Claims

1. A conjugate comprising a protein and a modified portion;

wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17; and

the modified portion is selected from an additional protein or polypeptide, a detectable label, and any combinations thereof;

for example, the modified portion is optionally connected to the N-terminus or C-terminus of the protein through a linker;

for example, the modified portion is fused to the N-terminus or C-terminus of the protein;

for example, the additional protein or polypeptide is selected from an epitope tag, a reporter gene sequence, a nuclear localization signal (NLS) sequence, a targeting moiety, a transcription activation domain (such as, VP64), a transcription repression domain (for example, KRAB domain or SID domain), a nuclease domain (for example, Fok1), a domain having an activity selected from: nucleotide deaminase, methylase activity, demethylase, transcription activation activity, transcription inhibition activity, transcription release factor activity, histone modification activity, nuclease activity, single-stranded RNA cleavage activity, double-stranded RNA cleavage activity, single-stranded DNA cleavage activity,

double-stranded DNA cleavage activity and nucleic acid binding activity; and any combinations thereof;

for example, the conjugate comprises an epitope tag;

for example, the conjugate comprises an NLS sequence;

for example, the NLS sequence is shown in SEQ ID NO: 81;

for example, the NLS sequence is located at, near or close to the end of the protein (e.g., N-terminal or C-terminal).

2. A fusion protein comprising a protein and an additional protein or polypeptide;

wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17; and

the additional protein or polypeptide is optionally linked to the N-terminus or C-terminus of the protein through a linker;

wherein the additional protein or polypeptide is selected from an epitope tag, a reporter gene sequence, a nuclear localization signal (NLS) sequence, a targeting moiety, a transcription activation domain (such as, VP64), a transcription repression domain (for example, KRAB domain or SID domain), a nuclease domain (for example, Fok1), a domain having an activity selected from: a nucleotide deaminase, methylase activity, a demethylase, transcription activation activity, transcription inhibition

activity, transcription release factor activity, histone modification activity, nuclease activity, single-stranded RNA cleavage activity, double-stranded RNA cleavage activity, single-stranded DNA cleavage activity, double-stranded DNA cleavage activity and nucleic acid binding activity ; and any combinations thereof；

for example, the fusion protein comprises an epitope tag;

for example, the fusion protein comprises an NLS sequence;

for example, the NLS sequence is shown in SEQ ID NO: 81;

for example, the NLS sequence is located at, near, or close to the end of the protein (for example, the N-terminus or the C-terminus);

for example, the fusion protein has an amino acid sequence selected from: SEQ ID NOs: 98.

3. A complex comprising:

(a) a protein component, wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17;; and

(b) a nucleic acid component, which is a guide RNA, which contains a direct repeat sequence and a guide sequence capable of hybridizing to the target sequence from 5’ to 3’ direction,

wherein the protein component and the nucleic acid component

combine with each other to form a complex;

for example, the guide sequence is attached to the 3' end of the nucleic acid molecule;

for example, the guide sequence comprises the complementary sequence of the target sequence;

for example, the complex does not comprise trans-acting crRNA (tracrRNA).

4. The complex of claim 3, comprising:

(a) a protein component selected from: a protein, a conjugate or fusion protein comprising the protein; and

(b) a nucleic acid component, which is a guide RNA, which contains a direct repeat sequence and a guide sequence capable of hybridizing to the target sequence from 5’ to 3’ direction;

wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17.

5. A vector comprising an isolated nucleic acid molecule comprising a nucleotide sequence containing:

(a) a nucleotide sequence encoding a protein,

wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17; and

(b) a nucleotide sequence encoding a guide RNA, which contains a direct repeat sequence and a guide sequence capable of hybridizing to the target sequence from 5’ to 3’ direction.

6. A host cell comprising the vector of claim 5, or an isolated nucleic acid molecule comprising a nucleotide sequence containing:

(a) a nucleotide sequence encoding a protein,

wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17; and

(b) a nucleotide sequence encoding a guide RNA, which contains a direct repeat sequence and a guide sequence capable of hybridizing to the target sequence from 5’ to 3’ direction.

7. A composition comprising:

(a) a first component, which is selected from: a protein, a nucleotide sequence encoding the protein;

wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17; and

(b) a second component, which is a nucleotide sequence containing a guide RNA, or a nucleotide sequence encoding the nucleotide sequence containing a guide RNA;

wherein the guide RNA includes a direct repeat sequence and a guide sequence from the 5' to 3', and the guide sequence can hybridize with the target sequence;

the guide RNA can form a complex with the protein as described in (a);

for example, the guide sequence is connected to the 3' end of the direct repeat sequence;

for example, the guide sequence comprises the complementary sequence of the target sequence;

for example, the composition does not contain trans-acting crRNA (tracrRNA);

for example, the composition is non-naturally occurring or modified;

for example, at least one component in the composition is non-naturally occurring or modified;

for example, the first component is non-naturally occurring or

modified; and/or, the second component is non-naturally occurring or modified.

8. The composition of claim 7, wherein:

the first component is selected from: a protein, or a conjugate or fusion protein comprising the protein, or a nucleotide sequence encoding the protein or fusion protein, and any combinations thereof;

wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17;when the target sequence is DNA, the target sequence is located at the 3' end of the original spacer sequence adjacent motif (PAM), and the PAM has the sequence shown by 5'-ATG.

9. A composition comprising one or more vectors, the one or more vectors comprising:

(a) a first nucleic acid, which is a nucleotide sequence encoding a protein;

wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17;

optionally, the first nucleic acid is operably linked to a first regulatory element; and

(b) a second nucleic acid, which encodes a nucleotide sequence comprising a guide RNA; optionally the second nucleic acid is operably linked to a second regulatory element;

wherein:

the first nucleic acid and the second nucleic acid are present on the same or different vectors;

the guide RNA comprises a direct repeat sequence and a guide sequence from the 5' to 3', and the guide sequence can hybridize with the target sequence;

the guide RNA can form a complex with the protein as described in (a);

for example, the guide sequence is connected to the 3' end of the direct repeat sequence;

for example, the guide sequence comprises the complementary sequence of the target sequence;

for example, the composition does not contain trans-acting crRNA (tracrRNA);

for example, the composition is non-naturally occurring or modified;

for example, at least one component in the composition is non-naturally occurring or modified;

for example, the first regulatory element is a promoter, such as an inducible promoter;

for example, the second regulatory element is a promoter, such as an inducible promoter.

10. The composition of claim 9, wherein:

the first nucleic acid is a nucleotide sequence encoding a protein or a fusion protein containing the protein;

wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17;

when the target sequence is DNA, the target sequence is located at the 3' end of the original spacer sequence adjacent motif (PAM), and the PAM has the sequence shown by 5'-ATG.

11. The composition of any one of claims 7-10, wherein when the target sequence is RNA, the target RNA sequence does not have PAM domain restrictions, and/or

wherein the target sequence is a DNA or RNA sequence derived from a prokaryotic cell or a eukaryotic cell; or the target sequence is a non-naturally occurring DNA or RNA sequence.

12. The composition of any one of claims 7-11, wherein the target sequence is present in a cell;

for example, the target sequence is present in the cell nucleus or in the cytoplasm (e.g., organelles);

for example, the cell is a eukaryotic cell;

for example, the cell is a prokaryotic cell.

13. The composition of any one of claims 7-12, wherein the protein is linked to one or more NLS sequences, or the conjugate or fusion protein comprises one or more NLS sequences;

for example, the NLS sequence is linked to the N-terminus or C-terminus of the protein;

for example, the NLS sequence is fused to the N-terminus or C-terminus of the protein.

14. A delivery composition comprising a delivery vehicle and one or more selected from the group consisting of: a protein or an isolated nucleic acid molecule encoding the protein;

wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17.

15. A delivery composition comprising a delivery vehicle and one or more selected from the group consisting of:the conjugate of claim 1, the fusion protein of claim 2, the complex of claim 3 or 4, an isolated nucleic acid molecule, the vector of claim 5, the composition of any one of claims 7-13;

wherein the isolated nucleic acid molecule comprises a nucleotide

sequence encoding the conjugate of claim 1, or the fusion protein of claim 2;

for example, the delivery vehicle is a particle;

for example, the delivery vehicle is selected from a lipid particle, sugar particle, metal particle, protein particle, liposome, exosome, microvesicle, gene gun, or viral vector (e.g., replication defective retrovirus, lentivirus, adenovirus or adeno-associated virus).

16. A non-diagnostic and non-therapeutic method for modifying a target gene, comprising: contacting the complex of claim 3 or 4, or the composition of any one of claims 7-13 with the target gene, or delivering that to a cell containing the target gene; the target sequence is present in the target gene;

for example, the target gene is present in the cell;

wherein the cell is a prokaryotic cell or a eukaryotic cell or a plant cell;

for example, the target gene is present in a nucleic acid molecule (e.g., a plasmid) *in vitro*;

for example, the modification refers to a break in the target sequence, such as a double-strand break in DNA or a single-strand break in RNA;

for example, the modification further includes inserting an exogenous nucleic acid into the break.

17. A non-diagnostic and non-therapeutic method for altering the expression of a gene product, comprising: contacting the complex of claim 3 or 4, or the composition of any one of claims 7-13 with a nucleic acid molecule encoding the gene product, or delivering that to a cell containing the nucleic acid molecule, the target sequence is present in the nucleic acid molecule;

for example, the nucleic acid molecule is present in the cell;

wherein the cell is a prokaryotic cell or a eukaryotic cell or a plant cell;

for example, the nucleic acid molecule is present in a nucleic acid molecule (e.g., a plasmid) *in vitro*;

for example, the expression of the gene product is altered (e.g., enhanced or decreased);

for example, the gene product is a protein.

18. The method of claim 16 or 17, wherein the protein, conjugate, fusion protein, complex, vector or composition is contained in a delivery vehicle;

for example, the delivery vehicle is selected from a lipid particle, sugar particle, metal particle, protein particle, liposome, exosome, viral vector (such as replication-defective retrovirus, lentivirus, adenovirus or adeno-associated virus).

19. The method of any one of claims 16-18, which is used to change one or more target sequences in a target gene or a nucleic acid molecule encoding a target gene product to modify a cell, cell line, or organism.

20. Use of a protein or an isolated nucleic acid molecule encoding the protein in the preparation of a formulation for:

(1) the *in vitro* gene or genome editing;

(2) the detection of an isolated single-stranded DNA;

(3) editing the target sequence in the target locus to modify a biological or non-human organism;

wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a

sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17.

21. Use of a protein or an isolated nucleic acid molecule encoding the protein in the preparation of a medicament for the treatment of the disease caused by defects in the target sequence in the target locus;

wherein, the protein comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in SEQ ID NO: 17;

(ii) compared with the sequence as shown in SEQ ID NO: 17, a sequence having 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions; or

(iii) a sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence as shown in SEQ ID NO: 17.

22. Use of the conjugate of claim 1, the fusion protein of claim 2, the complex of claim 3 or 4, an isolated nucleic acid molecule, the vector of claim 5, the composition of any one of claims 7-13 in the preparation of a formulation for:

(1) *in vitro* gene or genome editing;

(2) detection of an isolated single-stranded DNA;

(3) editing the target sequence in the target locus to modify a biological or non-human organism;

wherein the isolated nucleic acid molecule comprises a nucleotide

sequence encoding the conjugate of claim 1, or the fusion protein of claim 2.

23. Use of the conjugate of claim 1, the fusion protein of claim 2, the complex of claim 3 or 4, an isolated nucleic acid molecule, the vector of claim 5, the composition of any one of claims 7-13 in the preparation of a medicament for the treatment of the disease caused by defects in the target sequence in the target locus;

wherein the isolated nucleic acid molecule comprises a nucleotide sequence encoding the conjugate of claim 1, or the fusion protein of claim 2.