyoensis, R. ulzamae, R. vanillica, R. vuilleminii, R. yarrowii, R. yunnanensis* and *R. zsoltii*. Suitable examples of *Phaffia* species include *P. rhodozyma*. Suitable
15 examples of *Sporobolomyces* species include *S. alborubescens, S. bannaensis, S. beijingensis, S. bischofiae, S. clavatus, S. coprosmae, S. coprosmicola, S. corallinus, S. dimmenae, S. dracophylli, S. elongatus, S. gracilis, S. inositophilus, S. johnsonii, S. koalae, S. magnisporus, S. novozealandicus, S. odorus, S. patagonicus, S. productus, S. roseus, S. sasicola, S. shibatanus, S. singularis, S.*
20 *subbrunneus, S. symmetricus, S. syzygii, S. taupoensis, S. tsugae, S. xanthus* and *S. yunnanensis*. Suitable examples of *Pachysolen* and *Moniliella* species include *P. tannophilus* and *M. pollinis*, respectively. Still other examples of non-conventional yeasts herein include *Pseudozyma* species (e.g., *S. antarctica*), *Thodotorula* species (e.g., *T. bogoriensis*), *Wickerhamiella* species (e.g., *W. domercqiae*), and
25 *Starmerella* species (e.g., *S. bombicola*).

Yarrowia lipolytica is preferred in certain embodiments disclosed herein. Examples of suitable *Y. lipolytica* include the following isolates available from the American Type Culture Collection (ATCC, Manassas, VA): strain designations ATCC #20362, #8862, #8661, #8662, #9773, #15586, #16617, #16618, #18942,
30 #18943, #18944, #18945, #20114, #20177, #20182, #20225, #20226, #20228, #20327, #20255, #20287, #20297, #20315, #20320, #20324, #20336, #20341, #20346, #20348, #20363, #20364, #20372, #20373, #20383, #20390, #20400, #20460, #20461, #20462, #20496, #20510, #20628, #20688, #20774, #20775,

#20776, #20777, #20778, #20779, #20780, #20781, #20794, #20795, #20875, #20241, #20422, #20423, #32338, #32339, #32340, #32341, #34342, #32343, #32935, #34017, #34018, #34088, #34922, #34922, #38295, #42281, #44601, #46025, #46026, #46027, #46028, #46067, #46068, #46069, #46070, #46330, #46482, #46483, #46484, #46436, #60594, #62385, #64042, #74234, #76598, #76861, #76862, #76982, #90716, #90811, #90812, #90813, #90814, #90903, #90904, #90905, #96028, #201241, #201242, #201243, #201244, #201245, #201246, #201247, #201249, and/or #201847.

A fungal cell herein can be a yeast (e.g., as described above) or of any other fungal type such as a filamentous fungus. For instance, a fungus herein can be a Basidiomycetes, Zygomycetes, Chytridiomycetes, or Ascomycetes fungus. Examples of filamentous fungi herein include those of the genera *Trichoderma, Chrysosporium, Thielavia, Neurospora* (e.g., *N. crassa, N. sitophila*), *Cryphonectria* (e.g., *C. parasitica*), *Aureobasidium* (e.g., *A. pullulans*), *Filibasidium, Piromyces, Cryptococcus, Acremonium, Tolypocladium, Scytalidium, Schizophyllum, Sporotrichum, Penicillium* (e.g., *P. bilaiae, P. camemberti, P. candidum, P. chrysogenum, P. expansum, P. funiculosum, P. glaucum, P. marneffei, P. roqueforti, P. verrucosum, P. viridicatum*), *Gibberella* (e.g., *G. acuminata, G. avenacea, G. baccata, G. circinata, G. cyanogena, G. fujikuroi, G. intricans, G. pulicaris, G. stilboides, G. tricincta, G. zeae*), *Myceliophthora, Mucor* (e.g., *M. rouxii, M. circinelloides*), *Aspergillus* (e.g., *A. niger, A. oryzae, A. nidulans, A. flavus, A. lentulus, A. terreus, A. clavatus, A. fumigatus*), *Fusarium* (e.g., *F. graminearum, F. oxysporum, F. bubigenum, F. solani, F. oxysporum, F. verticillioides, F. proliferatum, F. venenatum*), and *Humicola*, and anamorphs and teleomorphs thereof. The genus and species of fungi herein can be defined, if desired, by morphology as disclosed in Barnett and Hunter (Illustrated Genera of Imperfect Fungi, 3rd Edition, Burgess Publishing Company, 1972). A fungus can optionally be characterized as a pest/pathogen of a plant or animal (e.g., human) in certain embodiments.

Trichoderma species in certain aspects herein include *T. aggressivum, T. amazonicum, T. asperellum, T. atroviride, T. aureoviride, T. austrokoningii, T. brevicompactum, T. candidum, T. caribbaeum, T. catoptron, T. cremeum, T. ceramicum, T. cerinum, T. chlorosporum, T. chromospermum, T. cinnamomeum, T. citrinoviride, T. crassum, T. cremeum, T. dingleyeae, T. dorotheae, T. effusum, T.*

erinaceum, T. estonicum, T. fertile, T. gelatinosus, T. ghanense, T. hamatum, T. harzianum, T. helicum, T. intricatum, T. konilangbra, T. koningii, T. koningiopsis, T. longibrachiatum, T. longipile, T. minutisporum, T. oblongisporum, T. ovalisporum, T. petersenii, T. phyllostahydis, T. piluliferum, T. pleuroticola, T. pleurotum, T. 5 *polysporum, T. pseudokoningii, T. pubescens, T. reesei, T. rogersonii, T. rossicum, T. saturnisporum, T. sinensis, T. sinuosum, T. spirale, T. stramineum, T. strigosum, T. stromaticum, T. surrotundum, T. taiwanense, T. thailandicum, T. thelephoricolum, T. theobromicola, T. tomentosum, T. velutinum, T. virens, T. viride* and *T. viridescens*. A *Trichoderma* species herein can be cultivated and/or manipulated as 10 described in Trichoderma: Biology and Applications (P.K. Mukherjee et al., Eds., CABI, Oxfordshire, UK, 2013), for example, which is incorporated herein by reference.

A microbial cell in certain embodiments is an algal cell. For example, an algal cell can be from any of the following: Chlorophyta (green algae), Rhodophyta 15 (red algae), Phaeophyceae (brown algae), Bacillariophycaeae (diatoms), and Dinoflagellata (dinoflagellates). An algal cell can be of a microalgae (e.g., phytoplankton, microphytes, or planktonic algae) or macroalgae (kelp, seaweed) in other aspects. As further examples, an algal cell herein can be a *Porphyra* (purple laver), *Palmaria* species such as *P. palmata* (dulse), *Arthrospira* species such as *A.* 20 *platensis* (*spirulina*), *Chlorella* (e.g., *C. protothecoides*), a *Chondrus* species such as *C. crispus* (Irish moss), *Aphanizomenon*, *Sargassum*, *Cochayuyo*, *Botryococcus* (e.g., *B. braunii*), *Dunaliella* (e.g., *D. tertiolecta*), *Gracilaria*, *Pleurochrysis* (e.g., *P. carterae*), *Ankistrodesmus*, *Cyclotella*, *Hantzschia*, *Nannochloris*, *Nannochloropsis*, *Nitzschia*, *Phaeodactylum* (e.g., *P. tricornutum*), *Scenedesmus*, *Stichococcus*, 25 *Tetraselmis* (e.g., *T. suecica*), *Thalassiosira* (e.g., *T. pseudonana*), *Crypthecodinium* (e.g., *C. cohnii*), *Neochloris* (e.g., *N. oleoabundans*), or *Schiochytrium*. An algal species herein can be cultivated and/or manipulated as described in Thompson (Algal Cell Culture. Encyclopedia of Life Support System (EOLSS), Biotechnology Vol 1, available at eolss.net/sample-chapters internet site), for example, which is 30 incorporated herein by reference.

In one embodiment, the method comprises a method of delivering a protein component of an RNA-guided endonuclease (RGEN) into a microbial cell, said method comprising: contacting the microbial cell with a composition comprising the

protein component of the RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP), wherein said protein component and CPP are covalently, or non-covalently, linked to each other in an RGEN protein-CPP complex, wherein said RGEN protein-CPP complex traverses (i) a cell membrane, or (ii) a cell wall
5 and cell membrane, of the cell, thereby entering the microbial cell. Microbial cells useful for the methods and composition described herein include cells selected from *Phytophtora* species such as *Phytophtora capsici* (Lamour et al. 2012. The oomycete broad-host-range pathogen *Phytophthora capsici*. Mol. Plant Pathol. May 13(4) : 329-337), *Zymoseptoria* species such as *Septoria tritici (*Testa et al. 2015.
10 Overview of genomic and bioinformatics resources for Zymoseptoria tritici. Fungal Genet. Biol. Jun. 79:13-16) and *Botrytis* species *such as Botrytis cinerea* (Hahn M. 2014. The rising threat of fungicide resistance in plant pathogenic fungi: Botrytis as a case study. J. Chem. Biol 7:133-141).

A protist cell herein can be selected from the class Ciliata (e.g., the genera
15 *Tetrahymena, Paramecium, Colpidium, Colpoda, Glaucoma, Platyophrya, Vorticella, Potomacus, Pseudocohnilembus, Euplotes, Engelmaniella*, and *Stylonichia*), the subphylum Mastigophora (flagellates), the class Phytomastigophorea (e.g., the genera *Euglena, Astasia, Haematococcus*, and *Crypthecodinium*), the class Zoomastigophorea, the superclass Rhizopoda, the class Lobosea (e.g., the genus
20 *Amoeba*), and the class Eumycetozoea (e.g., the genera *Dictyostelium* and *Physarum*), for example. Certain protist species herein can be cultivated and/or manipulated as described in ATCC® Protistology Culture Guide: tips and techniques for propagating protozoa and algae (2013, available at American Type Culture Collection internet site), for example, which is incorporated herein by reference. A
25 protist can optionally be characterized as a pest/pathogen of a plant or animal (e.g., human) in certain embodiments.

A bacterial cell in certain embodiments can be those in the form of cocci, bacilli, spirochetes, spheroplasts, protoplasts, etc. Other non-limiting examples of bacteria include those that are Gram-negative and Gram-positive. Still other non-
30 limiting examples of bacteria include those of the genera *Salmonella* (e.g., *S. typhi, S. enteritidis*), *Shigella* (e.g., *S. dysenteriae*), *Escherichia* (e.g., *E. coli*), *Enterobacter, Serratia, Proteus, Yersinia, Citrobacter, Edwardsiella, Providencia, Klebsiella, Hafnia, Ewingella, Kluyvera, Morganella, Planococcus, Stomatococcus,*

Micrococcus, Staphylococcus (e.g., *S. aureus, S. epidermidis*), *Vibrio* (e.g., *V. cholerae*), *Aeromonas, Plessiomonas, Haemophilus* (e.g., *H. influenzae*), *Actinobacillus, Pasteurella, Mycoplasma* (e.g., *M. pneumonia*), *Ureaplasma, Rickettsia, Coxiella, Rochalimaea, Ehrlichia, Streptococcus* (e.g., *S. pyogenes, S. mutans, S. pneumoniae*), *Enterococcus* (e.g., *E. faecalis*), *Aerococcus, Gemella, Lactococcus* (e.g., *L. lactis*), *Leuconostoc* (e.g., *L. mesenteroides*), *Pedicoccus, Bacillus* (e.g., *B. cereus, B. subtilis, B. thuringiensis*), *Corynebacterium* (e.g., *C. diphtheriae*), *Arcanobacterium, Actinomyces, Rhodococcus, Listeria* (e.g., *L. monocytogenes*), *Erysipelothrix, Gardnerella, Neisseria* (e.g., *N. meningitidis, N. gonorrhoeae*), *Campylobacter, Arcobacter, Wolinella, Helicobacter* (e.g., *H. pylori*), *Achromobacter, Acinetobacter, Agrobacterium* (e.g., *A. tumefaciens*), *Alcaligenes, Chryseomonas, Comamonas, Eikenella, Flavimonas, Flavobacterium, Moraxella, Oligella, Pseudomonas* (e.g., *P. aeruginosa*), *Shewanella, Weeksella, Xanthomonas, Bordetella, Franciesella, Brucella, Legionella, Afipia, Bartonella, Calymmatobacterium, Cardiobacterium, Streptobacillus, Spirillum, Peptostreptococcus, Peptococcus, Sarcinia, Coprococcus, Ruminococcus, Propionibacterium, Mobiluncus, Bifidobacterium, Eubacterium, Lactobacillus* (e.g., *L. lactis, L. acidophilus*), *Rothia, Clostridium* (e.g., *C. botulinum, C. perfringens*), *Bacteroides, Porphyromonas, Prevotella, Fusobacterium, Bilophila, Leptotrichia, Wolinella, Acidaminococcus, Megasphaera, Veilonella, Norcardia, Actinomadura, Norcardiopsis, Streptomyces, Micropolysporas, Thermoactinomycetes, Mycobacterium* (e.g., *M. tuberculosis, M. bovis, M. leprae*), *Treponema, Borrelia* (e.g., *B. burgdorferi*), *Leptospira, and Chlamydiae*. A bacteria can optionally be characterized as a pest/pathogen of a plant or animal (e.g., human) in certain embodiments. Bacteria can be comprised in a mixed microbial population (e.g., containing other bacteria, or containing yeast and/or other bacteria) in certain embodiments.

An archaeal cell in certain embodiments can be from any Archaeal phylum, such as Euryarchaeota, Crenarchaeota, Nanoarchaeota, Korarchaeota, Aigarchaeota, or Thaumarchaeota. Archaeal cells herein can be extremophilic (e.g., able to grow and/or thrive in physically or geochemically extreme conditions that are detrimental to most life), for example. Some examples of extremophilic archaea include those that are thermophilic (e.g., can grow at temperatures between 45-122

°C), hyperthermophilic (e.g., can grow at temperatures between 80-122 °C), acidophilic (e.g., can grow at pH levels of 3 or below), alkaliphilic (e.g., can grow at pH levels of 9 or above), and/or halophilic (e.g., can grow in high salt concentrations [e.g., 20-30% NaCl]). Examples of archaeal species include those of the genera *Halobacterium* (e.g., *H. volcanii*), *Sulfolobus* (e.g., *S. solfataricus*, *S. acidocaldarius*), *Thermococcus* (e.g., *T. alcaliphilus*, *T. celer*, *T. chitonophagus*, *T. gammatolerans*, *T. hydrothermalis*, *T. kodakarensis*, *T. litoralis*, *T. peptonophilus*, *T. profundus*, *T. stetteri*), *Methanocaldococcus* (e.g., *M. thermolithotrophicus*, *M. jannaschii*), *Methanococcus* (e.g., *M. maripaludis*), *Methanothermobacter* (e.g., *M. marburgensis*, *M. thermautotrophicus*), *Archaeoglobus* (e.g., *A. fulgidus*), *Nitrosopumilus* (e.g., *N. maritimus*), *Metallosphaera* (e.g., *M. sedula*), *Ferroplasma*, *Thermoplasma*, *Methanobrevibacter* (e.g., *M. smithii*), and *Methanosphaera* (e.g., *M. stadtmanae*).

Examples of insect cells herein include *Spodoptera frugiperda* cells, *Trichoplusia ni* cells, *Bombyx mori* cells and the like. *S. frugiperda* cells include Sf9 and Sf21, for instance. *T. ni* ovary cells include HIGH FIVE cells (alias BTI-TN-5B1-4, manufactured by Invitrogen), for example. B. mori cells include N4, for example. Certain insect cells herein can be cultivated and/or manipulated as described in Growth and Maintenance of Insect cell lines (2010, Invitrogen, Manual part no. 25-0127, MAN0000030), for example, which is incorporated herein by reference. In other aspects, an insect cell can be a cell of a plant pest/pathogen such as an armyworm, black cutworm, corn earworm, corn flea beetle, corn leaf aphid, corn root aphid, European corn borer, fall armyworm, granulate cutworm, Japanese beetle, lesser cornstalk borer, maize billbug, melanotus communis, seedcorn maggot, sod webworms, sorghum midge, sorghum webworm, southern corn billbug, southern corn rootworm, southern cornstalk borer, southern potato wireworm, spider mite, stalk borer, sugarcane beetle, tobacco wireworm, white grub, aphid, boll weevil, bollworm complex, cabbage looper , tarnished plant bug, thrip, two spotted spider mite, yellow striped armyworm, alfalfa weevil, clover leaf weevil, clover root curculio, fall armyworm, grasshopper, meadow spittlebug, pea aphid, potato leafhopper, sod webworm, variegated cutworm, lesser cornstalk borer, tobacco thrip, wireworm, cereal leaf beetle, chinch bug, English grain aphid, greenbug, hessian fly, bean leaf beetle, beet armyworm, blister beetle, grape colaspis, green cloverworm, Mexican

bean beetle, soybean looper, soybean stem borer, stink bug, three-cornered alfalfa hopper, velvetbean caterpillar, budworm, cabbage looper, cutworm, green june beetle, green peach aphid, hornworm, potato tuberworm, southern mole cricket, suckfly, tobacco flea beetle, vegetable weevil, or whitefringed beetle. Alternatively, 5 an insect cell can be a cell of a pest/pathogen of an animal (e.g., human).

A nematode cell, for example, can be of a nematode from any of the following genera: *Meloidogyne* (root-knot nematode), *Pratylenchus* (lesion nematode), *Heterodera* (cyst nematode), *Globodera* (cyst nematode), *Ditylenchus* (stem and bulb nematode), *Tylenchulus* (citrus nematode), *Xiphinema* (dagger nematode), 10 *Radopholus* (burrowing nematode), *Rotylenchulus* (reniform nematode), *Helicotylenchus* (spiral nematode), or *Belonolaimus* (sting nematode). A nematode can optionally be characterized as a pest/pathogen of a plant or animal (e.g., human) in certain embodiments. A nematode can be *C. elegans* in other aspects.

A fish cell herein can be any of those as disclosed in U.S. Patent Nos. 15 7408095 and 7217564, and Tissue Culture of Fish Cell Lines (T. Ott, NWFHS Laboratory Procedures Manual – Second Edition, Chapter 10, 2004), for example, which are incorporated herein by reference. These references also disclose information regarding cultivating and/or manipulating fish cells. Non-limiting examples of fish cells can be from a teleost such as zebrafish, medaka, Giant rerio, 20 or puffer fish.

A plant cell herein can be, for example, a monocot plant cell or dicot plant cell. Examples of monocot plants herein include corn (*Zea mays*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet, *Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail 25 millet (*Setaria italica*), finger millet (*Eleusine coracana*)), wheat (*Triticum aestivum*), sugarcane (*Saccharum spp.*), oats (*Avena*), barley (*Hordeum*), switchgrass (*Panicum virgatum*), pineapple (*Ananas comosus*), banana (*Musa spp.*), palm, ornamentals, and turfgrasses. Examples of dicot plants herein include soybean (*Glycine max*), canola (*Brassica napus* and *B. campestris*), alfalfa (*Medicago* 30 *sativa*), tobacco (*Nicotiana tabacum*), *Arabidopsis* (*A. thaliana*), sunflower (*Helianthus annuus*), cotton (*Gossypium arboreum*), peanut (*Arachis hypogaea*), tomato (*Solanum lycopersicum*), and potato (*Solanum tuberosum*). A plant cell may be from any part of a plant and/or from any stage of plant development.

Plant cells herein may be grown or regenerated into plants using conventional conditions, see for example, McCormick *et al.*, (1986) *Plant Cell Rep* 5:81-4. Regenerated plants may then be grown, and either pollinated with the same strain or with a different strain, and resulting progeny having the desired characteristic (e.g., alteration) and/or comprising an introduced polynucleotide or polypeptide identified. Two or more generations may be grown to ensure that an alteration is stably maintained and inherited, and seeds harvested.

Mammalian cells in certain embodiments can be human, non-human primate (e.g., monkey, ape), rodent (e.g., mouse, rat, hamster, guinea pig), rabbit, dog, cat, cow, pig, horse, goat, or sheep cells. Other examples of mammalian cells herein include primary epithelial cells (e.g., keratinocytes, cervical epithelial cells, bronchial epithelial cells, tracheal epithelial cells, kidney epithelial cells, retinal epithelial cells); established cell lines (e.g., 293 embryonic kidney cells, HeLa cervical epithelial cells, PER-C6 retinal cells, MDBK, CRFK, MDCK, CHO, BeWo, Chang cells, Detroit 562, Hep-2, KB, LS 180, LS 174T, NCI-H-548, RPMI 2650, SW-13, T24, WI-28 VA13, 2RA, WISH, BS-C-I, LLC-MK2, Clone M-3, RAG, TCMK-1, LLC-PK1, PK-15, GH1, GH3, L2, LLC-RC 256, MH1C1, XC, MDOK, VSW, TH-I, B1 cells); any epithelial, mesenchymal (e.g., fibroblast), neural, or muscular cell from any tissue or organ (e.g., skin, heart; liver; kidney; colon; intestine; esophagus; stomach; neural tissue such as brain or spinal cord; lung; vascular tissue; lymphoid tissue such as lymph gland, adenoid, tonsil, bone marrow, or blood; spleen); and fibroblast or fibroblast-like cell lines (e.g., TRG-2, IMR-33, Don cells, GHK-21, citrullinemia cells, Dempsey cells, Detroit 551, Detroit 510, Detroit 525, Detroit 529, Detroit 532, Detroit 539, Detroit 548, Detroit 573, HEL 299, IMR-90, MRC-5, WI-38, WI-26, MiCl1, CV-1, COS-1, COS-3, COS-7, Vero, DBS-FrhL-2, BALB/3T3, F9, SV-T2, M-MSV-BALB/3T3, K-BALB, BLO-11, NOR-10, C3H/IOTI/2, HSDM1C3, KLN205, McCoy cells, Mouse L cells, SCC-PSA1, Swiss/3T3 cells, Indian muntjac cells, SIRC, Jensen cells). Methods of culturing and manipulating mammalian cells lines are known in the art.

In certain embodiments, a microbial cell can be of any pathogen and/or pest of an animal or plant. Examples of such pathogens/pests include various types of bacteria, fungi, yeast, protists, nematodes, and insects. Those skilled in the art would recognize examples of such pathogens/pests disclosed above.

As described herein (see Example 10), cell-penetrating peptides were able to deliver cargo to different eukaryotic species including *Phytophthora capsici*, *Septoria tritici*, and *Botrytis cinerea*.

In one embodiment, the method described herein is a method of delivering a protein component of an RNA-guided endonuclease (RGEN) into a microbial cell selected from the group consisting of *Phytophthora capsici*, *Septoria tritici*, and *Botrytis cinerea*, said method comprising: contacting the microbial cell with a composition comprising the protein component of the RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP), wherein said protein component and CPP are covalently, or non-covalently, linked to each other in an RGEN protein-CPP complex, wherein said RGEN protein-CPP complex traverses (i) a cell membrane, or (ii) a cell wall and cell membrane, of the cell, thereby entering the microbial cell.

A composition in certain embodiments herein can comprise at least one protein component of a guide polynucleotide/ Cas endonuclease complex and at least one cell-penetrating peptide (CPP), wherein the protein component and CPP are covalently, or non-covalently, linked to each other in a polynucleotide/endonuclease protein-CPP complex, and wherein the polynucleotide/endonuclease protein-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a cell (such as a microbial cell). The guide polynucleotide and Cas endonuclease are capable of forming a complex, referred to as a "guide polynucleotide/Cas endonuclease complex", that enables the Cas endonuclease to introduce a double-strand break at a DNA target site.

The disclosed invention also concerns a method of delivering a protein component of an RNA-guided endonuclease (RGEN) into a cell (such as a microbial cell). This method comprises contacting a cell with a composition comprising the RGEN protein component and at least one cell-penetrating peptide (CPP), wherein the RGEN protein component and CPP are covalently, or non-covalently, linked to each other in an RGEN protein-CPP complex. As a result of this contacting step, the RGEN protein-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of the cell, and thereby gain entry to the cell. In certain embodiments in which an RGEN protein component is associated with an RNA component (thereby forming an RGEN), the disclosed method is directed to

delivering an RGEN-CPP complex into a cell. Additionally, since an RGEN can be used in RGEN-mediated DNA targeting in certain embodiments, this method can optionally be characterized as a method of targeting DNA in a cell.

This method can be practiced using any of the above-disclosed embodiments or below Examples regarding each of the method features (e.g., cell type, RGEN protein component, CPP, organelle-targeting sequence, etc.), for example. Thus, any of the features disclosed above or in the Examples, or any combination of these features, can be used appropriately to characterize embodiments of a delivery method herein. The following delivery method features are examples.

Embodiments of a delivery method herein comprise contacting a cell (such as a microbial cell) with a composition comprising an RGEN protein-CPP complex. It is believed that such contacting results in interaction of the complex with the outer surface of the cell (e.g., cell membrane, cell wall), thereby allowing the CPP component of the complex to initiate traversal of the complex across (i) a cell membrane, or (ii) a cell wall and cell membrane.

Contacting a composition comprising an RGEN protein-CPP complex with a cell (such as a microbial cell) can be done at a temperature that allows the complex to enter the cell. Such contacting can be done at any temperature between about 4 and 45 °C, for example. The contacting temperature can be about 4, 15, 20, 30, 37, or 42 °C in non-limiting embodiments. The same temperature or temperature range can be maintained during the contacting step, or modified appropriately (e.g., two or more different temperatures).

Contacting a composition comprising an RGEN protein-CPP complex with a cell can be done for an amount of time that is adequate for allowing the complex to enter the cell. For example, cells can be incubated with an RGEN protein-CPP complex for at least about 15, 30, 45, 60, 75, 90, 105, 120, 135, 150, 165, 180, 240, 300, 360, 420, 480, 540, 600, 660, or 720 minutes.

The milieu (e.g., buffer, water and salt concentrations, pH, purity of RGEN protein-CPP complex) in which the contacting is performed may be any of those conditions disclosed above regarding a composition comprising an RGEN protein-CPP complex. For example, cells can be incubated with a complex in a HEPES buffer (e.g., ~25 mM HEPES, such as 25 mM HEPES/KOH pH 7.5, 200 mM KCl, 20% glycerol, 1 mM DTT) or PBS (e.g., 1X PBS, pH 7).

One or more cells (such as microbial cells) may be contacted with a composition comprising an RGEN protein-CPP complex. A cell herein may be as it exists (i) in an organism/tissue *in vivo*, (ii) in a tissue or group of cells *ex vivo*, or (iii) in an *in vitro* state (e.g., cultured cells).

5 Entry of an RGEN protein-CPP complex into a cell herein typically refers to when a complex has completely traversed (i) a cell membrane, or (ii) a cell wall and cell membrane, and is comprised within at least the cell cytoplasm. Though not intending to be held to any particular theory or mechanism, it is believed that an RGEN protein-CPP complex held together by non-covalent linkage either remains in 10 a complete or partial complex, or the RGEN protein component separates from the CPP component(s) of the complex, after the RGEN protein-CPP complex gains cell entry. In either case, the RGEN protein component is able to associate with a suitable RNA component herein; such association can occur in the cytoplasm, nucleus, or mitochondria, for example. This capability likewise applies to an RGEN 15 protein-CPP complex held together by covalent linkage.

In certain embodiments of an RGEN protein delivery method, a composition herein further comprises at least one RNA component that is associated with the RGEN protein component of the RGEN protein-CPP complex (i.e., the composition comprises an RGEN-CPP complex). The RNA component in this embodiment can 20 be as disclosed herein, comprising a sequence complementary to a target site sequence on a chromosome or episome in the microbial cell. The RGEN can bind to the target site sequence, and optionally cleave one or both DNA strands at the target site sequence. Such an embodiment can also be characterized as a method of delivering an RGEN-CPP complex into a microbial cell, or alternatively as a 25 method of delivering an RNA into a microbial cell.

An RNA component (e.g., gRNA) for use in this embodiment can be prepared using any number of means known in the art. For example, an *in vitro* transcription process can be used to prepare an RNA component herein. Bacterial RNA polymerases (e.g., T7, T3, SP6) can be used to transcribe an RNA component from 30 a suitable DNA construct encoding the RNA component in certain non-limiting embodiments. An RNA component may be processed to at least about 70%, 80%, 90%, or 95% purity with respect to other biomolecules (e.g., protein, saccharides, lipids), if desired.

To prepare a composition comprising an RNA component and an RGEN protein-CPP complex, the RNA component can be dissolved in a composition in which an RGEN protein-CPP complex is already dissolved, or vice versa (or these components can be dissolved at the same time). A molar ratio of RNA component to RGEN protein-CPP complex of at least about 0.5:1, 1.0:1, 1.5:1, 2.0:1, 2.5:1, 3.0:1, 3.5:1, or 4.0:1, for example, can be used when mixing these elements together. In certain aspects, the molar ratio of RNA component to RGEN protein-CPP complex can be about 3.0:1, or can range from about 2.5:1 to 3.5:1, 2.75:1 to 3.25:1, or 2.9:1 to 3.1:1. In these and other aspects, the concentration of an RGEN protein-CPP complex with which an RNA component is mixed can be at least about 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, 7.0, 8.0, 9.0 or 10.0 μM, or about 0.5 to 5.0 μM, 0.5 to 2.5 μM, 1.0 to 5.0 μM, 1.0 to 2.5 μM, or 2.5 to 5.0 μM. The amount of time allowed for RNA association with an RGEN protein-CPP complex to form an RGEN-CPP complex can be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 45, or 60 minutes, for example. Other conditions (e.g., temperature, buffer, water and salt concentrations, pH, purity of RGEN protein-CPP complex) in which an RNA component can be associated with an RGEN protein-CPP complex may be any of those conditions disclosed above regarding (i) a composition comprising an RGEN protein-CPP complex, or (ii) contacting an RGEN protein-CPP complex with a cell. For example, an RNA component such as a gRNA can be contacted with an RGEN protein-CPP complex in a HEPES buffer (e.g., ~25 mM HEPES, such as 25 mM HEPES/KOH pH 7.5, 200 mM KCl, 20% glycerol, 1 mM DTT), or PBS (e.g., 1X PBS, pH 7), at room temperature (e.g., about 20-25 °C) for about 15 minutes. In those embodiments in which an RGEN protein-CPP complex is held together by non-covalent linkage, association of an RNA component to an RGEN protein can comprise adding an RNA component before, at the same time of, or after incubating a CPP with the RGEN protein component.

After associating an RNA component with an RGEN protein-CPP complex the resulting composition comprising an RGEN-CPP complex (e.g., CPP-Cas9/gRNA) can be immediately contacted with cells, for example. Contact can be made in the milieu in which the RNA component and RGEN protein-CPP complex were associated (e.g., see above), for example. A composition comprising an RGEN-CPP complex can be stored at about room temperature, 4 °C, or frozen (e.g.,

-20 or -80 °C) for later use, if desired. RGEN-CPP complex stability, and/or ability to enter cells and effect DNA targeting, can remain unchanged, or can have at least about 50%, 60%, 70%, 80%, 90%, or 95% of either respective activity, even if the complex is in a composition that has been through one, two, or more freeze-thaw cycles.

A composition comprising an RGEN protein-CPP complex or RGEN-CPP complex, for contacting with a cell, may optionally comprise one or more volume exclusion agents, which are contemplated to enhance contact points between the cell and complexes. Examples of suitable volume exclusion agents herein include glycerol and polyethylene glycol (PEG). Other examples include anionic polymer such as polyacrylate, polymethylacrylate, or anionic polysaccharidic polymers (e.g., dextran sulfate). Still other examples of volume exclusion agents are disclosed in U.S. Patent No. 4886741, which is incorporated herein by reference.

In certain embodiments of an RGEN protein delivery method, a cell (such as a microbial cell) comprises an RNA component that associates with an RGEN protein component of an RGEN protein-CPP complex after the RGEN protein-CPP complex enters the cell (i.e., thereby forming an RGEN-CPP complex in the cell). The RNA component in this embodiment can be as disclosed herein, comprising a sequence complementary to a target site sequence on a chromosome or episome in the cell. The RGEN can bind to the target site sequence, and optionally cleave one or both DNA strands at the target site sequence.

One or more RNA components herein can be stably or transiently expressed in a cell (such as a microbial cell) to which an RGEN protein-CPP complex is introduced, for example. As examples of transient expression, an RGEN protein-CPP complex can be (i) delivered into a cell that has previously been modified to transiently express an RNA component, (ii) co-delivered into a cell with an RNA component, or (iii) delivered into a cell afterwhich the cell is modified for transient RNA component expression.

A DNA polynucleotide sequence comprising (i) a promoter operably linked to (ii) a nucleotide sequence encoding an RNA component can typically be used for stable and/or transient RNA component expression herein. Such a polynucleotide sequence can be comprised within a plasmid, yeast artificial chromosome (YAC), cosmid, phagemid, bacterial artificial chromosome (BAC), virus, or linear DNA (e.g.,

linear PCR product), for example, or any other type of vector or construct useful for transferring a polynucleotide sequence into a cell. This polynucleotide sequence can be capable of existing transiently (i.e., not integrated into the genome) or stably (i.e., integrated into the genome) in a cell. Also, this polynucleotide sequence can comprise, or lack, one or more suitable marker sequences (e.g., selection or phenotype marker).

A suitable promoter comprised in a polynucleotide sequence for expressing an RNA component herein can be constitutive or inducible, for example. A promoter in certain aspects can comprise a strong promoter, which is a promoter that can direct a relatively large number of productive initiations per unit time, and/or is a promoter driving a higher transcription level than the average transcription level of the genes in a cell comprising the strong promoter.

Examples of strong promoters useful in certain aspects herein (e.g., fungal and/or yeast cells) herein include those disclosed in U.S. Patent Appl. Publ. Nos. 2012/0252079 (DGAT2), 2012/0252093 (EL1), 2013/0089910 (ALK2), 2013/0089911 (SPS19), 2006/0019297 (GPD and GPM), 2011/0059496 (GPD and GPM), 2005/0130280 (FBA, FBAIN, FBAINm), 2006/0057690 (GPAT) and 2010/0068789 (YAT1), which are incorporated herein by reference. Other examples of strong promoters include those listed in Table 2, which also may be useful in fungal and/or yeast cells, for example.

Table 2. Strong Promoters

Promoter Name	Native Gene	Reference[a]
XPR2	alkaline extracellular protease	U.S. Pat. No. 4937189; EP220864
TEF	translation elongation factor EF1-α (*tef*)	U.S. Pat. No. 6265185
GPD, GPM	glyceraldehyde-3-phosphate-dehydrogenase (*gpd*), phosphoglycerate mutase (*gpm*)	U.S. Pat. Nos. 7259255 and 7459546
GPDIN	glyceraldehyde-3-phosphate-dehydrogenase (*gpd*)	U.S. Pat. No. 7459546
GPM/FBAIN	chimeric phosphoglycerate mutase (*gpm*)/ fructose-bisphosphate aldolase (*fba1*)	U.S. Pat. No. 7202356
FBA, FBAIN, FBAINm	fructose-bisphosphate aldolase (*fba1*)	U.S. Pat. No. 7202356
GPAT	glycerol-3-phosphate O-acyltransferase (*gpat*)	U.S. Pat. No. 7264949

YAT1	ammonium transporter enzyme (*yat1*)	U.S. Pat. Appl. Publ. No. 2006/0094102
EXP1	export protein	U.S. Pat. No. 7932077

[a] Each reference in this table is incorporated herein by reference.

Other examples of strong promoters useful in certain embodiments herein include PGK1, ADH1, TDH3, TEF1, PHO5, LEU2, and GAL1 promoters, as well as strong yeast promoters disclosed in Velculescu et al. (*Cell* 88:243-251), which is
5 incorporated herein by reference.

A promoter for stable and/or transient expression of an RNA component herein can be an RNA polymerase II (Pol II) promoter, for example. It is believed that all the above-listed strong promoters are examples of suitable Pol II promoters. Transcription from a Pol II promoter may involve formation of an RNA polymerase II
10 complex of at least about 12 proteins (e.g., RPB1-RPN12 proteins), for example. RNA transcribed from a Pol II promoter herein typically is 5'-capped (e.g., contains an m^7G group at the 5'-end) and/or has a polyadenylate (polyA) tail, for example. Means for removing a 5'-cap and/or polyA tail from an RNA component can be employed, if desired, when expressing an RNA component from a Pol II promoter.
15 Suitable means for effectively removing a 5'-cap and/or polyA tail from a Pol II-transcribed RNA component herein include appropriate use of one or more ribozymes (see below), group 1 self-splicing introns, and group 2 self-splicing introns, for example.

Alternatively, a promoter for stable and/or transient expression of an RNA
20 component herein can be an RNA polymerase III (Pol III) promoter, for example. Such a promoter typically allows for expressing an RNA component with defined 5'- and 3'-ends, since initiation and termination of transcription with an RNA polymerase III can be controlled. Examples of Pol III promoters useful herein include U6 and H1 promoters. Other suitable Pol III promoters are disclosed in U.S.
25 Appl. Publ. No. 2010/0160416, for example, which is incorporated herein by reference.

One or more ribozyme sequences may be used to create defined 5' and/or 3' transcript ends, such as in those embodiments in which a Pol II promoter is used for expressing an RNA component in a cell. For example, a nucleotide sequence
30 herein encoding an RNA component may further encode a ribozyme that is upstream of the sequence encoding the RNA component. Thus, a cell in certain

embodiments further comprises a DNA polynucleotide sequence comprising (i) a promoter operably linked to (ii) a nucleotide sequence encoding, in 5'-to-3' direction, a ribozyme and an RNA component. Transcripts expressed from such a polynucleotide sequence autocatalytically remove the ribozyme sequence to yield
5 an RNA with a defined 5'-end (without a 5'-cap) but which comprises the RNA component sequence. This "autoprocessed" RNA can comprise a crRNA or gRNA, for example, and can complex with an RGEN protein component such as a Cas9, thereby forming an RGEN.

A ribozyme herein can be a hammerhead (HH) ribozyme, hepatitis delta virus
10 (HDV) ribozyme, group I intron ribozyme, RnaseP ribozyme, or hairpin ribozyme, for example. Other non-limiting examples of ribozymes herein include Varkud satellite (VS) ribozymes, glucosamine-6-phosphate activated ribozymes (glmS), and CPEB3 ribozymes. Lilley (*Biochem. Soc. Trans.* 39:641-646) discloses information pertaining to ribozyme structure and activity. Examples of ribozymes that should be
15 suitable for use herein include ribozymes disclosed in EP0707638 and U.S. Patent Nos. 6063566, 5580967, 5616459, and 5688670, which are incorporated herein by reference. Further information regarding using ribozymes to express RNA components with defined 5' and/or 3' ends is disclosed in U.S. Patent Appl. No. 62/036,652 (filed August 13, 2014).

20 In certain embodiments, a DNA polynucleotide comprising a cassette for expressing an RNA component comprises a suitable transcription termination sequence downstream of the RNA component sequence. Examples of transcription termination sequences useful herein are disclosed in U.S. Pat. Appl. Publ. No. 2014/0186906, which is herein incorporated by reference. Such embodiments
25 typically comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more residues following the end of the RNA component sequence, depending on the choice of terminator sequence. These additional residues can be all U residues, or at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% U residues, for example, depending on the choice of
30 terminator sequence. Alternatively, a ribozyme sequence (e.g., hammerhead or HDV ribozyme) can be 3' of (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotides downstream) the RNA component sequence, for example. A 3' ribozyme sequence can be positioned accordingly such that it cleaves itself from the RNA component

sequence; such cleavage would render a transcript ending exactly at the end of the RNA component sequence, or with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more residues following the end of the RNA component sequence, for example.

An RNA component in other examples can be provided in the nucleus and/or cytoplasm of a cell into which an RGEN protein-CPP complex is delivered. For example, an RNA component expressed from a Pol II promoter without use of a 5'-located ribozyme sequence can be expected to exist in both the nucleus and cytoplasm. An RNA component expressed from any type of promoter (e.g. Pol II or III promoter) and using a 5'-located ribozyme sequence can be expected to exist mostly in the nucleus in other embodiments. An RNA component expressed from a Pol III promoter in certain aspects can be expected to exist mostly in the nucleus. In certain aspects, an RNA component is uncapped (e.g., by virtue of being expressed from a Pol III promoter, and/or by ribozyme autoprocessing) and typically is located in the nucleus, while in other aspects is capped and located in nuclear and cytoplasmic locations. In general, the RGEN protein component of an RGEN protein-CPP complex, once delivered into a cell, can associate with an RNA component (thereby forming an RGEN) in the cytoplasm and/or nucleus (depending on RNA component location). Such association in the nucleus is generally due to the ability of an RGEN protein component herein to localize to the nucleus as directed by an NLS.

An RGEN herein is useful for RGEN-mediated DNA targeting. Any of the above embodiments regarding delivering an RGEN protein component into a cell can be applied to a DNA targeting method. For example, an RGEN protein-CPP complex can be contacted with at least one RNA component outside of a microbial cell to form an RGEN-CPP complex for delivery into a cell for DNA targeting therein. As another example, an RGEN protein-CPP complex, after its delivery into a microbial cell, can be contacted with at least one RNA component inside a microbial cell to form an RGEN-CPP complex therein that can then mediate DNA targeting. The following disclosure regarding targeting methods refers to an "RGEN", as opposed to referring to an "RGEN-CPP complex". It would be understood that, depending on whether a covalent or non-covalent RGEN-CPP complex is used in an RGEN delivery method herein (and depending on how strong a non-covalent

linkage is in embodiments employing a non-covalent RGEN-CPP complex), reference to an RGEN below refers to such an RGEN-CPP complex, accordingly.

An RGEN herein that can cleave one or both DNA strands of a DNA target sequence can be used in a DNA targeting method, for example. Such DNA 5 targeting methods can involve HR-mediated DNA targeting if a suitable donor DNA is provided in the method. Thus, in certain embodiments, a microbial cell in a targeting method herein can comprise a donor polynucleotide comprising at least one sequence homologous to a sequence at or near a target site sequence (a sequence specifically targeted by an RGEN herein). Such embodiments can 10 optionally be characterized in that the targeting method further comprises a step of providing a suitable donor polynucleotide to the microbial cell.

A donor polynucleotide herein can undergo HR with a sequence at or near a DNA target site if the target site contains a SSB or DSB (such as can be introduced using an RGEN herein). A "homologous sequence" within a donor polynucleotide 15 herein can, for example, comprise or consist of a sequence of at least about 25, 50, 75, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 or 10000 nucleotides, or about 50-500, 50-550, 50-600, 50-650, or 50-700 nucleotides, that have 100% identity with a sequence at or near the target site sequence, or at least about 90%, 91%, 92%, 93%, 94%, 95%, 20 96%, 97%, 98%, or 99% identity with a sequence at or near the target site sequence, for example.

A donor polynucleotide herein can have two homologous sequences (homology arms), for example, separated by a sequence that is heterologous to sequence at or near a target site sequence. HR between such a donor 25 polynucleotide and a target site sequence typically results in the replacement of a sequence at the target site with the heterologous sequence of the donor polynucleotide (i.e., a target site sequence located between target site sequences homologous to the homology arms of the donor polynucleotide is replaced by the heterologous sequence of the donor polynucleotide). In a donor polynucleotide with 30 two homology arms, the arms can be separated by at least about 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 75, 100, 250, 500, 1000, 2500, 5000, 10000, 15000, 20000, 25000, or 30000 nucleotides (i.e., the heterologous sequence in the donor polynucleotide can be at least about 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 75, 100, 250, 500, 1000, 2500,

5000, 10000, 15000, 20000, 25000, or 30000 nucleotides in length), for example. The length (e.g., any of the lengths disclosed above for a homologous sequence) of each homology arm may be the same or different. The percent identity (e.g., any of the % identities disclosed above for a homologous sequence) of each arm with 5 respective homologous sequences at or near the target site can be the same or different.

A DNA sequence at or near (alternatively, in the locality or proximity of) the target site sequence that is homologous to a corresponding homologous sequence in a donor polynucleotide can be within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 10 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 450, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, or 60000 (or any integer between 1 and 60000) nucleotides (e.g., about 1-1000, 100-1000, 500-1000, 1-500, or 100-500 nucleotides), for example, from the predicted RGEN cut site (DSB or nick) in the target sequence. These nucleotide 15 distances can be marked from the cut site to the first nucleotide of the homologous sequence, going either in the upstream or downstream direction from the cut site. For example, a sequence near a target sequence that is homologous to a corresponding sequence in a donor polynucleotide can start at 500 nucleotide base pairs downstream the predicted RGEN cut site in a target sequence. In 20 embodiments herein employing a donor polynucleotide with two homology arms (e.g., first and second homology arms separated by a heterologous sequence), a homologous sequence (corresponding in homology with the first homology arm of a donor) can be upstream the predicted RGEN cut site, and a homologous sequence (corresponding in homology with the second homology arm of a donor) can be 25 downstream the predicted RGEN cut site, for example. The nucleotide distances of each of these upstream and downstream homologous sequences from the predicted cut site can be the same or different, and can be any of the nucleotide distances disclosed above, for example. For instance, the 3' end of a homologous sequence (corresponding in homology with the first homology arm of a donor) may be located 30 600 nucleotide base pairs upstream a predicted RGEN cut site, and the 5' end of a homologous sequence (corresponding in homology with the second homology arm of a donor) may be located 400 nucleotide base pairs downstream the predicted RGEN cut site.

A donor polynucleotide in various aspects can be delivered into a cell (such as a microbial cell) at or near (e.g., within 1, 2, 3 or more hours) the time when an RGEN protein-CPP complex is delivered into the cell. Such delivery can be via by any means known in the art suitable for the particular type of cell being used. These techniques include transformation (e.g., lithium acetate transformation [*Methods in Enzymology*, 194:186-187]), transfection, biolistic impact, electroporation, and microinjection, for example. As examples, U.S. Patent Nos. 4880741 and 5071764, and Chen et al. (*Appl. Microbiol. Biotechnol.* 48:232-235), which are incorporated herein by reference, describe DNA transfer techniques for *Y. lipolytica*. Examples of delivery modes useful in plants include *Agrobacterium*-mediated transformation and biolistic particle bombardment.

An RGEN that cleaves one or both DNA strands of a DNA target sequence can be used to create an indel in other non-limiting embodiments of DNA targeting herein. A method of forming an indel in a cell can be performed as disclosed above for HR-mediated targeting, but without further providing a donor DNA polynucleotide that could undergo HR at or near the target DNA site (i.e., NHEJ is induced in this method). Examples of indels that can be created are disclosed herein. The size of an indel may be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more bases, for example. An indel in certain embodiments can be even larger such as at least about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 bases. In still other embodiments, insertions or deletions can be at least about 500, 750, 1000, or 1500 bases. When attempting to create an indel in certain embodiments, a single base substitution may instead be formed in a target site sequence. Thus, a targeting method herein can be performed for the purpose of creating single base substitution, for example.

In certain embodiments of a targeting method herein aimed at indel formation, the frequency of indel formation in a non-conventional yeast (e.g., *Y. lipolytica*) is significantly higher than what would be observed using the same or similar targeting strategy in a conventional yeast such as *S. cerevisiae*. For example, while the frequency of indel formation in a conventional yeast may be about 0.0001 to 0.001 (DiCarlo et al., *Nucleic Acids Res.* 41:4336-4343), the frequency in a non-conventional yeast herein may be at least about 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, or 0.80. Thus, the frequency of indel formation in a non-conventional yeast herein may be at least

about 50, 100, 250, 500, 750, 1000, 2000, 4000, or 8000 times higher, for example, than what would be observed using the same or similar RGEN-mediated targeting strategy in a conventional yeast.

A targeting method in certain embodiments can be performed to disrupt one or more DNA polynucleotide sequences encoding a protein or a non-coding RNA. An example of such a sequence that can be targeted for disruption is one encoding a marker (i.e., a marker gene). Non-limiting examples of markers herein include screenable markers and selectable markers. A screenable marker herein can be one that renders a cell visually different under appropriate conditions. Examples of screenable markers include polynucleotides encoding beta-glucuronidase (GUS), beta-galactosidase (lacZ), and fluorescent proteins (e.g., GFP, RFP, YFP, BFP). A selectable marker herein can be one that renders a cell resistant to a selective agent or selective environment. Examples of selectable markers are auxotrophic markers such as HIS3, LEU2, TRP1, MET15, or URA3, which allow cells such as yeast cells to survive in the absence of exogenously provided histidine, leucine, tryptophan, methionine, or uracil, respectively. Other examples of selectable markers are antibiotic- or antifungal-resistance markers such as those rendering a cell resistant to ampicillin, chloramphenicol, hygromycin B, nourseothricin, phleomycin, puromycin, or neomycin (e.g., G418). Examples of these methods can optionally be characterized as marker recycling methods.

At least one purpose for disrupting a marker in certain embodiments can be for marker recycling. Marker recycling is a process, for example, comprising (i) transforming a cell with a marker and heterologous DNA sequence, (ii) selecting a transformed cell comprising the marker and the heterologous DNA sequence (where a marker-selectable cell typically has a higher chance of containing the heterologous DNA sequence), (iii) disrupting the marker, and then repeating steps (i)-(iii) as many times as necessary (using the same [or different] marker, but each cycle using a different heterologous DNA sequence) to transform cells with multiple heterologous DNA sequences. One or more heterologous sequences in this process may comprise the marker itself in the form of a donor polynucleotide(e.g., marker flanked by homology arms for targeting a particular locus). Examples of marker recycling processes herein include those using URA3 as a marker, such as

in certain methods employing a yeast (e.g., a non-conventional yeast such as *Y. lipolytica*).

An RGEN herein that can bind to a DNA target site sequence, but does not cleave any strand at the target site sequence, can be used in a DNA targeting method in other embodiments. Any RGEN disclosed herein that has only dysfunctional nuclease domains, but retains specific DNA-binding activity, can be used in this type of targeting method.

In certain embodiments of DNA targeting with an RGEN having no functional nuclease domains, an RGEN can bind to a target site and modulate transcription of a polynucleotide sequence (i.e., gene transcription). Typically, an RGEN is targeted to a regulatory sequence such as a promoter (e.g., within 1-1000, 1-500, 1-250, 1-125, or 1-50 bases upstream a transcription start site), a sequence encoding a 5'-untranslated RNA sequence, or an intron (e.g., first intron) to effect transcriptional modulation of a polynucleotide sequence.

As a non-limiting example, an RGEN linked or fused to a repressor transcription factor or repressor domain thereof can be used to repress, or silence, expression of one or more polynucleotide sequences. An RGEN in certain alternative embodiments can, by itself (without a repressor or domain thereof), inhibit gene expression; such an RGEN can be targeted such that it inhibits binding and/or movement of RNA transcriptional machinery necessary for transcription. A method incorporating any repressing RGEN can optionally be characterized as a gene silencing or transcriptional silencing method. The level of transcriptional down-regulation in a silencing method can be about 100% (gene completely silenced), or at least about 30% (gene moderately silenced), 40%, 50%, 60%, 70%, 80%, 90%, or 95% (gene substantially silenced), for example, compared to the transcription level before application of a repressing RGEN.

An RGEN linked or fused to an activator transcription factor or activator domain thereof can be used to upregulate expression of one or more polynucleotide sequences. A method incorporating such an activating RGEN can optionally be characterized as a transcriptional up-regulation or activation method. The level of transcriptional up-regulation in such a method can be at least about 25%, 50%, 75%, 100%, 250%, 500%, or 1000%, for example, compared to the transcription level before application of an activating RGEN.

In certain embodiment, an RGEN that can bind to a DNA target site sequence, but preferably does not cleave any strand at the target site sequence, can be used as a diagnostic tool (e.g., probe for detecting a DNA sequence). An RGEN protein component in DNA probe can be linked to a reporter agent such as a reporter protein (e.g., fluorescent protein such as GFP), for example. Specific DNA binding of the RGEN-reporter protein, as specified by the RNA component of the RGEN, can be incorporated in a detection system accordingly, taking advantage of the activity of the reporter agent. Flow cytometry (e.g., flow-activated cell sorting [FACS]) and fluorescence *in situ* hybridization (FISH) are examples of suitable detection systems herein that use a fluorescent reporter.

A targeting method herein can be performed in such a way that two or more DNA target sites are targeted in the method, for example. Such a method can optionally be characterized as a multiplex method. Two, three, four, five, six, seven, eight, nine, ten, or more target sites can be targeted at the same time in certain embodiments. A multiplex method is typically performed by a targeting method herein in which multiple different RNA components are provided, each designed to guide an RGEN to a unique DNA target site. For example, two or more different RNA components can be used to prepare a mix of RGEN-CPP complexes *in vitro* (e.g., following a procedure disclosed herein for associating an RNA component with an RGEN protein-CPP complex), which mix is then contacted with a cell.

Another aspect of multiplex targeting herein can comprise providing two or more different RNA components in a cell which associate with the RGEN protein components of RGEN protein-CPP complexes that have traversed into the cell. Such a method can comprise, for example, providing to the cell (i) individual DNA polynucleotides, each of which express a particular RNA component that, and/or (ii) at least one DNA polynucleotide encoding two or more RNA components (e.g., see below disclosure regarding tandem ribozyme-RNA component cassettes).

A multiplex method can optionally target DNA sites very close to the same sequence (e.g., a promoter or open reading frame, and/or sites that are distant from each other (e.g., in different genes and/or chromosomes). A multiplex method in other embodiments can be performed with (for HR) or without (for NHEJ leading to indel and/or base substitution) suitable donor DNA polynucleotides, depending on the desired outcome of the targeting (if an endonuclease- or nickase-competent

RGEN is used). In still other embodiments, a multiplex method can be performed with a repressing or activating RGEN as disclosed herein. For example, multiple repressing RGENs can be provided that down-regulate a set of genes, such as genes involved in a particular metabolic pathway.

5 A multiplex method in certain embodiments can comprise providing to a cell a DNA polynucleotide comprising (i) a promoter operably linked to (ii) a sequence comprising more than one ribozyme-RNA component cassettes (i.e., tandem cassettes). A transcript expressed from such a DNA polynucleotide can have, for example, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more cassettes. A 3' ribozyme sequence can 10 optionally be included following all or some RNA component sequences to allow cleavage and separation of the RNA component from downstream transcript sequence (i.e., tandem cassettes may comprise one or more ribozyme-RNA component-ribozyme cassettes). A DNA polynucleotide herein for expressing tandem ribozyme-RNA component-ribozyme cassettes can be designed such that 15 there are about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more nucleotides between each cassette (e.g., non-coding spacer sequence). The distances between each cassette may be the same or different.

Any construct or vector comprising a DNA polynucleotide encoding an RNA component described herein can be introduced into a cell by any means known in 20 the art suitable for the particular type of cell being used. For example, any of the means disclosed above for delivering a donor DNA into a cell can be employed.

Certain embodiments herein concern a method of modifying or altering a target site in the genome of a microbial cell, wherein the method comprises contacting the microbial cell with a guide polynucleotide and Cas endonuclease 25 covalently or non-covalently linked to a CPP, wherein the guide polynucleotide and CPP-Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double-strand break at the target site in the genome of the microbial cell. The modification or alteration of the target site can include (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) 30 an insertion of at least one nucleotide, or (iv) any combination of (i)-(iii).

Certain embodiments herein concern a polynucleotide sequence comprising a nucleotide sequence encoding a fusion protein that comprises a protein component of an RNA-guided endonuclease (RGEN) and at least one cell-

penetrating peptide (CPP). Any fusion protein as disclosed herein, for example, can be encoded by the nucleotide sequence. The nucleotide sequence may optionally be in operable linkage with a promoter sequence. Certain embodiments include, for example, a polynucleotide (e.g., vector or construct) comprising at least one open reading frame encoding any RGEN protein-CPP fusion disclosed herein. Such a coding region can optionally be operably linked to a promoter sequence suitable for expressing an RGEN protein-CPP fusion in a cell (e.g., bacteria cell; eukaryotic cell such as a yeast, insect, or mammalian cell) or in an *in vitro* protein expression system, for example. Examples of a vector or construct include circular (e.g., plasmid) and non-circular (e.g., linear DNA such as an amplified DNA sequence) polynucleotide molecules.

Certain embodiments herein concern a method of producing an RGEN protein-CPP fusion protein comprising the steps of: providing a polynucleotide sequence having a nucleotide sequence encoding the RGEN protein-CPP fusion protein, and expressing the RGEN protein-CPP fusion protein from the polynucleotide sequence, thereby producing the RGEN protein-CPP fusion protein. The expression step in such a method can optionally be performed in a cell (e.g., bacteria cell such as *E. coli*; eukaryotic cell such as a yeast [e.g., *S. cerevisiae*], insect, or mammalian cell). Alternatively, expression of an RGEN protein-CPP fusion protein can be performed in an *in vitro* protein expression system (e.g., cell-free protein expression systems such as those employing rabbit reticulocyte lysate or wheat germ extract). Also, the RGEN protein-CPP fusion protein produced in the expression step can optionally be isolated. Such isolation can be performed in a manner that produces a composition having any of the above-disclosed features (e.g., purity, pH, buffer, and/or salt level), for example.

Non-limiting examples of compositions and methods disclosed herein include:

1. A composition comprising at least one protein component of an RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP),
wherein the protein component and CPP are covalently, or non-covalently, linked to each other in an RGEN protein-CPP complex, and
wherein the RGEN protein-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a cell.

2. The composition of embodiment 1, wherein the protein component of the RGEN is associated with at least one RNA component that comprises a sequence complementary to a target site sequence on a chromosome or episome in the cell, wherein the RGEN can bind to the target site sequence, and optionally cleave one or both DNA strands at the target site sequence.
3. The composition of embodiment 2, wherein the RNA component comprises a guide RNA (gRNA) comprising a CRISPR RNA (crRNA) operably linked to a trans-activating CRISPR RNA (tracrRNA).
4. The composition of embodiment 2, wherein the RGEN can cleave one or both DNA strands at the target site sequence.
5. The composition of embodiment 1, wherein the RGEN comprises a CRISPR-associated (Cas) protein-9 (Cas9) amino acid sequence.
6. The composition of embodiment 1, wherein the RGEN protein component and CPP are covalently linked.
7. The composition of embodiment 1, wherein the RGEN protein component and CPP are non-covalently linked.
8. The composition of embodiment 1, wherein the CPP is cationic or amphipathic.
9. The composition of embodiment 1, wherein the CPP comprises:
   (i) a CPP from an Epstein-Barr virus Zebra trans-activator protein,
   (ii) a CPP having 6 or more contiguous arginine residues,
   (iii) a transportan-10 (TP10) CPP, or
   (iv) a CPP from a vascular endothelium cadherin protein.
10. The composition of embodiment 1, wherein the RGEN protein-CPP complex can traverse a cell wall and cell membrane of a cell.
11. A cell comprising the composition according to embodiment 1.
12. A method of delivering a protein component of an RNA-guided endonuclease (RGEN) into a cell, the method comprising:
    contacting the cell with a composition comprising the protein component of the RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP),
    wherein the protein component and CPP are covalently, or non-covalently, linked to each other in an RGEN protein-CPP complex,

wherein the RGEN protein-CPP complex traverses (i) a cell membrane, or (ii) a cell wall and cell membrane, of the cell, thereby entering the cell.

13. The method of embodiment 12, wherein:
    (i) the composition further comprises at least one RNA component that is associated with the protein component of the RGEN; or
    (ii) the cell comprises the RNA component, wherein the RNA component associates with the protein component of the RGEN after the RGEN protein-CPP complex enters the cell;

    wherein the RNA component comprises a sequence complementary to a target site sequence on a chromosome or episome in the cell, wherein the RGEN can bind to the target site sequence, and optionally cleave one or both DNA strands at the target site sequence.
14. The method of embodiment 13, wherein the RGEN can cleave one or both DNA strands at the target site sequence.
15. The method of embodiment 14, wherein the cell further comprises a donor polynucleotide comprising at least one sequence homologous to a sequence at or near the target site sequence.
16. The method of embodiment 12, wherein the cell is a non-mammalian cell.
17. A composition comprising at least one protein component of a guide polynucleotide/Cas endonuclease complex and at least one cell-penetrating peptide (CPP), wherein the protein component and CPP are covalently, or non-covalently, linked to each other in a guide polynucleotide/Cas endonuclease-CPP complex, and wherein the guide polynucleotide/Cas endonuclease-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a cell, wherein the cell is optionally a plant cell.
18. The composition of embodiment 17, wherein the Cas endonuclease is a plant-optimized Cas9 endonuclease.
19. The composition of embodiment 17, wherein the guide polynucleotide comprises
    (i) a first nucleotide sequence domain that is complementary to a nucleotide sequence in a target DNA, and
    (ii) a second nucleotide sequence domain that interacts with a Cas endonuclease,

wherein the first nucleotide sequence domain and the second nucleotide sequence domain are composed of deoxyribonucleic acids (DNA), ribonucleic acids (RNA), or a combination thereof.

20. The composition of embodiment 17, wherein the guide polynucleotide/Cas endonuclease-CPP complex can traverse the cell wall of a plant cell.

21. The composition of embodiment 17, wherein the CPP comprises:
    (i) a CPP from an Epstein-Barr virus Zebra trans-activator protein,
    (ii) a CPP having 6 or more contiguous arginine residues,
    (iii) a transportan-10 (TP10) CPP,
    (iv) a CPP from a vascular endothelium cadherin protein, or
    (vi) a CPP selected from the group consisting of a synthetic nona-arginine CPP, a histidine-rich nona-arginine CPP, and a Pas nona-arginine CPP.

22. The composition of embodiment 20, wherein the plant cell is a monocot or a dicot cell.

23. The composition of embodiment 22, wherein the monocot is selected from the group consisting of maize, rice, sorghum, rye, barley, wheat, millet, oats, sugarcane, turfgrass, and switchgrass.

24. The composition of embodiment 22, wherein the dicot is selected from the group consisting of soybean, canola, alfalfa, sunflower, cotton, tobacco, peanut, potato, tobacco, Arabidopsis, and safflower.

25. A method for modifying a target site in the genome of a cell, the method comprising providing a guide polynucleotide, a cell-penetrating peptide (CPP) and a Cas endonuclease to the cell, wherein the guide polynucleotide, Cas endonuclease and CPP are covalently, or non-covalently, linked to each other in a guide polynucleotide/Cas endonuclease-CPP complex, and wherein the guide polynucleotide/Cas endonuclease-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a cell, wherein the cell is optionally a plant cell.

26. The method of embodiment 25, further comprising identifying at least one plant cell that has a modification at the target site, wherein the modification at the target site is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

27. The method of embodiment 25, wherein the plant cell is a monocot or dicot cell.
28. A composition comprising at least one protein component of an RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP), wherein the protein component and CPP are covalently, or non-covalently, linked to each other in an RGEN protein-CPP complex, and wherein the RGEN protein-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a microbial cell.
29. The composition of embodiment 28, wherein the protein component of the RGEN is associated with at least one RNA component that comprises a sequence complementary to a target site sequence on a chromosome or episome in the microbial cell, wherein the RGEN can bind to the target site sequence, and optionally cleave one or both DNA strands at the target site sequence.
30. The composition of embodiment 28, wherein the RGEN protein-CPP complex can traverse a cell wall and cell membrane of a microbial cell.
31. A microbial cell comprising the composition according to embodiment 28.
32. A method of delivering a protein component of an RNA-guided endonuclease (RGEN) into a microbial cell, the method comprising:
contacting the microbial cell with a composition comprising the protein component of the RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP),
wherein the protein component and CPP are covalently, or non-covalently, linked to each other in an RGEN protein-CPP complex,
wherein the RGEN protein-CPP complex traverses (i) a cell membrane, or (ii) a cell wall and cell membrane, of the microbial cell, thereby entering the microbial cell.
33. The method of embodiment 32, wherein:
    (i) the composition further comprises at least one RNA component that is associated with the protein component of the RGEN; or
    (ii) the microbial cell comprises the RNA component, wherein the RNA component associates with the protein component of the RGEN after the RGEN protein-CPP complex enters the microbial cell;

wherein the RNA component comprises a sequence complementary to a target site sequence on a chromosome or episome in the microbial cell, wherein the RGEN can bind to the target site sequence, and optionally cleave one or both DNA strands at the target site sequence.

5 34. The method of embodiment 33, wherein the RGEN can cleave one or both DNA strands at the target site sequence.

34. The method of embodiment 34, wherein the microbial cell further comprises a donor polynucleotide comprising at least one sequence homologous to a sequence at or near the target site sequence.

10 36. The method of embodiment 32, wherein the microbial cell is a yeast cell.

37. A composition comprising at least one protein component of a guide polynucleotide/Cas endonuclease complex and at least one cell-penetrating peptide (CPP), wherein the protein component and CPP are covalently, or non-covalently, linked to each other in a guide polynucleotide/Cas
15 endonuclease-CPP complex, and wherein the guide polynucleotide/Cas endonuclease-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a microbial cell.

38. The composition of embodiment 37, wherein the guide polynucleotide/Cas endonuclease-CPP complex can traverse the cell wall of the microbial cell.

20 39. A method for modifying a target site in the genome of a microbial cell, the method comprising providing a guide polynucleotide, a cell-penetrating peptide (CPP) and a Cas endonuclease to the microbial cell, wherein the guide polynucleotide, Cas endonuclease and CPP are covalently, or non-covalently, linked to each other in a guide polynucleotide/Cas endonuclease-
25 CPP complex, and wherein the guide polynucleotide/Cas endonuclease-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a microbial cell.

23. The method of embodiment 39, further comprising identifying at least one microbial cell that has a modification at the target site, wherein the
30 modification at the target site is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

EXAMPLES

The disclosed invention is further defined in the following Examples. It should be understood that these Examples, while indicating certain preferred aspects of the invention, are given by way of illustration only. From the above
5 discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various uses and conditions.

EXAMPLE 1

10 Vectors for Expressing a Cas9-CPP (Cell-Penetrating Peptide) Fusion Protein in *E. coli*

In this example, vectors designed for inducible expression of translational fusion proteins comprising Cas9 protein and a cell-penetrating peptide (CPP) were produced and tested for expression in *E. coli*. Cas9-CPP fusion proteins were
15 shown to express in *E. coli* as expected, and subsequently purified.

The open reading frame of the Cas9 gene from *Streptococcus pyogenes* M1 GAS (SF370) was codon-optimized for expression in *Yarrowia* per standard techniques, yielding SEQ ID NO:1. DNA sequence encoding a simian virus 40 (SV40) monopartite nuclear localization signal (NLS) plus a short linker (4 amino
20 acids) was incorporated after the last sense codon of SEQ ID NO:1 to render SEQ ID NO:2. SEQ ID NO:2 encodes the amino acid sequence shown in SEQ ID NO:3. The last seven amino acids of SEQ ID NO:3 encode the added NLS, whereas residues at positions 1369-1372 of SEQ ID NO:3 encode the added linker. The *Yarrowia* codon-optimized Cas9-NLS sequence (SEQ ID NO:2) was linked to a
25 *Yarrowia* constitutive promoter, FBA1 (SEQ ID NO:4), by standard molecular biology techniques. A *Yarrowia* codon-optimized Cas9 expression cassette containing the constitutive FBA1 promoter, *Yarrowia* codon-optimized Cas9, and the SV40 NLS is set forth in SEQ ID NO:5. This Cas9 expression cassette (SEQ ID NO:5) was cloned into the plasmid pZUF rendering construct pZUFCas9 (Figure 1,
30 SEQ ID NO:6).

The *Yarrowia* codon-optimized Cas9-NLS sequence was PCR-amplified from pZUFCas9 (SEQ ID NO:6) using standard molecular biology techniques. Primers for the PCR reaction were SEQ ID NO:7 (Forward) and SEQ ID NO:8 (Reverse),

which added a 5' EcoRI site and 3' HindIII site, respectively, to the amplified DNA product. The added 5' EcoRI site replaced the ATG start codon of the Cas9-NLS open reading frame (ORF) in the amplified product. The amplified product (SEQ ID NO:9) was digested with EcoRI and HindIII, and then purified using Zymoclean™ and concentrator columns (Zymo Research, Irvine, CA). The purified DNA fragment was cloned into the EcoRI and HindIII sites of plasmid pBAD/HisB from Life Technologies (Carlsbad, CA) (Figure 2A, SEQ ID NO:10) to create plasmid construct pRF48 (Figure 2B, SEQ ID NO:11). Plasmid pRF48 is capable of expressing, in *E. coli*, a Cas9-NLS comprising a hexahistidine (6xHis) tag at its N-terminus.

To fuse a cell-penetrating peptide (CPP) sequence to Cas9-NLS, individual DNA polynucleotide sequences were prepared, each codon-optimized for expression in *E. coli* and comprising sequence encoding a 6xHis tag linked to a particular CPP amino acid sequence: Zebra peptide (ECDSELEIKRYKRVRVASRKCRAKFKQLLQHYREVAAAKSSENDRLRLLLKQMC, SEQ ID NO:12), from the Epstein-Barr virus Zebra trans-activator protein; pVEC peptide (LLIILRRRIRKQAHAHSK, SEQ ID NO:13), from a murine endothelial cadherin protein; TP10 peptide (AGYLLGKINLKACAACAKKIL, SEQ ID NO:14), from a neuropeptide galanin protein; and synthetic arginine-rich "PolyR" peptide (GGGGRRRRRRRRRLLLL, SEQ ID NO:15). Each DNA polynucleotide sequence included a 5'-end NcoI restriction site and a 3'-end EcoRI site to create cloning sequences structured as follows: NcoI-6xHis-CPP-EcoRI (SEQ ID NO:16-19). Each of SEQ ID NOs:16-19 was individually cloned into the NcoI and EcoRI sites of pRF48, thereby creating plasmid constructs capable of expressing certain 6xHis-CPP-Cas9-NLS fusion proteins in *E. coli*. In particular, plasmid construct pRF144 (Figure 3A, SEQ ID NO:20) was prepared for expressing a 6xHis-Zebra CPP-Cas9-NLS fusion; plasmid construct pRF145 (Figure 3B, SEQ ID NO:21) was prepared for expressing a 6xHis-PolyR CPP-Cas9-NLS fusion; plasmid construct pRF146 (Figure 3C, SEQ ID NO:22) was prepared for expressing a 6xHis-TP10 CPP-Cas9-NLS fusion, and plasmid construct pRF162 (Figure 3D, SEQ ID NO:23) was prepared for expressing a 6xHis-pVEC CPP-Cas9-NLS fusion.

Each of plasmids pRF48, pRF144, pRF145, pRF146 and pRF162 was individually transformed into TOP10 competent cells (Life Technologies). Cells

were grown overnight at 37 °C with shaking (220 rpm) in L broth (Miller) containing 0.4% (w/v) glucose and 100 µg/mL ampicillin. Each pre-culture was diluted 1:100 in 2X YT medium containing 100 µg/mL ampicillin and further grown at 37 °C with shaking (220 rpm). When cultures reached an OD_{600} of about 0.5, protein expression from each plasmid was induced by adding L-arabinose to a final concentration of 0.2% (w/v). The cultures were grown for an additional 18 hours at 18 °C with shaking (200 rpm). Cells were pelleted at 5000 x g for 15 minutes at 4 °C. Medium was disposed of and cell pellets were frozen at -80 °C for at least 4 hours. Cell pellets were thawed for 15 minutes on ice and resuspended in 15 mL of lysis buffer (20 mM tris pH 7.5, 500 mM NaCl, 1 mM $MgCl_2$, 10 mM imidazole, 120 units/mL DNaseI, 1 mM PMSF, 1 mM DTT) per liter of original culture. Cells were lysed by passage twice through a large French pressure cell at 16000 psi. Cell debris was pelleted at 20000 x g for 30 minutes at 4 °C. Supernatants were transferred to a 50-mL conical tubes, to which 2 ml of a 50% slurry of Ni-NTA resin (Qiagen) was added for binding the 6xHis Tag of each expressed fusion protein. Each tube was slowly rotated at 4 °C for 1 hour and then applied to an empty gravity column through which the supernatant was allowed to flow. Flow-through sample (75 µL) was taken, added to 25 µL of 4x-reduced Laemmeli buffer, and stored on ice. The resin was washed four times in each column with 5 ml of wash buffer (20 mM tris pH 7.5, 500 mM NaCl, 10 mM imidazole, 1 mM PMSF, 1 mM DTT). A sample (75 µL) was taken from each wash, added to 25 µL of 4x-reduced Laemmeli buffer, and stored on ice. 1-ml aliquots of elution buffer (20 mM Tris pH 7.5, 500 mM NaCl, 1 mM $MgCl_2$, 500 mM imidazole, 1 mM PMSF, 1 mM DTT) were applied to the resin in each column and allowed to incubate for 10 minutes. Protein elution was monitored by absorbance at 280 nm. A sample (75 µL) was taken from each elution, added to 25 µL of 4x-reduced Laemmeli buffer, and stored on ice. For each plasmid expression experiment, fractions containing eluted protein from the column were combined, loaded into 10000 MWCO dialysis membrane, and dialyzed against dialysis buffer (25 mM HEPES/KOH pH 7.5, 200 mM KCl, 20% glycerol, 1 mM DTT) at 4 °C for at least 14 hours. The protein concentration of each dialysate was determined using the Bradford assay and absorbance at 565 nm. Purified protein was split into two aliquots, one of which was frozen at -80°C and the other stored on ice at 4 °C. Samples taken during the column purification process for each plasmid

expression experiment were heated at 95 °C for 5 minutes and loaded onto an 8% (w/v) tris-glycine polyacrylamide resolving gel with a 4% (w/v) stacking gel. Proteins were electrophoretically separated at 200 volts for 30 minutes and stained with Coomassie blue. The gel for the 6xHis-Zebra-Cas9-NLS purification process is shown in Figure 4 as an example.

Thus, four different CPP-Cas9 fusion proteins were expressed and isolated. These fusion proteins represent examples of RGEN protein-CPP complexes herein.

EXAMPLE 2

Expressing Short Guide RNA (sgRNA) by *in vitro* Transcription

In this example, a DNA sequence was designed that encodes an sgRNA fused to ribozymes at its 5'- and 3'-ends (referred to as "RGR"), respectively. The RGR sequence allowed for *in vitro* transcription by T7 RNA polymerase of an sgRNA with precisely defined ends.

Figure 5 illustrates an sgRNA molecule, which is a single RNA molecule containing two regions, a variable targeting domain (VT) (guide sequence) and Cas endonuclease recognition (CER) domain (SEQ ID NO:24 represents an example of a CER). The VT region can be a 20mer of RNA polynucleotide that has identity to a targeted nucleic acid molecule, for example. The VT domain specifies a target site for cleavage in the target site that lies 5' of a PAM motif. The CER domain interacts with Cas9 protein and allows the VT domain to interact and direct the Cas9 protein cleavage (Jinek et al., *Science* 337:816-821). Both VT and CER domains are required for the function of an sgRNA.

The addition of 5' HammerHead (HH) and 3' Hepatitis Delta Virus (HDV) ribozymes to an sgRNA sequence allows expression of the sgRNA from any promoter without consideration for certain transcriptional requirements of some RNA polymerases (e.g., T7 RNA polymerase requires one transcribed G residue directly after initiation of transcription, but works best with three transcribed G residues). When such sgRNA is expressed, the ribozymes present in the pre-sgRNA transcript autocleave, thereby separating from the transcript leaving an unmodified sgRNA.

A DNA sequence encoding an sgRNA that targets the Can1-1 locus (SEQ ID NO:25) in *Yarrowia lipolytica* was prepared; this sgRNA comprises SEQ ID NO:24 as its CER domain. The sgRNA-encoding sequence was linked at its 5'-end to sequence encoding an HH ribozyme (SEQ ID NO:26) and at its 3'-end to a

sequence encoding an HDV ribozyme (SEQ ID NO:27), such that the first 6 bases of the HH ribozyme were a reverse compliment to the first 6 bases of the VT region of the sgRNA. This particular RGR sgRNA is encoded by SEQ ID NO:28. The RGR sgRNA of SEQ ID NO:28 was then linked to a T7 RNA polymerase promoter (SEQ ID NO:29) via standard molecular biology techniques to create plasmid pRF46 (SEQ ID NO:30).

T7-RGR sgRNA-encoding sequence was PCR-amplified from plasmid pRF46 (SEQ ID NO:30) using standard techniques. Primers for the PCR reaction were SEQ ID NO:31 (T7 forward primer) and SEQ ID NO:32 (gRNArev1 reverse primer). The PCR product was purified by ethanol precipitation and resuspended in ddH_2O; this DNA was used as template in an *in vitro* transcription reaction. Template DNA was added to a final concentration of 150 nM in 20-µL *in vitro* transcription reactions (MEGAshortscript™ T7 Kit, Life Technologies). Reactions were allowed to proceed for various times (2 hours, 4 hours, 6 hours, and overnight) to determine suitable conditions for *in vitro* transcription (Figure 6). The reactions were then treated with 10 units of DNaseI for 15 minutes at 37 °C to remove template DNA. RNA was precipitated using ethanol and standard protocols. Each 20-µl *in vitro* transcription reaction produced between 60 and 100 µg of RNA.

Thus, sgRNA with defined 5'- and 3'-ends was synthesized *in vitro*. As demonstrated in Example 3 below, *in vitro* transcribed sgRNA can be associated with a Cas9-CPP fusion protein to form an RGEN-CPP complex.

EXAMPLE 3

Specific *in vitro* Cleavage of Target DNA Sequence Using Cas9-CPP Fusion Protein Complexed with sgRNA

In this example, the targeting endonuclease function of Zebra CPP-Cas9 fusion protein (comprising SEQ ID NO:39) in complex with an sgRNA was tested to confirm that fusion with a CPP does not hinder Cas9 endonuclease activity.

An *in vitro* Can1 cleavage assay DNA polynucleotide (SEQ ID NO:35) containing the Can1-1 target sequence of SEQ ID NO:25 was PCR-amplified from *Y. lipolytica* cells (ATCC 20362) and purified using standard techniques. Primers for the PCR reaction were SEQ ID NO:33 (IV-up forward primer) and SEQ ID NO:34 (IV-down reverse primer).

Purified Zebra CPP-Cas9 fusion protein (600 ng, prepared in Example 1), sgRNA targeting the Can1-1 target site (250 ng, prepared in Example 2), NEBuffer 3.1 (New England BioLabs, Ipswich, MA), and Can1 cleavage assay DNA (150 ng, SEQ ID NO:35) were mixed in a 10-µL reaction (volume brought up to final volume with ddH_2O). As negative controls, reactions lacking either Zebra CPP-Cas9 fusion protein or sgRNA were also prepared. As a positive control, wild type Cas9 protein (PNA Bio, Thousand Oaks, CA) was used in a reaction instead of Zebra CPP-Cas9. The reactions were incubated at 37 °C for 60 minutes. RNaseI (4 µg) was then added to each reaction and incubated at 37 °C for 15 minutes to degrade the sgRNA. Stop solution (1 µL; 30% [w/v] glycerol, 1.2% [w/v] SDS, 250 mM EDTA, pH 8.0) was added to terminate the reactions, which were then further incubated for 15 minutes at 37 °C. Each reaction was loaded onto a 1.2% FlashGel™ (Lonza, Basel, Switzerland) and electrophoresed for 10 minutes at 200 volts (Figure 7). The target DNA cleavage pattern rendered by Zebra CPP-Cas9 was consistent with the cleavage pattern rendered by wild type Cas9 (Figure 7), thereby indicating that Zebra CPP-Cas9 functions normally *in vitro*. Furthermore, this activity was not inhibited using Zebra CPP-Cas9/sgRNA that had been subjected to two freeze-thaw cycles.

Thus, a CPP-Cas9 fusion protein complexed with a suitable sgRNA (i.e., an example of an RGEN-CPP complex) had specific DNA cleavage activity. This activity was shown to be similar with the activity of a wild type Cas9-sgRNA complex, thereby indicating that CPP fusion does not inhibit Cas9-sgRNA endonucleolytic function. While the CPP-Cas9 fusion protein in this example comprised SEQ ID NO:39 (Zebra CPP-Cas9), it is contemplated that a CPP-Cas9 fusion protein comprising SEQ ID NO:40, 41, or 42, for example, also has cleavage activity when associated with a suitable sgRNA as an RNA component.

EXAMPLE 4

Delivery of a CPP-Cas9/sgRNA Complex into Yeast Cells and Cleavage of Target DNA Therein

In this example, Zebra CPP-Cas9 fusion protein (comprising SEQ ID NO:39) in complex with an sgRNA (Zebra CPP-Cas9/sgRNA) was tested for the ability to enter yeast cells after simple contact with the cells. Zebra CPP-Cas9/sgRNA

specific for Can1-1 was able to enter cells and cleave the Can1 gene, thereby rendering cells to be canavanine-resistant.

Y. lipolytica yeast cells (ATCC 20362) were grown in YPD (2% glucose, 2% peptone, 1% yeast extract) liquid medium at 30 °C with shaking (220 rpm) to OD_{600} = 0.5 (approximately $5x10^6$ cells per mL of culture). Purified Zebra CPP-Cas9 fusion protein (prepared in Example 1) and sgRNA targeting the Can1-1 target site (prepared in Example 2) were mixed in a 1:3 molar ratio, respectively, in the dialysis buffer used in Example 1 and pre-incubated at room temperature for 15 minutes to allow the sgRNA to associate with the Zebra CPP-Cas9. $5x10^5$ *Y. lipolytica* cells were mixed into the Zebra CPP-Cas9/sgRNA preparation such that the final concentration of Zebra CPP-Cas9 was 1 µM, 2.5 µM, or 5 µM. Cells were also mixed with 5 µM final concentration Zebra CPP-Cas9 alone (no sgRNA as RNA component) as a negative control. All the cell-Cas9 preparations were incubated at 30 °C with shaking (220 rpm) for 2 hours. The cells were then serially diluted 1000- and 10000-fold. Each serial dilution (100 µL) was plated onto complete medium lacking arginine (CM-Arg) and allowed to recover for 48 hours at 30 °C.

Colonies of the 10^{-3}-dilution plates were counted to determine the total number of cells plated. Colonies were transferred to CM-Arg plates with canavanine (60 µg/mL) via replica-plating technique. Colonies were allowed to grow at 30 °C for 48 hours. The number of canavanine-resistant colonies were scored and divided by the total number of colonies (from plates without canavanine) to determine a mutation frequency for each case. Contacting cells with Zebra CPP-Cas9/sgRNA complexes yielded colonies that were resistant to canavanine at frequencies of about 2% to 10% of the total colonies (Figure 8). This canavanine-resistance is expected to be due to loss of Can1 gene function by indel formation at/near the predicted Cas9 cleavage site in the Can1 gene coding sequence. However, contacting cells with Zebra CPP-Cas9 alone (no sgRNA) did not yield canavanine-resistant colonies (Figure 8), indicating that canavanine-resistance in the experimental cells was dependent on sgRNA-based specificity given to CPP-Cas9 protein. Given the nature of yeast cells, the CPP-Cas9/sgRNA complexes likely had to traverse both cell wall and cell membrane structures to mediate specific DNA targeting.

Thus, a CPP-Cas9 fusion protein complexed with a suitable sgRNA (i.e., an example of an RGEN-CPP complex) is able to enter yeast cells (traverse cell wall and cell membrane) and target a specific DNA sequence therein. While the CPP-Cas9 fusion protein in this example comprised SEQ ID NO:39 (Zebra CPP-Cas9), it
5 is contemplated that a CPP-Cas9 fusion protein comprising SEQ ID NO: 40, 41, or 42, for example, also has cell-entry activity, and specific DNA targeting activity in cells, when associated with a suitable sgRNA as an RNA component.

EXAMPLE 5

CPP-Facilitated Cas9/sgRNA Complex Delivery into Plant Cells and Cleavage of
10 Target DNA Therein

CPP-facilitated protein delivery into soybean cells can be tested by incubating soybean callus cells with DS-RED fluorescent proteins fused to CPPs. Fluorescent signals are expected in CPP-DS-RED treatments, but not in controls incubated with DS-RED proteins only. Various CPPs can be tested in this manner
15 to help identify the most effective CPPs for plant cell penetration and delivery of protein cargo. Some examples of CPPs that can be tested include:

(i) a CPP from an Epstein-Barr virus Zebra trans-activator protein,

(ii) a CPP having 6 or more contiguous arginine residues,

(iii) a transportan-10 (TP10) CPP,

20 (iv) a CPP from a vascular endothelium cadherin protein, or

(vi) a CPP selected from the group consisting of a synthetic nona-arginine CPP, a histidine-rich nona-arginine CPP and a Pas nona-arginine CPP. Examples of a synthetic nona-arginine CPP, a histidine-rich nona-arginine CPP and a Pas nona-arginine CPP are disclosed in, for example, Liu et al. (*Advanced Studies in*
25 *Biology* 5(2):71-88, HIKARI Ltd).

In vitro translated Cas9 proteins and synthetic sgRNA can be mixed with CPPs, by themselves or in a fusion (e.g., CPP-DS-RED above), and incubated with soybean callus to test if Cas9/sgRNA can be transported into the cells. Once in the cells, the Cas9/sgRNA complex can recognize a genomic target specified by the
30 sgRNA targeting sequence to make DNA double strand breaks (DSBs). Spontaneous repair of the DSBs by cell machinery can result in mutations through non-homologous end joining (NHEJ), or gene integration through homologous recombination if appropriate donor DNA is present. CPPs can also be covalently

linked to Cas9 proteins for potentially better efficiency. The success of CPP-Cas9/sgRNA delivery into soybean cells, and thus the transfer of the CPP-Cas endonuclease complex across a plant cell wall and plant cell membrane, can be verified by the detection of mutations or gene integrations at the specific target site by PCR analysis, for example.

EXAMPLE 6

Expression and Purification of CPP-dsREDexpress proteins from *E. coli* cells.

To rapidly assess the ability of a given cell-penetrating peptide to enter a specific cell type CPP fusions to the dsREDexpress protein (SEQ ID NO: 85) were created, expressed in *E. coli* cells, and purified. The CPP-dsREDexpress protein fusions are a tool that allows rapid assessment of cargo delivery into a given cell type by a given CPP. This allows selection of a species, cell type, or strain specific CPP molecule to maximize delivery of cargo in a rapid and high-throughput manner by assessing cellular fluorescence by microscopic or flow cytometric analysis.

An *E. coli* codon optimized dsREDexpress gene (SEQ ID NO: 86) was synthesized (IDT DNA) and cloned into the NcoI/HinDIII sites of pBAD/HisB (SEQ ID NO: 87) creating pRF161 (SEQ ID NO: 88). The *E. coli* codon optimized dsREDexpress contained an internal EcoRI site such that digestion of the plasmid with NcoI/EcoRI would allow replacement of the his tag with various his tag-CPP sequences to create his tag-CPP-dsREDexpress fusion expression plasmids. Various his-tag-CPP fusions; TAT (SEQ ID NO: 89), TLM (SEQ ID NO: 90), MPG1(SEQ ID NO: 91), pep1 (SEQ ID NO: 92), and CFFKDEL (SEQ ID NO: 93); were codon optimized for *E. coli* and flanked with in frame 5' NcoI and 3' EcoRI sites (SEQ ID NO: 94-98 respectively) and cloned using standard techniques into the NcoI/EcoRI sites of pRF161 (SEQ ID NO: 88) replacing the his tag sequence with the corresponding his tag-CPP fusion and generating plasmids pRF224 (his-TAT-dsREDexpress SEQ ID NO: 99), pRF214 (his-TLM-dsREDexpress SEQ ID NO: 100), pRF213 (his-MPG1-dsREDexpress SEQ ID NO: 101), pRF217 (his-pep1-dsREDexpress SEQ ID NO: 102), pRF216 (his-CFFKDEL-dsREDexpress SEQ ID NO: 103). Sequences of the inserted fragments were verified using standard sequencing techniques and oligo 36 (SEQ ID NO: 104).

E. coli codon optimized His-Zebra (SEQ ID NO: 105), His-tp10 (SEQ ID NO: 106), and His-pVEC (SEQ ID NO: 107) were PCR amplified from pRF144 (SEQ ID

NO 108), pRF162 (SEQ ID NO 109), and pRF146 (SEQ ID NO: 110) respectively using oligo 36 (SEQ ID NO: 104) and oligo 153 (SEQ ID NO: 111) with standard PCR techniques. PCR fragments were cloned into the NcoI/EcoRI sites of pRF161 (SEQ ID NO: 88) creating plasmids pRF186 (his-Zebra-dsREDexpress SEQ ID
5 NO:112), pRF192 (his-tp10-dsREDexpress SEQ ID NO: 113), and pRF190 (his-pVEC-dsREDexpress SEQ ID NO: 114). Sequences were verified using oligo 36 (SEQ ID NO: 104).

His tagged CPP-dsREDexpress fusion proteins were expressed using standard techniques. In brief, cells were precultured in either 10ml ZYM-505 (1% N-
10 Z amine, 0.5% yeast extract, 5% glycerol, 1.0% dextrose, 25mM Na_2HPO_4, 25mM KH_2PO_4, 50mM NH_4Cl, 5mM Na_2SO_4, 1x trace metals (Teknova), $5x10^{-5}$% Thiamine, 2mM $MgCl_2$, 100μg/ml Ampicillin) or lysogeny broth (1% Tryptone, 0.5% yeast extract, 1% sodium chloride, 100μg/ml Ampicillin, 0.4% dextrose) in 125ml flasks for 12-16 hours at 37°C and 220 RPM. Precultures were diluted 1:1000(ZYM-
15 505) in 500ml ZYM-5052 (1% N-Z amine, 0.5% yeast extract, 5% glycerol, 0.5% dextrose, 2% L-arabinose, 25mM Na_2HPO_4, 25mM KH_2PO_4, 50mM NH_4Cl, 5mM Na_2SO_4, 1x trace metals (Teknova), $5x10^{-5}$% Thiamine, 2mM $MgCl_2$, 100μg/ml Ampicillin) or 1:100 (Lysis broth) in 500ml 2xYT (1.6% Tryptone, 1% Yeast extract, 0.5% NaCl, 100μg/ml ampicillin) and grown at 37°C 220RPM in 2.9L Fernbach
20 flasks to OD_{600} ~0.5. L-arabinose was added to a final concentration of 0.1% to 2x YT cultures and all cultures were shifted to 18°C 220R\PM for 20-30 hours for protein expression. Cells were harvested at 5000 RPM for 10 minutes, spent medium was discarded and cell pellets frozen at -80°C.

Cell pellets were thawed and resuspended in Denaturing lysis buffer (50mM
25 Tris pH8.0, 150mM NaCl, 8M Urea, 20mM Imidazole) and lysed via passage through a French pressure cell at 16,000 PSI twice. Solid precipitates were removed from the supernatant by centrifugation at 10,000g 4°C for 15 minutes. 20μl of clarified extract was mixed with 20μl of 2x Laemmli buffer (4% SDS, 20% Glycerol, 100mM DTT, 0.004% bromophenol blue, 125mM Tris pH 6.8), heated to 95°C for 5
30 minutes and frozen at -20°C to save for analysis. Clarified extract was mixed with 6ml of 50% (v/v) Nickel-NTA-agarose slurry for 1 hour at room temperature. Beads were pelleted from mixture at 2000 RPM for 5 minutes. Supernatant was removed and a 20μl sample was taken as for the clarified extract. The pelleted beads were

resuspended in 10ml of denaturing lysis buffer and applied to a gravity flow chromatography column. The liquid was allowed to flow out leaving a bed of packed beads. The bed was washed with a series washes using different ratios of wash buffer 1 (50mM Tris pH8.0, 150mM NaCl, 8M Urea, 20mM Imidazole) and wash buffer 2 (50mM Tris pH 8.0, 500mM NaCl, 20mM Imidazole) to step down the concentration of the denaturant (urea) and step up the concentration of NaCl and allow the protein to refold on the column. In brief the column was washed with (Buffer 1: Buffer 2): 10ml of 1:0 (8M urea 150mM NaCl), 10ml of 7:1 (7M Urea, 194mM NaCl), 10ml of 3:1 (6M Urea, 238mM NaCl) 10ml of 5:3 (5M Urea, 281mM NaCl), 10ml of 1:1 (4M Urea, 325mM NaCl), 20ml of 3:5 (3M Urea, 369mM NaCl), 20ml of 1:3 (2M Urea, 413mM NaCl), 20ml of 3:13 (1.5M Urea, 434mM NaCl), 20ml of 1:5 (1M urea, 456mM NaCl), 20ml of 1:15 (0.5M Urea, 478mM NaCl), and 30ml of 0:1 (0M Urea, 500mM NaCl). Protein was eluted in native elution buffer (50mM Tris pH8.0, 500mM NaCl, 10% Glycerol, 500mM Imidazole) in 10x 1ml fractions. Fractions containing the eluted dsREDexpress or CPP-dsREDexpress protein were red in color. Red fractions were combined and dialyzed in 10,000 MWCO regenerated cellulose dialysis membrane against 1000 volumes of dialysis buffer (50mM Tris pH 8.0, 10% glycerol) overnight at room temperature. Protein solution was removed from dialysis membrane and filter sterilized using a 0.22μM Tuffryn® membrane. 20μl of protein solution was processed as for the clarified cell extract.

Samples taken during the purification in Laemmli buffer were heated to 95°C for 5 minutes and loaded onto a 12.5% PAGE gel. The gel was run at 200 volts constant for 1 hour and stained using simply blue stain. An example of a representative PAGE gel for the purification of CPP-dsREDexpress tagged proteins is shown in Figure 9. Total protein concentration for each purified protein was determined using Pierce™ Coomassie Plus assay with bovine serum albumin as a standard. The concentration of each purified CPP-dsREDexpress fusion is given in Table 3.

Table 3: Concentration of purified dsREDexpress protein fusions.

Protein	mg/ml	μM
dsREDexpress (SEQ ID NO: 700)	3.8	137
MPG1-dsREDexpress (SEQ ID NO: 751)	0.5	17

pVEC-dsREDexpress (SEQ ID NO: 752)	2.0	68
CFFKDEL-dsREDexpress (SEQ ID NO: 753)	1.5	54
TLM-dsREDexpress (SEQ ID NO: 754)	2.5	86
Zebra-dsREDexpress (SEQ ID NO: 755)	0.5	18
pep1-dsREDexpress (SEQ ID NO: 756)	0.3	10
tp10-dsREDexpress (SEQ ID NO: 757)	0.9	33

EXAMPLE 7

Expression and purification of additional CPP-Cas9 proteins from *E. coli* cells.

The delivery of Cas9 into different cell types may require Cas9 tagged with different CPP molecules. In order to isolate various CPP-Cas9 fusion proteins different CPPs were fused to Cas9 in an *E. coli* expression vector. These proteins were expressed and purified from *E. coli* cells for use in CPP mediated delivery of Cas9/sgRNA ribonucleoprotein complex to cells.

In order to make His-CFFKDEL-Cas9 (SEQ ID NO: 115) and His-MPG1-Cas9 (SEQ ID NO: 116) fusion expression cassettes the NcoI/EcoRI fragments of pRF216 (CFFKDEL SEQ ID NO: 103) or pRF213 (MPG1 SEQ ID NO: 101) were cloned into the same sites of the Cas9 protein expression plasmid pRF48 (SEQ ID NO: 117) using standard techniques generating plasmids pRF243 (his-CFFKDEL-Cas9 SEQ ID NO: 118) and pRF238 (his-MPG1-Cas9, SEQ ID NO: 119) respectively. Correct construction of the MPG1-Cas9 or CFFKDEL-Cas9 fusion cassettes was confirmed via Sanger sequencing with oligo 36 (SEQ ID NO: 104).

His tagged CPP-Cas9 fusion proteins were expressed using standard techniques. In brief, cells were precultured in either 10ml ZYM-505 (1% N-Z amine, 0.5% yeast extract, 5% glycerol, 1.0% dextrose, 25mM Na_2HPO_4, 25mM KH_2PO_4, 50mM NH_4Cl, 5mM Na_2SO_4, 1x trace metals (Teknova), $5x10^{-5}$% Thiamine, 2mM $MgCl_2$, 100μg/ml Ampicillin) or lysogeny broth (1% Tryptone, 0.5% yeast extract, 1% sodium chloride, 100μg/ml Ampicillin, 0.4% dextrose) in 125ml flasks for 12-16 hours at 37°C and 220 RPM. Precultures were diluted 1:1000(ZYM-505) in 500ml ZYM-5052 (1% N-Z amine, 0.5% yeast extract, 5% glycerol, 0.5% dextrose, 2% L-arabinose, 25mM Na_2HPO_4, 25mM KH_2PO_4, 50mM NH_4Cl, 5mM Na_2SO_4, 1x trace metals (Teknova), $5x10^{-5}$% Thiamine, 2mM $MgCl_2$, 100μg/ml Ampicillin) or 1:100 (Lysis broth) in 500ml 2xYT (1.6% Tryptone, 1% Yeast extract, 0.5% NaCl,

100µg/ml ampicillin) and grown at 37°C 220RPM in 2.9L Fernbach flasks to OD_{600} ~0.5. L-arabinose was added to a final concentration of 0.1% to 2x YT cultures and all cultures were shifted to 18°C 220R\PM for 20-30 hours for protein expression. Cells were harvested at 5000 RPM for 10 minutes, spent medium was discarded
5 and cell pellets frozen at -80°C. Proteins were purified as described in Example 1. The final concentrations of the purified CPP-Cas9 proteins as determined by Coomasie Plus assay (Pierce™) are listed in Table 4.

Table 4: Concentration of purified CPP-Cas9 proteins.

Protein	mg/ml	µM
Zebra-Cas9 (SEQ ID NO: 758)	1.5	9
CFFKDEL-Cas9 (SEQ ID NO: 730)	4.6	28
MPG1-Cas9 (SEQ ID NO: 731)	3.8	23
pVEC-Cas9 (SEQ ID NO: 759)	2.5	15

10

EXAMPLE 8

CPP-Cas9/gRNA mediated gene targeting in *E. coli* cells.

This example demonstrates the treatment of *Escherichia coli* cells with CPP-Cas9/sgRNA ribonucleoprotein complexes with sgRNAs targeting the *galK* gene of
15 *E. coli*. The entry of the CPP-Cas9/sgRNA into the cell allows targeting and cleavage to occur within the *galK* gene leading to gene inactivation by error-prone DNA repair mechanisms which can be phenotypically monitored as resistance to galactose. This method depends on delivery of Cas9/sgRNA cargo to the cells via CPP-mediated delivery.

20 The *galK* gene of *E. coli* (SEQ ID NO: 120) is responsible for a galactose sensitive phenotype seen in *galE* mutants in the presence of the sugar galactose. As galactose enters the cell it is phosphorylated by galactokinase, the product of the *galK* gene (SEQ ID NO: 120). Galactose phosphate is toxic to the cell. In wild-type cells the galactose phosphate is further metabolized by the products of the *gale*
25 (SEQ ID NO: 121) and *galT* (SEQ ID NO: 122) genes and used as a carbon source. In *galE* or *galT* loss-of-function mutants galactose phosphate accumulates leading to cell death. Therefore, loss of function mutations in the *galK* gene can be selected

in the background of a *galE* mutant as allowing colony formation in the presence of galactose.

In order to produce sgRNA (SEQ ID NO: 135) targeting the *galK* gene (SEQ ID NO: 120) at the galK2-1 target site (SEQ ID NO: 134) an *in vitro* transcription
5 template (SEQ ID NO: 131) was produced. First a PCR product of the DNA encoding the CER domain (SEQ ID NO: 123) was amplified from pRF291 (SEQ ID NO: 125) using CER forward (SEQ ID NO: 126) and universal reverse primers (SEQ ID NO: 127) in a standard PCR reaction (SEQ ID NO: 124). The CER encoding PCR product (SEQ ID NO: 124) was purified using Zymo™ clean and concentrate
10 25 columns and eluted in 35 µl of ddH_2O. Amplification of the sgRNA *in vitro* transcription template used a multiplex PCR containing 4 primers, a universal forward primer containing the T7 promoter (SEQ ID NO: 128), a target specific forward primer containing some of the T7 promoter and some of the target site (SEQ ID NO: 129), a target reverse primer containing some of the target site and
15 overlap with the CER domain (SEQ ID NO: 130), and the universal reverse primer (SEQ ID NO: 127). A PCR reaction was run using Phusion flash master mix containing 15 nM CER domain PCR product (SEQ ID NO: 124), 1 µM each the universal forward (SEQ ID NO: 128) and reverse primers (SEQ ID NO: 127) and 300 nM each target forward (SEQ ID NO: 129) and target reverse (SEQ ID NO: 130)
20 primers. The PCR reaction was cycled as for a standard reaction. sgRNA *in vitro* transcription template (SEQ ID NO: 131) was purified using Zymo clean and concentrate 25 columns and eluted in 35 µl of ddH_2O. The sgRNA in vitro transcription template (SEQ ID NO: 131) contained the T7 promoter (SEQ ID NO: 132), the DNA encoding the galK2-1 variable targeting domain (SEQ ID NO: 133),
25 and the DNA encoding the CER domain (SEQ ID NO: 125) The *in vitro* transcription reaction to create the galK2-1 sgRNA (SEQ ID NO: 135) was performed as described in Example 2.

CPP delivery of Cas9/sgRNA nucleoprotein complexes was performed by growing a strain of *E. coli* deleted for *galE* in lysogeny broth (1% Tryptone, 0.5%
30 Yeast Extract, 1% NACl) overnight at 37°C, 220RPM. The culture was diluted 1:100 in fresh lysogeny broth and grown at 37°C, 220RPM for 2 hours to obtain cells in exponential growth phase. CPP-Cas9 (pvEC-Cas9 (SEQ ID NO: 144), Zebra-Cas9 (SEQ ID NO: 143), MPG1-Cas9 (SEQ ID NO: 116), CFFKDEL-Cas9 (SEQ ID NO:

115)) were incubated at 10µM final concentration either in the presence or absence of 10µM galK2-1 sgRNA (SEQ ID NO: 135) in a 50µl volume for 30 minutes at room temperature. For the treatment 1.2 ml of cells were pelleted at 3000 RPM for 3 minutes, supernatant was discarded and cells were resuspended in 600µl of LB
5 containing 2x nuclease buffer (200mM NaCl, 100mM Tris-HCl, 20 mM $MgCl_2$, 200 µg/ml BSA pH 7.9). 50µl of the cell suspension was mixed with each reaction as well as gRNA only control and no treatment. Samples were incubated at 37°C, 220 RPM for 4 hours. 100 µl of 10^{-3}, 10^{-4}, and 10^{-5} dilutions of the samples were plated on lysogeny broth plates to obtain a viable cell count at the end of the treatment, the
10 remainder of the reaction was plated onto lysogeny broth plates and incubated overnight at 37°C. Viable cells were counted from the 10^{-5} dilution to determine the number of viable colony forming units (CFU) plated on the sample lysogeny broth plate. The sample plates were replica plated via standard techniques to minimal A medium (1g/L $(NH_4)_2SO_4$, 4.5g/L KH_2PO_4, 10.5g/L K_2HPO_4, 0.5g/L sodium
15 Citrate·$2H_2O$, 1mM $MgSO_4·7H_2O$, 5 x 10^{-5} % Thiamine) solidified with 1.5% (w/v) Bacto agar containing 0.2% (w/v) glycerol and 0.2% (w/v) galactose as carbon sources. The plates were incubated at 37°C for 24 hours and then scored for formation of colonies. Each CFU from a *galE* strain on a plate containing galactose represents a gene inactivation event of the *galK* gene. The results of the replica
20 plating are shown in Table 5.

Table 5: Frequency of *galK* gene inactivation in *galE* mutant *E. coli* cells treated with CPP-Cas9/sgRNA.

Cas9 protein	sgRNA	CFU on galactose	CFU plated on galactose	Frequency of Gal^R CFU	Fold Frequency Gal^R/untreated Gal^R frequency
None	None	21	1.65 x 10^8	1.27 x 10^{-7}	1.00
pVEC-Cas9	None	21	1.18 x 10^8	1.78 x 10^{-7}	1.39
pVEC-Cas9	galK2-1	15	1.23 x 10^8	1.22 x 10^{-7}	0.96
MPG1-Cas9	None	22	1.34 x 10^8	1.65 x 10^{-7}	1.29
MPG1-Cas9	galK2-1	16	1.11 x 10^8	1.44 x 10^{-7}	1.13
Zebra-Cas9	None	29	1.89 x 10^8	1.53 x 10^{-7}	1.20

Zebra-Cas9	galK2-1	25	8.88×10^{7}	2.82×10^{-7}	2.21
CFFKDEL-Cas9	None	29	1.24×10^{8}	2.34×10^{-7}	1.84
CFFKDEL-Cas9	galK2-1	63	1.24×10^{8}	5.10×10^{-7}	4.00
None	galK2-1	31	1.42×10^{8}	2.19×10^{-7}	1.72

The treatment of *E. coli* cells with CPP-Cas9/sgRNA ribonucleoprotein complexes in some cases enhanced the frequency of *galK* inactivation around 4 fold over the background of untreated cells. This enhancement was not seen in cells
5 treated with only CPP-Cas9 or sgRNA only suggesting that the increased inactivation of the *galK* gene was due to the CPP-Cas9/sgRNA ribonucleoprotein entering the cell and making DNA double-stranded breaks at the galK2-1 target site within the *galK* gene.

EXAMPLE 9

10 Delivery of CPP-dsREDexpress protein to Archeal cells

In order to test the delivery of cargo using cell-penetrating peptides to Archeal cells and determine candidate CPPs that cross the archeal cell wall which includes elements that are similar to bacterial and eukaryotic cell walls (*eg.* phospholipids) and membranes and elements that are distinctly archeal (*eg.* S-
15 layer) archeal cells were treated with CPP-dsREDexpress protein fusions. The CPPs identified in this screen could be used to deliver other cargo (*eg.* Cas9/sgRNA ribonucleoprotein complex) to Archeal cells.

The archeon *Halobacterium salinarum* ATCC19700 was grown on medium 213 (250 g/L NaCl, 10 g/L $MgSO_4 \cdot 7H_2O$, 5 g/L KCl, 0.2g/L $CaCl_2 \cdot 6H_2O$, 10 g/L Yeast
20 extract, 2.5 g/L Tryptone) solidified with 1.5% Bacto agar at 37°C until colonies formed (4 days). A single colony was used to inoculate 50 ml of medium 213 in a 250ml flask. The culture was grown at 37°C 220 RPM until the OD_{600} reached approximately 0.5 indicating exponential growth phase. 100µl of cells were mixed with either No protein, 5 µM dsREDexpress (SEQ ID NO: 85), 5µM MPG1-
25 dsREDexpress (SEQ ID NO: 136), 5 µM pVEC-dsREDexpress (SEQ ID NO: 137), 5 µM CFFKDEL-dsREDexpress (SEQ ID NO: 138), 5 µM TLM-dsREDexpress (SEQ ID NO: 139), 5 µM pep1-dsREDexpress (SEQ ID NO: 141), or 5 µM tp10 dsRED-

express (SEQ ID NO: 142) in a 24 well block. Mixtures were incubated for 4 hours at 37°C 220 RPM. Cells were washed twice with medium 213 lacking tryptone and yeast extract and resusepended in 100µl of medium 213 lacking tryptone and yeast extract. Cells were analyzed for flourecense in the red channel of an Accuri C5 flow
5 cytometer to determine which CPP tags had delivered the dsREDexpress cargo to *H. salinarum* cells. The untreated cells were used to create an analysis gate for the flow cytometry data between non-red and red cells such that the gate created a false positive frequency of 0.2% of the untreated cells falling in the red gate (Table 6).

10 Table 6: CPP delivery of dsREDexpress to *H. salinarum*.

Treatment	Percent of population in red cell gate ± standard deviation[1]	Fold increase in red population over dsREDexpress alone
No dsREDexpress	0.21±0.06	0.73
dsREDexpress	0.29±0.21	1.00
MPG1-dsREDexpress	0.37±0.08	1.27
pVEC-dsREDexpress	16.87±9.90	57.50
CFFKDEL-dsREDexpress	0.33±0.14	1.14
TLM-dsREDexpress	2.03±1.02	6.93
pep1-dsREDexpress	0.36±0.18	1.23
tp10-dsREDexpress	0.91±0.27	3.09

[1]Data represents three replicates ± standard deviation.

The delivery of the dsREDexpress cargo into archeal cells demonstrates that at least three of the cell-penetrating peptides (pVEC, TLM, tp10) are capable of delivering a protein cargo to the archeal cells with an efficiency as high as more
15 than 50 fold that of the delivery of the dsREDexpress protein alone suggesting that these three CPP motifs can be used to deliver other cargo to archeal cells (*eg*. Cas9 ribonucleoprotein complex). Additionally the CPP motifs deliver cargo to as much 16% of the entire cell population suggesting that deliver of cargo by CPP to archeal cells is an efficient process.

20 EXAMPLE 10

Delivery of CPP-dsREDexpress protein to eukaryotic cells

To test the ability of cell-penetrating peptides to deliver cargo to different eukaryotic species a panel of three species, *Phytophthora capsici* (Oomycete),

Septori tritici (True Fungus), and *Botrytis cinerea* (True Fungus) was treated with various CPP-dsREDexpress fusions. The delivery of dsREDexpress cargo was monitored for various CPP moieties by FACS analysis to determine the percentage of cells to which the cargo was delivered. CPPs that are capable of delivering the
5 dsREDexpress cargo to these cells which suggests that the CPPs would be capable of delivering other cargos to these classes of eukaryotic cells (*eg*. Cas9/sgRNA ribonucleoprotein complex).

P. capsici was grown on V8 medium (20% V8 juice, 4.5g/L $CaCO_3$) solidified with 1.8% Bacto Agar at 23°C in the dark for 3 days. The plate was then placed in
10 the light at 23°C for an additional 7 days. Plates were chilled at 4°C for 30 minutes. Water was placed on the plate to just cover the surface and allowed to incubate for 30 minutes at room temperature. Liquid was removed to harvest zoospores. Zoospores were confirmed via microscopic analysis. An equal volume of 2x encystment medium (40g/L Tryptone, 10g/L Yeast extract, 200ml/L 10x SOC salts
15 [5.84g/L NaCl, 1.86g/L KCl, 20.3g/L $MgCl_2 \cdot 6H_2O$, 24.6g/L $MgSO_4 \cdot 7H_2O$, 36g/L Dextrose], 36.4g/L Sorbitol, 1.47g/L $CaCl_2 \cdot 2H_2O$) was added to the zoospores and gently mixed. Zoospores in enzystment medium were incubated for 20 minutes at room temperature. Encystment was confirmed microscopically. Spores were pelleted and resuspended in an equal volume of YMA medium (2g/L Yeast extract,
20 4g/L Malt extract) and counted using a hemocytometer. Zoospores were diluted to $3x\ 10^7$ spores/ml in YMA. 100µl of Zoospores in YMA were mixed with various dsREDexpress fusion proteins (New example 5, table N1) to a final concentration of 5µM protein. Mixtures were incubated at 25°C 400 RPM for 2 hours. Cells were washed twice with phosphate buffered saline (PBS) (8g/L NaCl, 0.2g/L KCl, 1.44g/L
25 $Na_2HPO_4 \cdot 2H_2O$, 0.24g/L KH_2PO_4 pH 6.8) and resuspended in a final volume of 200µl PBS. Uptake of dsREDexpress fusion proteins was monitored using flow cytometry as for *Halobacterium salinarium* (Example 9). The percent of cells to which the cargo was successfully delivered was determined by drawing an arbitrary gate in the dsREDexpress treated cells such that 0.1% of the population scored as a
30 false positive red event (1:1000 cells). The results of this treatment can be seen in Table 7. pVEC, pep1, and tp10 produce 5.8, 5.5, and 1.8 fold more red cells than the dsREDexpress treated cells alone suggesting that these CPP moieties might be

candidates for delivering other cargo to Oomycetes (*eg*. Cas9/sgRNA ribonucleoprotein complex)

Table 7: CPP delivery of dsREDexpress to *Phytophora capsici.*

Treatment	Percent of population in red cell gate ± standard deviation[1]	Fold increase in red population over dsREDexpress alone
dsREDexpress	0.10±0.03	1.00
pVEC-dsREDexpress	0.56±0.16	5.79
CFFKDEL-dsREDexpress	0.01±0.01	0.07
TLM-dsREDexpress	0.00±0.00	0.00
pep1-dsREDexpress	0.53±0.29	5.52
Tp10-dsREDexpress	0.17±0.14	1.76
MPG-dsREDexpress	0.00±0.00	0.00
Zebra-dsREDexpress	0.03±0.05	0.34

5 [1]Data represents three biological replicates ± standard deviation

B. cinerea was grown on PDA medium (24g/L potato dextrose broth) solidified with 1.8% Bacto agar in the dark for 5 to 10 days. Conidia were harvested in water with a sterile plastic spreader and filtered through 2 layers of cheesecloth.
10 Conidia were counted on a hemocytometer and diluted to 3 x 10^7conidia per ml in YMA medium. 100µl of conidia in YMA were mixed with various dsREDexpress fusion proteins (New example 5, table N1) to a final concentration of 5µM protein. Mixtures were incubated at 25°C 400 RPM for 2 hours. Cells were washed twice with phosphate buffered saline (PBS) (8g/L NaCl, 0.2g/L KCl, 1.44g/L
15 $Na_2HPO_4{\cdot}2H_2O$, 0.24g/L KH_2PO_4 pH 6.8) and resuspended in a final volume of 200µl PBS. Uptake of dsREDexpress fusion proteins was monitored using flow cytometry as for *Halobacterium salinarium* (Example 8). The percent of cells to which the cargo was successfully delivered was determined by drawing an arbitrary gate in the dsREDexpress treated cells such that 0.1% of the population scored as a
20 false positive red event (1:1000 cells). The results of this treatment can be seen in Table8.

Table 8: CPP delivery of dsREDexpress to *Botrytis cinerea*

Treatment	Percent of population in red cell gate ± standard deviation[1]	Fold increase in red population over dsREDexpress alone
dsREDexpress	0.12±0.04	1.00
pVEC-dsREDexpress	0.08±0.10	0.68
CFFKDEL-dsREDexpress	0.03±0.01	0.22
TLM-dsREDexpress	0.01±0.01	0.05
pep1-dsREDexpress	0.01±0.01	0.05
Tp10-dsREDexpress	0.03±0.02	0.24
MPG-dsREDexpress	0.01±0.02	0.11
Zebra-dsREDexpress	0.01±0.02	0.11

[1]Data represents three biological replicates ± standard deviation

S. tritici was grown on YMA medium solidified with 1.8% Bacto agar at 23°C
5 in light. Conidia were harvested after 5 to 10 days with a sterile plastic spreader and water. Conidia was counted on a hemocytometer and diluted to 3 x 10^7 conidia in YMA medium. 100µl of conidia in YMA were mixed with various dsREDexpress fusion proteins (New example 5, table N1) to a final concentration of 5µM protein. Mixtures were incubated at 25°C 400 RPM for 2 hours. Cells were washed twice
10 with phosphate buffered saline (PBS) (8g/L NaCl, 0.2g/L KCl, 1.44g/L $Na_2HPO_4 \cdot 2H_2O$, 0.24g/L KH_2PO_4 pH 6.8) and resuspended in a final volume of 200µl PBS. Uptake of dsREDexpress fusion proteins was monitored using flow cytometry as for *Halobacterium salinarium* (Example 9). The percent of cells to which the cargo was successfully delivered was determined by drawing an arbitrary
15 gate in the dsREDexpress treated cells such that 0.1% of the population scored as a false positive red event (1:1000 cells). The results of this treatment can be seen in Table 9. pVEC, TLM, pep1, and tp10 increased the delivery of dsREDexpress 25, 4, 3, and 5 fold respectively compared to dsREDexpress alone. This suggests that these CPPs would be good candidates for the delivery of other cargo to True fungi
20 (*eg.* Cas9/sgRNA ribonucleoprotein complex).

Table 9: CPP delivery of dsREDexpress to *Septoria tritici*

Treatment	Percent of population in red cell gate ± standard deviation[1]	Fold increase in red population over dsREDexpress alone
dsREDexpress	0.12±0.03	1.00
pVEC-dsREDexpress	3.02±0.91	25.2
CFFKDEL-dsREDexpress	0.00±0.01	0.03
TLM-dsREDexpress	0.48±0.14	4.03
pep1-dsREDexpress	0.37±0.21	3.06
Tp10-dsREDexpress	0.71±0.69	5.94
MPG-dsREDexpress	0.14±0.05	1.17
Zebra-dsREDexpress	0.00±0.00	0.00

[1]Data represents three biological replicates ± standard deviation

EXAMPLE 11

5 Delivery of seven CPPs-dsRED and two CPPs-tagRFP into seven gut bacteria

In this example, the efficiency of CPPs in delivering two cargo proteins, dsRED and tag RFP, into 7 gut bacterial species (whose beneficial effects on host physiology have been demonstrated) was tested .

Bacterial cells were grown in appropriate media (see Table 10) overnight at 37˚C in
10 a rotary shaker at 150 rpm in an anaerobic tent (80% N_2, 15% CO_2, and 5% H_2). For the assay, $1x10^8$ bacterial cells were mixed with a final concentration of 5 uM of CPPs-dsRED and CPPs-tagRFP proteins in a 96 well plate, followed by two hours outgrowth at 37˚C. To measure the dsRED and RFP fluorescence signals in cells, bacterial cells were harvested by centrifugation (3,500 × g, 4°C, 20 min) and
15 washed twice in phosphate buffered saline (100 ul per well). Fluorescence intensities were quantitated with Tecan Spark 10M plate reader (Tecan, Männedorf, Switzerland) equipped with 554 nm excitation and 586 nm emission filters with 10 nm bandwidth. Raw fluorescence values were subtracted from that of the untreated cells (background). The fluorescence intensity values of 7000 as a minimum cutoff
20 was taken for delivery of CPPs inside the cells.

Table 10. Culture medium of 7 bacterial species

Bacteria	Phylum	Culture medium
Bacteroides thetaiotaomicron	Bacteroidetes	Brain and Heart Infusion supplemented with 10% bovine blood (Blood BHI)
Eubacterium hallii	Firmicutes	Blood BHI
Faecalibacterium prausnitzii	Firmicutes	Blood BHI
Blautia hydrogenotrophica	Firmicutes	YCFA
Bacteroides fragilis	Bacteroidetes	Blood BHI
Prevotella histicola	Bacteroidetes	Blood BHI
Clostridium scindens	Firmicutes	YCFA

As shown in Table 11, these results indicate that five CPPs including MPG, pVEC, TLM, ZEBRA, and pep1 were effectively delivered into the anaerobic gut bacteria
5 belonging to the phyla Firmicutes and Bacteroidetes, thereby indicating that the CPP's can traverse through the cell membrane of these (Table 9).

Table 11. Differential delivery efficiencies of CPPs in different bacterial strains as demonstrated by the fluorescence intensity above the cutoff value of 7000

	MPG-1-dsRED	pVEC-dsRED	TLM-dsRED	ZEBRA-dsRED	pep1-dsRED
Bacteroides thetaiotaomicron	-	-	-	10230	16657
Eubacterium hallii	10015	17156	-	16894	7004
Faecalibacterium prausnitzii	-	40525	14998	17014	12696
Blautia hydrogenotrophica	-	11770	14612	9623	-

Bacteroides fragilis	-	14783	-	15026	-
Prevotella histicola	-	-	-	22416	-
Clostridium scindens	-	17677	32492	-	-

CLAIMS

What is claimed is:

1. A composition comprising at least one protein component of an RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP), wherein said protein component and CPP are covalently, or non-covalently, linked to each other in an RGEN protein-CPP complex, and wherein said RGEN protein-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a microbial cell.

2. The composition of claim 1, wherein the protein component of the RGEN is associated with at least one RNA component that comprises a sequence complementary to a target site sequence on a chromosome or episome in the microbial cell, wherein the RGEN can bind to the target site sequence, and optionally cleave one or both DNA strands at the target site sequence.

3. The composition of claim 2, wherein the RNA component comprises a guide RNA (gRNA) comprising a CRISPR RNA (crRNA) operably linked to a trans-activating CRISPR RNA (tracrRNA).

4. The composition of claim 2, wherein the RGEN can cleave one or both DNA strands at the target site sequence.

5. The composition of claim 1, wherein the RGEN comprises a CRISPR-associated (Cas) protein-9 (Cas9) amino acid sequence.

6. The composition of claim 1, wherein the RGEN protein component and CPP are covalently linked.

7. The composition of claim 1, wherein the RGEN protein component and CPP are non-covalently linked.

8. The composition of claim 1, wherein the CPP is cationic or amphipathic.

9. The composition of claim 1, wherein the CPP comprises:
   (i) a CPP from an Epstein-Barr virus Zebra trans-activator protein,
   (ii) a CPP having 6 or more contiguous arginine residues,
5 (iii) a transportan-10 (TP10) CPP, or
   (iv) a CPP from a vascular endothelium cadherin protein.

10. The composition of claim 1, wherein said RGEN protein-CPP complex can traverse a cell wall and cell membrane of a microbial cell.
10

11. A cell comprising the composition according to claim 1.

12. A method of delivering a protein component of an RNA-guided endonuclease (RGEN) into a microbial cell, said method comprising:
15 contacting the microbial cell with a composition comprising the protein component of the RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP),
wherein said protein component and CPP are covalently, or non-covalently, linked to each other in an RGEN protein-CPP complex,
20 wherein said RGEN protein-CPP complex traverses (i) a cell membrane, or (ii) a cell wall and cell membrane, of the cell, thereby entering the microbial cell.

13. The method of claim 12, wherein:
25 (i) the composition further comprises at least one RNA component that is associated with the protein component of the RGEN; or
   (ii) the microbial cell comprises the RNA component, wherein the RNA component associates with the protein component of the RGEN after the RGEN protein-CPP complex enters the microbial cell;
30 wherein the RNA component comprises a sequence complementary to a target site sequence on a chromosome or episome in the cell, wherein the RGEN can bind to the target site sequence, and optionally cleave one or both DNA strands at the target site sequence.

14. The method of claim 13, wherein the RGEN can cleave one or both DNA strands at the target site sequence.

5 15. The method of claim 14, wherein the microbial cell further comprises a donor polynucleotide comprising at least one sequence homologous to a sequence at or near the target site sequence, and wherein the donor polynucleotide integrates at or near the target site sequence by homologous recombination.

10 16. A polynucleotide sequence comprising a nucleotide sequence encoding an RGEN protein-CPP fusion protein that comprises a protein component of an RNA-guided endonuclease (RGEN) and at least one cell-penetrating peptide (CPP), wherein optionally, said nucleotide sequence is operably linked to a promoter sequence.

15

17. A method of producing an RGEN protein-CPP fusion protein comprising:
(a) providing the polynucleotide sequence of claim 16;
(b) expressing the RGEN protein-CPP fusion protein from the polynucleotide sequence, thereby producing the RGEN protein-CPP
20 fusion protein, wherein said expressing is optionally performed in a cell; and
(c) optionally, isolating the RGEN protein-CPP fusion protein produced in step (b).

25 18. A composition comprising at least one protein component of a guide polynucleotide/Cas endonuclease complex and at least one cell-penetrating peptide (CPP), wherein said protein component and CPP are covalently, or non-covalently, linked to each other in a guide polynucleotide/Cas endonuclease-CPP complex, and wherein said guide polynucleotide/Cas
30 endonuclease-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of a microbial cell.

19. A method for modifying a target site in the genome of a cell, the method comprising providing a guide polynucleotide, a cell-penetrating peptide (CPP) and a Cas endonuclease to the cell, wherein said guide polynucleotide, Cas endonuclease and CPP are covalently, or non-covalently, linked to each
5 other in a guide polynucleotide/Cas endonuclease-CPP complex, and wherein said guide polynucleotide/Cas endonuclease-CPP complex can traverse (i) a cell membrane, or (ii) a cell wall and cell membrane, of the microbial cell

FIGURE 1

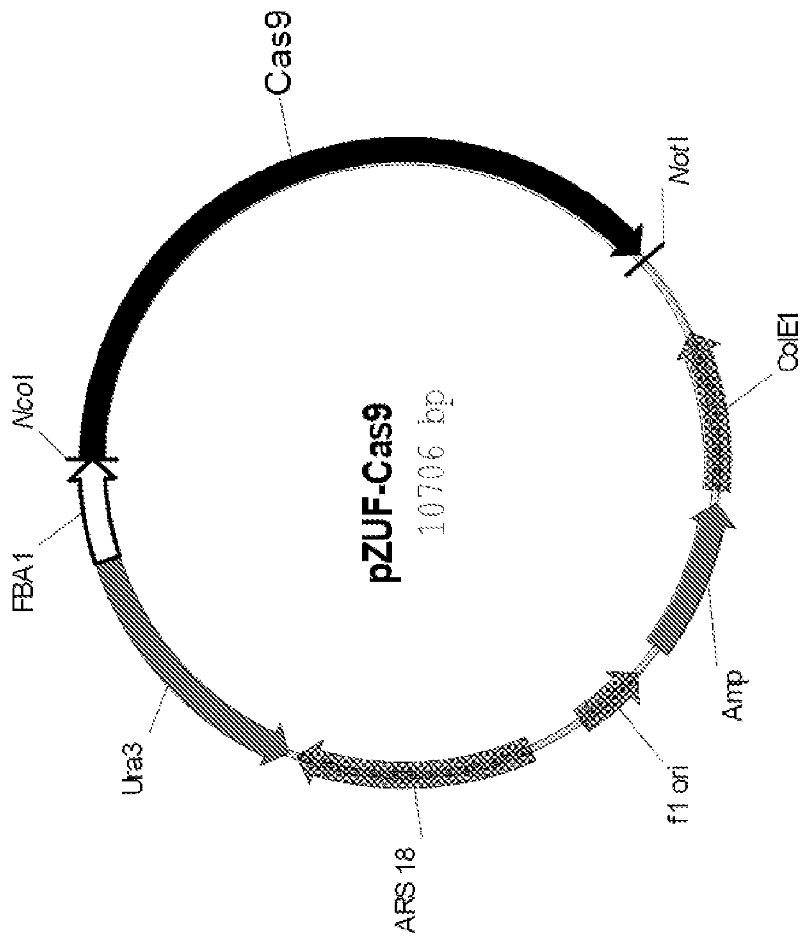


FIGURE 2B

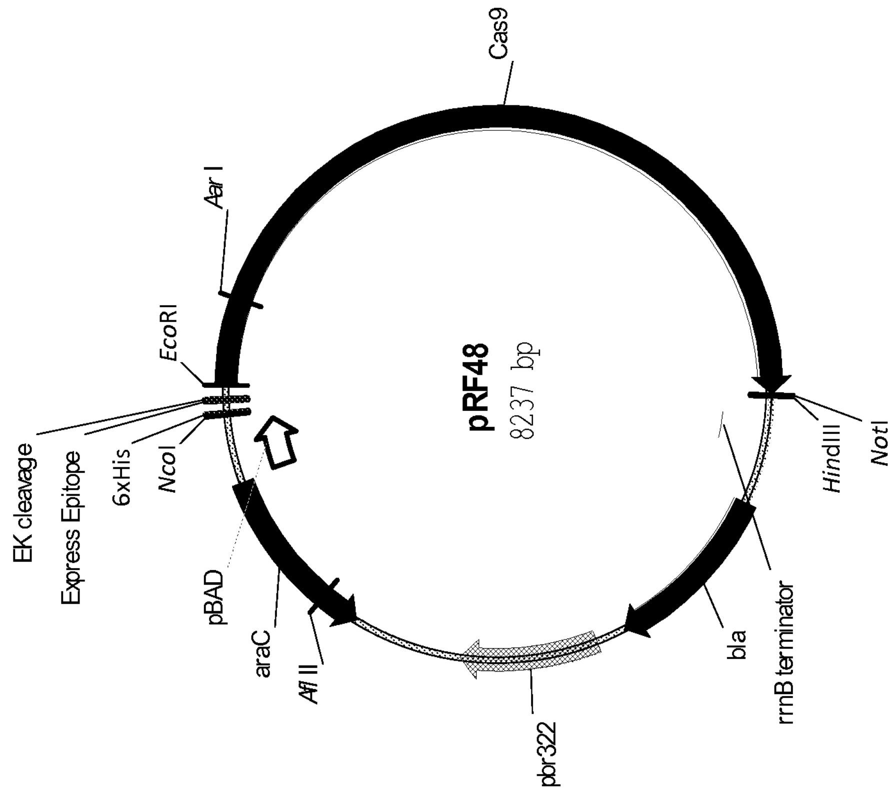
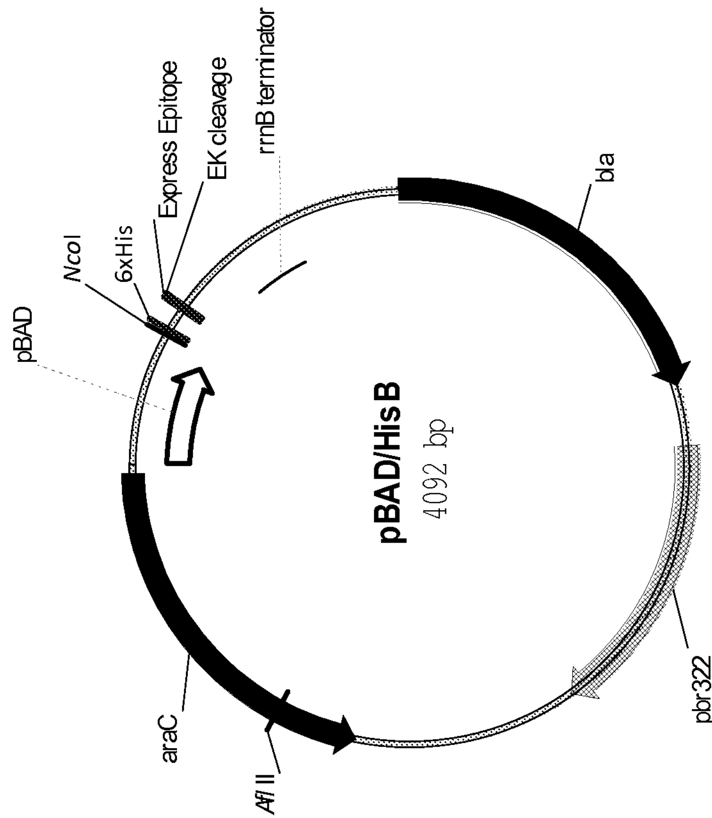


FIGURE 2A



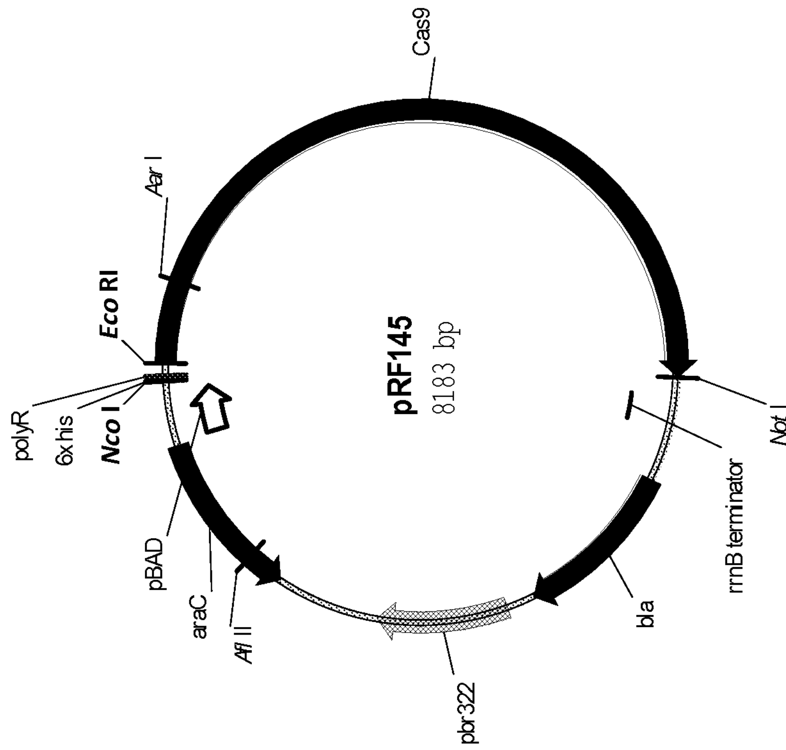


FIGURE 3B

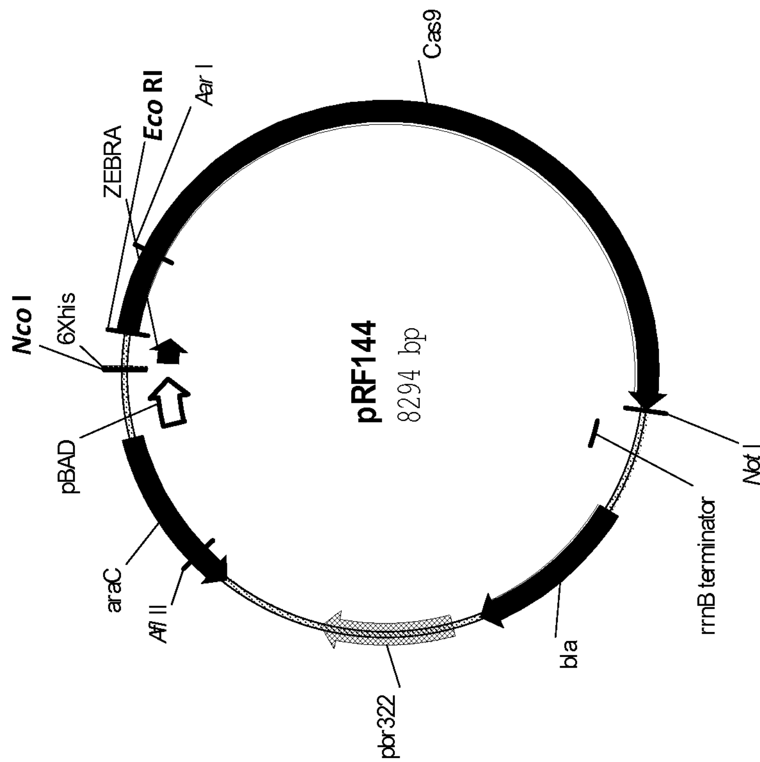


FIGURE 3A

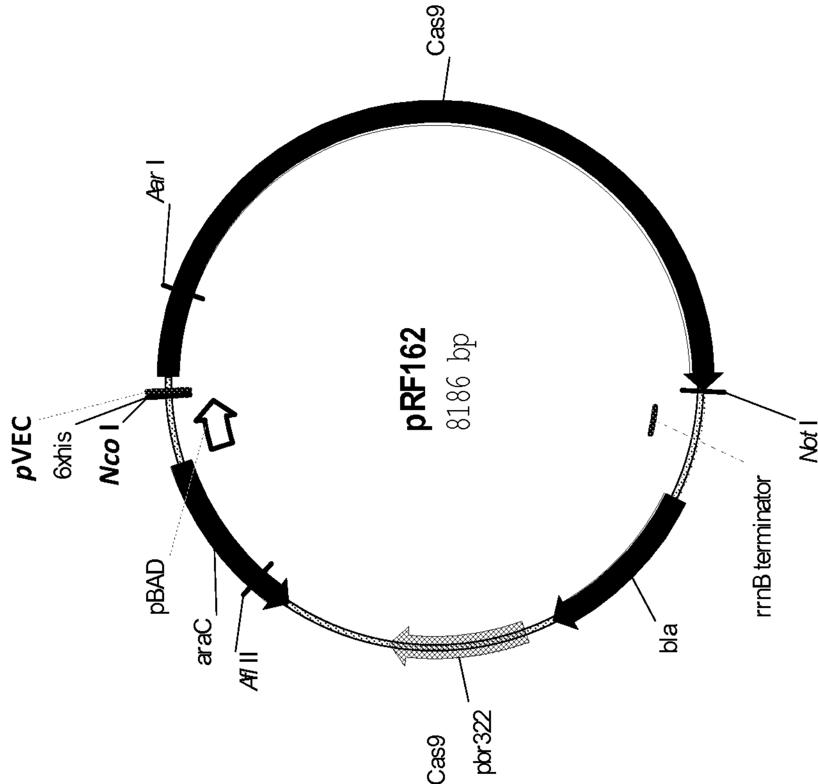


FIGURE 3D

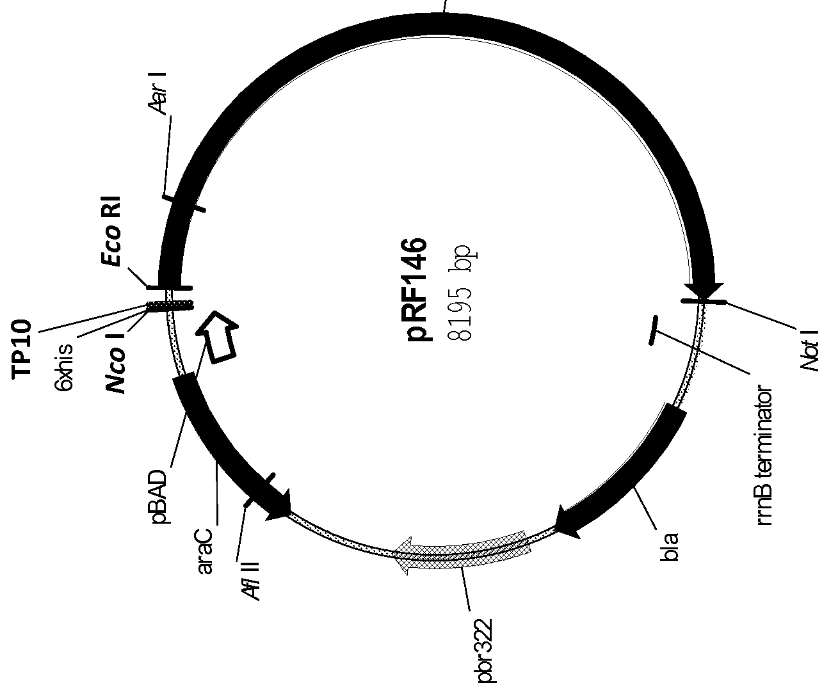


FIGURE 3C

FIGURE 4

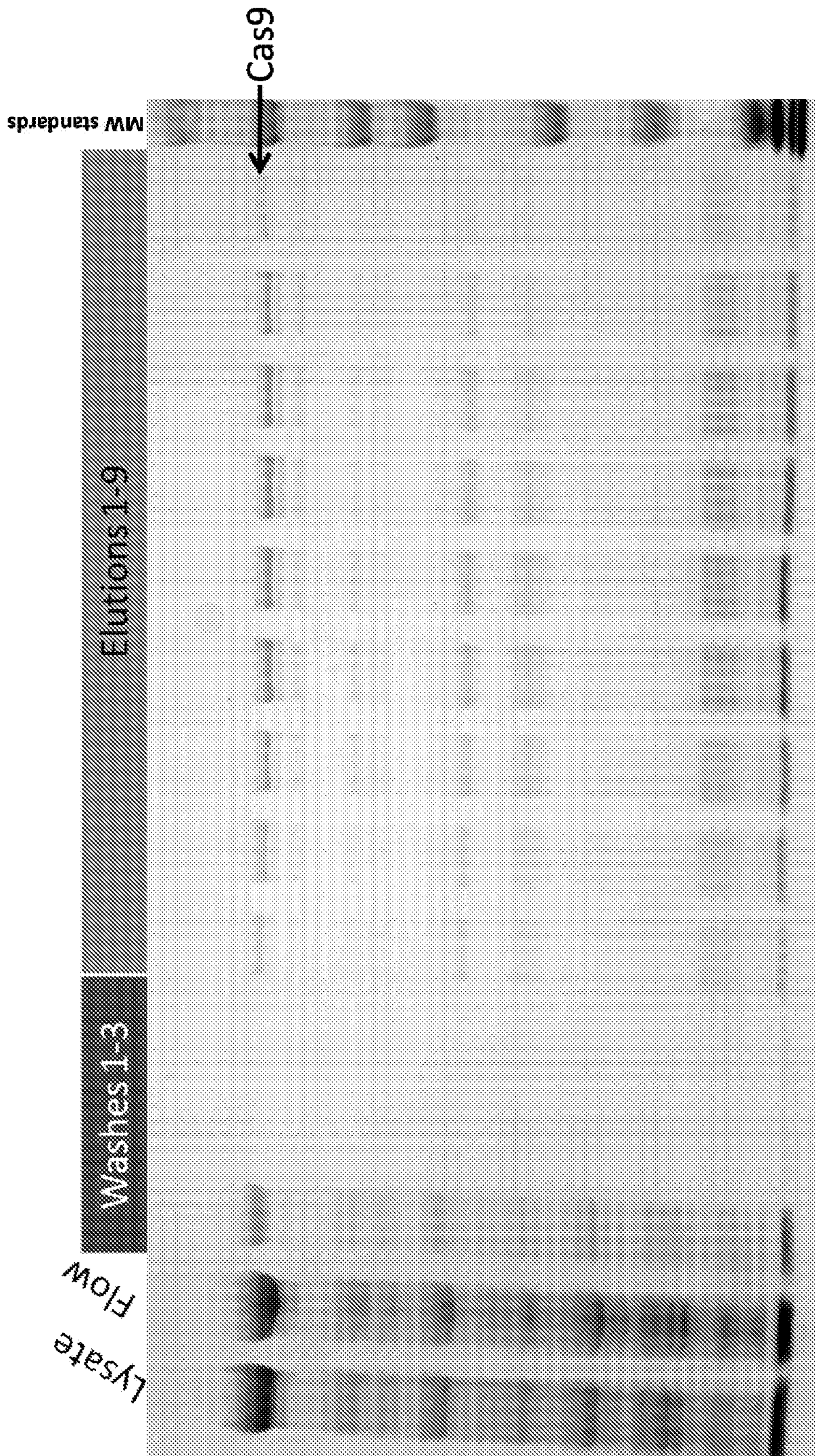


FIGURE 5

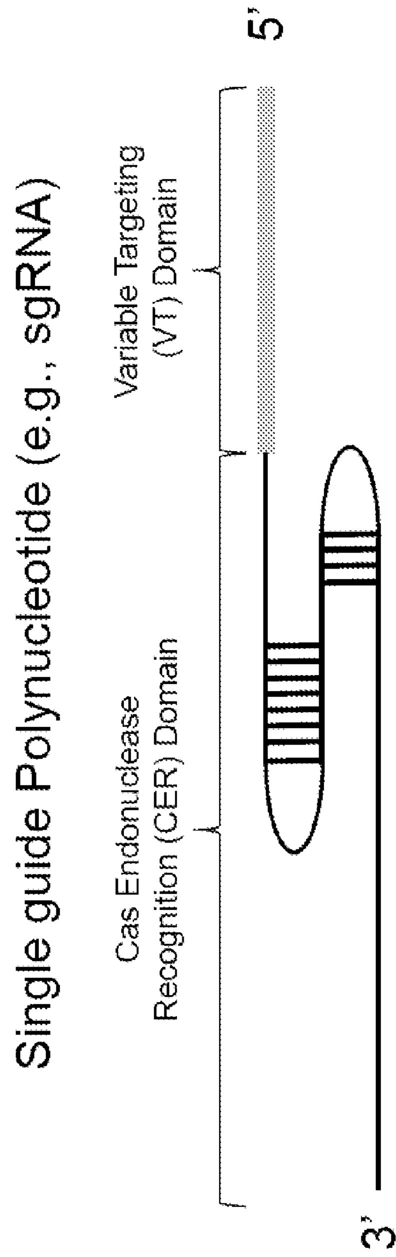


FIGURE 6

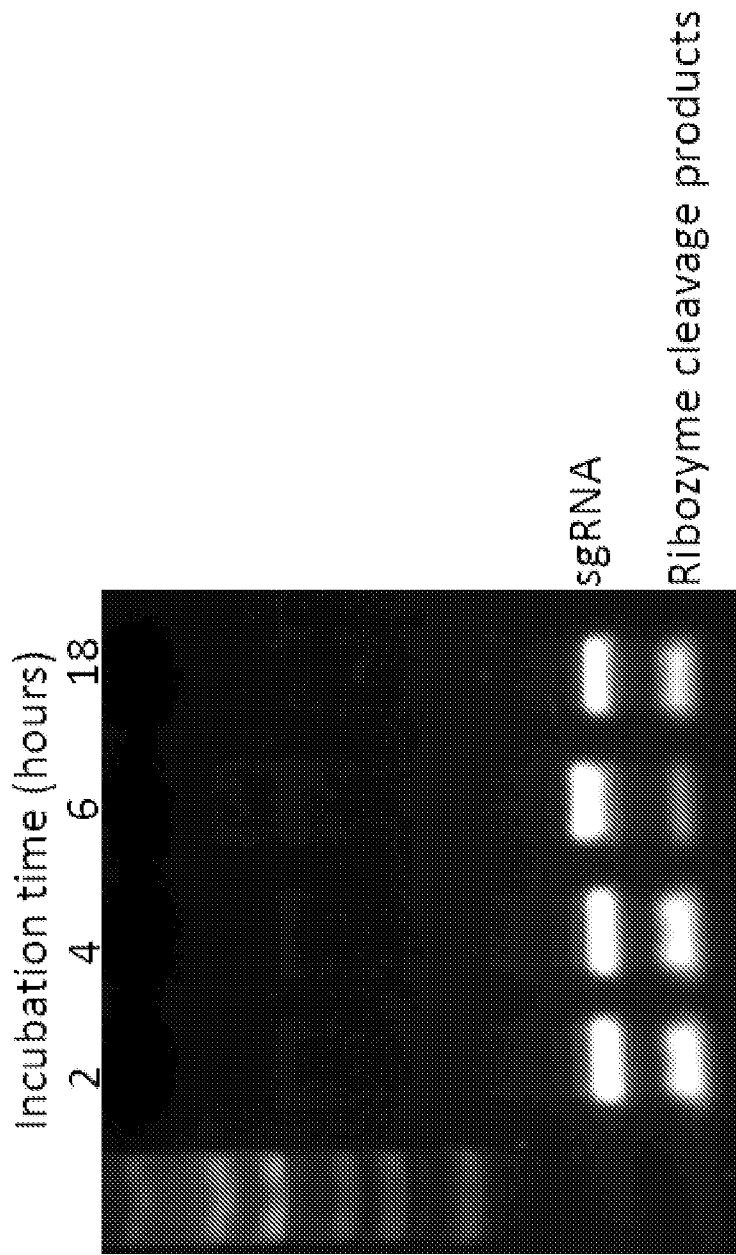


FIGURE 7

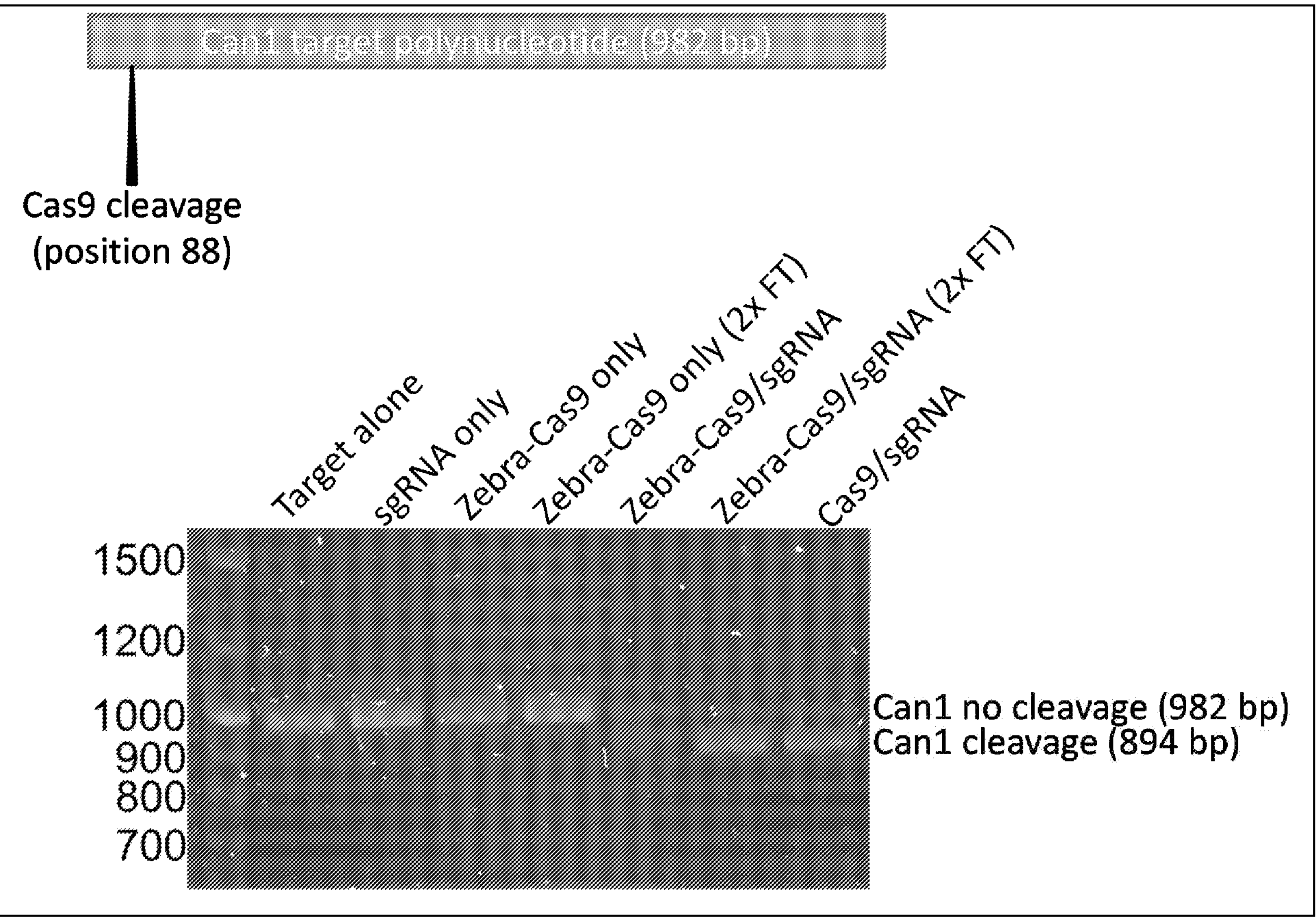


FIGURE 8

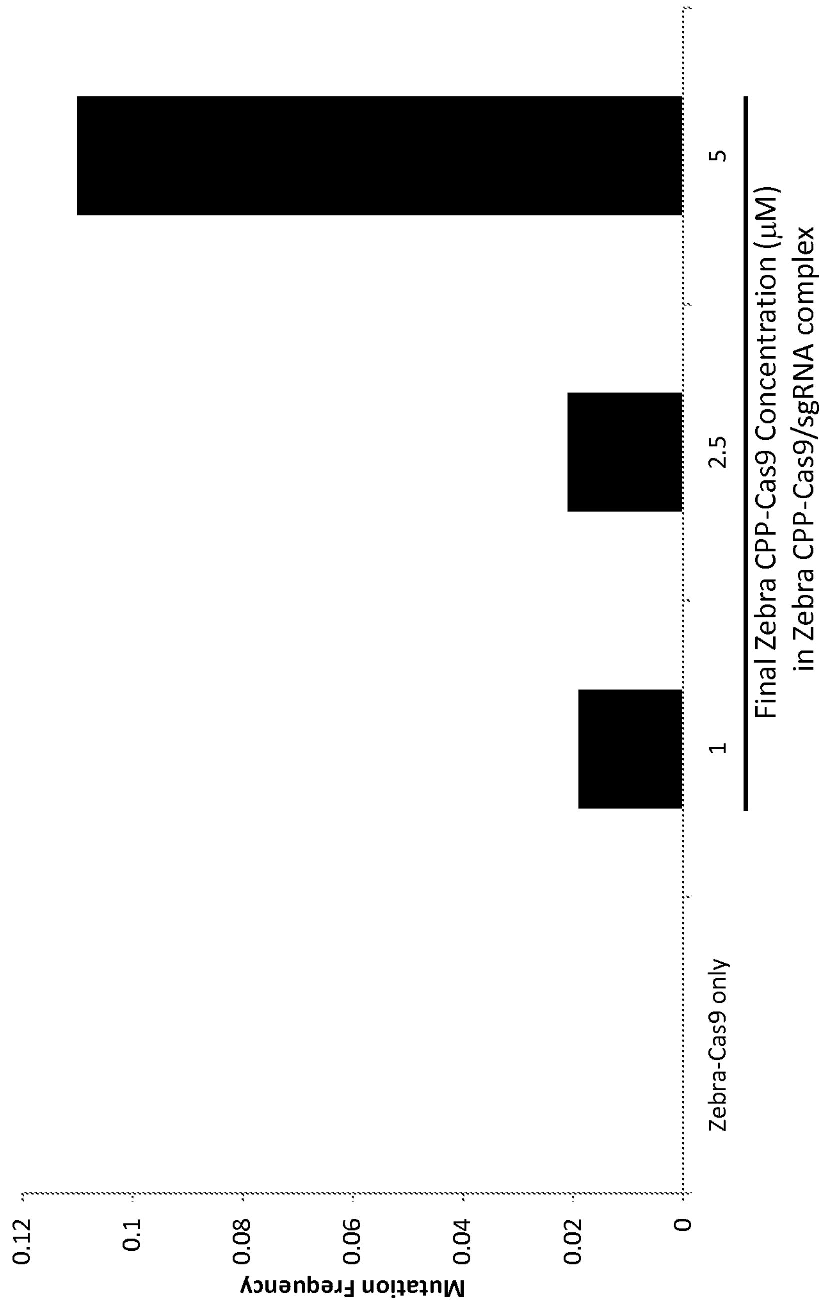
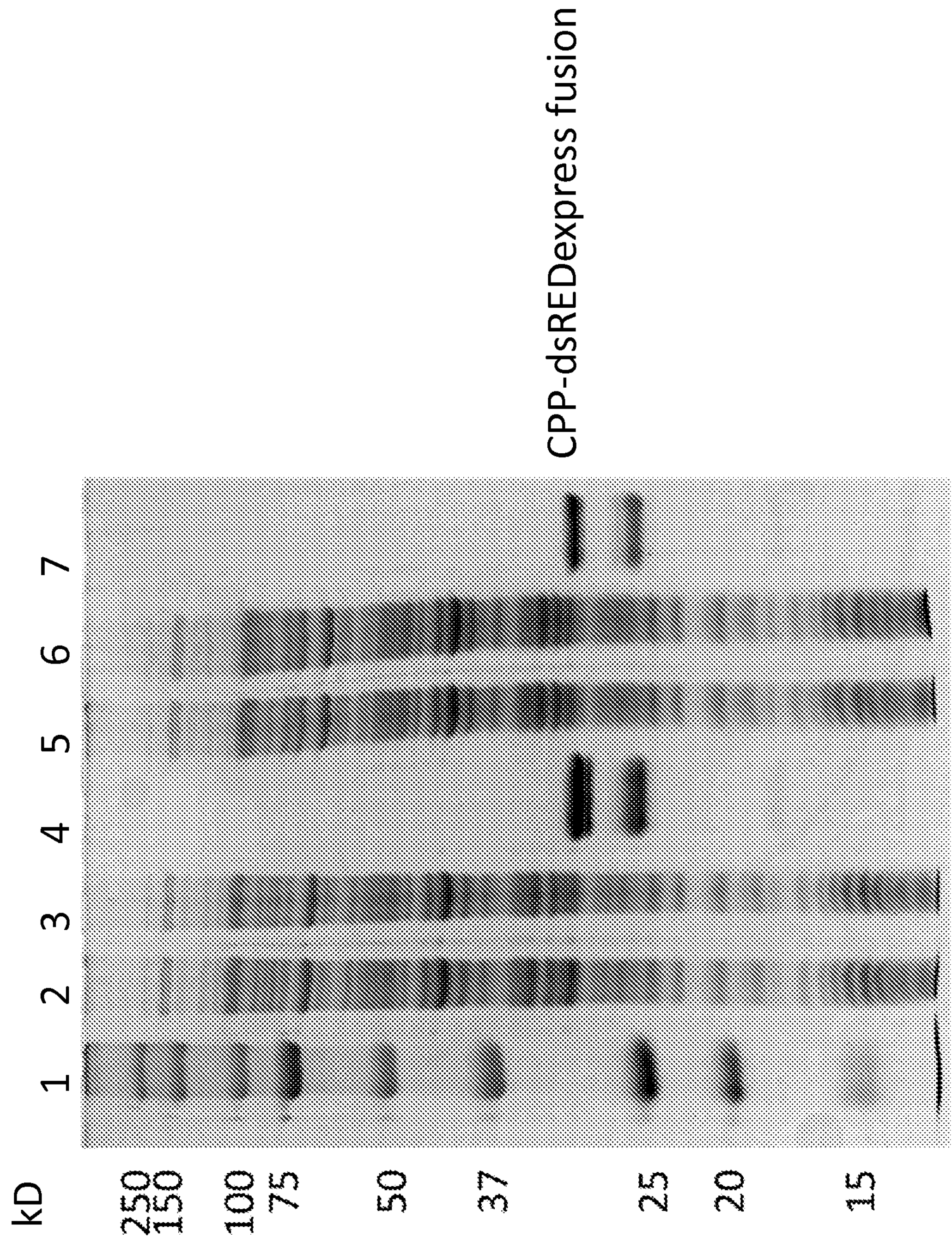


FIGURE 9



INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/058760

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/22 C12N15/62
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data, EMBASE, BIOSIS, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	S. RAMAKRISHNA ET AL: "Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA", GENOME RESEARCH, vol. 24, no. 6, 2 April 2014 (2014-04-02), pages 1020-1027, XP055128944, ISSN: 1088-9051, DOI: 10.1101/gr.171264.113 the whole document -----	1-19
X	WO 2014/065596 A1 (TOOLGEN INC [KR]) 1 May 2014 (2014-05-01) paragraphs [0154], [0217]; claim 18 figure 13; example 5 -----	1-19
	-/--	

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 January 2016	01/02/2016

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Wiame, Ilse

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/058760

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2014/089290 A1 (SIGMA ALDRICH CO LLC [US]; CHEN FUQIANG [US]; DAVIS GREGORY D [US]; KA) 12 June 2014 (2014-06-12) paragraphs [0016], [0024], [0060], [0068]; claim 4 -----	1,5,6, 8-11, 16-18
X	WO 2014/165825 A2 (HARVARD COLLEGE [US]; CHILDRENS MEDICAL CENTER [US]) 9 October 2014 (2014-10-09) paragraph [0152] - paragraph [0155] -----	1,5,6,8, 9,11, 16-18
X	WO 2013/176772 A1 (UNIV CALIFORNIA [US]; UNIV VIENNA [AT]; DOUDNA JENNIFER A [US]; JINEK) 28 November 2013 (2013-11-28) cited in the application paragraph [0289] paragraph [0265] paragraph [0254] - paragraph [0256] -----	1,5,6,8, 9,11, 16-18
A	ROBERTA MARCHIONE ET AL: "ZEBRA cell-penetrating peptide as an efficient delivery system in Candida albicans", BIOTECHNOLOGY JOURNAL, vol. 9, no. 8, 12 August 2014 (2014-08-12), pages 1088-1094, XP055243411, DE ISSN: 1860-6768, DOI: 10.1002/biot.201300505 the whole document -----	9
A	NEKHOTIAEVA N ET AL: "Cell entry and antimicrobial properties of eukaryotic cell-penetrating peptides", THE FASEB JOURNAL, FEDERATION OF AMERICAN SOCIETIES FOR EXPERIMENTAL BIOLOGY, BETHESDA; US, vol. 18, 1 February 2004 (2004-02-01), pages 394-396, XP003024027, ISSN: 0892-6638 the whole document -----	9
A	J. E. DICARLO ET AL: "Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems", NUCLEIC ACIDS RESEARCH, vol. 41, no. 7, 4 March 2013 (2013-03-04), pages 4336-4343, XP055086617, ISSN: 0305-1048, DOI: 10.1093/nar/gkt135 abstract -----	19
X,P	WO 2015/086795 A1 (CELLECTIS [FR]) 18 June 2015 (2015-06-18) the whole document -----	1-19
	-/--	

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	DATABASE WPI Week 201535 Thomson Scientific, London, GB; AN 2015-20198F XP002753224, -& CN 104 328 138 A (SHANGHAI DIDA BIOLOGICAL TECHNOLOGY CO) 4 February 2015 (2015-02-04) abstract -----	1-6,8, 11-19

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2015/058760

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2014065596 A1	01-05-2014	AU 2013335451 A1	07-05-2015
		AU 2015218519 A1	17-09-2015
		CA 2888190 A1	01-05-2014
		CN 104968784 A	07-10-2015
		EP 2912175 A1	02-09-2015
		KR 20150101446 A	03-09-2015
		KR 20150101476 A	03-09-2015
		KR 20150101477 A	03-09-2015
		KR 20150101478 A	03-09-2015
		US 2015284727 A1	08-10-2015
		US 2015322457 A1	12-11-2015
		US 2015344912 A1	03-12-2015
		WO 2014065596 A1	01-05-2014
WO 2014089290 A1	12-06-2014	AU 2013355214 A1	04-06-2015
		CA 2891347 A1	12-06-2014
		CN 105142669 A	09-12-2015
		EP 2928496 A1	14-10-2015
		KR 20150091052 A	07-08-2015
		WO 2014089290 A1	12-06-2014
WO 2014165825 A2	09-10-2014	EP 2981617 A2	10-02-2016
		US 2015071889 A1	12-03-2015
		US 2015176013 A1	25-06-2015
		WO 2014165825 A2	09-10-2014
WO 2013176772 A1	28-11-2013	AU 2013266968 A1	20-11-2014
		CA 2872241 A1	28-11-2013
		CN 104854241 A	19-08-2015
		EA 201401319 A1	29-05-2015
		EP 2800811 A1	12-11-2014
		GB 2518764 A	01-04-2015
		HK 1204003 A1	06-11-2015
		JP 2015523856 A	20-08-2015
		KR 20150016588 A	12-02-2015
		PE 03362015 A1	25-03-2015
		PH 12014502574 A1	21-01-2015
		SG 11201407702X A	30-12-2014
		US 2014068797 A1	06-03-2014
		WO 2013176772 A1	28-11-2013
WO 2015086795 A1	18-06-2015	NONE	
CN 104328138 A	04-02-2015	NONE	