

CLAIMS

What is claimed is:

1. A system comprising a CasX variant protein and a guide nucleic acid (gNA), wherein the CasX variant protein comprises a sequence having at least 70% sequence identity to SEQ ID NO: 138 and exhibits one or more improved characteristics relative to a CasX reference protein of any one of SEQ ID NOS: 1-3, wherein the CasX variant protein is capable of forming a ribonuclear protein complex (RNP) with the gNA, and wherein the gNA comprises a targeting sequence complementary to a target nucleic acid sequence comprising a *chromosome 9 open reading frame 72 (C9orf72)* gene.
2. The system of claim 1, wherein the *C9orf72* gene comprises one or more mutations, or wherein the *C9orf72* gene mutation comprises more than 30, more than 100, more than 500, more than 700, more than 1000, or more than 1600 copies of a hexanucleotide repeat sequence GGGGCC in a hexanucleotide repeat sequence expansion (HRS).
3. The system of claim 1, wherein:
 - (a) the gNA is a single-molecule guide RNA (gRNA) comprising a scaffold sequence and a targeting sequence, and wherein the scaffold sequence comprises a scaffold stem loop sequence of CCAGCGACUAUGUCGUAGUGG (SEQ ID NO: 32) or a sequence with at least 1, 2, 3, 4, or 5 mismatches thereto;
 - (b) the targeting sequence of the gNA comprises a sequence selected from the group consisting of SEQ ID NOS: 309-343, 363-2100 and 2295-21835, or a sequence having at least about 65%, at least about 75%, at least about 85%, or at least about 95% identity thereto;
 - (c) the gNA has a scaffold comprising a sequence having at least 70% sequence identity to the sequence of SEQ ID NO: 2238; and/or
 - (d) wherein the gNA is chemically modified.
4. The system of claim 1, further comprising a second gNA, wherein the second gNA has a targeting sequence complementary to a different or overlapping portion of the target nucleic acid sequence compared to the targeting sequence of the gNA.
5. The system of claim 4, wherein the gNA is complementary to a sequence that is 5' to the HRS and the targeting sequence of the second gNA is complementary to a sequence that is 5' or 3' to the HRS.

6. The system of claim 1, wherein the CasX variant protein further comprises one or more nuclear localization signals (NLS), wherein the one or more NLS are selected from the group of sequences consisting of PKKKRKV (SEQ ID NO: 165), KRPAATKKAGQAKKKK (SEQ ID NO: 166), PAAKRVKLD (SEQ ID NO: 167), RQRRNELKRSP (SEQ ID NO: 168), NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 169), RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: 170), VSRKRPRP (SEQ ID NO: 171), PPKKARED (SEQ ID NO: 172), PQPKKKPL (SEQ ID NO: 173), SALIKKKKKMAP (SEQ ID NO: 174), DRLRR (SEQ ID NO: 175), PKQKKRK (SEQ ID NO: 176), RKLKKKIKKL (SEQ ID NO: 177), REKKKFLKRR (SEQ ID NO: 178), KRKGDEV DGVDEVAKKK SKK (SEQ ID NO: 179), RKCLQAGMNLEARKTKK (SEQ ID NO: 180), PRPRKIPR (SEQ ID NO: 181), PPRKKRTVV (SEQ ID NO: 182), NLSKKKKRKREK (SEQ ID NO: 183), RRPSRPFRKP (SEQ ID NO: 184), KRPRSPSS (SEQ ID NO: 185), KRGINDRNFWRGENERKTR (SEQ ID NO: 186), PRPPKMARYDN (SEQ ID NO: 187), KRSFSKAF (SEQ ID NO: 188), KLKIKRPVK (SEQ ID NO: 189), PKTRRRPRRS QRKRPT (SEQ ID NO: 191), RRKKRRPRRKRR (SEQ ID NO: 194), PKKKS RKP KKS RKP (SEQ ID NO: 195), HKKKHPDASVNFSEFSK (SEQ ID NO: 196), QRP GPYDRP QRP GPYDRP (SEQ ID NO: 197), LSPSLSP LSPSLSP L (SEQ ID NO: 198), RGKGGKGLGKGGA KRHRK (SEQ ID NO: 199), PKRGRGRPKRGRGR (SEQ ID NO: 200), MSRRRKANPTKLS ENAKKLAK EVEN (SEQ ID NO: 192), PKKKRKVPPPPAAKRVKLD (SEQ ID NO: 190), PKKKRKVPPPPKKKRKV (SEQ ID NO: 201), PAKRARRGYKC (SEQ ID NO: 202), KLGPRKATGRW (SEQ ID NO: 203), and PRRKREE (SEQ ID NO: 204), and wherein:

- (i) the one or more NLS are at or near a C-terminus of the CasX variant protein,
 - (ii) the one or more NLS are at or near a N-terminus of the CasX variant protein,
- and/or
- (iii) the CasX variant protein comprises at least two NLS that are at or near the N-terminus and the C-terminus of the CasX variant protein.

7. The system of claim 1, wherein the Cas X variant protein is capable of forming a ribonuclear protein complex (RNP) with the gNA.

8. The system of claim 7, wherein the CasX variant protein and the gNA exhibit at least one or more improved characteristics as compared to the reference CasX protein of any one of SEQ ID NOS: 1-3 and a reference gNA of any one of SEQ ID NOS: 4-16, and

wherein the improved characteristic is selected from the group consisting of improved folding of the CasX variant protein, improved binding affinity to the gNA, improved binding affinity to a target DNA, improved ability to utilize a greater spectrum of one or more protospacer adjacent motif (PAM) sequences including ATC, CTC, GTC or TTC in the editing of target DNA, improved unwinding of the target DNA, increased editing activity, improved editing efficiency, improved editing specificity, increased nuclease activity, increased target strand loading for double strand cleavage, decreased target strand loading for single strand nicking, decreased off-target cleavage, improved binding of non-target DNA strand, improved protein stability, improved protein solubility, improved protein:gNA RNP stability, improved protein:gNA complex solubility, improved protein yield, improved protein expression, and improved fusion characteristics, and

wherein the improved characteristic of the RNP of the CasX variant protein is at least 1.1 to 100,000-fold improved relative to the reference CasX protein of any one of SEQ ID NOS: 1-3 and the reference gRNA comprising any one of SEQ ID NOS: 4-16.

9. The system of claim 8, wherein:

(a) the RNP comprising the CasX variant protein and the gNA exhibits greater editing efficiency and/or binding of a target sequence in the target DNA when any one of the PAM sequences TTC, ATC, GTC or CTC is located 1 nucleotide 5' to the non-target strand sequence having identity with the targeting sequence of the gNA in a cellular assay system compared to the editing efficiency and/or binding of an RNP comprising a reference CasX protein and a reference gNA in a comparable assay system, and

(i) the PAM sequence is TTC and the targeting sequence of the gNA comprises a sequence selected from the group consisting of SEQ ID NOS: 5427-12893,

(ii) the PAM sequence is ATC and the targeting sequence of the gNA comprises a sequence selected from the group consisting of SEQ ID NOS: 363-2100 and 2295-5426,

(iii) the PAM sequence is CTC and the targeting sequence of the gNA comprises a sequence selected from the group consisting of SEQ ID NOS: 16203-21835,

(iv) the PAM sequence is GTC and the targeting sequence of the gNA comprises a sequence selected from the group consisting of SEQ ID NOS: 12894-16202 ; and/or

(b) the RNP has at least a 5%, at least a 10%, at least a 15%, or at least a 20% higher percentage of cleavage-competent RNP compared to an RNP of the reference CasX protein of any one of SEQ ID NOS: 1-3 and the reference gNA comprising any one of SEQ ID NOS: 4-16, or the RNP has at least a 2-fold, at least 5-fold, or at least a 10-fold increased cleavage rate in an in vitro assay compared to a RNP of the reference CasX protein of any one of SEQ ID NOS: 1-3.

10. The system of claim 1, wherein the CasX variant protein comprises a nuclease domain having double-stranded cleavage activity.

11. A nucleic acid comprising a sequence encoding:

(a) a gNA comprising a targeting sequence complementary to a target nucleic acid sequence comprising a chromosome 9 open reading frame 72 (*C9orf72*) gene; and

(b) a CasX variant protein comprising a sequence having at least 70% sequence identity to SEQ ID NO: 138 and exhibits one or more improved characteristics relative to a CasX reference protein of any one of SEQ ID NOS: 1-3, wherein the CasX variant protein is capable of forming a ribonuclear protein complex (RNP) with the gRNA, and wherein the CasX variant protein further comprises one or more nuclear localization signals (NLS).

12. A vector comprising the nucleic acid of claim 11, wherein the vector is selected from the group consisting of a retroviral vector, a lentiviral vector, an adenoviral vector, an adeno-associated viral (AAV) vector, a herpes simplex virus (HSV) vector, a virus-like particle (VLP), a plasmid, a minicircle, a nanoplasmid, and an RNA vector.

13. The vector of claim 12, wherein the vector is an AAV vector selected from the group consisting of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAV44.9, AAV-Rh74, and AAVRh10.

14. A host cell comprising the vector of claim 12, wherein the host cell is selected from the group consisting of BHK, HEK293, HEK293T, NS0, SP2/0, YO myeloma cells, P3X63 mouse myeloma cells, PER, PER.C6, NIH3T3, COS, HeLa, CHO, and yeast cells.

15. A method of modifying a *chromosome 9 open reading frame 72 (C9orf72)* gene target nucleic acid sequence in a population of cells, the method comprising the steps of,

introducing into the population of cells:

- a. the system of claim 1;
- b. the nucleic acid of claim 11;
- c. the vector of claim 12; or
- d. a combination thereof; and

modifying the *C9orf72* gene target nucleic acid sequence in the population of cells targeted by the gNA and the CasX variant protein, wherein the modifying comprises:

(i) introducing a double-stranded break in the *C9orf72* target nucleic acid sequence; and/or

(ii) introducing an insertion, deletion, substitution, duplication, or inversion of one or more nucleotides in the *C9orf72* target nucleic acid sequence.

16. The method of claim 15, wherein the cell is a eukaryotic cell selected from the group consisting of a rodent cell, a mouse cell, a rat cell, a pig cell, a primate cell, and a non-human primate cell.

17. The method of claim 16, wherein the population of cells is selected from the group consisting of Purkinje cell, frontal cortex neuron, motor cortex neuron, hippocampus neuron, cerebellum neuron, upper motor neuron, spinal cord neuron, spinal cord motor neuron, glial cell, and astrocytes.

18. A population of cells modified by:

- a. the system of claim 1;
- b. the nucleic acid of claim 11;
- c. the vector of claim 12; or
- d. a combination thereof,

wherein the population of cells once modified include at least one of the following characteristics relative to a population of unmodified cells:

- (i) at least 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, or at least about 90% do not express a detectable level of dipeptide repeat protein (DPR),
- (ii) expression of a functional *C9orf72* protein is increased by at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, or at least about 90 %, and/or
- (iii) mutation of the *C9orf72* gene is corrected, resulting in expression of a functional *C9orf72* protein,

wherein cells of the population of cells are selected from the group consisting of Purkinje cell, frontal cortex neuron, motor cortex neuron, hippocampus neuron, cerebellum neuron, upper motor neuron, spinal cord neuron, spinal cord motor neuron, glial cell, and astrocytes.

19. A composition for use in treating a *chromosome 9 open reading frame 72 (C9orf72)*-related disorder in a subject in need thereof, wherein the use comprises:

modifying a *C9orf72* gene in cells of the subject by contacting said cells with a therapeutically effective dose of:

- a. the system of claim 1;
- b. the nucleic acid of claim 11;
- c. the vector of claim 12; or
- d. combinations thereof

wherein the *C9orf72* gene of the cells targeted by the gNA is modified by the CasX variant protein.

20. Use of a composition in the manufacture of a medicament for treating a *chromosome 9 open reading frame 72 (C9orf72)*-related disorder in a subject in need thereof, wherein the use comprises:

modifying a *C9orf72* gene in cells of the subject by contacting said cells with a therapeutically effective dose of:

- a. the system of claim 1;
- b. the nucleic acid of claim 11;
- c. the vector of claim 12; or
- d. combinations thereof

wherein the *C9orf72* gene of the cells targeted by the gNA is modified by the CasX variant protein.