

manipulation by the effector protein(s). The crRNA consists of a direct repeat responsible for protein binding to the crRNA and a spacer sequence that is complementary to the desired nucleic acid target sequence. CRISPR systems can be reprogrammed to target alternative DNA or RNA targets by modifying the spacer sequence of the crRNA.

CRISPR-Cas systems can be broadly classified into two classes: Class 1 systems are composed of multiple effector proteins that together form a complex around a crRNA, and Class 2 systems consists of one effector protein that complexes with the RNA guide to target nucleic acid substrates. The single-subunit effector composition of the Class 2 systems provides a simpler component set for engineering and application translation and have thus far been an important source of programmable effectors. Nevertheless, there remains a need for additional programmable effectors and systems for modifying nucleic acids and polynucleotides (*i.e.*, DNA, RNA, or any hybrid, derivative, or modification) beyond the current CRISPR-Cas systems, such as smaller effectors and/or effectors having unique PAM sequence requirements, that enable novel applications through their unique properties.

SUMMARY

In one aspect, provided herein is an engineered, non-naturally occurring Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) - Cas system comprising: (a) a CRISPR-associated protein or a nucleic acid encoding the CRISPR-associated protein, wherein the CRISPR-associated protein comprises an amino acid sequence of SEQ ID NO: 241; and (b) an RNA guide comprising a direct repeat sequence and a spacer sequence capable of hybridizing to a target nucleic acid, or a nucleic acid encoding the RNA guide; wherein the CRISPR-associated protein is capable of binding to the RNA guide and of modifying the target nucleic acid sequence complementary to the spacer sequence.

In another aspect, provided herein is a cell, wherein the cell comprises: (a) a CRISPR-associated protein or a nucleic acid encoding the CRISPR-associated protein, wherein the CRISPR-associated protein comprises the amino acid sequence of SEQ ID NO: 241; and (b) an RNA guide comprising a direct repeat sequence and a spacer sequence capable of hybridizing to a target nucleic acid, or a nucleic acid encoding the RNA guide.

In another aspect, provided herein is an *in vitro* method of binding the system of any previous claim to a target nucleic acid in a cell comprising: (a) providing the system; and (b) delivering the system to the cell, wherein the cell comprises the target nucleic acid, wherein the CRISPR-associated-protein binds to the RNA guide, and wherein the spacer sequence binds to the target nucleic acid.

In another aspect, provided herein is an *in vitro* method of modifying a target nucleic acid, the method comprising delivering to the target nucleic acid an engineered, non-naturally occurring CRISPR-Cas system comprising: (a) a CRISPR-associated protein or a nucleic acid encoding the CRISPR-associated

protein, wherein the CRISPR-associated protein comprises the amino acid sequence of SEQ ID NO: 241; and (b) an RNA guide comprising a direct repeat sequence and a spacer sequence capable of hybridizing to the target nucleic acid, or a nucleic acid encoding the RNA guide; wherein the CRISPR-associated protein is capable of binding to the RNA guide; and wherein recognition of the target nucleic acid by the CRISPR-associated protein and RNA guide results in a modification of the target nucleic acid.

In yet another aspect, provided herein is the system as described herein or the cell as described herein for use in medicine.

This disclosure provides non-naturally-occurring, engineered systems and compositions for novel single-effector Class 2 CRISPR-Cas systems, which were first identified computationally from genomic databases and subsequently engineered and experimentally validated. In particular, identification of the components of these CRISPR-Cas systems allows for their use in non-natural environments, *e.g.*, in bacteria other than those in which the systems were initially discovered or in eukaryotic cells, such as mammalian cells. These new effectors are divergent in sequence and function compared to orthologs and homologs of existing Class 2 CRISPR effectors.

In one aspect, the disclosure provides engineered, non-naturally occurring Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) - Cas systems of CLUST.091979 including: a CRISPR-associated protein, wherein the CRISPR-associated protein includes an amino acid sequence that is at least 80% (*e.g.*, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%) identical to an amino acid sequence set forth in any one of SEQ ID NOs: 1-56; and an RNA guide including a direct repeat sequence and a spacer sequence capable of hybridizing to a target nucleic acid; wherein the CRISPR-associated protein is capable of binding to the RNA guide and of modifying the target nucleic acid sequence complementary to the spacer sequence. In one aspect, the disclosure provides engineered, non-naturally occurring Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) - Cas systems of CLUST.091979 including: a CRISPR-associated protein or a nucleic acid encoding the CRISPR-associated protein, wherein the CRISPR-associated protein includes an amino acid sequence that is at least 80% (*e.g.*, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%,

CLAIMS

What is claimed is:

1. An engineered, non-naturally occurring Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) - Cas system comprising:
 - (a) a CRISPR-associated protein or a nucleic acid encoding the CRISPR-associated protein, wherein the CRISPR-associated protein comprises an amino acid sequence of SEQ ID NO: 241; and
 - (b) an RNA guide comprising a direct repeat sequence and a spacer sequence capable of hybridizing to a target nucleic acid, or a nucleic acid encoding the RNA guide;wherein the CRISPR-associated protein is capable of binding to the RNA guide and of modifying the target nucleic acid sequence complementary to the spacer sequence.
2. The system of claim 1, wherein the CRISPR-associated protein comprises an amino acid sequence that is at least 80% (*e.g.*, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%) identical to an amino acid sequence set forth in SEQ ID NO: 4, SEQ ID NO: 10, SEQ ID NO: 12, or SEQ ID NO: 14.
3. The system of claim 1, wherein the CRISPR-associated protein comprises an amino acid sequence that is at least 95% identical to an amino acid sequence set forth in SEQ ID NO: 4.
4. The system of claim 1, wherein the CRISPR-associated protein comprises an amino acid sequence that is at least 95% identical to an amino acid sequence set forth in SEQ ID NO: 10.
5. The system of any previous claim, wherein the CRISPR-associated protein comprises at least one RuvC domain or at least one split RuvC domain.
6. The system of any previous claim, wherein the spacer sequence of the RNA guide comprises between 15 nucleotides to 55 nucleotides.
7. The system of any previous claim, wherein the spacer sequence of the RNA guide comprises between 20 and 45 nucleotides.

8. The system of any previous claim, wherein the CRISPR-associated protein further comprises a peptide tag, a fluorescent protein, a base-editing domain, a DNA methylation domain, a histone residue modification domain, a localization factor, a transcription modification factor, a light-gated control factor, a chemically inducible factor, or a chromatin visualization factor.
9. The system of any previous claim, wherein the nucleic acid encoding the CRISPR-associated protein is codon-optimized for expression in a cell.
10. The system of any previous claim, wherein the nucleic acid encoding the CRISPR-associated protein is operably linked to a promoter.
11. The system of any previous claim, wherein the nucleic acid encoding the CRISPR-associated protein is in a vector.
12. The system of claim 11, wherein the vector comprises a retroviral vector, a lentiviral vector, a phage vector, an adenoviral vector, an adeno-associated vector, or a herpes simplex vector.
13. The system of any previous claim, wherein the system does not comprise a tracrRNA.
14. The system of any previous claim, wherein the system is present in a delivery composition comprising a nanoparticle, a liposome, an exosome, a microvesicle, or a gene-gun.
15. The system of any previous claim, within a cell.
16. A cell, wherein the cell comprises:
 - (a) a CRISPR-associated protein or a nucleic acid encoding the CRISPR-associated protein, wherein the CRISPR-associated protein comprises the amino acid sequence of SEQ ID NO: 241; and
 - (b) an RNA guide comprising a direct repeat sequence and a spacer sequence capable of hybridizing to a target nucleic acid, or a nucleic acid encoding the RNA guide.

17. The cell of claim 16, wherein the CRISPR-associated protein comprises one or more of the following sequences:
 - (a) $PX_1X_2X_3X_4F$ (SEQ ID NO: 216), wherein X_1 is L or M or I or C or F, X_2 is Y or W or F, X_3 is K or T or C or R or W or Y or H or V, and X_4 is I or L or M;
 - (b) $RX_1X_2X_3L$ (SEQ ID NO: 217), wherein X_1 is I or L or M or Y or T or F, X_2 is R or Q or K or E or S or T, and X_3 is L or I or T or C or M or K;
 - (c) NX_1YX_2 (SEQ ID NO: 218), wherein X_1 is I or L or F and X_2 is K or R or V or E;
 - (d) $KX_1X_2X_3FAX_4X_5KD$ (SEQ ID NO: 219), wherein X_1 is T or I or N or A or S or F or V, X_2 is I or V or L or S, X_3 is H or S or G or R, X_4 is D or S or E, and X_5 is I or V or M or T or N;
 - (e) LX_1NX_2 (SEQ ID NO: 220), wherein X_1 is G or S or C or T and X_2 is N or Y or K or S;
 - (f) $PX_1X_2X_3X_4SQX_5DS$ (SEQ ID NO: 221), wherein X_1 is S or P or A, X_2 is Y or S or A or P or E or Y or Q or N, X_3 is F or Y or H, X_4 is T or S, and X_5 is M or T or I;
 - (g) $KX_1X_2VRX_3X_4QEX_5H$ (SEQ ID NO: 222), wherein X_1 is N or K or W or R or E or T or Y, X_2 is M or R or L or S or K or V or E or T or I or D, X_3 is L or R or H or P or T or K or Q or P or S or A, X_4 is G or Q or N or R or K or E or I or T or S or C, and X_5 is R or W or Y or K or T or F or S or Q; and
 - (h) $X_1NGX_2X_3X_4DX_5NX_6X_7X_8N$ (SEQ ID NO: 223), wherein X_1 is I or K or V or L, X_2 is L or M, X_3 is N or H or P, X_4 is A or S or C, X_5 is V or Y or I or F or T or N, X_6 is A or S, X_7 is S or A or P, and X_8 is M or C or L or R or N or S or K or L.
18. An *in vitro* method of binding the system of any previous claim to a target nucleic acid in a cell comprising:
 - (a) providing the system; and
 - (b) delivering the system to the cell,
 wherein the cell comprises the target nucleic acid, wherein the CRISPR-associated-protein binds to the RNA guide, and wherein the spacer sequence binds to the target nucleic acid.
19. An *in vitro* method of modifying a target nucleic acid, the method comprising delivering to the target nucleic acid an engineered, non-naturally occurring CRISPR-Cas system comprising:
 - (a) a CRISPR-associated protein or a nucleic acid encoding the CRISPR-associated protein, wherein the CRISPR-associated protein comprises the amino acid sequence of SEQ ID NO: 241; and

- (b) an RNA guide comprising a direct repeat sequence and a spacer sequence capable of hybridizing to the target nucleic acid, or a nucleic acid encoding the RNA guide; wherein the CRISPR-associated protein is capable of binding to the RNA guide; and wherein recognition of the target nucleic acid by the CRISPR-associated protein and RNA guide results in a modification of the target nucleic acid.
20. The method of claim 19, wherein the CRISPR-associated protein comprises one or more of the following sequences:
- (a) $PX_1X_2X_3X_4F$ (SEQ ID NO: 216), wherein X_1 is L or M or I or C or F, X_2 is Y or W or F, X_3 is K or T or C or R or W or Y or H or V, and X_4 is I or L or M;
 - (b) $RX_1X_2X_3L$ (SEQ ID NO: 217), wherein X_1 is I or L or M or Y or T or F, X_2 is R or Q or K or E or S or T, and X_3 is L or I or T or C or M or K;
 - (c) NX_1YX_2 (SEQ ID NO: 218), wherein X_1 is I or L or F and X_2 is K or R or V or E;
 - (d) $KX_1X_2X_3FAX_4X_5KD$ (SEQ ID NO: 219), wherein X_1 is T or I or N or A or S or F or V, X_2 is I or V or L or S, X_3 is H or S or G or R, X_4 is D or S or E, and X_5 is I or V or M or T or N;
 - (e) LX_1NX_2 (SEQ ID NO: 220), wherein X_1 is G or S or C or T and X_2 is N or Y or K or S;
 - (f) $PX_1X_2X_3X_4SQX_5DS$ (SEQ ID NO: 221), wherein X_1 is S or P or A, X_2 is Y or S or A or P or E or Y or Q or N, X_3 is F or Y or H, X_4 is T or S, and X_5 is M or T or I;
 - (g) $KX_1X_2VRX_3X_4QEX_5H$ (SEQ ID NO: 222), wherein X_1 is N or K or W or R or E or T or Y, X_2 is M or R or L or S or K or V or E or T or I or D, X_3 is L or R or H or P or T or K or Q or P or S or A, X_4 is G or Q or N or R or K or E or I or T or S or C, and X_5 is R or W or Y or K or T or F or S or Q; and
 - (h) $X_1NGX_2X_3X_4DX_5NX_6X_7X_8N$ (SEQ ID NO: 223), wherein X_1 is I or K or V or L, X_2 is L or M, X_3 is N or H or P, X_4 is A or S or C, X_5 is V or Y or I or F or T or N, X_6 is A or S, X_7 is S or A or P, and X_8 is M or C or L or R or N or S or K or L.
21. The system of any one of claims 1-15 or the cell of claim 16 or 17 for use in medicine.