

ENDONUCLEASE VARIANTS AND METHODS OF USE

RELATED APPLICATIONS

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. provisional application number 63/297,182, filed January 6, 2022, and U.S. provisional application number 63/297,189, filed January 6, 2022 which is incorporated by reference herein in its entirety.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

The contents of the electronic sequence listing (T085570001WO00-SEQ-HJD.xml; Size: 634,084 bytes; and Date of Creation: January 6, 2023) is herein incorporated by reference in its entirety.

BACKGROUND

Prokaryotes have developed an adaptive immune system called Clustered regularly interspaced short palindromic repeats (CRISPR) that associate with Cas proteins to constitute an adaptive immune system that can combat attacks by foreign mobile genetic elements such as plasmids and phages. The CRISPR-Cas systems are classified into two classes (Classes 1 and 2) that are subdivided into six types (types I through VI). Class 1 (types I, III and IV) systems use multiple Cas proteins in their CRISPR ribonucleoprotein effector nucleases and Class 2 systems (types II, V and VI) use a single Cas protein. Class 2 type V is further classified into 4 subtypes (V-A, V-B, V-C, V-U). At present, V-C and V-U remain widely uncharacterized and no structural information on these systems is available. V-A encodes the protein Cas12a (also known as Cpf1) and recently several high-resolution structures of Cas12a have provided an insight into its working mechanism.

Class 2 type V CRISPR-Cas12a is an RNA-guided endonuclease that has been harnessed as a genome editing tool. Broader use of these enzymes for gene and epigenetic editing requires improvement of certain properties.

SUMMARY

The present disclosure provides, in some aspects, variant Cas12a endonucleases with improved properties, such as hyperactivity and low indiscriminate single strand DNA degradation activity. Broad use of wild-type Cas12a has been limited, in part, due to its lower editing efficiency, relative to Cas9, and its indiscriminate single strand DNA

degradation activity. The lid region, which is involved in the checkpoints for accurate target recognition, is responsible for this indiscriminate ssDNA degradation activity displayed by all wild-type Cas12a orthologs. Surprisingly, the data described herein demonstrate that certain modifications to the lid region of Cas12a can impact not only indiscriminate single strand deoxyribonuclease (ssDNase) activity but also targeted cleavage activity – both double strand and single strand cleavage activity.

Engineered variant endonucleases, in some embodiments, exhibit more efficient cleavage activity, relative to their wild-type reference Cas12a endonuclease. In other embodiments, engineered variant endonucleases of the present disclosure exhibit low to no indiscriminate single strand DNase activity. Also provided herein, in some embodiments, are variant Cas12a endonucleases that exhibit a preference for cleavage of one strand over the other strand of a double strand DNA.

From structural studies of the LbCas12a ND2006 endonuclease, Applicants have identified a particular domain, referred to herein as the “LID-hub domain,” that is involved in a subset of catalytic events. For example, certain substitutions made at positions K932, N933, and V936 increase cleavage efficiency (“hyperactivity”) and certain substitutions made at positions K932, N933, V936, Q944, F983, and M986 reduce indiscriminate ssDNase activity. Certain amino acid substitutions within the vicinity of the LID and LID-hub domains also impact activity (e.g., V938 or Q941).

Further still, Applicants have also unexpectedly shown that modifications to the LID stabilizing charge network (defined by its three-dimensional structure to include at least positions E835, R836, R935, and/or K940, with reference to amino acid position numbering of LbCas12a ND2006) shift Cas12a cleavage preferences (e.g., from double strand DNA cleavage activity).

In some embodiments, variant Cas12a endonucleases comprise amino acid mutations at one or more amino acid positions within the lid region. For example, in some embodiments, a variant Cas12a endonuclease comprises one or more mutations at an amino acid position corresponding to positions 925 to 937 of *Lachnospiraceae* bacterium ND2006 (e.g., SEQ ID NO: 1). In some embodiments, a variant Cas12a endonuclease comprises one or more mutations at an amino acid position corresponding to positions 936 to 948 of *Lachnospiraceae* bacterium COE1 (e.g., SEQ ID NO: 47).

As described herein, the term “variant Cas12a endonuclease(s)” is interchangeable with the term “Cas12a variant.”

Some aspects relate to an engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising a mutation at an amino acid position corresponding to position E95, E125, N256, R747, H759, N813, K932, N933, S934, V936, S982, or K984 with reference to amino acid position numbering of LbCas12a ND2006. In some
5 embodiments, as any one or more of the foregoing variant Cas12a endonucleases exhibits hyperactivity.

Other aspects relate to an engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising a mutation at an amino acid position corresponding to position E95R, E95Y, E125A, E125W, N256A, R747Y, H759V, H759D, N813R, N813H,
10 K932L, N933E, N933V, S934Q, V936E, V936M, V936K, S982N, or K984R with reference to amino acid position numbering of LbCas12a ND2006. In some embodiments, as any one or more of the foregoing variant Cas12a endonucleases exhibits hyperactivity.

Some aspects relate to an engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising a mutation at an amino acid position corresponding to position N256, I831, K932, N933, S934, V936, Q944, S982, F983, K984, M986, or T988
15 with reference to amino acid position numbering of LbCas12a ND2006. In some embodiments, as any one or more of the foregoing variant Cas12a endonucleases exhibits hypoactivity.

Other aspects relate to an engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising a mutation at an amino acid position corresponding to position N256K, I831A, I831Y, K932A, K932F, K932H, K932M, K932N, K932Q, K932R,
20 K932S, K932T, K932W, K932Y, N933L, S934W, V936G, Q944D, Q944E, Q944K, Q944M, S982T, S982W, F983G, F983L, K984F, M986G, M986L, M986S, or T988F with reference to amino acid position numbering of LbCas12a ND2006. In some embodiments, as any one
25 or more of the foregoing variant Cas12a endonucleases exhibits hypoactivity.

Yet other aspects relate to an engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising mutations at an amino acid positions corresponding to positions: K932F and F983L; K932F and T988F; K932R and Q944D; K932R and F983L;
K932R and T988F; K932Y and F983L; K932Y and T988F; N933L and Q944M; V936G and
30 Q944D; V936G and S982W; V936G and M986G; V936G and T988F; Q944D and S982W; Q944D and F983L; Q944D and T988F; S982W and F983L; S982W and T988F; or F983G and M986G with reference to amino acid position numbering of LbCas12a ND2006. In some
embodiments, as any one or more of the foregoing variant Cas12a endonucleases exhibits hypoactivity.

Some aspects relate to an engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising a mutation at an amino acid position corresponding to position N813, I831, K932, N933, S934, V936, Q944, S982, F983, K984, M986, or T988 with reference to amino acid position numbering of LbCas12a ND2006. In some
5 embodiments, as any one or more of the foregoing variant Cas12a endonucleases exhibits low (or no) ssDNase activity, such as low (or no) indiscriminate ssDNase activity.

Other aspects relate to an engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising a mutation at an amino acid position corresponding to position N813H, N813R, N813W, I831A, I831Y, K932A, K932F, K932H, K932M, K932N,
10 K932Q, K932R, K932S, K932T, K932W, K932Y, N933E, N933L, S934K, S934Q, V936E, V936G, Q944D, Q944E, Q944K, S982W, F983G, F983L, K984F, M986F, M986G, or T988F with reference to amino acid position numbering of LbCas12a ND2006. In some embodiments, as any one or more of the foregoing variant Cas12a endonucleases exhibits low (or no) ssDNase activity, such as low (or no) indiscriminate ssDNase activity.

Yet other aspects relate to an engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising mutations at an amino acid positions corresponding to positions: N933L and Q944M; or F983G and M986G with reference to amino acid position numbering of LbCas12a ND2006. In some embodiments, as any one or more of the foregoing variant Cas12a endonucleases exhibits low (or no) ssDNase activity, such as low (or no)
15 indiscriminate ssDNase activity.
20

In some embodiments, an engineered variant Cas12a endonuclease is fused to an effector protein.

In some embodiments, an engineered variant Cas12a endonuclease provided herein comprises an amino acid sequence having at least 85%, at least 90%, or least 95%, but less
25 than 100% identity with the amino acid sequence of a wild-type Cas12a endonuclease selected from *Acidaminococcus* sp., *Lachnospiraceae* sp., and *Francisella* sp.

In some embodiments, an engineered variant Cas12a endonuclease further comprises no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 additional amino acid substitutions relative to a wild-type reference Cas12a endonuclease. In some embodiments, a variant Cas12a
30 endonuclease further comprises no more than 5 additional amino acid substitutions relative to a wild-type reference Cas12a endonuclease.

The present disclosure also provides an engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising the amino acid sequence of any one of SEQ ID NOs: 48-119 and 367-387, or an ortholog thereof.

Also provided herein are polynucleotides encoding an engineered variant Cas12a endonuclease of the present disclosure.

Further provided herein are cells comprising (a) an engineered variant Cas12a endonuclease of the present disclosure or a polynucleotide endonuclease of the present disclosure and (b) a guide RNA or a polynucleotide encoding a guide RNA.

Some aspects herein relate to a method comprising introducing into a cell (a) an engineered variant Cas12a endonuclease of the present disclosure or a polynucleotide of the present disclosure and optionally (b) a guide RNA or a polynucleotide encoding a guide RNA.

The present disclosure also provides uses of an engineered variant Cas12a endonuclease of the present disclosure for cleaving a nucleic acid.

In some embodiments, a method for introducing a double strand break in a target nucleic acid comprises introducing into a cell comprising a target nucleic acid (a) an engineered variant Cas12a endonuclease of the present disclosure and (b) a guide RNA and incubating the cell to produce a double strand break in the target nucleic acid.

In other embodiments, a method for introducing a double strand break in a target nucleic acid comprises introducing into a cell comprising a target nucleic acid (a) an engineered variant Cas12a endonuclease of the present disclosure and (b) a guide RNA and incubating the cell to produce a double strand break in the target nucleic acid.

In some embodiments, the off-target single strand nucleic acid cleavage in the cell is reduced relative to off-target single strand nucleic acid cleavage in a control cell comprising a wild-type Cas12a endonuclease and a guide RNA.

In some embodiments, a method for introducing a single strand break in a target nucleic acid, comprises introducing into a cell comprising a target nucleic acid (a) an engineered variant Cas12a endonuclease of the present disclosure (b) a guide RNA, and incubating the cell to produce a single strand break in the target nucleic acid.

Some aspects relate to an engineered polypeptide comprising the amino acid sequence of any one of SEQ ID NOs: 48-119 and 367-387 or a variant thereof having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to the amino acid sequence of any one of SEQ ID NOs: 48-119 and 367-387, optionally wherein the engineered polypeptide is an endonuclease that exhibits hyperactive, hypoactivity, and/or low ssDNase activity (e.g., low indiscriminate ssDNase activity) relative to a naturally-occurring Cas12a endonuclease (e.g., SEQ ID NO: 1).

Some aspects relate to a fusion protein comprising an engineered variant Cas12a endonuclease of any one of the preceding aspects or embodiments and a base editing enzyme.

Some aspects relate to an engineered polypeptide comprising the amino acid sequence of any one of SEQ ID NOs: 163-185 and 388-408 or a variant thereof having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to the amino acid sequence of any one of SEQ ID NOs: 163-185 and 388-408.

Further aspects relate to fusion protein comprising an engineered variant Cas12a endonuclease an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the amino acid sequence of any one of SEQ ID NOs: 367-387.

In some embodiments, the base editing enzyme is capable of converting a purine into a different purine or a pyrimidine into a different pyrimidine.

In some embodiments, the base editing enzyme comprises a deaminase, a guanine oxidase, or a guanine methyltransferase.

In some embodiments, the deaminase is a cytidine deaminase or an adenosine deaminase.

In some embodiments, the deaminase comprises a rAPOBEC1 polypeptide, an evoAPOBEC1 polypeptide, a hAPOBEC3A polypeptide, an evoCDA polypeptide, an evoFERNY polypeptide, or a TadA polypeptide.

In some embodiments, fusion protein further comprises a uracil glycosylase inhibitor (UGI).

In some embodiments, fusion protein further comprises one or more nuclear localization signal (NLS), optionally selected from an SV40 NLS, a nucleoprotein (NP) NLS, and a bipartite (BP) NLS.

In some embodiments, fusion protein further comprises a uracil DNA glycosylase (UNG), optionally a human UNG (hUNG) or an Escherichia coli UNG (eUNG).

In some embodiments, fusion protein further comprises a N-methyl purine glycosylase (MPG), optionally wherein the MPG is positioned at or near the N-terminal or C-terminal ends of the fusion protein.

In some embodiments, fusion protein further comprises one or more linker.

In some embodiments, the linker comprises the sequence of SGSETPGTSESATPES (SEQ ID NO: 203).

In some embodiments, the linker comprises the sequence of SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 204)

In some embodiments, fusion protein further comprises a DNA binding domain (DBD).

In some embodiments, the DBD is a Rad51 DBD.

Some aspects relate to a polynucleotide encoding the fusion protein of any one of the preceding aspects or embodiments.

Other aspects relate to a cell comprising: a target nucleic acid comprising a target strand and a non-target strand; a guide RNA (gRNA) or a nucleic acid encoding a gRNA that binds to the target strand; and the fusion protein of any one of the preceding aspects or embodiments or the polynucleotide of any one of the preceding aspects or embodiments.

In some embodiments, the cell is a human cell.

In some embodiments, the human cell is from a human subject, wherein the human subject has a disease, disorder or condition associated with the target nucleic acid.

In some embodiments, the cell is a plant cell.

Some aspects relate to a method comprising introducing into a cell (a) the fusion protein of any one of the preceding aspects or embodiments or the polynucleotide of any one of the preceding aspects or embodiments and optionally (b) a guide RNA or a polynucleotide encoding a guide RNA.

Some aspects relate to a method of gene editing comprising (i) contacting a target nucleic acid sequence with the fusion protein of any one of the preceding aspects or embodiments and a guide RNA, wherein the target nucleic acid comprises a target nucleobase; and (ii) modifying the target nucleobase.

In some embodiments, the target nucleic acid is a target double-stranded DNA nucleic acid.

In some embodiments, the guide RNA directs the fusion protein to bind to a specific segment of the target nucleic acid and in proximity to the target nucleobase.

In some embodiments, the fusion protein cleaves the target nucleic acid.

In some embodiments, the fusion protein comprises a cytidine deaminase and the target nucleobase is a cytidine. In some embodiments, the fusion protein comprises a guanine oxidase and the target nucleobase is a guanosine. In some embodiments, the fusion protein comprises a guanine methyltransferase and the target nucleobase is a guanosine.

In some embodiments, the method is performed in a cell, optionally a human cell or a plant cell.

In some embodiments, the method is performed in vitro or ex vivo. In other embodiments, the method is performed in vivo.

In some embodiments, the target nucleic acid is a gene comprising a nucleobase mutation relative to a wild-type gene.

In some embodiments, the gene comprising a nucleobase mutation is associated with a disease or disorder.

5

BRIEF DESCRIPTION OF THE DRAWINGS

FIGs. 1A-1X show an alignment of various wild-type Cas12a endonuclease amino acid sequences using the Clustal Omega online multiple sequence alignment program (top to bottom SEQ ID NO: 29, 39, 12, 36, 8, 2, 43, 18, 37, 6, 46, 14, 28, 38, 45, 20, 13, 5, 21, 7, 42, 24, 40, 16, 35, 19, 41, 3, 10, 30, 1, 44, 22, 11, 27, 31, 4, 25, 26, 34, 32-33, 9, 17, and 23).

10

FIGs. 2A-2D show experimental data for various examples of the hyperactive Cas12a endonuclease variants of the present disclosure.

FIGs. 3A-3E show experimental data for various examples of the hypoactive Cas12a endonuclease variants of the present disclosure.

15

FIGs. 4A-4C show experimental data for various examples of the hypoactive Cas12a endonuclease variants of the present disclosure.

FIG. 5 shows experimental data for various examples of the Cas12a endonuclease variants of the present disclosure having low indiscriminate ssDNase activity.

FIGs. 6A-6C provide graphs of data comparing percent (%) of total reads having a C-to-T nucleotide edit at genomic positions corresponding to positions C8, C9, C10, C11, and C13 of the guide RNA (gRNA) using the LbBEv2 base editor in U2OS cells.

20

FIGs. 7A-7D provide graphs of data comparing percent (%) of total reads having a C-to-T nucleotide edit at genomic positions corresponding to positions C8 and C10 of the gRNA using the LbBEv2, LbBEv3, LbBEv4, or LbBEv5 base editor in U2OS cells.

25

FIGs. 8A-8D provide graphs of data comparing percent (%) of total reads having a C-to-T nucleotide edit at genomic positions corresponding to positions C9, C10, and C15 of the gRNA using the LbBEv2, LbBEv3, LbBEv4, or LbBEv5 base editor in U2OS cells.

FIGs. 9A-9F provide data showing increased efficiency and specificity of base editing with LbBEv5 C-to-T base editor containing TBN04 (LbCas12a) as compared to LbBEv5 base editor containing inactive LbCas12a in U2OS cells (top to bottom SEQ ID NOs: 214-263).

30

FIGs. 10A-10F provide data showing increased efficiency and specificity of base editing with the LbBEv5 C-to-T base editor (top to bottom SEQ ID NOs: 264-315).

FIGs. 11A-11F provide data showing increased efficiency and specificity of base editing with the LbABE8e A-to-G base editor (top to bottom SEQ ID NO: 316-365).

FIGs. 12A-12C provide data showing increased efficiency and specificity of base editing with A-to-G base editors comprising mutant Cas12a.

5 **FIGs. 13A-13B** provide data showing the ability of base editors comprising a N-methyl purine glycosylase (MPG) to perform A-to-C base editing (SEQ ID NO: 409 and 366).

DETAILED DESCRIPTION

10 **I. Cas12a Endonuclease**

Provided herein are variants of the Class II type V CRISPR-Cas12a endonuclease. An “endonuclease” is an enzyme capable of cleaving the phosphodiester bond within a polynucleotide chain. Some endonucleases are specific (i.e., they recognize a given nucleotide sequence which directs the site of cleavage), while some are non-specific. The
15 present disclosure provides specific variant Cas12a endonucleases. An endonuclease may cleave both strands of a double strand polynucleotide, or an endonuclease may demonstrate a preference for cleave cleaving one strand over the other strand of a double strand polynucleotide.

The recently discovered clustered regularly interspaced short palindromic repeats (CRISPR)-Cpf1 system, now reclassified as Cas12a, is a DNA-editing platform analogous to
20 the widely used CRISPR-Cas9 system. The Cas12a system exhibits several distinct features over the CRISPR-Cas9 system, such as increased specificity and a smaller gene size to encode the nuclease and the matching CRISPR guide RNA (crRNA), which could mitigate off-target and delivery problems, respectively, described for the Cas9 system. However, the
25 Cas12a system exhibits reduced gene editing efficiency compared to Cas9. Many of the variant Cas12a endonucleases provided herein exhibit increased gene editing efficiency compared to the wild-type Cas12a systems characterized to date.

RNA sequencing of small RNA molecules extracted from *Francisella novicida* U112 culture containing Cas12a-based CRISPR loci showed that mature crRNAs for Cas12a are
30 42–44 nucleotides (nt) in length, with the first 19/20 nt corresponding to the repeat sequence and the remaining 23-25 nt to the spacer sequence. Cas12a processes its own pre-crRNA into mature crRNAs, without the requirement of a tracrRNA, making it a unique effector protein with both endoribonuclease and endonuclease activities. After the pre-crRNA has been transcribed during the expression stage, Cas12a cuts it 4 nt upstream of the hairpin structures

formed by the CRISPR repeats, producing intermediate crRNA molecules that undergo further processing *in vivo* into mature crRNAs.

Type V (Cas12a) CRISPR-Cas systems possesses a characteristic Ruv-C like nuclease domain, which has been shown to be related to IS605 family transposon encoded TnpB proteins. Crystallographic and cryo-EM data reveal that Cas12a adopts a bilobed structure formed by the REC and Nuc lobes. The REC lobe is comprised of REC1 and REC2 domains, and the Nuc lobe is comprised of the RuvC, the PAM-interacting (PI) and the WED domains, and additionally, the bridge helix (BH). The RuvC endonuclease domain of this effector protein is made up of three discontinuous parts (RuvC I-III). The RNase site for processing its own crRNA is situated in the WED-III subdomain, and the DNase site is located in the interface between the RuvC and the Nuc domains. These structural studies have also shown that the only the 5' repeat region of the crRNA is involved in the assembly of the binary complex. The 19/20 nt repeat region forms a pseudoknot structure through intramolecular base pairing. The crRNA is stabilized through interactions with the WED, RuvC and REC2 domains of the endonuclease, as well as two hydrated Mg²⁺ ions. This binary interference complex is then responsible for recognizing and degrading foreign DNA. *See Paul, B. & Montoya, G. et al. Biomedical Journal 2020; 43(1): 8-17.*

Protospacer adjacent motif (PAM) recognition is a critical initial step in identifying a prospective DNA molecule for degradation because the PAM allows the CRISPR-Cas systems to distinguish their own genomic DNA from invading nucleic acids. Cas12a employs a multistep quality control mechanism to ensure the accurate and precise recognition of target spacer sequences. The WED II-III, REC1 and PAM-interacting domains are responsible for PAM recognition and for initiating the hybridization of the DNA target with the crRNA. After recognition of the dsDNA by WED and REC1 domains, the conserved loop-lysine helix-loop (LKL) region in the PI domain, containing three conserved lysines (K667, K671, K677 in FnCas12a), inserts the helix into the PAM duplex with assistance from two conserved prolines in the LKL region. Structural studies show the helix is inserted at an angle of 45° with respect to the dsDNA longitudinal axis, promoting the unwinding of the helical dsDNA. The critical positioning of the three conserved lysines on the dsDNA initiates the uncoupling of the Watson–Crick interaction between the base pairs of the dsDNA after the PAM. The target dsDNA unzipping allows the hybridization of the crRNA with the strand containing the PAM, the target strand, while the uncoupled DNA strand, non-target strand (NTS), is conducted towards the DNase site by the PAM-interacting domain. Cas12a has been shown to efficiently target spacer sequences following 5'T-rich PAM sequence. The

PAM for LbCas12a and AsCas12a has a sequence of 5'-TTTN-3' and for FnCas12a a sequence of 5'-TTN-3' and is situated upstream of the 5' end of the non-target strand [26,31,34]. It has also been shown that in addition to the canonical 5'-TTTN-3' PAM, Cas12a also exhibits relaxed PAM recognition for suboptimal C-containing PAM sequences by forming altered interactions with the targeted DNA duplex. See Paul, B. & Montoya, G. *et al.*

Exemplary, non-limiting, wild-type Cas12a protein sequences are provided in **Table 1**.

Table 1. Non-limiting Examples of Wild-type Cas12a Sequences

Name	Sequence	SEQ ID NO:
<i>LbCas12a-ND2006</i> <i>Lachnospiraceae</i> <i>bacterium ND2006</i>	MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLSFIND VLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNKGYKSLFKKDIETIL PEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAPFCINENLTRYISNMDIFEK VDAIFDKHEVQEIKEKIILNSDYDVEDFFEGEFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLN EYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKK LEKLFKNFDEYSSAGIFVKNPAISTISKDIFGEWNVIRDKWNAEYDDIHLKKAVVTEKYEDD RRKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIQKVDEIYKVYGSSEKLFDAFVLEKSLKK NDAVVAIMKDLLDSVKSFENYIAFFGEGKETNRDESFGDFVLAYDILLKVDHIYDAIRNYVT QKPYSKDKFLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQIDKDDVNG NYEKINYKLLPGPNKMLPKVFFSKKWMAYNPNSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFK DSISRYPKWSNAYDFNFSETEKYKDIAFGYREVVEQGYKVSFESASKEVDKLVVEGKLYMFQI YNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFMRRASLKKEELVVHPANSPAIANK NPDNPKKTTTSLSYDVYKDKRFSEDQYELHIPAIANKCPKNIFKINTEVRVLLKHDDNPVYIGID RGERNLLYIVVVDGKGNIVEQYSLNEIINNPNFIRIKTDYHSLDDKKEKERFEARQNWTSIENI KELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKLNMYMDK KSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKYTSIADS KKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNNVFDWEE VCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSFMAIAMSMLQMRNITGRTDQVDFLIS PVKNSDGIIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAISNK EWLEYAQTSVKH	1
<i>AsCas12a</i> <i>Acidaminococcus</i> <i>sp. BV3L6</i>	MTQFEGFTNLVQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNHYKELKPIIDRIYKTYADQ CLQLVQLDWENLSAIDSYRKEKTEETRNALIEEQATYRNAIHDFYIGRTDNLTDANKRHAET YKGLFKAELFNGKVLKQLGTVTTEHENALLRSFDKFTTYFSGFYENRNKRVSAEDISTAIPHR IVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIIEEVFSFPFYNQLLTQTQID LYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFIL EEFKSDEEVIQSFCKYKTLRNENVLETAELFNEIINLTHIFISHKKLETISSALCDHWD LRNALYERRISELTGKITSAKEKVQSLKHEDINLQEIISAAGKELSEAFKQKTEILSHAH ALDQPLPTTLKKQEEKEILKSQDLSLLGLYHLLDWFVAVDESNEVDPEFSARLTGKLEMEPSLS FYNKARNYATKKPYSVEKFKLNFQMPITLASGWDVNKEKNNGAILFVNGLYYLGIMPKQKGRYK ALSFEPTKTESEGFDKMYDYFPAAKMIPKSTQLKAVTAHFQTHHTPILLSNNFIEPLEITK EIYDLNNPEKEPKFQTAAYAKKTGQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPFSSQY KDLGEYAAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGL FSPENLAKTSIKLNGQAEFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIDTLYQELDYVNH RLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFHVPITLNYQAANSPSKFNQRVNAYL KEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSINTIQQFDYQKKLDNREKERVAAQWASV VGTIKDLKQGYLSQVIHEIVDLMIHYQAVVLENLNGFKSKRTGIAEKAVYQQFEKMLIDKLN CLVLKDYPAEKVGGVLNYPQLTDQFTSFAKMGTSQGFIFYVPAPYTSKIDPLTGFVDPFVWKT KNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRLPGFMPAWDIVFEKNETQFDAQGT FFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDSDHAIDTMV ALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPPEWPMADANGAYHIALKGQLLLNH LKESKDLKLQNGISNQDWLAYIQELRN	2

Name	Sequence	SEQ ID NO:
<i>RbCas12a</i> <i>Ruminococcus</i> <i>bromii</i>	MQERKKISHLTHRNSVKKTI RMQLNPVGKTMDFYQAKQILENDEKLKENYQKIKEIADRFRYRNL NEDVLSKTGLDKLDYAEIYYHCNTDADRKRINKCASELRKEIVKNFKNRDEYNKLFDRKRMIEI VLPKHLKNEDEKEVVASFKNFTTYFTGFFTNRKNMYSDGEESTAIAAYRCINENLPKHLDNVKA EKAISKLSKNAIDDLATYSGLCGTNLVDVFTVDYFNFLLPQSGITEYNKIIGGYTTNDGTVK GINEYINLYNQVSKRDKIPNLQILYKQILSESEKVSFIPPKFEDDNELLSAVSEFYANDET GMPKKAIIDETKLLFGNLDNSSLNGIYIQNDRSVTNLSNSMFGSWSVIEDLWNKNYDSVNSNR IKDIQKREDKRKKAYKAEKKLSLSFLQVLISNSENDEIRKKSIVDYKTSLSMQLTDLNLSDKYNE AAPLLNENYSNEKGLKNDKKSISLIKNFLDAIKEIEKFIKPLSETNITGEKNDLFYSQFTPLLD NISRIDILYDKVRNYVTQKPFSTDKIKLNFNGYQLLNGWDKDKEREYGAULLCKDEKYIYLAID KSNNRILENIDFQDCDESDCYEKIYKLLPTPNKMLPKVFFAKKKHKKLLSPSDEILKIYKNGTF KKGDKFSLDDCHKLIDFYKESFKKYPKWLIIYNFKFKKTNGYNDIREFYNDVALQGYNISKMKIP TSFIDKLVDEGKIYLFQLYNKDFSPHSKGTPLNHLTYFKMLFDERNLEDVYRLNGEAEFMYRP ASIKYDKPHTPKNTPIKKNKTLNDKRASTFPYDLIKDKRYTKWQFSLHFFITMNFKDPDKAMIN DDVRNLLKSCNNNFIIIGIDRGERNLLYVSVINSNGAIIYQHSNLIIGNKFKGKTYTNYREKLA TREKDRTEQRRNWKAIIESIKELKEGYISQAVHVICQLVVKYDAIIVMEKLTGDFKRGKRFKFEQ VYQKFEKMLIDKLNYYVDKKLDPDEEGLLHAYQLTNKLESFQSGGFIYVVRPDTSTKID PVTGEVNLLYPRYEKIDKAKDMISRFDDIRYNAGEDFFEFDDIDYDKFPKTASDYRKKWITCTNG ERIEAFRNPANNNEWSYRTIILAEEKFELFDNNSINYRSDDDLKAEILSQTGKGFEDFFKLLR LTLQMRNSNPETGEDRILSPVKDKNGNFYDSSKYDEKSKLPDADANGAYNIARKGLWIVEQPK KADNVSTVEPVIHNDKWLKFKVQENDMANN	3
<i>LiCas12a</i> <i>Leptospira</i> <i>ilyithenensis</i>	MKATSIWDNFTRKYSVSKTLRFELRPVGKTEENIVKKEIIDAEWISGKNIPKGTADARRDYKI VKLLNLQHLILFINGALSSENVKEFEKEDKKSKTFVAWSDLLATHFDNWIQYTRDKNSTVLKS LEKSKDLYSKLGLKLLNSKANAWKAEFISYHKIKSPDNKIRLSASNQVILFGNTSDPIQLLKY QIELDNIFLKDGGSEYTTKELADLLSTFEKFGTYFSGFNQNRANVYDIDGEISTSIAYRLFNQ NIEFFFNQIKRWEQFTSSIGHKEAKENLKLQWQDQSKLKELDMEIVQPRFNLKFEKLLTPQSF IYLLNQEGIDAFNTVLGGIPAEVKAEEKQGVNELINLTRQKLNEDKRKFPFLQIMYKQIMSEK TNFIDQYEDDVEMLEKEIQEFSDNDWNEKKKRHSASSKEIKESAIAIYQREFHETFDLEERATVK EDFYLSKSIQNLSDIFGGYNTIHNLYTEVEGMLKSGERPLTRVEKEKLLKQEIYISFAQIER LISKHSQQYLDSTPKEANDRSLFKEKWKKTFFKNGFKVSEYTNLKLNELISEGETFOKIDQETGK ETTIKIPGLFESYENAILVESIKNQSLGTNKKESVPSIKEYLDSCRLRSKFIESFLVNSKDLKE DQSLDGCSDFTQNTLTQWLNEEFDFVILYNKVRNHVTKKPGNTDKIKINFDNATLLDGGWVDK ANFGFLKKADNYYLGIADSSFNQDLKYFNEGERLDEIEKNRKNLEKEESKNISKIDQEKVKKY KEVIDDLKAISNLNKGYSKAFYKQSKFTTILIPKCTTQLNEVIEHFKKFDTDYRIENKKFAKPF IITKEVFLNNNTVYDTATKKFTLKIAGEDTKGLKKFQIGYRATDDKKGYESALRNWITFCIE FTKSYKSCLNYSNLKSVSEYKSLDEFYKDLNGIGYITDFVDISEEYINKKINEGKLYLFQY NKDFSEKSGKENLHTTYWKLFLDSDKNLEDVVIKLNQQAEEVFRPASIEHEKEKITHFKNQEIQN KNPNAVKTTSKFYDIIKDNRTKKNKFLHFCPITLNFKADGNPYVNVNEQENIAPNPVNIIGI DRGEKHLLYFTVINQGGQVILDAGSLNSIKSEYKDKNQQSVSFETPYHKILDKKESERKEAESW QEIENIKELKAGYLSHVHQLSNLIVKYNAIVVLEDLNKGFKRGRFKVEKQVYQKFEKSLIEKL NYLVFKDRKESNEPQGHHLNAYQLTNKFLSFERLGKQSGVLFYATASYSKVDVPTGFMQNIYDP YHKEKTREFYKNFTIIVYNGNYEFNYDLNSVKPDSEKRYRTNWTVCSCVIRSEYDSNSKTQK TYNVNDQVLVLFEDAKIKIENGNDLKSTILEQDDKFIRDLHFYFIAIQKMRVVDKIEKGEDSN DYIQSPVYPFYCSKEIQPNKKGFYELPSNGDSNGAYNIARKGIVILDKIRLRVQIEKLFEDGTG IDWQKLPNLISKVKDKLLMTVFEEWAEALTHQGEVQQGDLGKKMSKKGEQFAEFIKGLNVTK DWEIYTQNEKVVQKQIKTWKFLSNST	4
<i>FsCas12a-S85</i> <i>Fibrobacter</i> <i>succinogenes</i> subsp. <i>succinogenes</i> S85	MNLNTYFSQFTGLYPVSKTLRFELKPMGKTLKIKETGIIENDKKRHNDFDAKKIIDKYHKYF IDAALSKEPCIDWNPLKEAIESRLDRSDASKKKLEKTQTEFRKKIAKALTTHGHYKELTASTPK DLFLKVFPDHFGKQPAIDTFDGFSSYFTGFGQENRQNIYSDEAISTAIPYRLVHDNPPKFLSNIE VYNILKDNAPSVLSDAENELKDFLNGKPLANIFELNAYNDVLTQSGIDFFNQVIGGFSGEGGEK KTRGINEFSNLYRQGHPEFAQKRLATKMIPLYKQILSDRETKSFILSYSTDSQVQESVKEFFE SQILNCIDIAGRKNVNLKELSSLIKRITEFDLGSIIYVNEELSSISLELFSKSWNTINAILFKNAE NRIGSAEKAANKKKIDAWMKSNESFIATLNLIAAESDSEESIRVKIESYWNNEFAKVQSILCGD NRRNLDEFISATFNENNALREDSKVIEKLKAFDLALIEIMHSIKPLISDAENRDLsfYNELMPL YDQLSLVPLYNKIRNYATQKLTSESEKFLNFDNPTLADGWDQNKKEANTAILLLKNGLYLGI MNAKNPKIKDKFTSESDCYDKMYKLLPGPNKMLPKVFFSEKGLATFKPPKDILDGYNAGKH KKGDLFDIGFCHQLIDFFKESIAKHDPDWKKFDFKFSDTSSYEDISGFYKEVTDQGYKITFSKIP TPQIDEWVNEGKFLFLQIYNKDFAPGAKGSPNLHTLYWKSVFSPENLKDVVVKLNGEAEFYRP SSVKKPYSHKVGEKLVNRIGKDGFLPESVFGELFRYFNGKLDGELSDEAKRYLDVAVVKDVKH EIVKDRRYTQDKFEFHVPLTLNFKADSKNEYMNERVRHFLKDNPDVNIIGIDRGERHLLYMTLI NQKGEILKQKSFNIVESVNYQAKLVQREKERDTARRSWSSVGKIKDLKEGFLSQVIEHIEITMTMI ENNAIVVLEDLNFGRKGRFCVERQVYQKFEKMLIDKLNLYLVFNKPEGDVGVLKGYQLAEKF DSFQKLKQSGFLFYIPAAYTSKIDPTTGANLNFNMTELTSAEKKKEFLSHFEDITYDGKNDRF LFSFDYKKFKCFQTDYIKKWTVYSQGKRIVYDKESKSAKAIISPVEIIKAALAKQNIALTDLQLDV LSAINSVEASRETASFFGDIYAFKTLQMRNSIPNTDEDYLVSPVMNKKGEFYDSSRCGSDSLP KNADANGAYHIALKGLYLIKNVFDAGGDKLKISHEDWFKFAQSRNR	5

Name	Sequence	SEQ ID NO:
<i>CsCas12a-AM42-36</i> <i>Clostridium</i> sp. <i>AM42-36</i>	MGKNQNFQEFIGVSPLOKTLRNLIPETETTKKNIAQLDLLTEDEVRAQNREKLKEMDDYYRDV IDSTLRGELLIDWSYLFSCMRNHLSENSKESKRELERTQDSVRSQIHDKFAERADFDMFGASI ITKLLPTYIKQNSKYSEYDESVMKIMLYGKFTTSLTDYFETRKNIFSKEKISSAVGGRIVEEN AEIFLQONQAYDRICKIAGLDLHGLDNEITAYVDGKTLKEVCSDEGFAKVITQGGIDRYNEAIG AVNQYMNLLCQKNKALPKGQFKMKRLHKQILCKGTTSTFDIPKKFENDKQVYDAVNSFTETVTKN NDLKRLLNITQANDYDMNKIYVVADAYSMSISQFISKKNWNLIEECLDDYSDNLPKGKNKENK VKKAVKEETYRSVSQLNEVIEKYVEKTGQSVWVSESYSISLAEMIKLELCHEIDNDEKHNLE DDEKISEIKELLDMMYDVFHIIKVFRVNEVLNFDDETFYSEMDEIYQDMQEIIVLYNHVRNRYVTQ KPYKQEKYRLYFHTPTLANGWSKSKEYDNNAILVREDKYLYGILNAKKPKSKEIMAGKEDCSE HAYAKMNYLLPGANKMLPKVFLSKKGIQDYHPPSYIVEGYNEKKHIGKSKNFDIRFCRDLDY FKCEIKKHPDWNKNFNEFSATETYEDISVFYREVEKQGYRVEWTYINSEDIQKLEEDGQLFLPQ IYNKDFAVGSTGKPNLHLYLKNLFSEENLRDIVLKLNGEAEIFFRKS SVQKPVIIHKCGSILVN RTYEITESTGTRVQSIPESEYMELYRYFNSEKQIELSDEAKKYLDKVCQCNKAKTDIVKDYRYTM DKFFIHLPIITINFVKDGNVNAIAQQYIAEQEDLHVIGIDRGERNLIYVSVIDMYGRILEQKS FNLVEQVSSQGTGRYYDYKEKLQNRREEERDKARKSWKTIGKIKELKEGYLSSVIHEIAQMVMVKY NAI IAMEDLNYGFKRGRFVERQVYQKFETMLISKNLNLADSKQAVDEPGGILRGQMTYVVPDN IKNVGRQCGIIFYVPAAYTSKIDPTTGFINAFKRDVVSTNDAKENFLMKFDSIQYIEKGLFFK SFDYKNFATHKLTAKTKWDVYTNTRIQNMKVEGHWSMEVELTTMKKELDDSHIPYEEGQN ILDDLRMKDITTVNGILEIFWLTVQLRNSRIDNPDYDRIISPVLNNDGEFFDSDEYNSYIDA QKAPLPIDADANGAFICIALKGMYTANQIKENWVEGEKLPADCLKIEHASWLAFLAFMQGERG	6
<i>SlCas12a</i> <i>Saccharobes</i> <i>litoralis</i>	MSDRDLVTNQYPLSKTLRFELKPVGATADWIRKHNVIYRHNGKLVGKDAIRFQNYKYLKMKMLD EMHRLFLOQALVLEPNSNOAQELTALLRAIENNYCENNNDLLAGDYPSSLSTDKTIKISNGLSKLT TDLPDKKFEDWAYQYKEDMPNFWQDIAELEQKLOVSANAKDQKFYKGIKKLKNKIQKSELKA ETHKGLYSPTESLQLEWLVRRGDIKLYLEIGKENEKLNELVPLVELKDIHRNFNNFATLYLGG FSKNRENVYSTKFDRRSGYKATSVIARTFEQNLMFCLGNIKAKWHKVTEFINQANNYELLQEHGI DWNKQIAALEHKLVDVCLAEFFALNNFSQTLAQGGIEKYNQVLAGIAEIAQPKTQGLNELINLA RQKLSAKRSQGLTQLLYKQILSKGDKPFIIDFKSDQELIAELNEFVSSQIHGEHQAIAKLINHE LESFINEARAAQQQIYVPKDKLTLSLLTGSWQAINQWRYKLFQKQLDKQKQYFSFLAQVE RWLATEVEQQNFYQTEKERQQHKDTQPANVTSSDGHSLTAFEEQQVQTLTLLTNICVAAEKYRQL SDNLTAIDKQRESESSKGFQIAVTKTLLDACNELNHFLARFTVNKDKKLPEADRAEFWYEKLQA YIDAFPIYELYNKVRNYSKPFSTTEKVKINFDNHFLSGWTADYERHSALLFKFNENYLLGGV NENLSSEEEELKLVGGEEHAKRFIYDFQKIDNSNPPRVFIRSKGSSFAPAVEKYQLPIDIID IYDQGKFKTEHKKKNEAEFKDSLVRLLIDYFKLGFSRHDSYKHYPFKWKASHQYSDIAEFYAHTA SFCYTLKEENINFNVLRELSSAGKVYLFIEIYNKDFSKNKRGGQRDNLHTSYWKLFLSAENLKD VLKNGQAEIFYRPASLAETKAYTHKKGEVLKHKAYSKVWEALDSPIGTRLSDWDLAKPISITE KTNNHNNQRVVQYNGQEIIGRKAFAI IKNRRYSVDKFLFHCPIITLNFKANGQDNINARVNQFLAN NKKINIIGIDRGEKHLLYISVINQQGEVLHQESFNTITNSYQTANGEKQVVTDYHQKLDMSD KRDARKSWSTIENIKELKAGYLSHVVRHLAQLIIEFNAI VALEDNLHGFKGRFKIEKQVYQK FEKALIDKLSYLAFKDRTSCLTGHYLNALFQTSKFKGFNNLGKQSGILFYVNADYSTTDDPLT GYIKNVYKTYSSVKDSTEFWQRFNSIRYIASENRFEFSYDLADLKQKSLESKTKQTPAKTQWT VSSHVTRSYNNQQTQHELFVETARIQQLLSKAEISYQHNDLIPALASCQSKALHKELIWLFN SILTMVRVTDSSKPSATSENDILSPVAPYFDSRNLNKQLPENGANGAYNIARKGIMLLERIGD FVPEGNKKYPDLLIRNNDWQNFVQRPVEMVNGKQKKLVKLKTEYSNGSLFNDLAFK	7
<i>SdCas12a</i> <i>Succinivibrio</i> <i>dextrinosolvens</i>	MSSLTKFTNKYSKQTLTIKNEIPVGGKTLNENIKENGLIDGDEQLNENYQKAKIIVDDFLRDFINK ALNNTQIGNWRELADALNKEDEDNIEKLDQKIRGIIIVSKFETFDLFFSSYSIKKDEKIIDDDNDV EEEELDLGKKTSSFKYIFKKNLFLKLVLPSYLTNTQDKLKIISDFDNFSTYFRGFFENRKNIFT KKPISTSIAYRIVHDNFPKFLDNIRCFNVWQTECPQLIVKADNYLKSNNVIADKDSLANYFTVYG AYDYFLSQNGIDFYNNIIGGLPAFAGHEKIQGLNEFINQECQKDSSELKSKLKNRHAFKMAVLFK QILSDREKSFVIDEFESDAQVIDAVKNFYAEQCKDNNVIFNLLNLKNIATFLSDDELDGIFIEG KYLSSVSQKLYSDWSKLRNDIEDSANSKQGNKELAKKIKTNKGDVEKAI SKYEFSLSEKNSIVH DNTKFSDLLSCTLHKVASEKLVKVNEDGWPKHLKNNEEKQKIKEPLDALLEIYNTLLIFNCKSF NKNGNFYVDYDRICINELSSVYLYNKTRNYCTKKPYNTDKFLNLFNSPQLGEGFSKSKENDCLT LLFKKDDNYYVGIIRKGAKINFDDTQAIADNTDNCIFKMNYFLLDKAKKFIKPCSIQLKEVKAH FKKSEDDYILSDKEKFASPLVIKKSTFLLATAHVKGKGNIKKFKQKEYSKENPTEYRNSLNEWI AFCKEFLKTYKAATIFDITTLKAAEYADIVEFYKDVDNLCYKLEFCPIKTSFIENLIDNGDLY LFRINNKDFSSKSTGTGNLHLYLQAI FDERNLNNPTIMLNGGAELFYRKESIEQKNRITHKAG SILVNKVKCDGTSLDDKIRNEIYQYENKFIDTSLDEAKKVLPNVIKKEATHDITDKDRFTSDKF FFHCPLTINYKEGDTKQFNNEVLSFLRGNPDINIIGIDRGERNLIYVTVINQKGEILDVSFNT VTNKSSKIEQTVDYEEKLAVREKERIEAKRSWDSISKIATLKEGYLSAIVHEICLLMIKHNAIV VLENLNAGFKRIRGGLSEKSVYQKFEKMLINKLNYFVSKKESDWNKPSGLLNLQLSDQFESFE KLGIQSGFIFYVPAAYTSKIDPTTGANVLNLSKVRNVDAIKSFFSNFNEISYSKKEALFKFSF DLDSLKKGFSFVFKFSKSNVYTFGERI IKPNKQGYREDKRINLTFEMKLLNNEYKVSFDL ENNLIPLNLT SANLKDTFWKELEFFIFKTTQLRNSVTNGKEDVLI SPVKNKAGEFFVSGTHNKT PQDCDANGAYHIALKGLMILERNNLVREEKDTKKIMAINVDWFYVQKRRGV	8

Name	Sequence	SEQ ID NO:
<i>ScCas12a</i> <i>Sedimentisphaera cyanobacteriorum</i>	MKEFTNQYSLTKTLRFELRPVGETAEKIEDFKSGGLKQTV EKDRERTEAYKQLKEVIDSYHRDF IEQAFARQQTLSEEDFKQTYQLYKEAQKEKDGETLTQYEHRLRKKIAAMFSKATKEWAVMGENN ELIGKNKESKLYQWLEKNYRAGRIEKEEFDHAGLIEYFEKFSYFVGFDMKNRANMYSKEAKAT AISFRTINENMVKHFDNCORLEKIKSKYPDLAEELKDFEEFFKPSYFINCMNQSGIDYYNISAI GGKDEKDKQANMKINLFTQKNHLKGS DKPPFFAKLYKQLSDREKSVVVIDE FEKDS ELTEALKN VFSKDGLINEEFFTKLSALENFMLPEYQGGLYIRNAFLTKISANIWGSGSWGIIKDAVTQAAE NNFTRKSDKEYAKKDFYSIAELQQAIDEYIPTLENGVQNASLIEYFRKMNYKPRGSEEDAGLI EEINNLRQAGIVLNQAE LSGSGKQREENIEKIKNLLDSVLNLERFLKPLYLEKEKMRPKAANLN KDFCESFDPLYEKLKTFKLYNKVRNYATKKPYSKDKFKINFDATLLYGWSLDKETANLSVIF RKREKFYLGIIINRYSQIFNYKIAGSESEKGLERKRS LQQKVLAE EGEDYFEKVMVHLLLGASK TIPKCS TQLKEVKAHFQKSS EDYIIQSKSFAKSLTLTKEIFDLNLRNTETGEISSELSDTYP KKFKQGYLTQTGDVSGYKTALHKWIDFCKEFLRCYRNTEIFTFHFKDTKEYESLDEFLKEVDSS GYESFSDIKASYINEKVNAGELYLFEIYNKDFSEYSGKPNLHTIYWKSLFETQNL LDKTAKL NGKAEIFFRPRS IKHNDKIIHRAGETLKNKNP LNEKPPSSRFDYDITKDRRFTKDFKFLHCPITL NFKQDKPVRNFNEQVNL YLKDNDPDVNIIGIDRGERHLLYTLINQNGEILQQGSLNRI GEEESRP TDYHRLLDEREKQRRQARETWKAVEGIDK LKAGYLSRVVHKLAGLMVQNNAI VVLEDLNKGFKR GRFAVEKQVYQNFELKALIQKLNLYLVFKEVNSKDPAGHYLKAYQLTAPFISFEKLGTSQGLFV RAWNTSKIDPATGFTDQIKPKYKNQKQAKDFMSSSFSVRYNRKENYFEFEADFEKLAQPKGRT RTWICSYQGERYSYSPKERKFVKNVNTQNLAE LFNSEGISFDSGQCFKDEI LKVEDASFFKSI I FNLRLLLKL RHTCKNAEIERDFIISP VKGNSSSFFDSRIAEQENITSIPQ NADANGAYNIALKG LMNLHNISKDGKAKLIKDEDWIEFVQKRKF	9
<i>RsCas12a</i> <i>Ruminococcus sp.</i> <i>JE7A12</i>	MSININKFSDECRKIDFFTDLYNIQKTLRFS LIPIGATADNFEFKGRLSKEKD LLSAKRIKEY ISKYLADES DICLSQPVKLKHLEDEYEL YITKDRDEQKFKSVEEKLRLKELADLLKEILKRLNKK ILSDYLP EYLEDDEKALEDIANLSSSSTYFNYSYDNCNMYTDKEQSTAIPIYRCINDNLPKFID NMKAYEKALEELKPSDLEELRNNGFVYD TTVDDMFTLDYFNCVLSQS GIDSYNAIIGNDKVKG INEYINLHNQTAEQGHKVPNLKRLYKQIGSQKKTISFLPSKFE SDESNELLKAVDYFNTGDAEKN FTALKDITIEFEKIFDNLSEYNLDGVFVRNDISLTLNLSQSMFNDWSVFRNLWNDQYDKVKNNEPK AKDIDKYNDRKHVKYKKSSESFINQLQELIATLLEEDINSKKITDYFSCDFHRVTTEVENKYQL VKDLLSSDYPKNKNLKTS EEDVALIKDFLDSVKSLESFVKILTGTGKESGKDELFGSFTWKW QLRYIDKLYDKVRNYITEKPYSLDKIKLSFDNPQFLGGWQHSKETDYSQAQLFMKDGLYLGVMD KETKREFKTQYNTPENDSDTMVKIEYNQIPNPGRVIQNLMLVDGKIVKKNRKNADGVNAVLEE LKNQYLPENINRIKRTESYKTTSNFNKDDLKAYLEYI IARTKEYYCKYNFVFSKSADEYGSFNE FVDDVNNQAYQITKVKVSEKQLLSLVEQGLYLFKIYNKDFSEYSGKKNLHTMYFQMLFDDRN LENLVYKLQGGAE MFYRPASIKKDS EFKHDANVEI IKRCTEDKVNDDKNPTDDEKAKYYSKFYD DIVKNRFTKDQFSLHLTAMNCNQP DHYWLNNDVRELKKSNNKHIIGIDRGERNLIVYTIIN SDGVIQDQINFNI IENSYNGKKYKTDYQKKLNQREEDRQKARKTWKT IETIKELKDGYSIQVVH QICKLIVQYDAIVVMENLNGGFKRGRTKVEKQVYQKFETMLINKLNYVDKGTDYKECGGLLKA YQLTNKFETFERIGKQSGIIFYVDPYLT SKIDPVTGFANLLYPKYETIPKTHNFISNIDDIRYN QSEDYFEFDIDYDKFPQGSYNYRKKWTICSYGNRIKYYKDSRNKTASVVVDITEKFKETFTNAG IDFVNDNIKEKLLLVNSKELLKSFMDTLKLT VQLRNSEINSVDVYIISP I KDRNGNFYYS ENYK KSNNEVP SQPQDGDANGAYNIARKGLMIINKLKKADDVTNNELLKISKKEWLEFAQKGDLE	10
<i>PbCas12a</i> <i>Prevotella brevis</i>	MKQFTNLYQLSKTLRFELKPIGKTLLEHINANGFIDNDAHRAESYKKVKKLIDDYHKDYIENVLN NFKLNGEYLQAYFDLYSQDTKDKQFKDIQDKLRKSIASALKGDDRYKTIDKKELIRQDMKTF LK KDTDKALLDEFYEFTTYFTGYHENRKNMYSDEAKSTAIAYRLIHDNL PKFIDNIAVFKKIANTS VADNFS TIYKNFEEYLVNVSIDEIFSLDYINIVLTQTQIEVYNSIIGGRTLEDDTKIQGINEFV NLNYQQLANKKDRLPKLP LFKQILSDRVQLSWLQEEFTNGADV LNAVKEYCTSYFDNV EESVK VLLTGISDYDLSKIYITNDLALTDVSQRMFGEWSIIPNAIEQRLRS DNPKKTNEKKEYRNIT KLKLPKSYSLGYINECISELNGIDIADYYATLGAINTESKQEPSIPTSIQVHYNALPKILD TD YPREKNLSQDKLTVMQLKDLDDFKALQHF I KPLLNGGDEAEKDEK FYGELMQLWEVIDSITPL YNKVRNYCTRKPFS TEKIKVNFENALQLDGWDENKES TNASII LRKNGMYLIGIMKKEYRNIT KPMPSDGDYDKVYKFFKDIITMVPKCTTQMKSVKEHFSNSNDYTLFEKDKFIAPVVI TKEI FDLNNVLYNGVKKFQIGYLNNTGDSFGYNHAVEIWKSFCLKFLKAYKSTS IYDFSSIEKNIGCY NDLNSFYGAVNLLLYNLTYRKVSVDYIHQLVDEDKMYLFMIYNKDFSTYSKGTPNMHTLYWKML FDESNLNDVVYKLNQAEV FYRKKSITYQHPHPANKPIDNKNVNNPKKQSNFEYDLIKDKRYT VDKFMFHPITLNFKGMGNGDINMQVREYIKTTDDLHFI GIDRGERHLLYICVINGKGEIVEQY SLNEIVNNYKGT EYKTDYHTLLSERDKKRKEERSWQTIEG IKELKSGYLSQVIHKITQLMIKY NAIVLLEDLNMGFKRGRQKVESSVYQQFEKALIDKLNLYLVDKNKDANEIGGLLHAYQLTNDPKL PNKNSKQSGFLFYPAWNTSKIDPVTG FVNLLDTRYENVAKAQAFFKKFDSIRYNKEYDRFEFK FDYSNFTAKAEDTRTQWTLCTYGTRIETFRNAEKNSNWD SREIDLTT EWKTLFTQHNIPLNANL KEAILLQANKNFYTDILHLMKLTLMQRNSVTGTDIDYMVSPVANECGEFFDSRKVKEGLPVNAD ANGAYNIARKGLWLAQQIKNANDLSDVKLAITNKEWLQFAQKKQY LKD	11
<i>HoCas12a</i> <i>Helcococcus ovis</i>	MSFERLTNIAISIKTLRFRLKPVGKTLENLEKLGKLDKDFERNNFYPI LKNIADDYYRQYIRNR LTDNLNDWIKLYYAH ELLNSTDKESKKNLTTIQSEYRKILLNLSGELDKNGEKFSKDIVKKNK ELYGKLFKKEF ILEILPKFVETTNINKEYFNGINLYNKFTTRL SNFWEARKNIFTDKGDIATGI PFRVVNENFVYFYKNIQVFNKNIKYLEDKLDNLEKNL KSEGIMSDKSIKDFNFNPNMHTLYWKML KGIDTYQAIRGGFTKENG EKVGGINEILNLTQQLRRNPYTKNIKLGVLTKLRKQILEYSESTS FLIDQIEDNDLVD RINKFNVSFFESTEVSPSIFVQLENLYNSLRTANSEDIYIDARN TQKFQ MLFGQWDVIRRGYS LKITEGTKEEKKYKYYIELDETSKAKGYLTLMIEIQELVSSVEGYEIDV FNVLLEKFKINI IERLKVETPIYGSPMKLEAIKEYLEKHLEEYHKWKL L INNDEL DLEAFY LLNEV I SDYNI IQLYNLTRNYLTRKYSDEKIKINFDPFTLADGWSESISDNRSIILRKDGNY YLGLEDNKLLDNNITN FLENCYEIMKYNLFPDAAKMIPKCSISKEVKNHFENGEDKSIYLSN QFVGRLEISKELYELQNNLV DGGKKYQIDYLRNTDDKVGYNALNQWITFCCKFLNKYQGTQDF DYSKLKEAKYYDKLDQFYADVDSYGS LDFDTINEDLVNKAVEDGKLLLFQIYNKDFSPESK GK KNLHTLYWLSMFS DENL KARKLKLNGQAEIFYRKKLEKKPIIHKEGSI LLNKIDKDGNTIPENI	12

Name	Sequence	SEQ ID NO:
	YHECYRYLNKKIGRKDLSDKAITLFLNKDVLNYKEARFDI IKDRRYSESQFFHFVPITFNWDLKS NQNVNSIVQNMIKDREIKHIIIGDRGERHLLYYSVIDLEGNIVEQGSNLTLNQNRFDNSIVEVD YQDKLRTREEDRDARKNWTNINKIKELKDGYSHVHKLKSLIIDYEAIVILENLNQGFKRGR FKVERQVYQKFELALMNKLSALSFKETYDEGNLEPSGILNP IQACYPVDSYQDLQGGQNGIVFY LPAAYTSVIDPVTGFTNLFLRLNSINTTKYEEF IKGFKNIYFDNEDLDFKFI FDYKNFEKFNVS FKNKKSKWIVSTRGERISYNSKKKEYFYVKPTEILKNKLIELGINFEDKDKDIIISLDKINDS KKIKLLKVVFDAFKYYSVQLRNHDNIQDYIIISPVADENGNYNSNDVAIKNLKLPDNGDANGAFN IARKGLLLIERISNSDDSKVDLKIKNEDWIDFIIIS	
<i>FsCas12a-UWH8</i> <i>Fibrobacter sp.</i> <i>UWH8</i>	MQAITHQFCGQKNGYSRSITLNRRLIP IGKTEENIQKFLES DKNRADKYPGAKQLIDNLHRDFIA EVLSTHSFDWQPLADSIEKFQKTKDARDKKNLQTQQTNLRKQIAKAFSSSEKGGKFLSKELFTE LLPEYIKGKVDEKANEEIVKEFDRFTTYFTGFYDNRKNMYSDEEQATAISFRLVNFNFPKFLTN AKLFQEIKGKYPEIINDALKSLKNEKIDSFEVNGFNACLTQQGIDAYNQVLGGTAAEAGQKES KGLNECINLYKQQHSVDVIGKMSMLYKQILSDRDGSFIDAFEKDEEDVFKAVQSYHEILISQLSE IEKLFVDAEYDLDKIIVPVKKLTEYSQVCTGRWNVVEESIRQNF IAKHGEPKKKKDEDAKDDEL KDDQSLLELKNILASAPSMKGINIVDYLNNDNLVKQTFSSVELEVKNLEEGFVTLIQKISYKDG SDLKQKDDDDVEHIKIYLDALNLYHYLELVDYRGEAEKDGDFYSTYEKVIERLSGLFLFYLNKVR NYVTKKIDTEKKFKLNFDSPTLANGWDANKESANNAILLRKNGKYYLGFNPNDKPKIDNEATC DASDCYEMKMYKLLPGPNKMLPKVFFSKKGLTFNPPKEILEGYTKEQYKKGDTFPIIFCHKLI DWFKDAINQHPDWKKNFKFSKTESYADISEFYREITEQGYKISFTKIAESETQNLVDCGKFLF FQIYNKYDAENSCGSKNLHTLYWENLFSEENLKNLTVLKLNGEAELEFFRPQVikedKI IAHKKDS YLVNRIGKDGRIPEFSFYQEIYKKANGIIDKISDEAKEFEKNNAVVKKATHDIVKDRFFYQNVQY FHCPITMNFKAELTGKGFNERVQELLAKDPTVKVIGIDRGERHLLYLSLINQKGEIELQKTLN LVELNRNGQTVQVDYQKQLTLKEKERDNARKNWKTIINNIKEIKEGYLSAVVHEIARMMVEHNAI VVMEEYNGYKGRGRFVVERQYQKFELALIEKLNFLVFNKNVSEAGGVNLNAFLQKQKPDSLTD FGKQNGWIFYIPAAYSKIDPKTGFI DFFKLSKVATKNLTNMDAKKSFFKGSSTCVGGFATLFC	13
<i>CsCas12a-AF34-10BH Clostridium</i> <i>sp. AF34-10BH</i>	MNYKTGLEDFIGKESLSKTLRNLALIPTESTKIHMEEMGVIRDELRAEKQOELKEIMDDYYRAF IEEKLQGIQGIQWNSLQKMEETMEDISVRKDLDDKIQNEKRKEICCYFTSDKRFKDLFNAKLIT DILPNFIKDNKEYTEEEKAKEQTRVLFRQFATAFTNYFNQRRNFSEDNISTAISFRIVNENS EIHQNMRARFQRIEQQYPEEVCMEEEYKMDLQEWQMKHIYLVDFYDRVLTQPGIEYYNGICGK INEHNMQFCQKNRINKNDFRMKHLKQILCKSSYYEIPFRFESDQEVYDALNEFIKTMEKEEI ICRCVHLGQKCDYDLGKIYISSNKYEQISNALYGSWDTIRKCIKEEYMDALPGKEKKEEKA AAAKKEEYRSIADIDKIIISLYGSEMDRTISAKKCITEICDMAGQISTDPLVCNSDIKLLQNKEK TTEIKTILDSFLHVYQWGQTFIVSDIEKDSYFYSELEDVLEDFEGITTLYNHVRYSYVTKPY TVKFKLHFGSPTLANGWSQSKEYDNNAILLMRDQKFYLGIFNVRNPKDQKQIKGHEKEEKGDYK KMIYNLPLPGSKMLPKVFIKTSRSGQETYPKSKHILDGYNEKRHIKSSPKFDFLGYCWDLDIYYKE CIHKHPDWKNYDFHFSDDTKDYEDISGFYREVEMQGYQIKWTYISADEIQKLEKGGIFLFIQIYN KDFSVSHTGKDNLHTMYLKNLFSEENLKDIVLKLNGEAELEFFRKASIKTPVHVHKKGSVVLNRSY TQTVGDKEIRVSIPEEYYTEIYNYLNHIGRGLSTEAQRYLEERKIKSFTATKDIVKNRYRCCD HYFLHLPITINFKAKSDIAVNERTLAYIAKKEDIHIIIGIDRGERNLLYISVVDVHGNIREQRSD NIVNGDYDQQLDKREKSRDAARKNWEEIEKIKELKEGYLSMVIHYIAQLVVKYNAVVMEDLN YGFKTGRFKVERQVYQKFETMLIEKLHYLVFKDREVCEEGVLRGYQLTYIPESLKKVKGQCGF IFYVPAGYTSKIDPTTGFINLFSFKNLNRESRQDFVGKFDIEIRYDRDKKMFESFDYNNYIKK GTMLASTKWKVYTNTRILKRIVVNGKYTSQSMEVELTDAMEKMLQRAGIEYHDGDKLKGQIVEK GIEAEIIDIFRLTVQMRNSRSESEDREYDRILISPLVNDKGFEFFDTATADKTLFQDADANGAYCI ALKGLYEVKQIKENWKENEQFPFNKLVDNKTWDFDMQKKRYL	14
<i>BaCas12a</i> <i>Brumimicrobium</i> <i>aurantiacum</i>	MKNQINLFTNKFQLSKTLRFELKPQKKTLEHINSKGFITKNDKRAKADSYKKMKATIDAFHRDFID IAMSNNKLTNLIDFEEIYNASNADKKDEKYTKLSKIQEILRKEIAKGPKGEEVKDIFS KIDKK DLITKLEEWIIEENKIEDIHFDPEFKNFTTYFSGFHQNRKNMYTDQEQSTAIAAYRLIHENLPRF IDNINIFQKINKVPDLEENLKKLYQEIEEYLGINAINEAFELEYFNETLSQKGIDIIYNILGGR TAEKGKQIQGLNEYINLYNQKQDKKNRVPKLVLYKQILSDRTRTSFLPDTFEDDEESSASQK VLDSINNIFYLENLIDYLPNDKNSTINVLENLKLALLAELINFELDKVYIKNDTSITNISMKIFKN YSVIREALNYFYENKIDPNFAHNFNANNTDKKREKLEKEKAKITKQTYLSISFIEEAHLYINE NSNGNQYKNTYKPCIANIYFKDFFIAENKEGSNKEFDFISKIKARYNTIKGVLNTPFPDNKRLH QEKNNIDNIKHFSDIMEYLHFAKPLVLSGSAFAEKDEQFYTNFDELYNQLELIIPLYNKVRNY ATQKPYSTEKFKLNFNSTLLNGWDVNKEEANTSLFIKNGFYLLGIMDKNHNKIFRNP KSTN TDIYKKNYKLLPGASKMLPKVFFGKKNLDDYKPSKDILIRNHGHTHTKGGKQSGFDKDLFNL NDCHKLIDFFKDSIQKHDPWSKFKFSDTQIYESIDQFYRELEPQAYSITYTNIDSSFIEEQI NEGKLYLFQIYNKDFSFKFSNGKPNLHTLYWKALFDEQNLKDVYTKLNGEAEIFYRKKSQIHDRQ IIHKRNQPIINKNPNNEKESIFKYNIIDKRYTIDKFQFHVPTILNFKAKGTDYINVDVLDYL KENPDVKIIGLDRGERHLIYLTLDQKGIKLEQISLNEIVNKKHNITTSYHNLEETKEIERDKA RKNWGTVETIKELKEGYISQVVHKISKMMIEHNAIVVMEDLNMGFKRGRFKVEKQYVQKLEKML IDKLNLYLVKDRQPNPAGIYNALQLTNKFESFQKLKGQSGFLFYVPAWNTSKIDPTTGFINL HVKYESVRKSQEFFNKFNSIKYNPKEAIFEFDFDYNEFTTAEAGTKTNWTVCTYGDRIKTRNP EKLQWQDNKEINITTAFEDFFGRHNIYTGNGSDIKSOLISREEKDFSELIIHLFRLTLQMRNSK TNSEIDYLI SPVKNENGFFYDSRHADKNLPKADANGAYHIAKKGLQWIKELQSFEGNEWKKLK LDKTNKGWLKFVQENQ	15

Name	Sequence	SEQ ID NO:
<i>LbCas12a-MA2020</i> <i>Lachnospiraceae</i> <i>bacterium MA2020</i>	MYYESLTQYVPVSKTIRNELIPIGKTLDNIRQNNILES DVKRKQNYEHVKGILDEYHKQLINEA LDNCTLP SLKIAAEIYLKNQKEVSDREDFNKTQDLLRKEVVEKLKAHENFTKIGKKDILDLEK LPSISEDDYNALLESFRNFYTYFTSYNKVRENLYSDKEKSSVAYRLINENFPKFLDNVKS YRFV KTAGILADGLGEEEQDSLFI VETFNKTLTQDGIDTYNSQVGVKINSINLYNQKNQKANGFRKIP KMKMLYKQILSDREESFIDEFQSDDEVLDNVESYGSVLES LKSSKVS AFFDALRESKGNVYV KNDLAKTAMSNIVFENWRTFDDLLNQEYDLANENKKDDKYFEKRQKELKKNKSYSLHLCNLS EDSCNLIENYIHQISDDIENIIINNETFLRIVINEHDRSRKLAKNRKAVKAIKDFLDSIKVLER ELKLINSSGQELEKDLIVYSAHEELLVELKQVDSLYNMTRNYLTKKPFSTEKVKLNFNRSTLLN GWRNKETDNLGVLLKDGYYLGIMNTSANKAFVNP PVAKTEKVFKKVDYKLLPVPNQMLPKV FFAKSNIDFYNP SSEIYSNYKKGTHKKGNMF SLEDCHNLIDFFKESISKHEDWSKFGFKFSDTA SYNDISEFYREVEKQGYKLTITDIDETYINDLIERNELYLFIQIYNKDFSMYSGKLNHLTYLPM MLFDQRNIDDDVYKLNGEAEVFRPASISEDELI IHKAGEEIKNKNPNRARTKETSTFFSYD IVK DKRYSKDKFTLHIPITMNFQVDEVKRFNDVANSAIRIDENVNVI GIDGERNLLYVVVIDSKGN ILEQISLNSI INKEYD IETDYHALLDEREGRDKARKDWNVTENIRDLKAGYLSQVNVVNAKIV LKYNAIICLEDLNFGRGRQKVEKQVYQKFEKMLIDKLNLYVIDKSREQTSPKELGGALNALQ LTSKFKSFELGKQSGVIYYVPAYLTSKIDPTTG FANLFYMKCENVEKSKRFFDGDGDFIRFNAL ENVFEFGFDYRSFTQACGINSKWTVC TNGERI IKYRNPDKNNMDFEKVVVVTDEMKNLIEQYK IPYEDGRNVKDMI ISNEEA EFYRRLYRLLQQTLMRNSTSDGTRDYIISPVKNKREAYFNSELS DGSVPKADANGAYNIARKGLWVLEQIRQKSEGEKINLAMTNAEWLEYAQTHLL	16
<i>TsCas12a</i> <i>Thiomicrospira</i> sp. <i>XS5</i>	MTKTFDSEFFNLYSLQKTVRFELKPVGETASPVDFKNEGLKRVVSEDERAVDYQKVKEIID YHRDFIEESLNYPPEQVSKDALEQAFHLYQKLAAKVEEREKALKEWEALQKRLREKVVKCFSD SNKARFSRIDKKELIKEDLINWLVAQNREDDIPTVETFNFTTYFTGFGHENRKNIYSKDDHATA ISFRLIHENLPKFFDNVISFNKLKGGPELKFDPVKEDLEVYDLKHAFIEYFVNFVTQAGID QYNYLLGGKTLEDGTTKQGMNEQINLFKQQQTRDKARQIPKLIPLFKQILSERTESSQSPKPKF ESDQELFDSLQKLNHNCQDKFTVLQQA ILGLAEADLKKVFIKTSDLNALSNTIFGNYSVFSDAL NLYKESLKTKAQEA FEKLPAHSIHDLIQYLEQFNSSSLDAEQQSTDTVLNIFYIKTDELYSRFI KSTSEAFQVQPLFELEALSSKRRPPESEDEGAKGQEGFEQIKRIKAYLDTLMEAVHFPAKPLYL VKGKRMIEGLDKDQSFYEAFEMAYQELES LIPIY NKARSYLSRPFKADKFKINFNDNTLLSG WDANKETANASILFKKDGLYLGI MPKGKTF LFDYFVSSSESEKLQRRQKTAEEALADQGESY FEKIRYKLLPGASKMLPKVFFSNKNI GFYNPSDDILRIRNTASHTKNGTPQKGHSKVEFNLND HKMIDFFKSSIQKHP EWGSGFTFSDTSDFEDMSAFYREVENQGVVISFDKIKETYIQSQVEQG NLYLFQIYNKDFSPYSGKPNLHTLYWKALFEEANLNNVVAKLNGEAEIFFRRHSKASDKVVH PANQAIDNKNPHTEKTQSTFEYDLVKDKRYTQDKFFHVPISLNFKAQGVSKFNDKVNFGFLKGN PDVNIIGIDRGERHLLYFTVNVQKGEILVQESLNTLMSDKGHVNDYQQKLDKKEQERDAARKSW TTVENIKELKEGYLSHVVKLAHLIIKYNAIVCLEDLNFGFKRGRFKVEKQVYQKFEKALIDKL NYLVFKEKELGEVGHYLTAYQLTAPFESFKKLGKQSGILFYVPADYTSKIDPTTG FVNF DLRY QSVEKAKQLLSDFN AIFRNSVQNYFEFEIDYKKLTPKRKVGTSKQVWICTYGDVRYQNRNRNQK HWETE EVNVTEKLKALFASDSKTTTVIDYANDNLIDVILEQDKASFFKELLWLLKLTMTLRHS KIKSEDDF ILSPVKNEQGEFYDSRKAGEVWPKDADANGAYHIALKGLWNLQQINQWEKGKTLNL AIKNQDWFSFIQEKPYQE	17
<i>SvCas12a Sneathia</i> <i>vaginalis</i>	MTEEDTKSFVDEIILTPESVIKTIDNFIDSIMNDIEGLKEEF LKISLENFEGIYISNKKLNEI SNRKFGDYN SINMMIKQSMNEKGILSKKEINELIPDLENINKPKVKSFNLSFIFENLTKEHKEL IIDYIRENICNVIENVKITIEKYRNIDNKIEFKNNAEKVSKIKEMLESINELCKLIKEFNTDEI EKNNEFYNILKNFEIFESSYKVLNKVRNFVTKEVIENTKMMLNFSNYQLGNGWHKNKEKDCSI ILFRKRNNERWIYYLGILKHGTHKIKENDYLSVSDTG FYKMDYYAQNSLSKMIPKCSITVKNVKN APEDES VILND SKKFNEPLEITPEIRKLYGNNEHIGDKFKKESLVKWIDFCKEFLLYKYSFEK AKKEILKLKESNLYENLEEFYSDAEKAYFLEFINIDEDKIKKLVKEKNLYLFQIYNKDFSAYS TGNKNLHMTMYFEELFTDENLKKPVFKLNGNTEVFYRIASSPKPIVHNKGEKLVNKTYLDDGIK TIPDSVYEEISEKVKNNE DYSKLLEENNIKLEIKVATHEIVKDKRYFENKFLFYLPITLNNKV SNKNTNKNINKNVIDEIKDCNEYNVIGIDRGERNLISLCIINQNGEIILOKEMNIIQSSDKYNV DYNEKLEIKSKERD NAKKNWSEIGKIKDLKSGYLSAVVHEIVKLAIEYNAVIILEDLNNGFKN RKKVDKQIYQKFERALIEKLQFLIFKNYDKNEKGGLRNAFQLTPELKNITKVASQQGII IYTNP AYTSKIDPTTG YANIIKKSNNNEESIVKAIDKISYDKEKDMFYFDINLSNSFNLTVKNVLKKE WRIYTNGERIIYKDRKYITLNTIQEMKDILSKCGIDYLNIDNLKQDILKNKLHKKVYIYIFELAN KMRNENKDVYIISPVLNKDGKFFMTQEINELTPKDADLNGAYNIALKGLMIDNLNKKKEFV LSNEDWLNFIQGR	18

Name	Sequence	SEQ ID NO:
<i>BsCas12a-OAE603</i> <i>Bacillus sp.</i> <i>OAE603</i>	MNNMFKDFINKYVPTLKFELIPIGNTSENIKKYKIIETDRELEKGYEKVKLLIDEYHRSFI SRVLNNIEFGESLLKYEEFYSHNTDLKREKFEIHKKEMRKKISKAFKDAGAAELFKNTLITKLL PGLYEGKDEVNLNLFKFTTYFKNFHENRKNISSEKSTAIYRIINENLPFIHNIKTYE RICSLIDFDTALEDSFLNQIKKELNCQFFSEFFNIHTFNRVLSQEGINSYNLLGGKSEEDGTK IKGVNELLNLFQETKEKLPKFKFLKKQILSDMDSKSFVLDADFNSDSVLEAISSYHEYLMENI EGSEITLKEFIGQMKNNLDTIYIKDKQSLKSISQKVFSGSWSTITEAIYSFEYDEKNGGKGSN SVKYNEKKNEFNKYEQMAKSNTLKKIYSLSYINKCIKYFGKSEDICDYFIGMGQYGVKEEV PGENLIEAITSNYSIAKPNFIDKILSENELLIEKVRYLDSIKELQMFLLKPLNKEGDKNPLFYG EFDIFYGALESVTPLYNMVRNYVTKKPYSKDKLKNFSNAQLLNGWDKDKESDYLSLLFKKGSK YYLGVNNKIPKVGKCFDCQFEVDSYEEKMEYKQLSNVVANIPRIAFSDSNKSLFSPSNEIL KIQERGSYLKSSIDFIDKDLHQMIDFFKNGLKKKYSEYDFNFSNTSSYSNISDFYQEVIKATYK VKFRKVPKSFIDDLVDVGKFLFLQHLNKDYSIKSGKKNLHTIYWESLSEANLKNPVHKLNGE AEIFFRYSIQRHITHPKGQLIESKREKGFKNKFSYDLIKDKRYTEDKFFLHVPITLNRSANDK GTNPNTEVCNLSKFPPEPHIIGVDRGERHLIYLSVIDSRGKIIHQESLNTITNSYIDGSGKEQV TEINYHSLDKSEERSQSRKNWKKIENIKELKQGYLSHVVRITSLMPKYNGIVILEDLNFGF KRGRFHVEKQVYQKFEKALIDKLNLIIVSKNVNENELGGIRKPQYLTSTKFTSFKEKLGQGTFLFY VNPNTYTKLDPITGFSNQLLIKYESINKTKEFLKFDKIFKNKEENYFFEHVDFSKTQKKVGK TKWVICTKGDRIISFNRRQVYLTNELKELFNKYEIDYKNDIQEQFRQLELSKAFYESFSLYRL TLQMRNNDPNKKDENGNEIDYIISPVKNDNRRFDSRDVKNHEGLPVNGDANGAYNIALKGLML LNEIKATKEGRRPNLAIKNEDWFKFVQNKKEYNG	19
<i>TpCas12a</i> <i>Treponema</i> <i>porcinum</i>	MKISEEFCGQNGYSISKTLRFELPKGKTLNIEKEKLLSDFFKKSQDYKDVKIILDNYHKYF IDDLVQKVNLDWTCLAASITDYNKNEKEDSSVIKEQDFLRKEIVKIIISKDKRFACTASTPKDL FNSILLEWFEKSTEFSLNKKAVETFKRFSYKFGQENRKNMYKDEPIPTAVPYRIVNENFPKF LQNAESFKEIQKKKIDYEVIELVEKELSAYLGNDKLSDIFCVKNFNRYLCQTGAENQGDIDYNNQI IGGIVQKENDVKLRGINEFLNLFWQQHVDFAKDNRRIKFVPLYKQILSDRSLSLFSKIQTLESDE ELKEAVLSFAKKLDSKNKGKNIIDVLVMELTENINQYDLSQIYINQKDMNAVSKTITGDWAYLQ KRMNIFAEETLTKEEQKRWKELDDDTSKSGIFSFEEELNKVLEYSSENCASVSIKIQYFETT KRWFYEQGTGIFTKGEEIIEPSISGLCGQIKSNFDEVNKVFGNVSSENTLRENPEEVEKIKNYL DSVQNLHLRIKPLKVNIGIGDTSFYSEYDEIYSVLVEYISLYNKTRNYISKSGIPEKFKLNFDN PTLADGWDQNKQANTSVILIKDDEYFLGIMNANNKPKFLENYEGNTEKCYQKMIYKLLPGPNK MLPKVFFSTKGIEFTNPPKEILNGYNAGKHKKGDSFDFLDFCHSLIDWFKDAINRHEDWKKFDFK FSETSSYKDISSEFYREISEQGYKLTFTAIPESVVEKMTVDGNLFLFQIYNKDFAKGASGKPNMH TLYWKQLFSKENLSDTILKLNGEAEIFYREPGIKEPIVHKTGSKLVNKTGKGVSPAEIYNEI YKVQNGMQTELSETAQVFVKEHEVSVKTASHDITDKHFTAKFLFHVPITINFKAQGNLSLTMN ERVRFKLNNEPVNIIGLDRGERHLIYFSLINQGEILKQFTFNEVERKQNDRIIKVDVHEKID NREKERDAARKNWTAGIKIAELKEGYLSAVIHELTMMIQQYNAVIVMEDLNFQFKRGRFHVEKQ VYQKFERMLIDKLNLYLVFKDKGFTPEPGVLNGYQLAQGFESFQKLGKQSGFLFYVPAGYTSKID PKSGFADLNLRLDNLNRRKREFFSKFDSIKYDSETMSTSFADYKNFDGKGTETMAKTWTVF SKDKRIVYFPKNKSYSDVFPPTDELKQTFEQAEIKIHDDENLLDVIMEIGADLPKDEKPNQNVAS FWSLLRNFKLILQMRNSNAQTGEDYIISPVKADDGTFDSDRNQILSLGKEAKLPIDADANGAYH IALKGLELLRRFNETDEIKLKKADMKISNADWFKFVQEKQYLN	20
<i>SjCas12a</i> <i>Synergistes jonesii</i>	MANSLKDFTNIIYQLSKTLRFELPKIGKTEEHINRKLIMHDEKRGEDYKSVTKLIDDYHRKFIH ETLDPAHFDWNPALAEALIQSGSKNNKALPAEQKEMREKIIISMFTSQAVYKKLFKKELFSELLPE MIKSELVSDLEKQAQLDAVKSFDFKSTYFTGFHENRKNISYKDKTSTSIAFRIVHQNFPKFLAN VRAYTLIKERAPEVIDKAQKELSGILGGKTLDDIFSIESFNNVLTQDKIDYYNQIIGGVSGKAG DKKLRGVNEFSNLYRQQHPEVASLRKIMVPLYKQILSDRTTSLFVPEALKDDEQAINAVDGLRS ELERNDIFNRIKRLFGKNNLISLDKIWIKNSSISAFSNELFKNWSFIEDALKEFKENEFNGARS AGKKAELWLSKYFSFADIDAAVKSYSEQVSADISSAPSASYFAKFTNLIETAAENGRKFSYFA AESKAFRGDDGKTEIIKAYLDSLNDILHCLKPFETEDISDIDTEFYSAFAEIYDSVKDVIPVYN AVRNYTTQKPFSTEFKFLNFENPALAGWDKNKEQNNTAIILMKDGKYYLGVIDKNNKLRADDL ADDGSAYGYMKMNYKFIPTPHMELPKVFLPKRAPKRYNPSREILLIKENKTFIKDKNFNRDCH KLIDFFKDSINKHKDWRTFGDFSDTDSYEDISDFYMEVQDQGYKLTFTRLSAEKIDKVVEEGR LFLFQIYNKDFADGAQGSFNLHTLYWKAIFSEENLKDVLKLNGEAELEFFRRKSIDKPAVHAKG SMKVNRRDIDGNPIDEGTYVEICGYANGKRDMAASLNAGARGLIESGLVRITEVKHELVDKRYT IDKYFFHVFPFTINFKAQGGNINSDVNLFLRNNKDVNIIGIDRGERNLVYVSLIDRDGHIKLQK DFNIIGMDYHAKLNQKEKERDTARKSWKTIGTIKELKEGYLSQVVHEIVRLAVDNNNAVIMED LINIGFKRGRFKVEKQVYQKFEKMLIDKLNLYLVFKDAGYDAPCGILKGLQLTEKFESFTKLKQKC GIIFYIPAGYTSKIDPTTGFVNLFNINDVSSKEKQKDFIGKLDSIRFDAKRMFTFEFDYDKFR TYQTSYRKKWAVWTNGKRIVREKDKDGKFRMNDRLLTEDMKNILNKYALAYKAGEDILPDVISR DKSLASEIFYVFKNTLQMRNSKRDTEGDFIISPVLNAGGRFFDSRKTDAALPIDADANGAYHIA LKGSILVDAIDEKLEKEDGRIDYKDMVSNPKWFEFQMTRKFD	21

Name	Sequence	SEQ ID NO:
<i>Sc2Cas12a Sulfurimonas crateris</i>	MLSNFTNQYQLSKTLRFELKPVGDTLKHIEKSGLIAQDEIRSQEYQEVKTIIDKYHKAFIGEAL QNVVLSNLEEYEAFFERNRDEKAFKQLQAVLRKEIVAHFKQHPQYKTLFKKELIKADLKNWQE LSDAEKELVSHFDNFTTYFTGFHENRANMYTDEAKHSSIAYRIIHENLPIFLTNKKLFETIKQK APHLAQETQDALLEYLSGAIVEDMFELSYFNHLLSQTHIDLQNMIGGVKQDSLKIQGLNEKIN IYRQANGLSKRELPLNKLPHKQILSDRETLSWLPESFESDEELMQGVQAYFESEVLAFECCDGK VNLEKLPPELLHQTDQDYDFSKVYFKNDLALTAASQAIFKDYRIIKEALWEVKNPKKSKDLVADE EKFFNKKNYSYFSIEQIDGALNSAQLSANMMHYFQSESTKVIEQIQLTYNDWKRNSNKKELLKAF LDALLSYQRLKPLNAPNDLEKDVAFYAYFDAYFTSLCGVVKLYDKVRNFMTKKPYSLKFKLIN FENSTLLDGWDVNKESDNTAILFRKEGLYYLIGIMNKYKYNKVFNRNISSSQDEGYQKIDYKLLPGA NKMLPKVFFSDKNKEYFKPNAKLLERYKAGEHKKGDNDLDFCHELIDFFKTSIEKHQDWKHFA YQFSPTESYEDLSGFYREVEQQGYKISYKNIAASFIDTLVAEGKLYFFQIYNKDFSPYSKGTNP MHTLYWRALFDEKNLADVYKLNQGAEIFFRKKSIEYSQEKLGHHHEMLKDFAYPIIKDRR FAFDKFQFHVPIITLNFKAEGNENITPKTFEYIRSNPDNIKVIGIDRGERHLLYLSLIDAEKGI EQFTLNQIINSYNGKDHVIDYHAKLDAKEKDRDKARKEWGTVENIKELKEGYLSHVIHKIATLI IEHGAVVAMEDLNGFGRGRFVKEQVYQKFEKALIDKLNLYLVDDKKEPHKLGGLNLAQLTSSK FQSFEKMGKQNGFLFYVPAWNTSKIDPVTGFINLFDTRYASVEKSKAFFTKFQSIYNKADYF ELVFDYNDFTEKAKETRSEWTLCTYGERIVSFRNAEKHQQWDSKTIHLTEPKNLNSELHGNV KEYILEQNSVEFFKSLIYLLKITLQMRNSITGTDIDYLVSPVADEAGNFYDSRKADTSLPKDAD ANGAYNIARKGLMLMHRIQNAEDLKKVNLAIISNRDLRNAQGLDK	22
<i>LsCas12a Limihaloglobus sulfuriphilus</i>	MKDFTHQYSLSKTLRFELKPVGETAERIEDFKNQGLKSIVEEDRQRAEDYKMKRILDDYHKEF IEEVLNDIF TANEMESAFEVYRKYMASKNDDKLKKEITEIFTDLRKKIAKAFENRSKEYCLYK GDFSGLINEKKTGDKGPGKLYWLKAKADAGVNEFGDQGTFEQAEALAKNNFTTYFTGFNQ NRDNITYDAEQQTAYSIRVINENMTRYFDNCIRYSSIENKYPELVKQLEPLSGKFAPGNKYDYL SQTADIDYNEAVGHKSDDINAKGINQF INEYQRNSIKGRELPIMSVLYKQILSDINKDLIDK FENAGELLDVAKTLHRELTDKKILLKIKQTLNEFLTEDNSEDIIYKSGTDLTAVSNAIWGEWSV IPKALEMYAENITDMNAKREKWLKREAYHLKTVQEAIEAYLKDNEEFETRNISEYFTNFKSGE NDLIQVVQSAAYAKMESIFGIEDFHKDRRPVTESGEPGEGFRQVELVREYLDLNLINVEHFIPKLIH MFRSGKPIELEDNCNSNFYDPLNEAYKELDVVFGIYNKVRNYVTQKPYSKDKFKINFQNSTLLDG WDVNKESANSVLLKNGKYLGVMKQASINILNYPPEPDSKKNINAKKQLSEIALAGATDDY YEKMIYKLLPDPAPKMLPKVFFSAKNIEFYNPQSQEIYIRENGLFKKDAGDKESLKKWIGFMKTS LLKHPEWGSYFNFEFEPADYQDISIFYKQVAEQGYSVTFDKIKTSYIEEKVASGELYFEIYN KDFSPHSKGRPNLHMTYMKSLFEKENLQNLVTKLNGEAEVFFRQHSIKRNEKVVRANRP IQNK NPLTEKKQSIFEYDLVKDRRTKDKFFLHCPITLNFKEAGPGRFNDKVNKYIAGNPDIRIGID RGERHLLYSLIDQSGRIVEQGTNLQITSTLNSGGREIPKTTDYRGLDTEKERDKARKSWSM IENIKELKSGYLSHIVHLAKLMVKNNAVVLLEDLNFGRGRFRKVEKQVYQKFEKALIEKLN LVFKDARPAEPGHYLNAYQLTAPLESFKKLKQSGFIYVPAWNTSKIDPVTGFVNQFYIEKNS MQYLKNFFGKFDISIRFNPDKNYFEFGFDYKNFNHKAASKWKTCTHGDKRSWYNRQKLEIHN VTENLASLLSGKGINFADGGSIKDKILSVDDASFFKSLAFNFKLTAQLRHTFEDNGEEDCIIS PVAADGTTFFCSETAKKLNMLPHDADANGAYNIARKGLMVLRLQIRESGKPKPISNADWLDFAQ QNED	23
<i>PxCas12a Pseudobutyrvibrio xylanivorans strain DSM 14809</i>	MIIGRDFNMYYQNLTKMYPISKTLRNLIPVGTLENIRKNGILEADIQRKADYEHVKKLMNDY HKQLINEALQGVHLSDSLSDAYDLYFNLSKEKNSVDAFSKCQDKLRKEIVSFLKNHENFPKIGNK EIIKLIQSLNDNDADNNALDSFSNFYTYFSSYNEVRKNLYSDEEKSSTVAYRLINENLPKSLDN IKAYAIKKAGVRAEGLSEEEQDCLFIETTFERTLTQDGIDNYNADIGKLNTAINLYNQNNKKQ EGFRKVPQMKCLYKQILSDREEAFIDEFSDDDELITNIESFAENMNVFLNSEIITDFKNALVES DGSLVYIKNDVSKTLFSNIVFGSWNAIDEKLSDEYDLANSKKKKDEKYIEKRQKELKKNKSYDL ETIIIGLFDDSIDVIGKYEKLESIDITAEAKNDFDEIVLRKHKDNKSLRKNTNAVEAIKSYLD TVKDFERDIKLINGSGQEVEKNLVVYAEQENILAEIKNVDSLYNMSRNYLTQKPFSTEKFKLNF ENPTLLNGWDRNKEKDYLGI LFEKEGMYLGIINNHRKIFENEKLTGKESCFNKIYVQKISN AAKYLSKQINPQNPPKEIAEILLKRRKADSSLSRKETELFIDYLDKDFLVNYPMIINSDGENF FNFHFKQAKDYGSLQEFPKEVEHQAYS LKTRPIDDSYIYRMIDEGKLYLFIHKNDFSPYSKGN INLHTIYLQMLFDQRNLNNVYKLNGEAEVYRPA SINDEEVI I HKAGEEIKNKNKRAVDKPT SKFGYDIIKDRRYSKDKFMLHIPVTMNFVDETRRFENDVNDALRNDEKVRVIGIDRGERNLLY VVVVDTDGTILEQISLNSIINNEYSIETDYHKLLEKEGDRDRARKNWTTIENIKELKEGYLSQ VVNVIAKLVLYNAIICLEDLNFGRGRQKVEKQVYQKFEKMLIDKLNLYLVIDKSRKQEKPEE FGGALNALQLTSKFTSFKDMGKQGTGIYVVPAYLTSKIDPTTGANLFYVYENVEKAKEFFSR FDSISYNNESGYFEFADYKKTDRACGARSQWTVCTYGERI IKYRNADKNNSFDDKTIIVLSEE FKELF'SIYGISYEDGAELKNKIMSVDADFFRCLTGLLQKTLQMRNSSNDGTRDYIISPIMNDR GEFFNSEACDASKPKDADANGAFNIARKGLWVLEQIRNTPSGDKLNLA MSNAEWLEYAQRNQIA S	24

Name	Sequence	SEQ ID NO:
<i>AsCas12a-RM50</i> <i>Anaerovibrio</i> sp. <i>RM50</i>	MVAFIDEFVGQYPVSKTLRFEARPVPETKKWLESQCSVLFNDOQRNEYYGVLKELDDYYRAY IEDALTSFTLDKALLENAYDLYCNRDNTAFSSCCEKLRKDLVKAFGNLKDYLKSGDQLKDLVKL KAKVDAPAGKGKKKIEVDSRLINWLNNNAKYSAEDREKYIKAIESFEGFVTYLTNYKQARENMF SSEDKSTAIAFRVIDQNMVTFYGNIRIYEKIKAKYPELYSALKGFEKFFSPTAYSEILSQSKID EYNYQCIGRPIDDAADFQGVNSLINEYRQKNGIKARELPVMSMLYKQILSDRDNFSMSEVINRNE EAIECAKNGYKVSIALFNELLQLYKKIFTEDNYGNIYVKTQPLTELSQALFGDWSILRNALDNG KYDKDIIINLAELEKYFSEYCKVLDADDAAKIQDKFNLKDYFIQKNALDATALPDLDKITQYKPHL DAMLQAIRKYKLFMSYNGRKKMDVPENGIDFSNEFNAIYDKLSEFSILYDRIRNFATKKPYSDE RMKLSFNMPMTLAGWDYNNETANGCFLFIKDGKYLGLVADSKSKNIFDFKKNPHLLDKYSSKDI YYKVYKQVSGSAKMLPKVVFAGSNEKIFGHLISKRIEIREKKLYTAAAGDRKAVAEWIDFMK SAIAIHPEWNEYFKFKNTAEYDNANKFYEDIDKQTYSEKVEIPT EYIDEMVSQHKLYLFQL YTKDFSDDKKKKGTDLNHTMYWHGVFSDENLKAVTEGTQPIIKLNGEAEFMFRNP SIEFQVTHE HNKPIANKNP LNTKKEVSFNVDLIKDKRYTERKFYFHCPIITLNFADKPIKYNEKINRNVENNP DVCIIIGIDRGERHLLYYTINQTDGIDLEQGSILNKISGSYTNCKEYKVNKETDYHDLDRKEKKG HVAQQAWETIENIKELKAGYLSQVYVYKLTQMLQYNNAVIVLENLNVGFRGRKVKVEQVYQKFE KAMIDKLNLYLVFKDRGYEMNGSYAKGLQTDKFESEFDKIGKQTCGIYYVIPSYTSHIDPKTGIV NLLNAKLYENITKAQDTRKFD S ISYNKADYFEFAFDYRSFGVDMARNEWVVC TCGDLRGV SAKTRETKAYSVTDRKELFKAHGIDYVGGENLVSHITEVADKHFLSTLLFYLRVLKMRVTVS GTENENDFILSPVEYAPGKFFDSREATSTEPMNADANGAYHIALKGLMTIRGIEDGKLNHYGKG GENAAWFKFMQNEQYKNNG	25
<i>AsCas12a-YH12106</i> <i>Acinetobacter</i> sp. <i>YH12106</i>	MNYQLLFQKQFVHLYPISKTLRFELIPQGATQKFI TEKQVLLQDEVRAKYPPEMKAIDGYHKDF IQRALGNIDSQSFEQALQTFQELFLRSQAERSTEAYKKEFETQTQKLRRELIVNSFEKGEFFKQY KSLPDKNLITNLLKPWEKQSQTGDNNYTHNDFNKFTTYFLGFHDNRKNIYSKEPHKTALAYR LIHENLPKFLNENKILRKIQNDHPALWEQLQALHHTMPQLFNGWDSQLLQVSFFSNTLTQTGI DQYNTIIGGISSEGENRQKIQQINELINLYNQKQDKKNRVAKLKQLYKQILSDRSTLSFLLPQQFA DDAELYHAINMFYLDHLHYQSMVNGHSYTLLEVRQLLINELANYDLSKVYLAFLNQLSAVSHQMF GDFGYISRALSYYYMQVDPDYELLASAKTTAKIEAEKLTAFLDAPHSLVVIQAAIDKYLQ LQPSKPHQTQTD FIIISLLKQYETVADDQS IKIINIFSDIEGKYSICIGLVNTESTSESKREIL QNEKLATDIKAFMDAINNVIKLLKPFALNEKLAASVEKDARFYSDFEEIYQALLVFPVLYNKVR NYITQKPYSTEKFKLNFNKPTLLSGWDANKEADNLSILLRKNGNYLAIMDTAGKANKAFEPKA LNQLKVDTTDCYKEMVYKLLPGPNKMFVKVFFSESRAQFNPPQHIIIESYNKKEHISSEAHFD LKCHALIDWFKQCIELHEDWKHFNFKFSPTSQYNSISDFYKEVSEQNYKVHFQIDPADIYIEQL VAEGKLYLFQIYNKDFSPHAKGKENLHTMYFKALFSEENLKQPVFKLSGEAEFMYRPASLQLEN TTIHKAGEAMVAKNPLTPDARTLAYDIIKDRRFTTDKYLHPIISLNFHAQESMSIKKHNDLV RQMIKHNHQLDHIIGIDRGEKHLLYVSVIDLKGNIVYQESLNSIKSEAQNFTPTPYHQLQHREE GRAQARTAWGKIEINIKELKDGYSQVVRHRIQQLILKYNAIVMLEDLNFGFKRGRFKIEKQIYQK FEKALIHKLNVVVDKSTQADELGGVRKAYQLTAPFSEFEKLKGQSGVLFYVPANWTSKIDPVTG FVDLLPKYENLDKAQAFFKTFDSIIFNAKDYFEFKVNLNQFAGLKAQAARAWEITCSYGPER HVYQKKNAAQQGETVIVNVTEELKALFAKNNIEVAEGVELKEMICAQTQVDFFKRLIWLLQVLLA LRYSSSKDKLDYILSPVANVLGEFFDSRHASTHLPQDSANGAYHIALKGLVWIEQLKTAANTE KVNLAISNDEWLRFAQEKL YLT	26
<i>BoCas12a</i> <i>Bacteroidetes</i> oral taxon 274 strain <i>F0058</i>	MRKFNEFVGLYPISKTLRFELPIGKTLLEHQRNKLLEHDAVRADDYVVKKKIDKYHKLCLIDE ALSGFTFDTEADGRSNNLSSEYYLYNLKRRNEQEQTFTKIQNNLRKQIVNKLTSQSEYKRID KKELITTDLPDFLTNESEKELVEKFNFTTYTFEFHKNRKNMYSKEEKSTAIAFRLINENLPKF VDNIAAFEKVVSPLAEKINALYEDFKEYLNVVEISRVFRLDYDELLTQKQIDLYNAIVGGRT EEDNKIQIKGLNQYINEYNQQQTDNRNRLPKLPLKLYKQILSDRESVSWLPPKFDSDKNLLIKIK ECYDALSEKEKVFDKLESILKSLSTYDLSKIYISNDSQLSYISQKMFGRWDIISKAIREDCAKR NPQKSRESLEKFAERIDKKLKTIDSISIGDVDECLAQLGETYVKRVEDYFVAMGESEIDDEQTD TTSFKKNIEGAYESVKELLNADNITDNNLMQDKGNVEIKITLLDAIKDLQRFIKP LLLKGDEA DKDGVFYGEFTSLWTKLDQVTPLYNMVRNYLTSKPYSTKKIKLNFNENSTLMDGWDLNKEPDNT VIFCKDGLYYLIGIMKKYNRVFDREDLPHDGECDKMEYKLLPGANKMLPKVFFSETGIQRFL PSEELLGKYERGTTHKKGAGFDLGD CRALIDFFKKSIERHDDWKKFDKFSDTSTYQDISSEFYRE VEQQGYKMSFRKVSVDYIKSLVEEGKLYLFQIYNKDFS AHSKGTNPMHTLYWKMLFDEENLKDV VYKLNGEAEVFFRKSSITVQSPTHPANSP IKNKNKDNQKKEKFEYDLIKDRRYTVDFKLFHVP ITMNFKSVGGSNINQLVKRHRSATDLHIIGIDRGERHLLYLTVIDSRGNIKEQFSLNEIVNEY NGNTYRTDYHELLDTREGERTARRNWQTIQNIRELKEGYLSQVVIHKKISELAIKYNAVIVLEDL NFGFMRSRQKVEKQVYQKFEKMLIDKLNLYVDKKKPVAE TGGLLRAYQLTGEFESFKTLGKQSG ILFYVPAWNTSKIDPVTGFVNLFDTYENIEKAKVFFDKFKSIRYNSDKDWFEEVDDYTRFSP KAEGTRRDWTICTQGRKRIQCRNHQRNNEWEGQ EIDLTKAFKEHFEAYGVDISKDLREQINTQN KKEFFEELLRLRLTLQMRNSMPSDDIDYILSPVANDTGCFFDSRKQAELENKAVLPMNADANG AYNIARKGLLAIRKMKQEENDSAKISLAISNKEWLKFAQTKPYLED	27

Name	Sequence	SEQ ID NO:
<i>ErCas12a</i> <i>Eubacterium rectale</i> strain 2789STDY5834884	MNNGTNNFQNFIFIGISSLQKTLRNALPTTETTQQFIVKNGI I KEDEL RGENRQILKDIMDDYYRG FISSETLSSIDDIDWTSLFKEKMEIQLKNGDNKDTLIKEQAEKRKAIYKKFADDDRFRKNMFSAKLI SDILPEFVIHNNNYSASEKEEKTQVIKLSRFATSFKDYFKNRANCFSSADDISSSSCHRIVNDN AEIFFSNALVYRRIVKNSNDDINKISGDMKDSLKKMSLEKIYSYEKYGFEITQEGISFYNDIC GKVNFSFMNLYCQKNKENKNLYKLRKLHKQILCIADTSYEVYPKFESDEEYVQSVNGFLDNISSK HIVERLRKIGDNYNGYNLDKIYIVSKFYESVSQKTYRDWETINTALEIHYNILPNGNGSKADK VKKAVKNDLQKSITEINELVSNYKLCPPDDNIKAETIHEISHILNFEAQELKYNPEIHLVESE LKASELKNVLDVIMNAFEWCSVFMTEELVDKDNFAYAELEEIYDEIYPVISLYNLVRNYVTQKP YSTKKIKLNFGIPTLADGWSKSKEYSNNAILMRDNLYYLGIFNAKNKPEKKIEGNTSENKGD YKKMIYNLLPGPNKMIPKVFLSSKTGVETYPKSAYILEGYKQNKHLKSSKDFDITFCRDLIDYF KNCIAIHPPEWKNFGFDFSDTSTYEDISGFYREVELQGYKIDWTYISEKDIDLLQEKQGLYLFQI YNKDFSCKSTGNDNLHMYLKNLFSEENLKDVLKLNGEAEIFFRKSSIKNPIIHKKGSILVNR TYEAEEDQFGNIQIVRKTIPIENIYQELYKYFNDKSDKELSDEAAKLKNAVGHEEAATNIVKDY RYTDKYFLHMPITINFKANTSFINDRILQYIAEKDLHVIGIDRGERNLIVYSVDLTCGNIV EQKSFNIVNGYDYQIKLKQQEGARQIARKWEKIGKIKEIKEGYLSLVIHEISKMVIKYNIAITA MEDLSYGFKKGRFVERQVYQKFETMLINKNLVLFKDISITENGGLLKGYQLTYIPEKLNKVG HQCICIFYVPAAYTSKIDPTTGfVNIFFKFDLTVDAREFICKFDSIRYDSKNLPCFTTFDYNF FITQNTVMSKSSWSVYTYGVRIKRRFVNGRFSNESDTIDITKMEKTLEMTDINWRDGHDLRQD IIDYEVQHFIEIFKLTVMRNSLSELEDNRNYDRLISPLVNNENFIYDSAKAGDALPKDADANG AYCIALKGLYEIKQITENWKEDGKFSRDKLLKSNKDWDFIQNKRYLTS	28
<i>CPbCas12a</i> <i>Candidatus</i> <i>Peregrinibacteria</i> <i>bacterium</i> GW2011_GWA2_33 _10	MSNFFKNFTNLIELSKTLRFELKPVGDTLTNMKDHLEYDEKLQTLKQNDIDDAYQALKPQFDE IHEEFITDSLESKKAKEIDFSEYLDLFQEKKEKLNDSKKLRNKIGETFNKAGEKWKKEYPQYE WKKGSKIANGADILSCQDMLQFIKYKNPEDEKIKNYIDTLKGFFTYFGGFNQNRANYETKKE ASTAVATRIVHENLPKFCNDVIQFKHIIKRKKDGTVEKTERKTEYLNAYQYLKNNKTIQIKDA ETEKMIESTPIAEKIFDVYFSSCLSQKQIEEYNRIGHYNLLINLYNQAKRSEGKHLKSANEKK YKDLPKFKTLYKQIGCGKKDLFYTIKCDTEEEANKSRNEGKESHVSEIINKAQEAINKYRYDL NNDNCENINTVPDFINYLTKENYEGVYWSKAAMNTISDKYFANYHDQLDRLEAKQVFKADKKS EDDIKIPAEIELSGLFGVLDLADWQTLTKSSILSNEDKLLIITDSQTPSEALLKMFINDIEK NMESFLKETNDIITLKKYKGNKEGTEKIKQWFDYTLAINRMLKYFLVKNKIKGNSLDTNISEA LKTLIYSDDAEWFKWYDALRNYLTQKPQDEAKENKLLNFDNPSLAGGWDVNKECSNFCVILKD KNEKKYLAIMKKGENTLFQKEWTEGRGNLTKKSNPLFEINNCEILSKMEYDFWADVSKMIPKC STQLKAVNVHFKQSDNEFIFPIGYKVTSGEKFRECKISKQDFELNNKVFKNELSVTAMRYDL SSTQEKQYIKAFQKEYWELLFKQEKRDTKLTNNEIFNEWINFCNKKYSELLSWERKYKDALTNW INFCKYFLSKYPKTTLFNYSFKESYNSLDEFYRDVDICSYKLNINTTINKSILDRIVEEGKL YLFEIKNQDSNDGKSGHKNLHTIYNNAIFENFDNRPKLNGEAEIFYRKAISKDKGLGVKGGK TKNGTEIINKYRFSKEKFIHVPITLNFCSNNEYVNDIVNTKFNFSNLHLFLGIDRGEKHLAYY SLVNKNGEIVDQGTNLNLPFTDKDGNQRSIKKEKYFYNKQEDKWEAKEVDCWYNDLLDAMASNR DMARKNWQRIGTIKEAKNGYVSLVIRKIADLAVNNERPFIIVLEDLNTGFKRSRQIDKSQVYQK FELALAKLNLFLVDKNAKRDEIGSPTKALQLTTPPVNNYGDIENTKKQAGIMLYTRANYSQTDP TGWRKTIYKLAGPEETTYKDKGKIKNKSVDQIETFTDIGFDGKDYFEYDKGEFVDEKTEGEI KPKKWRLYSGENGKSLDRFRGEREKDYEWKIDKIDIVKILDDLNVNFDKNISLLQLKEGVEL TRNNEHGTGESLRFAINLIQQIRNTGNNERDNDFILSPVRDENGKHFDSSREYWDKETKGEKISM PSSGDANGAFNIARKGIIMNAHILANSDSKDLSLFVDEEWDLHLNNKTEWKKQLNIFSSRKAM AKRKK	29
<i>LbCas12a-MC2017</i> <i>Lachnospiraceae</i> <i>bacterium MC2017</i>	MGLYDGVFNRYSVSKTLRFELIPQGRTRYETNGILSDDEERAKDYKTIKRLIDEYHKDYISR CLKNVNIISCLLEYYHLYNSSNRDKRHEELDALSQMRGEIASFLTGNDEYKEQKSRDIIINERI INFASDEELAAVKRFRKFTSYFTGFTNRENMYSAEKKSTAIHRIIDVNLPKYVDNIKAFNT AIEAGVFDIAEFESNFKAITDEHEVSDLLDITKYSRFRINEDIITYNTLLGGISMKDEKIQGLN ELINLHNQKHPGKKVPLLVLYKQILGDSQTHSFVDDQFEDDQVQVAVNAVAVTDTFSETLLGSL KIIINNIHGYDLDRYIKAGQDITTLKRALNDWHIITECLESEYDDKFPKNKKSSTYEMERNR YVKSFSFSIGRLNSLVTTYTEQACFLENYLGSGFGGDTDKNCLTDFTNLSLMEVEHLLNSEYPVT NRLITDYESVRILKRLDSEMEVIHFLKPLLGNGNESDKDLVYGEFEAEYKLLPVIKVYNRV RNYLTRKPFSTEKIKLNFNSPTLLCGWSQSKEKEYMGVILRKDGQYVYLGIMTPSNKKIFSEAPK PDEDCYKMLVRYIPHPYQMLPKVFFSKSNIAFFNPSDEILRIKKQESFKKGSFNRDDCHKFI DFYKDSINRHEEWRKFNFKSDTDSYEDISRFYKEVENQAFSMSFTKIPTVYIDSLVDEGLYL FKLHNKDFSEHSKGPNLHTVYWNALFSEYNLQNTVYQLNGSAEIFFRKASIPENERVHKKNV PITRKVAELNGKKEVSVPFYDIKNNRYTVDKFQFHVPLKMNFKADEKKRINDDVIEAIRSNKG IHVIGIDRGERNLILYSLINEEGRIIEQRSLNIDSSEGHGTQNYRDLDSREKDRKARENWQE IQEIKDLKTGYLSQAIHTITKWMKEYNAIIVLEDLNDRTNGRKKVEKQVYQKFEKMLIDKLNY YVDKDEEFDRMGGTHRALQLTEKFESFQKLGRQTGFIFYVPAWNTSKLDPTTGFDVLLYPKYKS VDATKDFIKKPDFIRFNSEKNYFEFGLHYSNFTERAIGCRDEWILCSYGNRIVNFRNAAKNNSW DYKEIDITKQLLDLFEKNGIDVKQENLIDSICEMKDKPFFKSLIANIKLILQIRNSASGTDIDY MISPAMNDRGEFFDTRKGLQQLPLDADANGAYNIAKKGLWIVDQIRNTTGNNVKMAINSREWMH FAQESRLA	30

Name	Sequence	SEQ ID NO:
<i>Pb2Cas12a</i> <i>Prevotella bryantii</i> <i>B14</i>	MQINNLIKIIYMKFTDFTGLYLSKTLRFELKPIGKTLENIKKAGLLEQDQHRADSYKKVKKIID EYHKAFIEKSLSNFELKYQSEDKLDSLEEYLMYYSMKRIEKTEDKFAKIQDNLRKQIADHLKG DESYKTIIFSKDLIRKNLPDFVKSDEERTLIKEFKDFTTYFKGFYENRENMYSAEDKSTAIHSHRI IHENLPKFVDNINAFSKIILPELREKLNQIYQDFEEYLNVESIDEIFHLDYFSMVMTOQKQIEV YNAIIGGKSTNDKKIQGLNEYINLYNQKHKDCCKLPKLKLLFKQILSDRIAISWLPDNFKDDQEA LDSIDTCYKNLLNDGNVLGEGNLKLLLENIDTYNLKGIFIRNDLQLTDISQKMYASWNVIQDAV ILDLLKQVSRKKKESAEDYNDRLKKLYTSQESFSIQYLNDCLRAYGKTENIQDYFAKLGAVNNE HEQTINLFAQVRNAYTSVQAAILTTPYPENANLAQDKETVALIKNLLDSLKRLQRFIKPLLGKGD ESDKDERFYGDFTPLWETLNQITPLYNMVRNYMTRKPYSQEKIKLNFENSTLLGGWDLNKEHDN TAIILRKNGLYLLAIMKKSANKIFDKDKLNSGDCYKEMVYKLLPGANKMLPKVFFSKSRIDEF KPSENIENYKKGTHKKGANFNADCHNLIDFFKSSISKHEDWSKFNHFHSDTSSYEDLSDFYR EVEQQGYSISFCDVSVEYINKMVEKGDLYLFQIYNKDFSEFSKGTNPMHTLYWNSLFSKENLNN IIYKLNQAEIFFRKKSILNYKRPTHAPHAQAIKNKNKCNEKESIFDYDLVKDKRYTVDFKQFHFV PITMNFKSTGNTNINQQVIDYLRTEDDTHIIGIDRGERHLLYLVLVIDSHGKIVEQFTLNEIVNE YGGNIYRTNYHDLDTREQNREKARESWQTIENIKELKEGYISQVIHKITDLMQKYHAVVLED LNMGMFRGRQKVEKQVYQKFEMLINKNLNVLNKKADQNSAGGLLHAYQLTSKFESFQKLGKQS GFLYIPAWNTSKIDPVTGTVNLFDTRYESIDKAKAFFGKFDISIRYNADKDWFEAFDYNFTT KAEGTRTNWTICTYGSRTFRNQAKNSQWDNEEDLTAKYAKFAKHGINIYDNKEAIAMET EKSFFEDLLHLLKLTQMRNSITGTTDYILSPVHDSKGNFYDSRICDNLSPANADANGAYNIA RKGLMLIQQIKDSTSSNRKFSPIITNKDWLIFAQEKPYLND	31
<i>Mb2Cas12a</i> <i>Moraxella bovoculi</i> <i>AAX08_00205</i>	MLFQDFTHLYPLSKTVRFELKPIGRTLEHIAKNFLSQDETMDMYQKVKVILDDYHRDFIADM MGEVKLTCLAIFYDVYLKFRKNPKDDGLQKQLKDLQAVLRKESVKGIGSGGKYKTGYDRFLGAK LFKDGKELGDLAKFVIAQEGESSPKLAHLAHFEKFSYTYTGFHDNRKNMYSDEDEKHTAIAYRLI HENLPRFIDNLQIILTTIKQKHSALYDQIINELTASGLDVSLASHLDGYHKLTLQEGITAYNRII GEVNGYTNKHNQICHKSERIAKLRLPHKQILSDGMGVSFPLPSKFADDSEMCQAVNEFYRHYTDV FAKVQSLFDGDDHQKDGIIYVEHKNLNELSKQAFGDFALLGRVLDGYYVDVVPNEFNERFAKAK TDNAKAKLTKEKDKFIKGVHSLASLEQAIEHHTARHDDSVQAGKLGQYFKHGLAGVDNP IQKI HNNHSTIKGFLERERPAGERALPKIKSGKNPEMTQLRQLKELLDNALNVAHFAKLLTTKTTLDN QDGNFYGEFGVLYDELAKIPTLYNKVRDYLKQKPFSTEKYKLNFGNPTLLNGWDLNKEKDNFV ILQKDGCCYLLALLDKAHKKVFDNAPNTGKNVYQKMYKLLPGPNKMLPKVFFAKSNLDYYNPSA ELLDKYAKGTHKKGDNFNLKDCHALIDFFKAGINKHPEWQHFGFKFSPTSSYRDLSDFYREVEP QGYQVKFVDINADYIDELVEQGLYLFQIYNKDFSPKAHGKPNLHTLYFKALFSEDNLADPIYK LNGEAQIFYRKASLDMNETTIHRAGEVLENKNPDNPKKRQFVYDIIDKRYTQDKFMLHVPITM NFGVQGMTIKEFNKKVNQSIQQYDEVNVIIGIDRGERHLLYLTVINSGKEILEQSRSLNDITTASA NGTQVTPPYHKILDKREIERLNARVGWGEIETIKELKSGYLSHVHVIQINQLMLKYNIAIVVLEDL NFGFKRGRFKVEKQIYQNFENALIKKLNHLVLKDKADDEIGSYKNALQLTNNFTDLKSIGKQTG FLFYVPAWNTSKIDPETGFVDLLKPRYENIAQSQAFFGKFDKICYNTRDKGYFEFHIDYAKFTDK AKNSRQKWAICSHGDKRYVYDKTANQNKGAAGKINVNDELKSLFARYHINDKQPNLVMIDICQNN DKEFHKSLMCLLKTLLALRYSNASSDEDFILSPVANDEGVFFNSALADDTQPQNADANGAYHIA LKGLWLLNELKNSDDLNVKLAIDNQTWLNFAQNR	32
<i>Mb3Cas12a</i> <i>Moraxella bovoculi</i> <i>AAX11_00205</i>	MLFQDFTHLYPLSKTVRFELKPIGKTLEHIAKNFLNQDETMDMYQKVKAILDDYHRDFIADM MGEVKLTCLAIFYDVYLKFRKNPKDDGLQKQLKDLQAVLRKEIVKPIGNNGKYKAGYDRFLGAK LFKDGKELGDLAKFVIAQEGESSPKLAHLAHFEKFSYTYTGFHDNRKNMYSDEDEKHTAIAYRLI HENLPRFIDNLQILATIKQKHSALYDQIINELTASGLDVSLASHLDGYHKLTLQEGITAYNTLL GGISGEAGSRKIQGINELINSHHNQCHKSERIAKLRLPHKQILSDGMGVSFPLPSKFADDSEVC QAVNEFYRHYADVFAKVQSLFDGDDYQKDGIIYVEYKNLNELSKQAFGDFALLGRVLDGYYVDV VNPEFNERFAKAKTDNAKAKLTKEKDKFIKGVHSLASLEQAIEHYTARHDDSVQAGKLGQYFK HGLAGVDNP IQKIHNNHSTIKGFLERERPAGERALPKIKSDKSPEIRQLKELLDNALNVAHFAK LLTTKTTLHNQDGNFYGEFGALYDELAKIATLYNKVRDYLKQKPFSTEKYKLNFGNPTLLNGW LNKEKDNFVILQKDGCCYLLALLDKAHKKVFDNAPNTGKSVYQKMIYKLLPGPNKMLPKVFFAK SNLDYYNPSAELLDKYAQGTTHKKGDNFNLKDCHALIDFFKAGINKHPEWQHFGFKFSPTSSYQD LSDFYREVEPQGYQVKFVDINADYINELVEQGLYLFQIYNKDFSPKAHGKPNLHTLYFKALFS EDNLVNP IYKLNGEAEIFYRKASLDMNETTIHRAGEVLENKNPDNPKKRQFVYDIIDKRYTQD KFMLHVPITMNFVQGMTIKEFNKKVNQSIQQYDEVNVIIGIDRGERHLLYLTVINSGKEILEQ SLNDITTASANGTQMTTPYHKILDKREIERLNARVGWGEIETIKELKSGYLSHVHVIISQLMLK YNAIVVLEDLNFGRGRFKVEKQIYQNFENALIKKLNHLVLKDKADDEIGSYKNALQLTNNFT DLKSIGKQTGFLFYVPAWNTSKIDPETGFVDLLKPRYENIAQSQAFFGKFDKICYNADRGYFEF HIDYAKFNDKAKNSRQIWKICSHGDKRYVYDKTANQNKGATIGVNVNDELKSLFTRYHINDKQ NLVMDICQNNDEKFKSLMYLLKTLLALRYSNASSDEDFILSPVANDEGVFFNSALADDTQPQN ADANGAYHIALKGLWLLNELKNSDDLNVKLAIDNQTWLNFAQNR	33

Name	Sequence	SEQ ID NO:
<i>MCas12a</i> <i>Moraxella lacunata</i>	MLFQDFTHLYPLSKTVRFELKPIGKTLEHIHAKNFLSQDETMADMYQKVKAILLDDYHRDFITKM MSEVTLTKLPEFYEVYLLALRKNPKDDTLQKQLTEIQTALREEVVKPIDSGGKYKAGYERLFGAK LFKDGKELGDLAKFVIAQEGESSPKLPQIAHFEKFTSYFTGFHDNRKNMYSDDKHTAIAYRLI HENLPRFIDNLQILVTIKQKHSVLYDQIVNELNANGLDVSLASHLDGYHKLLTQEGITAYNRII GEVNSYTNKHNQICHKSERIAKLRLPHKQILSDGMGVSFLPSKFADDSEMCQAVNEFYRHYAHV FAKVQSLFDRFDDYQKDGIIYVEHKNLNELSKQAFGDFALLGRVLDGYVYDVVNPEFNDKFAKAK TDNAKEKLTKEKDKFIKGVHSLASLEQAIEHYIAGHDDSVQAGKLGQYFKHGLAGVDNPIQKI HNSHSTIKGFLERERPAERTLPKIKSDKSLEMTQLRQLKELLDNALNVVHFAKLLTTKTTLDN QDGNFYGEFGALYDELAKIATLYNKVRDYLQKPFSTEKYKLNFGNPTLLNGWDLNKEKDNFV ILQKDGCCYLLALLDKAHKKVFDNAPNTGKSVYQKMYKLLPGSNKMLPKVFFAKSNLDVYNP SAILDKYAQGTHKKGDNFNLKDCHALIDFFKASINKHPEWQHFGFEFSLTSSYQDLSDFYREVEP QGYQVKFVDIDADYIDELVEQGQLYLFQIYNKDFSPKAHGKPNLHTLYFKALFSEDNLANPIYK LNGEAEIFYRKASLDMNETTIRHAGEVLENKNPNPKERQFVYDIKDKRYTQDKFMLHVPITM NFGVQGMTIKEFNKKVNSIQQYDEVNVIIDRGERHLLYLTVINSGEILEQRSNDIITTS NGTQMTTPYHKILDKRIERLNRVGVGEIETIKELKSGYLSHVHQLISQLMLKYNAIVLEPDL NFGFKRGRFKVEKQIYQNFENALIKKLNHLVLKDKADNEIGSYKNALQLTNNFTDLKSGKQTK FLFYVPWNTSKIDPVTGFDLLKPRYENIAQSQAFFDKFDKICYNADKGYFEFHDYAKFTG AKNSRQIWTICSHGDKRYVYDKTANQNKGATIGINVDELKSLFARYRINDKQPNLVMDICQNN DKEFHKSLTYLLKALLALRYSNASDEDFILSPVANDKGVFFNSALADDTQPQNADANGAYHIA LKLWLNLNELKNSDDLDKVKLAIDNQTLWLNFAQNR	34
<i>BsCas12a</i> <i>Butyrivibrio sp.</i> <i>NC3005</i>	MYQNLTKKYPVSKTIRNELIPIGKTLENIRKNNILES DVKRRQDYEHVKIMDEYHKQLINEA LDNYMLPSLNQAAEITLKKHVDVEDREEFKKTQDILLREVTGRLEHENYTKIGKDKILDLEK LPSISEEDYNALSFNRYTYFTSYNKNVRENLYSDEEKSSTVAYRLINENLPKFLDNKISYAFV KAAGVLADCIIEEEQDALFMVETFNMTLTQEGIDMYNYQIGKVNSAINLYNQKHNVEEFKYP KMKVLYKQILSDREEVFIGEFKDDETLLSSIGAYGNVMTYKSEKINIFFDALRESEGNVYV KNDLSKTTMSNIVFGSWSAFDELLNQEYDLANENKKDDKYFEKQKELKKNKSYTLEQMSNLS KEDISPIENYIERISEIEKICINYGEFEKIVVNEHDSRRLSKNIKAVKVIKDYIDLSIKELEH DIKLINGSQGELEKNLVVYVGQEEALEQLRPVDSLYNLTRNYLTKKPFSTEKVKLNFNKSTLLN GWDKNKETDNLGILFFKDKGYLIGIMNTTANKAFVNPPAAKTENVFKKVDYKLLPGSNKMLPKV FFAKSNIGYYNPFSTELYSNYKKGTHKKGPSFSIDDCNHLIDFFKESIKKHEDWSKFGFEFSDTA DYRDISSEFYREVEKQGYKLTFTDIDESYINDLIEKNELYLFIYNKDFSEYSGKLNHLHTLYFM MLFDQRNLNDNVYKLNGEAEVYRPAEIAENELVIHKAGEGKKNPNRAKVETSTFSYDVIK DKRYSKYKFTLHIPITMNFVDEVRRFNDVINNALRTDDNVNVIIDRGERNLLYVVVINSEGG ILEQISLNSIINKEYDIETNYHALDEREDDRNKARKDWNTIENIKELKTGYLSQVNVVAVKLV LKYNAIICLEDLNFGRGRQKVEKQVYQKFEKMLIEKLNLYVIDKSREQVSPEKMGALNALQ LTSKFKSFAELGKQSGIIYYPAYLTSKIDPTTGfVNLFIKYENIEKAKQFFDGFDFIRFNKK DDMFEFSFDYKSFTQKACGIRSKWIVYTNGERI IKYPNEKNNLFDEKVINVTDEIKGLFKQYR IPYENGEDIKEIIISKAEDFYKRLFRLLHQLTQMRNSTSDGTRDYIISPVKNDRGEEFCSEFS EGTMPKADADANGAYNIARKGLWVLEQIRQKDEGEKVNLSMTNAEWLKYAQLHLLAS	35
<i>HkCas12a</i> <i>Helcococcus kunzii</i> <i>ATCC 51366</i>	MFEKLSNIVSISKTIRFKLIPVGKTLENIEKLGKLEKDFERSDFYPIKNIISDDYYRQYIKEKL SDNLNDWQKLYDAHELLDSSKESQKNLEMIQAQYRKVLFNLSGELDKSGEKNKSDLIKNNKA LYGKLFKKQFILEVLPDFVNNNDSSYSEEDLEGLNLYSKFTTRLKNFWETRKNVFTDKDIVTAIP FRAVNENFGFYDNKIFNKNIEYLENKIPNLENELKEADILDDNRSVKDYFTPNFGNFYVITQD GIDVYQAIRGGFTKENGEKQVGINEILNLTOQQLRRKPKETKNVKLGVLTKLRKQILEYSESTSF LIDQIEDDNDLVDRINKFNVSFESTEVSPSLFEQIERLYNALKSIIKKEEVYIDARNTQKFSQM LFGQWDVIRRGYTVKITEGSKEEKKYKEYLELDETSKAKRYLNIREIEELVNLVEGFEEDVDF SVLLEKFKMNNIERSEFEAPYIGSPIKLEAIKEYLEKHLEEYHKWKLILLIGNDLDTDETfYPL INEVISDYIIIPLYNLTRNYLTRKHSDDKIKVNFDPPTLADGWSESKISDNRSIIILRKGYYY LGILIDNKLLINKKNKSKKIYEILYNYQIPEFSKSI PNYPFTKKVKEHFKNVSDFQLIDGYVS PLIITKEIYDIKKEKKYKDKFYKDNNTNKNLYTYIKWIEFCKQFLYKYKGNKESYKEMYDFS TLKDTSLYVNLNDFYADVNSCAYRVLFNKIDENTIDNAVEDGKLLLFQIYNKDFSPESKGGKNL HTLYWLSMFSEENLRTRKLKLNQAEIFYRKLEKKPIIHKEGSILLNKIDKEGNTIPENIYHE CYRYLNKKIGREDLSDEAIALFNKDVLYKEARFDI IKDRRYSESQFFHVPITFNWDIKTNKN VNQIVQGMIDKEIKHIIIGDRGERHLLYYSVIDLEGNIVEQGSNLNTLQNRFDNSTVKVDYQN KLRTREEDRDARKNWTNINKIKELKDGYSHVHKLRLI IKYEAIVIMENLNQGFKRGRFKV ERQVYQKFEALMKNLSALSFKKYDERKNLEPSGILNPIQACYPVDAQELQGGQNGIVFYLPA AYTSVIDPVTGFTNLFRLKSINSSKYEEFIKFKNIYFDNEEDFKFIPNYKDFAKANLVILNN IKSKDWKISTRGERISYNSKKKEYFYVQPTFLINKLKLNDIYENIDIIPIDNLEEKAKRKI LKALFDTFKYSVQLRNYDFENDYIISPTADDNGNYNSNEIDIDKTNLPNNGDANGAFNIARKG LLLKDRIVNSNESKVDLKIKNEDWINFIISAS	36

Name	Sequence	SEQ ID NO:
<i>LpCas12a</i> <i>Lachnospira</i> <i>pectinoschiza</i> strain 2789STDY5834836	MIMNNVTGDFSEFVAISKVQKTLRNLRLPTPLTMKHKQKGIITEDEYKQQSLELKRIADGYY RDYITHKLNDTNLDFRNLFEAIEEKYKKNDKDNRDKLDLVEKSKRGEIAKLLSADDNFKSMFE AKLITQLLPVYVEQNYIGEDKEKALETIALFKGFTTYFTDYFNIRKNMFKENGGAASSICYRIVN VNASIFYDNLKTFMCIKEKAETETIAIEEELTELLDSWRLEHIFSEDYYNELLAQKQIDYNNQI CGDVNKHMLNYCQQNKLKANVFKMTKLQKQIMGISEKAFEPMPYQNDEEVYAAFNGFISRLEE VKLIDRLGNVLQNSNIYDTAKIYINARCYTNVSSYVYGGWGVIESAIERYWYNTIAGKGQSKAK KIEKAKKDNKFMSVKELDSIVSDYEPDYFNASNMDDDNSGRAFSGHGVLGYFNKMSKLLANMSL HTITYDSGDSLIENTALNIKKDLDLIMSIIYHWLQTFIDEVVEKDNAFYAELEDIYYELENV VTLYDRIRNYVTRKPYSTQKFKNLNFASPTLASGWSRSKEFDNNAIILLRNNKYIIAIFNVNPKP DKQIIKGSEEQRLSTDYKKMVYNLLPGPNKMLPWVFIKSNITGKRDPNPSYILEGVEKNRHKIS SGNFDINychDLIDYYKACINKHPEWKNYGKFKETTQNDIGQFYKDVKEKQGSISWAYISEA DINRDEEGKIYLFETIYNKDLSSHSTGKDNLHTMYLKNIFSEDNLKNICIELNGNAELFYRKSS MKRNIITHKDDTVLVNKTINEAGVRVSLTDEDYIKVYNYNNNDYVIDVEKDKKLVEILERIGHR KNPIDIIKDKRYTEDKYFLHFPITINYGVDENINAKMIEYIAKHNNMNVIGIDRGERNLIIYIS VINNKGNIIEQKSFNLVNNYDYKNKLKNMEKTRDNARKNWQIEGKIKDVKNGLYLSGVISKIARM VVDYNAIIVMEDLNRGFKRGRFVVERQVYQKFENMLISKNLNLFVFEKKADENGGLIKGYQLTY LPKSALQIGKQCGICIFYVPAAYTSKIDPATGFINIFDFKKYSGSAINAKVVKDKKEFLMSMNSR YVNEGSAYEYKIGHRQLFAFSFDYNNFKTYNVSIPVNEWTTYTYGERIKKLYKDGWGSSEVLN LTEDLIELMEQYGIYKDGHDHIREDSHMDERNADFCINLFEKFKYTQVLRSKSEAGDDDD RLVSPVLNSHNGFFDSSDYKENEKSDDIIDDKQIMPKADANGAYCIAKGLYEINKIKENWSD DKKLESELYIGVTEWLDYIQNRRFEAS	37
<i>CmaCas12a</i> <i>Candidatus</i> <i>Methanomethylophil</i> <i>us alvus</i> Mx1201	MDAKEFTGQYPLSKTLRFELRPIGRWTWNLEASGYLAEDRHAECYPRAKELDDNHRFAFLNRV LPQIDMDWHPIAEAFCKVHKNPNGKELAQDYNLQLSKRRKEISAYLQDADGYKGLFAKALDEA MKIAKENGNESEDIVLEAFNGFSVYFTGYHESRENIYSDMDVSAVYRITEDNPPRFVSNALIF DKLNESEHPDIISEVSGNLGVDDIGKYFDVSNYNNFLSQAGIDDYNNHIIIGGHTEDGLIQAFNVV LNLRHQKDPGEKIQFKQLYKQILSVRTSKSYIPKQFONSKEMVDCICDYVSKIEKSETVERAL KLVRNIISSFDLRGIFVNNKRLILSNKILGWDIAETALMHSSSENDKKSVDYSAEAFITLDDI FSSVKKFSDASAEIDGNRAEDICRVISETAPFINDLRVDDLSLNDGGEAAVSKIRSELEPYM DLFHELEIFSVGDEFPKCAAFYSELEEVSEQLIEIIPLFNKARSFCTRKRYSITDKIKVNLKFFT IADGWDLNRKERNKAAIIRKDGKYYLAILDMKKDLSSIRTSDEDESSFEKMEYKLLPSPVKMLP KIFVKSAAKEKYGLTDRMLECYDKGMHKSGSADFGLGFCHELIDYKRCIAEYPGWDFDFKFR ETSDYGSMEFNEVDAGAGYMSLRKIPCEVYRLLDEKSIYLFQIYNKYDSENAGHNKMMHTM YWEGLFSPQNLSESPVKLSGGAEFFRKSSIPNDAKTVHPKGSVLVPRNDVNGRRIPDPSIYREL TRYFNRGDCRISDEAKSYLDKVKTKKADHDIVKDRRFTVDMKMFHVP IAMNFKAISKPNLNKKV IDGIIIDDQDLKIIGIDRGERNLIIYVTMVDKGNILYQDSLNIILNGYDYRKALDVRREYDNKEARR NWTKEGIRKMKEGYLSLAVSKLADMIENNNAIIVMEDLNHGFKAGRSKIEKQVYQKFSMLIN KLGYMVLKDKSIDQSGGALHGYQLANHVTTLASVGKQCGVIFYIPAAFTSKIDPTTGFAADLFA SNVKNVASMREFFSKMKSVIYDKAEGKFAFTFDYLDYNVKSECGRTLWTVTYVTEGFEFTYSRVNR EYVRKVPTDIIYDALQKAGISVEGDLRDRIAESDGDITLKSIFYAFKYALDMRVENREEDYIQSP VKNASGEFFCCKNAGKSLPQSDANGAYNIALKGLILQLRMLSEQYDPNASEIRLPLITINKAWLT FMQSGMKTWKN	38
<i>PgCas12a</i> <i>Parcubacteria</i> group bacterium GW2011_GWC2_44 _17	MENIFDQFIGKYSLSKTLRFELKPVGKTEDFLKINKVFEKDTIDDSYNQAKFYFDSLHQKFID AALASDKTSELSEFQNFADVLEKQNKIILDKKREMGALRRKDKNAVIGIDRLQKEINDAEDIIQKE KEKIKYDVRTLFDNEAESWKTYQEREVDGKKITFSKADLKQKGADFLTAAGILKVLKYEFPPE KEKEFQAKNQPSLFVEEKENPGQKRYIFDSFDKFAGYLTKFQQTCKNLYAADGTSTAVATRIAD NFIIFHQNTKVFDRKYKNNHTDLGDEENIFIEIERYKNCLLQREIEHIIKNENSYNKIIGRINKK IKEDYDQKAKDTKLTSDFPFFKKNLQKILGEVEKEKQLEKTRKTEEDVLIERKEFTENNE ERFTAACKLMNAFCNGEFSEYEGIYLKNKAINTISRWFVSDRDFELKLPQKSKNKKSEKNEP KVKKFISIAEIKNAVEELDGDIFKAVFYDKKIIAQGGSKLEQFLVIWKYEFYLFYLRDIERENGE KLLGYDSCCLKIAKQIGIFFQKEAREKATAVINKYADAGLGIFQMMKYFSLDDKDRKNTPGQLS TNFYAEYDGYKDFEFIKYNEFRNFITKKPFDEDKIKLNFENGALLKGWDENKEYDFMGVILK KEGRILYLGIMHKNHRKLFQSMGNAKGDNANRYQKMIYKQIADASKDVPRLLLTSKKAMEKFKPS QEILRIKKEKTFKRESKNFSLRDLHALIEYYRNCIPQYSNWSFYDFQFQDTGKYQNIKEFTDDV QKYGYKISFRDIDDEYINQALNEGKMYLFEVVNKDIYNTKNGSKNLHTLYFEHILSAENLNDPV FKLSGMAEIFQRQPSVNEREKIITQKNQCILDKGDRAYKYRRYTEKKIMFHMSLVLTNGKGEIK QVQFNKIIINQRISSSDNEMRVNVIIGDRGEKNLLYYSVVKQNGEIIIEQASLNEINGVNYRDKLI EREKERLKNRQSWKPVVKIKDLKKGYSISHVIHKICQLIEKYSATVLEDLNMRFKQIRGGIERS VYQQFEKALIDKGLYLVFKDNRLRAPGGVLNGYQLSAPFVSFEKMRKQTGLFYTQAEYTSKT DPITGFRKNVYISNSASLDKIEAVKKFDAIGWDGKEQSYFFKYNPYNLADEKYKNSTVSKWEA IFASAPRIRROKGEDGYWKYDRVKVNEEFKLLKVWNFVNPATDIKQEIIEKKEAGDLQGEKE LDGRLRNFWHSFIYLFNLVLELRNSFLQIKIKAGEVIAVDEGVDFIASPVKFFFTTPNPYIPS NLCWLAVENADANGAYNIARKGMILKKIREHAKKDPEFKKLPNLFISNAEWEAARDWGKYAG TTALNLDH	39
<i>PrCas12a</i> <i>Pseudobutyrvibrio</i> <i>ruminis</i> CF1b	MIIGRDFNMYYQNLTKMYPISKTLRNLIPVGTLENIRKNGILEADIQRKADYEHVKKMLMDNY HKQLINEALQGVHLSDLSDAYDLYFNLSKEKNSVDAFSKQDKLRKEIVSLKNHENFPKIGNK EIIKLLQSLYDNDTDYKALDSFSNFYTYFSSYNEVRKNLYSDEEKSSTVAYRLINENLPKFLDN IKAYAIKAGVRAEGLSEEDQDCLFIETFTERTLTQDGIDNNAIIGKLNAINLNFQNNKKQ EGFRKVPQMKCLYKQILSDREAFIDEFSDDDELITNIESFAENMMNVFLNSEIITDFKIALVES DGSIVYIKNDVSKTSFSNIVFGSNAIDEKLSDEYDLANSKKKKKDEKYYEKRKELKKNKSYDL ETIIGLFDNDSVIGIKYEKLESDITAEAKNDPDEIVLRKHDKNKSLRKNTNAVETAKSYLD TVKDFERDIKLINGSQGEVEKNLVVYAEQENILAEIKNVDSLNNMSRNYLTQKPFSTEFKFLN NRATLLNGWDKNKETDNLGILFEKDGMYLIGIMNTKANKIFVNPICKATSDNVYHKVNYKLLPGP NKMLPKVFFAQSNDLYYKPSSELLAKYKAGTHKKGDNFSLEDCHALIDFFKASIEKHPDWSSFG FEFSETCTYEDLSGFYREVKEQGYKITYTDVDADYITSLVERDELYLFQIYNKDFSPYSKGNLN	40

Name	Sequence	SEQ ID NO:
	LHTIYQLMLFDQRNLNNVYKLNGEAEVFRPASINDEEVI IHKAGEEIKNKNKSKRAVDKPTSK FGYDI IKDRRYSKDKFMLHIPVTMNFVDETRRFNDVVDALRNDEKVRVIGIDRGERNLLYVV VVDTDGTILEQISLNSI INNEYSIETDYHKLDEKEGDRDRARKNWTTIENIKELKEGYLSQVV NVI AKLVLYKNAI ICLEDLNF GFKRGRQKVEKQVYQKFEKMLIDKLNLYLVDKSRKQDKPEEFG GALNALQLTSKFTSFKDMGKQTGI IYYPAYLT SKIDPTTG FANLFYVKYENVEKAKEFFSRFD SISYNNESGYFEFAFDYKKFTDRACGARSQWTVCTYGERI IKFRNTEKNNSFDDKTI VLSEEFK ELFSIYGISYEDGAELKNKIMSVD EADF FRS LTRLFQQTMQMRNSSNDVTRDYI IISPIMNDRGE FFNSEACDASKPKDADANGAFNIARKGLWVLEQIRNTPSGDKNLNAMSNAEWLEYAQRNQIAS	
<i>FnCas12a</i> <i>Francisella</i> <i>tularensis</i> subsp. <i>novicida</i> strain <i>U112</i>	MSIYQEFVNKYSLSKTLRFELIPQGKTLLENIKARGLILDDDEKRAKDYKKAQI IDKYHQFFTEB ILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSADTIKKQISEYIKDSEKFNLFNQNL I DAKKGQESDLILWLKQSKDNGIELFKANSIDITDIDEALEI IKSFKGWTTFYFKGFHENRKNVYS NDIPTSI IYRIVDDNLPKFLENKAYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSSEVNQR VFSLDEVEFIANFNNYLNQSGITKFNTI IGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYK MSVLFKQLSDTESKSFVIDKLEDDSDVVTMTQSFYEQIAAFKTVEEKS IKETLSLFFDDLKAQ KLDLSKIYFNKDKSLTDLSSQVDDYSVIGTAVLEYITQIAPKNLDNPSKKEQELIAKTEKA KYL SLETIKLAL EEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQIS KYQNQGGK DLLQASAEDDVKA IKDLDLQTNLHLKLIKFIHSQSEDKANILDKDEHFYLVFEECYFELANIV PLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYLLGVMMKNKNKI FDDKAIKENKGEQYKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILIRINHSHTTKNGSPQ KGYEKFEPNIEDCRKFI DYFKQSI SKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLT FEN ISESYIDSVVNQGLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFY RKQSI PKKITHPAKEAIANKNKNPKKESVF EYDLIKDKRFTEDKFFHF CPTITINPKSSGANKF NDEINLLKKEKANDVHILSIDRGERHLAYITLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI EKDRDSARKDWKKINN IKEMKEGYLSQVVHEIAKLVI EYNAIVVFEDLNF GFKRGRFVKVEKQVY QKLEKMLIEKLNLYLVFKDNEFDKTGGVLRA YQLTAPFETFKKMGKQTGI IYYPAGFTSKICPV TGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASGFRI I NFRNSDKNNHNDTREVPYTKLEKLLKDYSEYGHGECIKAAICGESDKKFFAKLTISVLNTILQ MRNSKTGTELDYLISP VADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNQGEK KLNLI VIKNEEYFEFVQNRNNG	41
<i>BfCas12a</i> <i>Butyrivibrio</i> <i>fibrisolvens</i>	MYESLTKLYPIKKTIRNELVPIGKTLENIKKNNILEADED RKIAYIRVKAIMDDYHKRLINEA LSGFALIDLDKAANLYLSRSKSADTIESFSRFQDKLRKAIAKRLREHENFGKIGNDKI IPLLQK LSENEDDYNAL ESKNFYTYFESYNDVRLNLYSDKEKSSTVAYRLINENLRFPLDNIRAYDAVQ KAGITSELSSEAQDGLFVNTFN NVL IQDGINTYNEIDIGKLNVA INLNYQKNASQGRFVKVPK MKVLYKQILSDREESFIDEFESDTELLDSLESHYANLAKYFGSNKVLQFLTALRESKGVNVYVK NDIAKTSFSNVVFGSWSRIDELINGEYDDNNRKKDEKYIDKRQKELKKNKSYTIEKIITLSTE DVDVIGKYIEKLESDIDDIRFGKNFYEA V LCGHDSRKKLSKNKGAVEAIKGYLD SVKDFERDL KLINGSGQELEKNLVVYGEQEA V LSELSGIDSLYNMTRNYLT KKPFTSTEKIKLNFNKPFTFLDGW DYGNEEAYLGFFMIKEGNYFLAVMDANWNKEFRNIPSVDKSDCYKKVIYKQISSPEKSIQNLNV IDGKTVKKNRKEKEGIHSGENLILEELKNTYLPKKINDIRKRRSYLNGDTSFKDLTEF IGYY KQRVIEYNGYSFYFKSDDDYASFKEFQEDVGRQAYQISYVDVVPVSFVDDLINSGLKLYLFRVYN KDFSEYSGKRLNLHTLYFKMLFDERNLKNVYKLNQAEVFRPSSIKKEELIVHRAGEIKKN NPKRAAQKPTRRLDYD IVKDRRYSQDKFMLHTSIIMNFGAEENVSFNDIVNGVLNEDKVNIVIG IDRGERNLLYVVVIDPEGKILEQRSINCIITDSNLDIETDYHRLDEKESDRKIARRDWTTIENI KELKAGYLSQVVHIVAEVLVLYKNAI ICLEDLNF GFKRGRQKVEKQVYQKFEKMLIDKLNLYVMD KSREQLSPEKISGALNALQLTPDFKSFKVLGKQTGI IYYPAYLT SKIDPMTGFANLFYVKYEN VDKAKEFFSKFDSIKYKNDGKNWNTKGYFEFAFDYKKFTDRAYGRVSEWTVCTVGERI IKFNK EKNNSYDDKVIDLTNSLKELFDSYKVTYSEVDLKDAILAIDDPAFYRDLTRRLQQLTQMRNSS CDGSRDYIISP VKNSKGEFFCS DNDDTTPNADANGAFNIARKGLWVLNEIRNSEEGSKINLA MSNAQWLEYAQDNTI	42
<i>SrCas12a</i> <i>Succiniclasicum</i> <i>ruminis</i>	MGNFGEFTHKYQVSKTLRFELIPQGKTLNVAKYIGVDDDKRRSENYKKLKPVIDRIYKYFIDE SLKNVSI DWQPLYEAI IAYRKEQTANVVRLEEQEA CRKATAAWFEGKVPDKGSKDLKEFNKT QSKLFKELFGKELFTESVTQLPLGLSLTEEEKELLASFNKFTSYFKGFYVNRKNVFSADDISTS IPHRLVQENFPKFMNDNCEAYRRIVEEYPELKAKEGTAQATGIFIGFKLDNIFKVSFYNHLLQ SQIDL YNQLFCGIAGEEGTMRVQGLNVTLNLAMQDKVLGQKLKSMPHRFIPLYKQILSDRTTL SF IPEAFQND EEVLLTVEEYRKSLEAERTTGAVSDIFNSLQAADLRHVYNPAKLTAFSQMLFE DWSLCRESLRNWKLSYGKAATKKVREEIESWLKESAI SLD ELQAALADGTL SVIINQVQSVI TTLEQELAKPLPKKLKTAEEKESLSLLDSVQEA CHSLEMFAVGENMDTDP CFYVPLREAMEAI QP I IPLYNKVRNFATQKPY SIEKFKLNF SNPILASGWDENRERQTCAILFRKGEKYLLGIYNAK VKPDFSIIKAVKGKCNFEKVYRQFPDFSKMMPKCTTQLKEVQQHFASSEDDYLVNKKFIKPL TITKEIYDLNVLDFGKKKQIDYLRKTKDEDDGYHALHTWINF AKEFVASYESTSIYDTSTVL STEQYVKLNDFYGDLDNLFYRIKFESVSEETISEFVDEGKFLFQIYNKDFAEAGATGAPNLHTI YWKAVDFDENMKNVVVKLNQAE LFYRPKSAMDIVRHKVGEKLVNRRLDGTSLTEELHEELY YANGKLKKKLSEAAAALPQAVIYDVHHEIVKDRRFTEDKFFFHVPLTLN YKCDKNAVQFNASV QEY LKENPDYIIGIDRGERNLIYAVVIDPQGNIVEQKSFNVINGFDYHNKLEQREKERNKARQ DWTTVGKIKELKQGYLSLVVHEITSMVMKYNAIVVLENLNVGFKRIRSGIAEKAVYQQFEKMLI NKLNYLMFKDVEGAKPGSVLNAYQLTDRFESFASMRNQIGFLFYIPAAFTSKIDPATGFVDPFC WSA IKTLDKKTIFSGFDTLKYDNVTGNF I LHFEMKKNKDFQKKLEGFMP EWDIVVEANKDRRD AEGKTFISGRIEFVRENNGHGHYEDYLPCKKLVEILRQYDILFEDGKDVPLIMKNGSKLH EVFKVIRLSLQMRNSNAESGEDFISSPVENNEGICFDSRLGVETLPKADANGAYHIALKGLLL LEKIRHDERKLGISNSEWLNHIQSLRG	43

Name	Sequence	SEQ ID NO:
<i>LbCas12a-MD335</i> <i>Lachnospiraceae</i> <i>bacterium MD335</i>	MHENNGKIADNFIGIYPVSKTLRFELKPVGKTQEYIEKHGILDEDLKRAGDYKSVKKIIDAYEHK YFIDEALNGIQLDGLKNYYELYEKKRDNNEEKEFQKIQMSLRKQIVKRFSEHPQYKYLKFKELI KNVLPFTKDNAAEQTLVKSQFEFTTYFEGFHQNRKNMYSDEEKSTAIAYRVVHQNLPKYIDNM RIFSMILNTDIRSDLTELFNNLKTMDITIVEEYFAIDGFNKVVNQKIDVYNTILGAFSTDDN TKIKGLNEYINLYNQKNKAKLPKLPKLFKQILSDRDKISFIEPQFSDSTEVLAVDMFYNRLLQ FVIENEGQITISKLLTNFSAYDLNKIYVKNDDTISAINDLFDWWSYISKAVRENYDSENVNKN KRAAYEKKKEKALSKIKMYSIEELNFFVKKYSCNECHIEGYFERRILEILDKMRYAYESCKIL HDKGLINNISLCQDRQAISELKDFLDSIKEVQWLLKPLMIGQEADKEEAFYTELLRIWEELEP ITLLYNKVRNYVTKKPYTLEKVKLNFYKSTLLDGDWKNKEKDNLGIILLKDGQYYLIGIMNRRNN KIADDAPLAKTDNVYRKMEYKLLTKVSANLPRIFLKDKNPSEEMLEKYEKGTHLKGNFICID CRELIDFFKKGIKQYEDWQDFKFSDESYYDISAFYKEVEHQGYKITFRDIDETYIDSLVNE GKLYLFQIYNKDFSPYSKGTKNLHTLYWEMLFSSQNLQNIYKVLNGNAEIFYRKASINQKDVVV HKADLPKKNKDPQNSKKESMFYDIIKDKRFTCDKYQFHVPTMNFKALGENHFNKRVNRLIHD AENMHIIIGIDRGERNLIYLCMIDMKGNIVKQISLNEIISYDKNKLHKKRNYHQLLKTREDENKS ARQSWQTIHTIKELKEGYLSQVIHVIDLMVEYNAIVVLEDLNFQKQGRQKFERQVYQKFEKM LIDKLNLYLVDSKGMDEDDGGLLHAYQLTDEFKSFQKLGKQSGFLYYIPAWNTSKLPTTGFVNL FYTKYESVEKSEFINNFTSILYNQEREYFEFLDYSAFTSKAEGSRLEKWTVCSEKGERVETYRN PKKNNEWDTQKIDLTFFELKKLFNDYSISLLDGLREQMGKIDKADFYKKFMKLFALIVQMRNSD EREDKLI SPVLNKYGAFFETGKNERMPLDADANGAYNIARKGLWIEIKNTDVEQLDKVKLTI SNKEWLQYAEHIL	44
<i>CMtCas12a</i> <i>Candidatus</i> <i>Methanoplasma</i> <i>termitum</i>	MNNYDEFTKLYPIQKTIREFELKPVQGRTEHLETFNFFFEEDRDRAEKYKILKEAIDEYHKKFIDE HLTNMSLDWNLSKQISEKYYKSREEKDKKVFLEQKMRQEIVSEFFKKDDRFKDLFSKFLFSEL LKEEYKKGNGHQEIDALKSFDKFSGYFIGHLHENRKNMYSGDGEITAISNRIVNENPKFLDNLQ KYQEARKKYPEWIIKAESALVAHNKIMDEVFSLEYFNKVLNQEGIQRYNLAALGGYVTKSGEKM GLNDALNLAHQSEKSSKGRIHMTPLFKQILSEKESFSYIPDVFTEDSLLPSIGGFFAQIENDK DGNIFDRALELISSYAEYDTERIYIRQADINRVSNVIFGEWGTLLGLMREYKADSNINDINLERT CKKVDKWLDSKEFALSDVLEAIKRTGNNDAFNEYISKMRAREKIDAARKEMKFISEKISGDEE SIHIIKTLLDSVQQLHFHFNLFKARQDIPLDGAFYAEFDEVHSEKLFVPLYNKVRNYLTNNL NTKKIKLNFKNPTLANGWDQNKVYDYASLIFLRDGNYYLGINPKRKNKIKFEQSGSGNGFFYRK MVYKQIPGPNKNLPRVFLTSTKGKKEYKPSKEIEGYEADKHIRGDKFDLDFCHKLIDFFKESI EKHKDWSKFNFYFSPTESYGDISEFYLDVEKQGYRMHFENISAETIDEYVEKGDLEFLFQIYNKD FVKAATGKDMHTIYWNAAFSPENLQDVVVKLNGEAEFLYRDKSDIKEIVHREGEILLVNTYNG RTPVPDKIHKKLTDYHNGRTKDLGEAKEYLDKVRYFAHYDITKDRRYLNDKIYFHVPLTLNFK ANGKKNLNMVIEKFLSDEKAHIIIGIDRGERNLLYYSIIDRSGKIIDQQSLNVIDGFDYREKLN QREIEMKDARQSWNAIGIKIDLKEGYLSKAVHEITKMAIQYNAIVVMEELNYGFKRGRFKVEQ IYQKFENMLIDKMNYLVFKDAPDESPPGVLNAYQLTNPLESFAKLGKQTGILFYVPAAYTSKID PTTGFVNLFTNTSSKTNAQERKEFLQKFESISYSAKDGIFAFADYRKFGTSKTDHKNVWTAYT NGERMRYIKEKKRNELFDPSEIKEALTSSGICYDGGQNILPDILRSNNNGLIYTMYSFIAAI QMRVYDGKEDYIISPKNKSGEFFRTDPKRRELPIADANGAYNIALRGELTMRAIAEKFPDPS EKMAKLELKHKDWFEFMQTRGD	45
<i>Cas12a uncultured</i> <i>Clostridium</i> sp.	MEDKQFLERYKEFIGLNSLSKTLRNSLIPVGSTLKHIEQYEGILEEDSLRAQKREELKGMDDYY RNYIEMHLRDVHIDWNELFEALTEVKKNQTDADAKKCLEKIQEKKRKEIYQYLSDDAVFSEMPK EKMISGILPDFIRCNEEYSEEEKEKLTVALFHRFTSSFNDFFLNRKNVFTKEAATAIGYRV VHENAEIFLENMVAFQNIQKSAESQISIERKNEHYFMEWKLSHIFTADYYMMLMTQKAIIEHYN EMCGVVNQHMKKEYCQKEKKNWNLRYMRKLHKQILSNASTSFKIPEKYENDAEVYESVNSFLQNV MEKTVMERIAVLKNNTDNFDLSKIYITAPYYEKISNYLCGSWNTIADCLTHYEEQQIAGKGARK DQKVKAADVADKWKSLSEIEQLLKEYARAEEVKRKPPEYIAEIEINIVSLKEVHLLEYHPEVNL ENEKYATEIKDVLNDYMELFHWMKWFYIEEAVEKEVNFYGELEDDLYEEIRDIVPLYNKVRNYVT QKPYSDTKIKLNFGTPTLANGWSKSKEYDYNAILLQKDGKYYMGIFNPNVQKPEKEIEIGHSHPL EGNEYKKMVYYYLP SANKMLPKVLLSKKGMEIYQPSYIINGYKERRHIKSEKFDLQFCHDLI DYFKSGIERNPDWKFVGFHSDTDYQDISGFYREVEDQGYKIDWTYIKEADIDRLNEEGKLYL FQIYNKDFSEKSTGRENLTMYLKNLFSEENIREQVLKLNGEAEIFFRKSSVKKPIIHKKGTM VNRTYMEEMHGESVKKNIPEKEYQEIYNYMNRHWKGELSAEAEKYLKAVCHETKKDIVKYRY SVDKFFIHLPTITINRASGKEALNSVAQRYIAHQNDMHVIGIDRGERNLIYVSVINMQGEIIEQ KSFNVVNKYNYKEKLKEREQNRDEARKNWKEIGQIKDLKEGYLSGVIHEIAKMMIKYHAIVAME DLNYGFKRGRFKVERQVYQKFENMLIQKLNLYLVFKDRSADEDGGVLRGYQLAYIPDSVKKLGRQ CGMIFYVPAFTSKIDPATGFVDIFNHKAYTTDQAKREFILSFDEICYDVERQLFRFTFDYANF ATHNVTLARNNWTIYTNGTRTQKEFVNRRVRDKKEVDPTEKMLKLELEGVEYQSGANLLPKL EKISDPHLFHELQRIVRFTVQLRNSKNEENDVDYDHVISPVLNNEGKFFDSSKYENKEKKESL LPVDADANGAYCIALKGLYIMQAIQKNWSEKALSPDVLRLNNNDWFDYIQNKRYR	46

Name	Sequence	SEQ ID NO:
<i>Lachnospiraceae</i> <i>bacterium COE1</i>	MHENNGKIADNFIGIYPVSKTLRFELKPVGKTQEYIEKHGILDEDLKRAGDYKSVKKIIDAYEK YFIDEALNGIQLDGLKNYYELYEKKRDNNEEKFQKIQMSLRKQIVKRFSEHPQYKYLKKELI KNVLPFTKDNAAEQTLVKSQFEFTTYFEGFHQNRKNMYSDEEKSTAIAYRVVHQNLPKYIDNM RIFSMILNTDIRSDLTELFNNLKTMDITIVEEYFAIDGFNKVVNQKIDVYNTILGAFSTDDN TKIKGLNEYINLYNQKNKAKLPKLPKLPKQILSDRDKISFIEQFDSDETEVLEAVDMFYNRLLQ FVIENEGQITISKLLTNFSAYDLNKIYVKNDDTISAIISNDLFDDWSYISKAVRENYDSENVKDN KRAAAYEEKKEKALSKIKMYSIEELNFFVKKYSCNECHIEGYFERRILEILDKMRYAYESCKIL HDKGLINNISLCQDRQAISELKDFLDSIKEVQWLLKPLMIGQEADKEEAFYTELLRIWEELEP ITLLYNKVRNYVTKKPYTLEKVKLNFKYSTLLDGWDKNKEKDNLGIILLKDGQYYLGIMNRRNN KIADDAPLAKTDNVYRKMEYKLLTKVSANLPRIFLKDKNPSEEMLEKYEGTHLKGNCFCIDD CRELIDFFKKGIKQYEDWQDFKFSDESDDISAFYKEVEHQYKITFRDIDETYIDSLVNE GKLYLFQIYNKDFSPYSGTKNLHTLYWEMLFSSQNLQNIYVKLNGNAEIFYRKASINQKDVVV HKADLPKKNKDP QNSKKESMFDYDIKDKRFTCDKYQFHVPIITMFKALGENHFNKVRNLIHDAENMHIIIGDRG ERNLIYLCMIDMKGNIVKQISLNEIISYDKNKLEHKRNYHQLLKTREDNKSARQSWQTIHTIK ELKEGYLSQVIHVITDLMVEYNAIVVLEDLNFQKQGRQKFERQVYQKFEKMLIDKLNLYVDKS KGMDDEGGLHAYQLTDEFKSFQKLGKQSGFLYYIPAWNTSKLDPTTGFVNLFYTKYESVEKSK EFINNFTSILYNQEREYFEFLFDYSAFTSKAEGSRLKWTVCSEKGERVETRYNPKKNNEWDTKI DLTFELKKLFNDYSISLLDGLDREOMGKIDKADFYKKFMKLFALIVQMRNSDEREDKLI SPVLN KYGAFETGKNERMPLDADANGAYNIARKGLWIEIKNTDVEQLDKVKLTISNKEWLQYAQEH IL	47

Guide RNA (crRNA)

A guide RNA (gRNA) is an RNA that functions to guide an RNA- or DNA-targeting enzyme to a specific target. Targeting requires a gRNA complementary to the target site as well as a 5' protospacer adjacent motif (PAM) on the DNA strand opposite the target sequence. The gRNA for a Cas12a endonuclease is relatively short, in some embodiments, about 35-50 nucleotides long (e.g., 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides long). In some embodiments, a gRNA is about 40-44 nucleotides long. The portion of the gRNA that base pairs to the protospacer may be about 15-30 nucleotides long (e.g., 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides long). In some embodiments, the portion of the gRNA that base pairs to the protospacer is about 20-24 nucleotides long, e.g., about 21 nucleotides long. There is also a constant portion that binds to Cas12a, which is about 15-25 nucleotides long. In some embodiments, the constant portion that binds to Cas12a is about 20 nucleotides long).

For Cas12a endonucleases, the target sequence to which a gRNA binds should be next to a PAM sequence – e.g., TTTV, where V can represent A, C, or G. The “V” of the TTTV is typically immediately adjacent to the most 5' base of the non-targeted strand side of the protospacer element. The PAM sequence may vary, dependent on the variant Cas12a endonuclease.

20

II. Variant Cas12a Endonucleases

Provided herein, in some aspects, are engineered variant Cas12a endonucleases that have altered activity relative to a wild-type Cas12 endonuclease. An “variant Cas12a

endonuclease” herein refers to a non-naturally occurring endonuclease obtained by mutation of a wild-type (e.g., naturally-occurring) *Cas12a* gene, for example, a *Cas12a* gene from Table 1 (e.g., any one of SEQ ID NOs: 1-47). Variants of other wild-type *Cas12* genes are contemplated herein. Thus, the variant *Cas12a* endonucleases provided herein are

5 “engineered.”

Mutations contemplated herein, with respect to an amino acid sequence, include, without limitation, substitutions, additions, and deletions. An amino acid “substitution” is a change in a single amino acid relative to a reference amino acid sequence. For example, with references to the Lb*Cas12a* ND2006 amino acid sequence of SEQ ID NO: 1 of FIGs. 1A-1X,
10 a substitution at position E95 would include any amino acid, other than E, at position 95 (counting from the methionine (M) start codon) – e.g., E95R (R substituted for E) and E95Y (Y substituted for E).

The variant *Cas12a* endonucleases provided herein, in some aspects, exhibit hyperactivity or low indiscriminate single strand deoxyribonuclease (DNase) activity,
15 described in more detail elsewhere herein.

The activity (e.g., hyperactivity and/or indiscriminate single strand DNase activity) of a variant *Cas12a* endonuclease may be assessed using any method known in the art. In some embodiments, the activity of a variant *Cas12a* endonuclease is determined with a gel-based assay. In some embodiments, the activity of a variant *Cas12a* endonuclease is determined
20 using fluorophores and/or a fluorophore-quencher system. In some embodiments, the activity of a variant *Cas12a* endonuclease may be assessed using short, labelled oligonucleotides, which measure the activity of Cas9 and Cas Φ , respectively (see, e.g., Jinek et al., Science, 2012, 337(6096): 816-821 and Pausche et al., Science, 2020, 396(6501): 333-337). In some embodiments, fluorophore-labeled short oligonucleotides are used to assess cleavage on both
25 strands (see, e.g., Stella et al., Cell, 2018, 175: 1856-1871). In some embodiments, nickase activity is determined using optical tweezers (see, e.g., Paul et al., bioRxiv, 2021, doi.org/10.1101/2021.06.09.447528). In some embodiments, longer fluorophore-labeled oligonucleotides are used to assess *Cas12a* cleavage on both strands (see, e.g., Yamano et al., Cell, 2016, 165(4): 494-962; Cofsky et al., eLife, 2020, 9:e55143). In some embodiments,
30 quencher-fluorophore-labeled single-stranded DNA is used to assess ssDNase activity (see, e.g., Chen et al., Science, 2018, 360(6387):436-439). In some embodiments, variant *Cas12a* endonuclease activity is assessed using single molecule fluorescence resonance energy transfer (FRET) (see, e.g., Son et al., PNAS, 2021, 118(49): e2113747118). Other methods are contemplated herein.

In embodiments in which an amino acid substitution is exemplified (e.g., E95R), the present disclosure contemplates alternative substitutions having an “equivalent” charge, polarity, and or chemical class (defined by the amino acid side chain). **Table 2** provides the 20 naturally-occurring amino acids with a description of corresponding charge, polarity, and chemical class. For example, arginine has an equivalent charge to histidine and lysine; an equivalent polarity to asparagine, glutamine, serine, threonine, tyrosine, aspartic acid, glutamic acid, arginine, histidine, and lysine; and an equivalent chemical class/side chain to histidine and lysine. Thus, using the E95R substitution as an example, E95H and E95K are examples of amino acid substitutions having an equivalent charge; E95N, E95Q, E95S, E95T, E95Y, E95D, E95E, E95H, and E95K are examples of amino acid substitutions having an equivalent polarity, and E95H and E95K are examples of amino acid substitutions having an equivalent chemical class. In some embodiments, a given amino acid substitution is equivalent in charge, polarity, and chemical class. Again, using the E95R substitution as an example, E95H and E95K are examples of amino acid substitutions having an equivalent charge (i.e., positive), an equivalent polarity (i.e., polar), and an equivalent chemical class (i.e., basic).

Table 2. Amino Acids

Amino acid	Abbreviation		Charge	Polarity	Chemical Class/Side Chain
Alanine	Ala	A	uncharged	nonpolar	aliphatic
Glycine	Gly	G	uncharged	nonpolar	aliphatic
Isoleucine	Ile	I	uncharged	nonpolar	aliphatic
Leucine	Leu	L	uncharged	nonpolar	aliphatic
Proline	Pro	P	uncharged	nonpolar	aliphatic
Valine	Val	V	uncharged	nonpolar	aliphatic
Phenylalanine	Phe	F	uncharged	nonpolar	aromatic
Tryptophan	Trp	W	uncharged	nonpolar	aromatic
Cysteine	Cys	C	uncharged	nonpolar	sulfur
Methionine	Met	M	uncharged	nonpolar	sulfur
Asparagine	Asn	N	uncharged	polar	amide
Glutamine	Gln	Q	uncharged	polar	acidic
Serine	Ser	S	uncharged	polar	hydroxyl
Threonine	Thr	T	uncharged	polar	hydroxyl
Tyrosine	Tyr	Y	uncharged	polar	aromatic
Aspartic acid	Asp	D	negative	polar	acidic
Glutamic acid	Glu	E	negative	polar	amide
Arginine	Arg	R	positive	polar	basic

Amino acid	Abbreviation		Charge	Polarity	Chemical Class/Side Chain
Histidine	His	H	positive	polar	basic
Lysine	Lys	K	positive	polar	basic

In some embodiments, a variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position E95, E125, N256, R747, H759, N813, K932, N933, S934, V936, S982, or K984 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1), optionally wherein the variant Cas12a endonuclease exhibits hyperactivity. In other embodiments, a variant Cas12a endonuclease comprises a polypeptide sequence that comprise a mutation at an amino acid position corresponding to position N256, I831, K932, N933, S934, V936, Q944, S982, F983, K984, M986, or T988 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1), optionally wherein the variant Cas12a endonuclease exhibits hypoactivity. In yet other embodiments, a variant Cas12a endonuclease comprises a polypeptide sequence that comprise a mutation at an amino acid position corresponding to position N813, I831, K932, N933, S934, V936, Q944, S982, F983, K984, M986, or T988 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1), optionally wherein the variant Cas12a endonuclease exhibits low (or no) indiscriminate ssDNase activity. It should be understood that a variant Cas12a endonuclease comprising a mutation at an amino acid position corresponding to a specific position with reference to amino acid position numbering of LbCas12a ND2006 encompasses variants of LbCas12a ND2006 (e.g., SEQ ID NO: 1), as well as variants of Cas12a orthologs of LbCas12a ND2006, including without limitation, variants of any one of the Cas12a endonucleases in **Table 1** (e.g., SEQ ID NOs: 2-47). Identification of such “corresponding” amino acid positions can be readily performed by aligning any Cas12a endonuclease amino acid sequence to those examples provided herein, and in particular, with an LbCas12a ND2006 sequence, such as the amino acid sequence of SEQ ID NO: 1, as shown in FIGs. 1A-1X.

FIG. 1A, for example, shows an alignment of various Cas12a homologs, highlighting that a variant Cas12a endonuclease comprising a mutation at an amino acid position corresponding to E95 with reference to amino acid position numbering of LbCas12a ND2006 includes: variant Cas12a endonucleases comprising a mutation at position I96 with reference to amino acid position numbering of AsCas12a BV3L6, and variant Cas12a endonucleases

comprising a mutation at position K99 with reference to amino acid position numbering of FnCas12a.

The variant Cas12a endonucleases of the present disclosure may share a certain percent identity relative to a wild-type Cas12a endonuclease. For example, a variant Cas12a endonuclease may comprise an amino acid sequence that includes any one or more mutation(s) (e.g., amino acid substitution(s)) described herein and has at least 80% (e.g., at least 80%, at least 85%, at least 90%, at least 95%, or at least 98%) identity to the amino acid sequence of any one of the Cas12a endonucleases in Table 1 (e.g., SEQ ID NOs: 1-47), an ortholog thereof, or other wild-type Cas12a protein sequence.

In some embodiments, a variant Cas12a endonuclease comprises a polypeptide sequence that comprise a mutation at an amino acid position corresponding to position E95, E125, N256, R747, H759, N813, K932, N933, S934, V936, S982, or K984 with reference to amino acid position numbering of LbCas12a and has at least 80% (e.g., at least 80%, at least 85%, at least 90%, at least 95%, or at least 98%) identity to the amino acid sequence of any one of the Cas12a endonucleases in Table 1 (e.g., SEQ ID NOs: 1-47), an ortholog thereof, or other wild-type Cas12a protein sequence. In some embodiments, any one or more of the foregoing variant Cas12a endonucleases exhibits hyperactivity.

In some embodiments, a variant Cas12a endonuclease comprises a polypeptide sequence that comprise a mutation at an amino acid position corresponding to position E95R, E95Y, E125A, E125W, N256A, R747Y, H759V, H759D, N813R, N813H, K932L, N933E, N933V, S934Q, V936E, V936M, V936K, S982N, or K984R with reference to amino acid position numbering of LbCas12a and has at least 80% (e.g., at least 80%, at least 85%, at least 90%, at least 95%, or at least 98%) identity to the amino acid sequence of any one of the Cas12a endonucleases in Table 1 (e.g., SEQ ID NOs: 1-47), an ortholog thereof, or other wild-type Cas12a protein sequence. In some embodiments, any one or more of the foregoing variant Cas12a endonucleases exhibits hyperactivity.

In some embodiments, a variant Cas12a endonuclease comprises a polypeptide sequence that comprise a mutation at an amino acid position corresponding to position N256, I831, K932, N933, S934, V936, Q944, S982, F983, K984, M986, or T988 with reference to amino acid position numbering of LbCas12a and has at least 80% (e.g., at least 80%, at least 85%, at least 90%, at least 95%, or at least 98%) identity to the amino acid sequence of any one of the Cas12a endonucleases in Table 1 (e.g., SEQ ID NOs: 1-47), an ortholog thereof, or other wild-type Cas12a protein sequence. In some embodiments, any one or more of the foregoing variant Cas12a endonucleases exhibits hypoactivity.

In some embodiments, a variant Cas12a endonuclease comprises a polypeptide sequence that comprise a mutation at an amino acid position corresponding to position N256K, I831A, I831Y, K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, K932Y, N933L, S934W, V936G, Q944D, Q944E, Q944K, Q944M, S982T, S982W, F983G, F983L, K984F, M986G, M986L, M986S, or T988F with reference to amino acid position numbering of LbCas12a and has at least 80% (e.g., at least 80%, at least 85%, at least 90%, at least 95%, or at least 98%) identity to the amino acid sequence of any one of the Cas12a endonucleases in Table 1 (e.g., SEQ ID NOs: 1-47), an ortholog thereof, or other wild-type Cas12a protein sequence. In some embodiments, any one or more of the foregoing variant Cas12a endonucleases exhibits hypoactivity.

In some embodiments, a variant Cas12a endonuclease comprises a polypeptide sequence that comprise mutations at amino acid positions corresponding to positions K932F and F983L; K932F and T988F; K932R and Q944D; K932R and F983L; K932R and T988F; K932Y and F983L; K932Y and T988F; N933L and Q944M; V936G and Q944D; V936G and S982W; V936G and M986G; V936G and T988F; Q944D and S982W; Q944D and F983L; Q944D and T988F; S982W and F983L; S982W and T988F; or F983G and M986G with reference to amino acid position numbering of LbCas12a and has at least 80% (e.g., at least 80%, at least 85%, at least 90%, at least 95%, or at least 98%) identity to the amino acid sequence of any one of the Cas12a endonucleases in Table 1 (e.g., SEQ ID NOs: 1-47), an ortholog thereof, or other wild-type Cas12a protein sequence. In some embodiments, any one or more of the foregoing variant Cas12a endonucleases exhibits hypoactivity.

In some embodiments, a variant Cas12a endonuclease comprises a polypeptide sequence that comprise a mutation at an amino acid position corresponding to position N813, I831, K932, N933, S934, V936, Q944, S982, F983, K984, M986, or T988 with reference to amino acid position numbering of LbCas12a and has at least 80% (e.g., at least 80%, at least 85%, at least 90%, at least 95%, or at least 98%) identity to the amino acid sequence of any one of the Cas12a endonucleases in Table 1 (e.g., SEQ ID NOs: 1-47), an ortholog thereof, or other wild-type Cas12a protein sequence. In some embodiments, any one or more of the foregoing variant Cas12a endonucleases exhibits low (or no) ssDNase activity.

In some embodiments, a variant Cas12a endonuclease comprises a polypeptide sequence that comprise a mutation at an amino acid position corresponding to position N813H, N813R, N813W, I831A, I831Y, K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, K932Y, N933E, N933L, S934K, S934Q, V936E, V936G, Q944D, Q944E, Q944K, S982W, F983G, F983L, K984F, M986F, M986G, or T988F with

reference to amino acid position numbering of LbCas12a and has at least 80% (e.g., at least 80%, at least 85%, at least 90%, at least 95%, or at least 98%) identity to the amino acid sequence of any one of the Cas12a endonucleases in Table 1 (e.g., SEQ ID NOs: 1-47), an ortholog thereof, or other wild-type Cas12a protein sequence. In some embodiments, any one or more of the foregoing variant Cas12a endonucleases exhibits low (or no) ssDNase activity.

In some embodiments, a variant Cas12a endonuclease comprises a polypeptide sequence that comprise mutations at amino acid positions corresponding to positions: N933L and Q944M; or F983G and M986G with reference to amino acid position numbering of LbCas12a and has at least 80% (e.g., at least 80%, at least 85%, at least 90%, at least 95%, or at least 98%) identity to the amino acid sequence of any one of the Cas12a endonucleases in Table 1 (e.g., SEQ ID NOs: 1-47), an ortholog thereof, or other wild-type Cas12a protein sequence. In some embodiments, any one or more of the foregoing variant Cas12a endonucleases exhibits low (or no) ssDNase activity.

“Identity” refers to a relationship between two or among three or more sequences (e.g., amino acid sequences or nucleotide sequences) as determined by comparing the sequences to each other. Identity also refers to the degree of sequence relatedness between or among sequences as determined by the number of matches between or among strings of amino acids or strings of nucleotides. Identity is a measure of the percent of identical matches between the smaller of two or more sequences with gap alignments (if any) addressed by a particular mathematical model or computer program (e.g., “algorithms”). Identity of related polypeptides and polynucleotides can be readily calculated by known methods. “Percent (%) identity” as it applies to proteins or genes, for example, such as the Cas12a endonucleases described herein, is defined as the percentage of residues (amino acid or nucleic acid residues) in a first protein or gene sequence that are identical with the residues in a second protein or gene sequence after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent identity.

Methods and computer programs for the alignment are well known in the art. It is understood that identity depends on a calculation of percent identity but may differ in value due to gaps and penalties introduced in the calculation. Generally, variants of a particular protein or gene have at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% but less than 100% sequence identity to that particular wild-type, native, or reference sequence as determined by sequence alignment programs and parameters described herein and known to those skilled in the art. Such tools for alignment include but are not limited to those of the BLAST suite (Altschul, S.F., et al. Nucleic Acids Res. 1997;25:3389-3402); and those based

on the Smith-Waterman algorithm (Smith, T.F. & Waterman, M.S. J. Mol. Biol. 1981;147:195-197). A general global alignment technique based on dynamic programming is the Needleman-Wunsch algorithm (Needleman, S.B. & Wunsch, C.D. J. Mol. Biol. 1920;48:443-453). A Fast Optimal Global Sequence Alignment Algorithm (FOGSAA) also
 5 has been developed that purportedly produces global alignment of nucleotide and amino acid sequences faster than other optimal global alignment methods, including the Needleman-Wunsch algorithm.

An alignment of the non-limiting examples of wild-type Cas12a endonuclease sequences is provided in **FIGs. 1A-1X**.

10 A Cas12 “homolog” refers to a Cas12a endonuclease that has at least some sequence identity (e.g., at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% identity) to a wild-type reference Cas12a endonuclease and exhibits at least one activity exhibited by the wild-type reference Cas12a endonuclease (e.g., cleavage of a double strand or single strand polynucleotide, binding to a crRNA, etc.). For example, the
 15 wild-type Cas12a endonuclease exhibits indiscriminate ssDNase activity, cuts ~14 bp away from the PAM, and possesses RNase activity to self-process pre-crRNA. Further, its cleavage activity results in 5’ staggered overhangs, and its PAM site is 3’ to the targeting binding site. By contrast, a wild-type Cas9 endonuclease does not exhibit indiscriminate ssDNase activity, cuts ~3-4 bp away from the PAM, and does not possess RNase activity to self-process pre-
 20 crRNA (it requires accessory proteins to mediate pre-crRNA processing). Further, Cas9 cleavage activity results in blunt ends, and its PAM site is 5’ to the targeting binding site.

A Cas12a “ortholog” refers to Cas12a genes (and proteins encoded by the genes) inferred to be descended from the same ancestral sequence separated by a speciation event: when a species diverges into two separate species, the copies of a single gene in the two
 25 resulting species are said to be orthologous. Orthologs, or orthologous genes, are genes in different species that originated by vertical descent from a single gene of the last common ancestor. Cas12a ortholog can be identified and characterized based on sequence similarities to the present Cas12a system, as has been described with Type II systems, for example. For example, orthologs of Cas12a include the Cas12a endonucleases of **Table 1**.

30

III. Cas12a Endonuclease Hyperactive Variants

Some aspects of the present disclosure relate to hyperactive variant Cas12a endonucleases, i.e., variant Cas12a endonucleases that exhibit hyperactivity. “Hyperactivity” herein refers to polynucleotide cleavage activity of a variant endonuclease that is at least 10%

greater than polynucleotide cleavage activity of the wild-type or other reference endonuclease. A hyperactive variant Cas12a endonuclease has a higher reaction speed or initiates a cleavage reaction faster than the corresponding wild-type Cas12a endonuclease. In some embodiments, a hyperactive variant Cas12a endonuclease exhibits cleavage activity that is at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, or at least 50% greater than polynucleotide cleavage activity of the wild-type or other reference endonuclease. *See, e.g., Zhang, L. et al. Nat Commun. 2021 Jun 23;12(1):3908.*

In some embodiments, a variant Cas12a endonuclease (a) comprises a mutation at an amino acid position corresponding to position E95, E125, N256, R747, H759, N813, K932, N933, S934, V936, S982, or K984 with reference to amino acid position numbering of LbCas12a and (b) exhibits hyperactivity, optionally wherein the variant Cas12a endonuclease has at least 85%, at least 90%, at least 95%, or at least 98% identity with a wild-type reference Cas12a endonuclease.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position E95 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is E95R or E95Y. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position E95 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position E95 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position E125 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is E125A or E125W. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position E125 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type

reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position E125 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position N256 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1).

10 In some embodiments, the mutation is N256A. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N256 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant

15 LbCas12a endonuclease comprises a mutation at position N256 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

20 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position R747 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is R747Y. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position R747 with reference to amino acid position

25 numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position R747 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no

30 more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position H759

with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is H759V or H759D. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position H759 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position H759 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position N813 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is N813R or N813H. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N813 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N813 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position K932 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is K932L. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K932 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K932 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no

more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide
5 sequence that comprises a mutation at an amino acid position corresponding to position N933 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is N933E or N933V. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N933 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at
10 least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N933 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference
15 LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position S934 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1).
20 In some embodiments, the mutation is S934Q. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position S934 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant
25 LbCas12a endonuclease comprises a mutation at position S934 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

30 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position V936 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is V936E, V936M, or V936K. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position V936 with reference to

amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position V936 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position S982 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is S982N. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position S982 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position S982 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position K984 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is K984R. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K984 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K984 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

Additional Engineered Variant Cas12a Endonucleases With Hyperactivity

In some embodiments, a variant LbCas12a endonuclease comprises an E95 (e.g., E95R or E95Y) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an E95 (e.g., E95R or E95Y) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an E95 (e.g., E95R or E95Y) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an E95 (e.g., E95R or E95Y) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an E95 (e.g., E95R or E95Y) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an E125 (e.g., E125A or E125W) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an E125 (e.g., E125A or E125W) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an E125 (e.g., E125A or E125W) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an E125 (e.g., E125A or E125W) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an E125 (e.g., E125A or E125W) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an N256 (e.g., N256A) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N256 (e.g., N256A) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N256 (e.g., N256A) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N256 (e.g., N256A) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N256 (e.g., N256A)

substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an R747 (e.g., R747Y) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO:

- 5 1. In some embodiments, a variant LbCas12a endonuclease comprises an R747 (e.g., R747Y) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an R747 (e.g., R747Y) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an R747 (e.g., R747Y) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an R747 (e.g., R747Y) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

- 15 In some embodiments, a variant LbCas12a endonuclease comprises an H759 (e.g., H759V or H759D) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an H759 (e.g., H759V or H759D) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an H759 (e.g., H759V or H759D) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an H759 (e.g., H759V or H759D) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an H759 (e.g., H759V or H759D) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

- 25 In some embodiments, a variant LbCas12a endonuclease comprises an N813 (e.g., N813R or N813H) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N813 (e.g., N813R or N813H) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N813 (e.g., N813R or N813H) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N813 (e.g., N813R or N813H) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a

endonuclease comprises an N813 (e.g., N813R or N813H) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g.,
5 K932L) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932L) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932L) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In
10 some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932L) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932L) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g.,
15 N933E or N933V) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g., N933E or N933V) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an
20 N933 (e.g., N933E or N933V) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g., N933E or N933V) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g., N933E or N933V) substitution and has at least 98%
25 identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an S934Q (e.g.,
S934Q) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S934Q (e.g.,
30 S934Q) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S934Q (e.g., S934Q) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S934Q (e.g., S934Q) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO:

1. In some embodiments, a variant LbCas12a endonuclease comprises an S934Q (e.g., S934Q) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936E, V936M, or V936K) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936E, V936M, or V936K) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936E, V936M, or V936K) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936E, V936M, or V936K) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936E, V936M, or V936K) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982N) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982N) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982N) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982N) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982N) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984R) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984R) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984R) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984R)

substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984R) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hyperactivity.

5

IV. Cas12a Endonuclease Hypoactive Variants

Other aspects of the present disclosure provide hypoactive variant Cas12a endonucleases, i.e., variant Cas12a endonucleases that exhibit hypoactivity. "Hypoactivity" herein refers to polynucleotide cleavage activity of a variant endonuclease that is at least 10% lower than polynucleotide cleavage activity of the wild-type or other reference endonuclease. In some embodiments, a hypoactive variant Cas12a endonucleases exhibits cleavage activity that is at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, or at least 50% lower than polynucleotide cleavage activity of the wild-type or other reference endonuclease.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position N256 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is N256K. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N256 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N256 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position I831 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is I831A or I831Y. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position I831 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type

reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position I831 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position K932 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K932 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K932 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position N933 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is N933L. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N933 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N933 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position S934 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is S934W. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position S934 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position S934 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position V936 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is V936G. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position V936 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position V936 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position Q944 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is Q944D, Q944E, Q944K, or Q944M. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In

some embodiments, a variant LbCas12a endonuclease comprises a mutation at position Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position S982 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is S982T or S982W. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position S982 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position S982 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position F983 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is F983G or F983L. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position K984 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1).

In some embodiments, the mutation is K984F. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K984 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference

- 5 LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K984 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant
- 10 Cas12a endonucleases may exhibit hypoactivity.

- In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position M986 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is M986G, M986L, or M986S. In some
- 15 embodiments, a variant LbCas12a endonuclease comprises a mutation at position M986 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position M986
- 20 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

- In some embodiments, the variant Cas12a endonuclease comprises a polypeptide
- 25 sequence that comprises a mutation at an amino acid position corresponding to position T988 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is T988F. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at
- 30 least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a

ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions K932 and F983 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are K932F and F983L. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932 and F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932 and F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions K932 and T988 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are K932F and T988F. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932 and T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932 and T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions K932 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are K932R and Q944D. In some

embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

- 5 In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit
- 10 hypoactivity.

- In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions K932 and F983 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are K932R and F983L. In some
- 15 embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932 and F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions
- 20 K932 and F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

- 25 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions K932 and T988 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are K932R and T988F. In some
- embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932 and
- 30 T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932 and T988 with reference to amino acid position numbering of LbCas12a ND2006

endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

5 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions K932 and F983 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are K932Y and F983L. In some
10 embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932 and F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions
15 K932 and F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide
20 sequence that comprises mutations at amino acid positions corresponding to positions K932 and T988 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are K932Y and T988F. In some
 embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932 and T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease
25 (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932 and T988 with reference to amino acid position numbering of LbCas12a ND2006
30 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions N933

and Q944 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are N933L and Q944M. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions N933 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions N933 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions V936 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are V936G and Q944D. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions V936 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions V936 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions V936 and S982 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are V936G and S982W. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions V936 and S982 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions V936 and S982 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions V936 and M986 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are V936G and M986G. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions V936 and M986 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions V936 and M986 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions V936 and T988 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are V936G and T988F. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions V936 and T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions V936 and T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions Q944 and S982 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are Q944D and S982W. In some
5 embodiments, a variant LbCas12a endonuclease comprises mutations at positions Q944 and S982 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions
10 Q944 and S982 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

15 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions Q944 and F983 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are Q944D and F983L. In some
20 embodiments, a variant LbCas12a endonuclease comprises mutations at positions Q944 and F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions
25 Q944 and F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

30 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions Q944 and T988 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are Q944D and T988F. In some
embodiments, a variant LbCas12a endonuclease comprises mutations at positions Q944 and T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease

(e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions Q944 and T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions S982 and F983 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are S982W and F983L. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions S982 and F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions S982 and F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions S982 and T988 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are S982W and T988F. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions S982 and T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions S982 and T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID

NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions F983 and M986 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are F983G and M986G. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions F983 and M986 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions F983 and M986 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

Additional Engineered Variant Cas12a Endonucleases With Hypoactivity

In some embodiments, a variant LbCas12a endonuclease comprises an N256 (e.g., N256K) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N256 (e.g., N256K) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N256 (e.g., N256K) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N256 (e.g., N256K) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N256 (e.g., N256K) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an I831 (e.g., I831A or I831Y) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an I831 (e.g., I831A or I831Y) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an I831 (e.g., I831A or I831Y) substitution and has at least 90% identity to the amino acid

sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an I831 (e.g., I831A or I831Y) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an I831 (e.g., I831A or I831Y) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g., N933L) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g., N933L) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g., N933L) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g., N933L) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g., N933L) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an S934 (e.g., S934W) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S934 (e.g., S934W) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S934 (e.g., S934W) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S934 (e.g., S934W) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S934 (e.g., S934W) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936G) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936G) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936G) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936G) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936G) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an Q944 (e.g., Q944D, Q944E, Q944K, or Q944M) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944 (e.g., Q944D, Q944E, Q944K, or Q944M) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944 (e.g., Q944D, Q944E, Q944K, or Q944M) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944 (e.g., Q944D, Q944E, Q944K, or Q944M) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944 (e.g., Q944D, Q944E, Q944K, or Q944M) substitution and has at least

98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982T or S982W) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982T or S982W) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982T or S982W) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982T or S982W) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982T or S982W) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an F983 (e.g., F983G or F983L) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983 (e.g., F983G or F983L) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983 (e.g., F983G or F983L) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983 (e.g., F983G or F983L) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983 (e.g., F983G or F983L) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984F) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984F) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984F) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984F) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In

some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984F) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an M986 (e.g.,
5 M986G, M986L, or M986S) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an M986 (e.g., M986G, M986L, or M986S) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an M986 (e.g., M986G, M986L, or M986S) substitution
10 and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an M986 (e.g., M986G, M986L, or M986S) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an M986 (e.g., M986G, M986L, or M986S) substitution and has at least 98% identity to the amino acid
15 sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an T988 (e.g., T988F) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an T988 (e.g., T988F)
20 substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an T988 (e.g., T988F) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an T988 (e.g., T988F) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In
25 some embodiments, a variant LbCas12a endonuclease comprises an T988 (e.g., T988F) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an K932F and F983L substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO:
30 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932F and F983L substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932F and F983L substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932F and F983L

substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932F and F983L substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

- 5 In some embodiments, a variant LbCas12a endonuclease comprises an K932F and T988F substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932F and T988F substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932F and T988F substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932F and T988F substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932F and T988F substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

- 15 In some embodiments, a variant LbCas12a endonuclease comprises an K932R and Q944D substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932R and Q944D substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932R and Q944D substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932R and Q944D substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932R and Q944D substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

- 25 In some embodiments, a variant LbCas12a endonuclease comprises an K932R and F983L substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932R and F983L substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932R and F983L substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932R and F983L substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In

some embodiments, a variant LbCas12a endonuclease comprises an K932R and F983L substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an K932R and
5 T988F substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932R and T988F substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932R and T988F substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In
10 some embodiments, a variant LbCas12a endonuclease comprises an K932R and T988F substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932R and T988F substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an K932Y and
15 F983L substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932Y and F983L substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932Y and F983L
20 substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932Y and F983L substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932Y and F983L substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any
25 one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an K932Y and
T988F substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932Y and T988F
30 substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932Y and T988F substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932Y and T988F substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932Y and T988F

substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an N933L and Q944M substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO:

- 5 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933L and Q944M substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933L and Q944M substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933L and Q944M substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933L and Q944M substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an V936G and

- 15 Q944D substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and Q944D substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and Q944D substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and Q944D substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and Q944D substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

- 25 In some embodiments, a variant LbCas12a endonuclease comprises an V936G and S982W substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and S982W substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and S982W substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and S982W substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and S982W
- 30

substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an V936G and M986G substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO:

- 5 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and M986G substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and M986G substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and M986G substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and M986G substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an V936G and

- 15 T988F substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and T988F substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and T988F substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and T988F substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936G and T988F substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

25 In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and S982W substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO:

1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and S982W substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and S982W substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and S982W substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and S982W

substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and F983L substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO:

- 5 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and F983L substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and F983L substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and F983L substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and F983L substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and

- 15 T988F substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and T988F substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and T988F substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and T988F substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944D and T988F substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

- 25 In some embodiments, a variant LbCas12a endonuclease comprises an S982W and F983L substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982W and F983L substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982W and F983L substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982W and F983L substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982W and F983L
- 30

substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an S982W and T988F substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO:

- 5 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982W and T988F substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982W and T988F substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982W and T988F
- 10 substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982W and T988F substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

In some embodiments, a variant LbCas12a endonuclease comprises an F983G and

- 15 M986G substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983G and M986G substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983G and M986G substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In
- 20 some embodiments, a variant LbCas12a endonuclease comprises an F983G and M986G substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983G and M986G substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit hypoactivity.

25

V. Cas12a Endonuclease Low Indiscriminate ssDNase Variants

In addition to high-specific double strand DNA (dsDNA) cleavage, Cas12a has also been shown to exhibit indiscriminate single strand DNA (ssDNA) degradation activity upon activation with a ssDNA complementary to the crRNA guide as well as with dsDNA

- 30 complementary to the crRNA guide. This activity is displayed by all Cas12a orthologs and degrades any available ssDNA molecule into single/double nucleotides. Comparisons of the structures of Cas12a before, during and after cleavage reveal the structural changes that result in such an indiscriminate activity. The lid region, which is involved in the checkpoints for accurate target recognition is responsible for this action. Before the crRNA-DNA hybrid is

formed, the lid occludes the cleft where the catalytic residues reside. Upon formation of the hybrid, the lid changes conformation to form an α helix, thus interacting with the crRNA of the hybrid assembly, thus dissociating the polar interactions and making available the catalytic pocket. In the R-loop structure after cleavage, this region appears disordered
5 indicating that the catalytic site is accessible after the distal part of the dsDNA substrate dissociates from the complex. Therefore, the catalytic cleft is open and able to sever ssDNA indiscriminately. This molecular mechanism would explain how ssDNA molecules are degraded by Cas12a after being activated by the presence of the RNA-DNA hybrid. In addition, recent studies have reported non-specific nicking of target sequences bearing
10 mismatches in distal regions of the target DNA, suggesting that this could be a problem for potential applications. See Paul, B. & Montoya, G. *et al.*

Several of the variant Cas12a endonucleases provided herein, surprisingly, exhibit low to no indiscriminate ssDNA degradation activity, also referred to as indiscriminate single strand deoxyribonuclease (ssDNase) activity. This activity was unexpected in part because
15 the mutations made in the parent wild-type enzyme are outside of the lid region – the region thought to be responsible for indiscriminate ssDNase activity. “Low ssDNase activity” herein refers to indiscriminate ssDNA degradation activity of a variant endonuclease that is at least 10% lower than indiscriminate ssDNA degradation activity of the wild-type or other reference endonuclease. In some embodiments, a variant Cas12a endonucleases exhibits
20 indiscriminate ssDNase activity that is at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% lower than indiscriminate ssDNA degradation activity of the wild-type or other reference endonuclease. In some embodiments, a variant Cas12a endonucleases exhibits no (no measurable)
25 indiscriminate ssDNase activity.

In some embodiments, a variant Cas12a endonuclease (a) comprises a mutation at an amino acid position corresponding to position N813, I831, K932, N933, S934, V936, Q944, S982, F983, K984, M986, or T988 with reference to amino acid position numbering of LbCas12a and (b) exhibits low (or no) single indiscriminate ssDNase, optionally wherein the
30 variant Cas12a endonuclease has at least 85%, at least 90%, at least 95%, or at least 98% identity with a wild-type reference Cas12a endonuclease.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position N813 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1).

In some embodiments, the mutation is N813H, N813R, or N813W. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N813 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N813 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position I831 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is I831A or I831Y. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position I831 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position I831 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position K932 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K932 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K932 with reference to amino acid

position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

5 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position N933 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is N933E or N933L. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N933 with reference to amino acid
10 position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position N933 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1)
15 and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position S934
20 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is S934K or S934Q. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position S934 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type
25 reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position S934 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing
30 variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position V936 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is V936E or V936G. In some embodiments, a variant

LbCas12a endonuclease comprises a mutation at position V936 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position V936 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position Q944 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is Q944D, Q944E, or Q944K. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position S982 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is S982W. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position S982 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position S982 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a

ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position F983 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is F983G or F983L. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position F983 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position K984 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is K984F. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K984 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K984 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position M986 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is M986F or M986G. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position M986 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1)

and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position M986 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position T988 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is T988F. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position T988 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions N933 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are N933L and Q944M. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions N933 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions N933 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID

NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions F983 and M986 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are F983G and M986G. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions F983 and M986 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions F983 and M986 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) indiscriminate ssDNase activity.

Additional Engineered Variant Cas12a Endonucleases With Low ssDNase Activity

In some embodiments, a variant LbCas12a endonuclease comprises an N813 (e.g., N813H, N813R, or N813W) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N813 (e.g., N813H, N813R, or N813W) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N813 (e.g., N813H, N813R, or N813W) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N813 (e.g., N813H, N813R, or N813W) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N813 (e.g., N813H, N813R, or N813W) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

In some embodiments, a variant LbCas12a endonuclease comprises an I831 (e.g., I831A or I831Y) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an I831

(e.g., I831A or I831Y) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an I831 (e.g., I831A or I831Y) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease
5 comprises an I831 (e.g., I831A or I831Y) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an I831 (e.g., I831A or I831Y) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

10 In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y)
15 substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y)
20 substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K932 (e.g., K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any
25 one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g., N933E or N933L) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933
30 (e.g., N933E or N933L) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g., N933E or N933L) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g., N933E or N933L) substitution and has at least 95% identity to the

amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933 (e.g., N933E or N933L) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

- 5 In some embodiments, a variant LbCas12a endonuclease comprises an S934 (e.g., S934K or S934Q) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S934 (e.g., S934K or S934Q) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an
- 10 S934 (e.g., S934K or S934Q) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S934 (e.g., S934K or S934Q) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S934 (e.g., S934K or S934Q) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

- In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936E or V936G) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936
- 20 (e.g., V936E or V936G) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936E or V936G) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936E or V936G) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an V936 (e.g., V936E or V936G) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

- In some embodiments, a variant LbCas12a endonuclease comprises an Q944 (e.g.,
- 30 Q944D, Q944E, or Q944K) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944 (e.g., Q944D, Q944E, or Q944K) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944 (e.g., Q944D, Q944E, or Q944K) substitution

and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944 (e.g., Q944D, Q944E, or Q944K) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an Q944 (e.g., Q944D, Q944E, or Q944K) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982W) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982W) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982W) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982W) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an S982 (e.g., S982W) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

In some embodiments, a variant LbCas12a endonuclease comprises an F983 (e.g., F983G or F983L) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983 (e.g., F983G or F983L) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983 (e.g., F983G or F983L) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983 (e.g., F983G or F983L) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983 (e.g., F983G or F983L) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984F) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984F)

substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984F) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984F) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an K984 (e.g., K984F) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

10 In some embodiments, a variant LbCas12a endonuclease comprises an M986 (e.g., M986F or M986G) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an M986 (e.g., M986F or M986G) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an M986 (e.g., M986F or M986G) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an M986 (e.g., M986F or M986G) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an M986 (e.g., M986F or M986G) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

In some embodiments, a variant LbCas12a endonuclease comprises an T988 (e.g., T988F) substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an T988 (e.g., T988F) substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an T988 (e.g., T988F) substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an T988 (e.g., T988F) substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an T988 (e.g., T988F) substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

In some embodiments, a variant LbCas12a endonuclease comprises an N933L and Q944M substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933L and Q944M substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933L and Q944M substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933L and Q944M substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an N933L and Q944M substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

In some embodiments, a variant LbCas12a endonuclease comprises an F983G and M986G substitution and has at least 80% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983G and M986G substitution and has at least 85% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983G and M986G substitution and has at least 90% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983G and M986G substitution and has at least 95% identity to the amino acid sequence of SEQ ID NO: 1. In some embodiments, a variant LbCas12a endonuclease comprises an F983G and M986G substitution and has at least 98% identity to the amino acid sequence of SEQ ID NO: 1. Any one or more of the foregoing variant Cas12a endonucleases may exhibit low (or no) ssDNase activity.

25

Table 3. Variant Cas12a Endonucleases

Variant	Sequence	SEQ ID NO:
E95R	MSKLEKFTNC YSLSKTLRFE AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELRNLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDEF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNV VEQYSLNEII NNFNGIRIKT DYHSLDKKE KERFEARQNW TSINIKEIK AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNMYVDK	48

Variant	Sequence						SEQ ID NO:
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFIS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIQGFKK	
	AEDEKLDKVK	IAISNKWELE	YAQTSVKH				
E95Y	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	49
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELYNLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL	
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTSERGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKN	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQYADADLSV	VEKLEKIIQ	
	KVDEIYKVG	SSEKLFDAF	VLEKSLKKN	AVVAIMKDLL	DSVKSFEYNI	KAFFGEGKET	
	NRDESFGYDF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK	
	KWMAAYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKTTTTL	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNGIRIKT	DYHSLDDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVKH	ICELVEKYDA	VIALEDLNSG	FKNSRVKVEK	QVYQKFEKML	IDKLNMYVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFIS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIQGFKK	
	AEDEKLDKVK	IAISNKWELE	YAQTSVKH				
E125A	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	50
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIATILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL	
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTSERGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKN	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQYADADLSV	VEKLEKIIQ	
	KVDEIYKVG	SSEKLFDAF	VLEKSLKKN	AVVAIMKDLL	DSVKSFEYNI	KAFFGEGKET	
	NRDESFGYDF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK	
	KWMAAYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKTTTTL	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNGIRIKT	DYHSLDDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVKH	ICELVEKYDA	VIALEDLNSG	FKNSRVKVEK	QVYQKFEKML	IDKLNMYVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFIS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIQGFKK	
	AEDEKLDKVK	IAISNKWELE	YAQTSVKH				
E125W	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	51
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIWTILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL	
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTSERGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKN	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQYADADLSV	VEKLEKIIQ	
	KVDEIYKVG	SSEKLFDAF	VLEKSLKKN	AVVAIMKDLL	DSVKSFEYNI	KAFFGEGKET	
	NRDESFGYDF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK	
	KWMAAYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKTTTTL	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNGIRIKT	DYHSLDDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVKH	ICELVEKYDA	VIALEDLNSG	FKNSRVKVEK	QVYQKFEKML	IDKLNMYVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFIS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIQGFKK	
	AEDEKLDKVK	IAISNKWELE	YAQTSVKH				
N256A	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	52
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL	
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTSERGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKN	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQYADADLSV	VEKLEKIIQ	

Variant	Sequence						SEQ ID NO:
	KVDEIYKVYG	SSEKLFDA	VLEKSLKND	AVVAIMKDLL	DSVKSFENYI	KAFFGEGKET	
	NRDESFYGF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGPN	KMLPKVFFSK	
	KWMAYYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYKV	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTTLT	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVHK	ICELVEKYDA	VIALEDLNSG	FKNSRVKVEK	QVYQKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFISS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIGQFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
N256K	MSKLEKFTNC	YSLSKTLRFF	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	53
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL	
	TRYISNMDIF	EKVDALFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTESGE	KIKGLKEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKV	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQADADLSV	VEKLKEIIQ	
	KVDEIYKVYG	SSEKLFDA	VLEKSLKND	AVVAIMKDLL	DSVKSFENYI	KAFFGEGKET	
	NRDESFYGF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGPN	KMLPKVFFSK	
	KWMAYYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYKV	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTTLT	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVHK	ICELVEKYDA	VIALEDLNSG	FKNSRVKVEK	QVYQKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFISS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIGQFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
R747Y	MSKLEKFTNC	YSLSKTLRFF	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	54
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL	
	TRYISNMDIF	EKVDALFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTESGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKV	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQADADLSV	VEKLKEIIQ	
	KVDEIYKVYG	SSEKLFDA	VLEKSLKND	AVVAIMKDLL	DSVKSFENYI	KAFFGEGKET	
	NRDESFYGF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGPN	KMLPKVFFSK	
	KWMAYYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYKV	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTTLT	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVHK	ICELVEKYDA	VIALEDLNSG	FKNSRVKVEK	QVYQKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFISS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIGQFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
H759V	MSKLEKFTNC	YSLSKTLRFF	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	55
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL	
	TRYISNMDIF	EKVDALFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTESGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKV	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQADADLSV	VEKLKEIIQ	
	KVDEIYKVYG	SSEKLFDA	VLEKSLKND	AVVAIMKDLL	DSVKSFENYI	KAFFGEGKET	
	NRDESFYGF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGPN	KMLPKVFFSK	
	KWMAYYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYKV	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTTLT	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVHK	ICELVEKYDA	VIALEDLNSG	FKNSRVKVEK	QVYQKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFISS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	

Variant	Sequence	SEQ ID NO:
	ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	
H759D	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEKIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVDP ANSPIANKNP DNPKKTITLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFRSDADYI KKWLYSYGN RIRIFRNPCK NNVFDWEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	56
N813W	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEKIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTITLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFRSDADYI KKWLYSYGN RIRIFRNPCK NNVFDWEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	57
N813R	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEKIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTITLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFRSDADYI KKWLYSYGN RIRIFRNPCK NNVFDWEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	58
N813H	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEKIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK	59

Variant	Sequence	SEQ ID NO:
	KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYPPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	
1831A	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAYEDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKKEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNIEK INYKLLPGN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG ADR GERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYPPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	60
1831Y	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAYEDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKKEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNIEK INYKLLPGN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG YDR GERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYPPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	61
K932L	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAYEDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKKEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNIEK INYKLLPGN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYPPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	62
K932I	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS	63

Variant	Sequence	SEQ ID NO:
	FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTLT YDVVKDRKFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNV VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FINSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	
K932V	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTLT YDVVKDRKFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNV VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FINSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	64
K932M	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTLT YDVVKDRKFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNV VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FINSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	65
K932F	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTLT	66

Variant	Sequence	SEQ ID NO:
	YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG F ~~H~~ NSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYPVEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	
K932R	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESEGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIR KWNAYEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG F ~~H~~ NSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYPVEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	67
K932A	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESEGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIR KWNAYEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG F ~~H~~ NSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYPVEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	68
K932H	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESEGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIR KWNAYEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG F ~~H~~ NSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYPVEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	69
K932N	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI	70

Variant	Sequence	SEQ ID NO:
	IGGFVTESGE KIKGLNEYIN LYNQKTQKQL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	
K932Q	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTQKQL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	71
K932S	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTQKQL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	72
K932T	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTQKQL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FNSRVKVEK QVYQKFEKML IDKLNVMVDK	73

Variant	Sequence	SEQ ID NO:
	KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	
K932Y	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKIKEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYK SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSIENTIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FVNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	74
K932W	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKIKEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYK SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSIENTIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FVNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	75
N933E	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKIKEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYK SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSIENTIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FVNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	76
N933V	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKIKEIIQ	77

Variant	Sequence	SEQ ID NO:
	KVDEIYKVYG SSEKLFDAF VLEKSLKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNIGIRIKT DYHSLDDKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FK ~~V~~ SRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	
N933L	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIR KWNAYEDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNIGIRIKT DYHSLDDKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FK ~~L~~ SRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	78
S934Q	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIR KWNAYEDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNIGIRIKT DYHSLDDKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FK ~~N~~ SRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	79
S934K	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIR KWNAYEDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNIGIRIKT DYHSLDDKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FK ~~N~~ SRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS	80

Variant	Sequence	SEQ ID NO:
	ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	
S934W	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAQSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEKIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYK SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTITLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNRSRKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDFWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	81
V936E	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAQSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEKIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYK SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTITLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNRSRKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDFWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	82
V936M	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAQSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEKIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYK SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTITLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNRSRKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDFWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	83
V936K	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAQSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEKIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK	84

Variant	Sequence	SEQ ID NO:
	KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRKKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYPPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	
V936G	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAYDDIH LKKKAVVTEK YEDDRKSFK KIGSFSLEQL QEYADADLSV VEKKEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNIEK INYKLLPGN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRKKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYPPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	85
Q944D	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAYDDIH LKKKAVVTEK YEDDRKSFK KIGSFSLEQL QEYADADLSV VEKKEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNIEK INYKLLPGN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRKKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYPPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	86
Q944E	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAYDDIH LKKKAVVTEK YEDDRKSFK KIGSFSLEQL QEYADADLSV VEKKEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNIEK INYKLLPGN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRKKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYPPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	87
Q944K	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLLDRYYLS	88

Variant	Sequence	SEQ ID NO:
	FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTLT YDVVKDRKFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRXKVEK QVYKFEKML IDKLNYMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTADAYI KWKLYSYGN RIRIFRNPCK NNVDWEVVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	
Q944M	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTLT YDVVKDRKFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRXKVEK QVYKFEKML IDKLNYMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTADAYI KWKLYSYGN RIRIFRNPCK NNVDWEVVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	89
S982W	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTLT YDVVKDRKFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYKFEKML IDKLNYMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTADAYI KWKLYSYGN RIRIFRNPCK NNVDWEVVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	90
S982T	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTLT	91

Variant	Sequence	SEQ ID NO:
	YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ET FKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYPPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEDEV LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	
S982N	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESE KIKGLNEYIN LYNQKTQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAYEYDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVG SSEKLFADAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF EN FKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYPPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEDEV LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	92
F983L	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESE KIKGLNEYIN LYNQKTQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAYEYDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVG SSEKLFADAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ES LKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYPPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEDEV LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	93
F983G	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESE KIKGLNEYIN LYNQKTQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAYEYDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVG SSEKLFADAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ES GKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYPPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEDEV LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	94
K984F	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI	95

Variant	Sequence	SEQ ID NO:
	IGGFVTESGE KIKGLNEYIN LYNQKTQKQL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEII IQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNQVRVVEK QVYQKFEKML IDKLNMYMDK KSNPCATGGA LKGYQITNKF ESFESMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	
K984R	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTQKQL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEII IQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNQVRVVEK QVYQKFEKML IDKLNMYMDK KSNPCATGGA LKGYQITNKF ESFESMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	96
M986G	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTQKQL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEII IQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNQVRVVEK QVYQKFEKML IDKLNMYMDK KSNPCATGGA LKGYQITNKF ESFESMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	97
M986F	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTQKQL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEII IQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNQVRVVEK QVYQKFEKML IDKLNMYMDK	98

Variant	Sequence	SEQ ID NO:
	KSNPCATGGA LKGYQITNKF ESFKS F STQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	
M986L	MSKLEKFTNC YLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGFEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKIKEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYK SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSIENTIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNQVRVVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKS L STQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	99
M986S	MSKLEKFTNC YLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGFEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKIKEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYK SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSIENTIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNQVRVVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKS S STQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	100
T988F	MSKLEKFTNC YLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGFEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKIKEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYK SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSIENTIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNQVRVVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESF F SMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	101
K932F, F983L	MSKLEKFTNC YLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGFEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKIKEIIQ	102

Variant	Sequence								SEQ ID NO:
	KVDEIYKVYG	SSEKLFDA	VLEKSLKKN	AVVAIMKDL	DSVKSFENY	KAFFGEGKET			
	NRDESFYGF	VLAYDILLK	DHIYDAIRN	VTQKPYSKD	FKLYFQNPQ	MGGWDDKDK			
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK			
	KWMAYYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET			
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH			
	TMFYKLLFDE	NNHGQIRLSG	GAELEFMRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTTLT			
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY			
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK			
	AGYISQVVHK	ICELVEKYDA	VIALEDLNSG	F ~~FN~~ NSRVKVEK	QVYQKFEKML	IDKLNVMVDK			
	KSNPCATGGA	LKGYQITNKF	ES ~~L~~ KSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS			
	IADSKKFISS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK			
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS			
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIQGFKK			
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH						
K932F, T988F	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS		103	
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK			
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL			
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI			
	IGGFVTESGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV			
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVK	GPAISTISKD	IFGEWNVIRD			
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QYADADLSV	VEKLKEIIQ			
	KVDEIYKVYG	SSEKLFDA	VLEKSLKKN	AVVAIMKDL	DSVKSFENY	KAFFGEGKET			
	NRDESFYGF	VLAYDILLK	DHIYDAIRN	VTQKPYSKD	FKLYFQNPQ	MGGWDDKDK			
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK			
	KWMAYYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET			
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH			
	TMFYKLLFDE	NNHGQIRLSG	GAELEFMRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTTLT			
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY			
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK			
	AGYISQVVHK	ICELVEKYDA	VIALEDLNSG	F ~~FN~~ NSRVKVEK	QVYQKFEKML	IDKLNVMVDK			
	KSNPCATGGA	LKGYQITNKF	ESF ~~K~~ SMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS			
	IADSKKFISS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK			
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS			
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIQGFKK			
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH						
K932R, Q944D	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS		104	
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK			
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL			
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI			
	IGGFVTESGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV			
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVK	GPAISTISKD	IFGEWNVIRD			
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QYADADLSV	VEKLKEIIQ			
	KVDEIYKVYG	SSEKLFDA	VLEKSLKKN	AVVAIMKDL	DSVKSFENY	KAFFGEGKET			
	NRDESFYGF	VLAYDILLK	DHIYDAIRN	VTQKPYSKD	FKLYFQNPQ	MGGWDDKDK			
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK			
	KWMAYYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET			
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH			
	TMFYKLLFDE	NNHGQIRLSG	GAELEFMRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTTLT			
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY			
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK			
	AGYISQVVHK	ICELVEKYDA	VIALEDLNSG	F ~~FN~~ NSRVKVEK	QVYQKFEKML	IDKLNVMVDK			
	KSNPCATGGA	LKGYQITNKF	ESF ~~K~~ SMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS			
	IADSKKFISS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK			
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS			
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIQGFKK			
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH						
K932R, F983L	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS		105	
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK			
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL			
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI			
	IGGFVTESGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV			
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVK	GPAISTISKD	IFGEWNVIRD			
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QYADADLSV	VEKLKEIIQ			
	KVDEIYKVYG	SSEKLFDA	VLEKSLKKN	AVVAIMKDL	DSVKSFENY	KAFFGEGKET			
	NRDESFYGF	VLAYDILLK	DHIYDAIRN	VTQKPYSKD	FKLYFQNPQ	MGGWDDKDK			
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK			
	KWMAYYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET			
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH			
	TMFYKLLFDE	NNHGQIRLSG	GAELEFMRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTTLT			
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY			
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK			
	AGYISQVVHK	ICELVEKYDA	VIALEDLNSG	F ~~FN~~ NSRVKVEK	QVYQKFEKML	IDKLNVMVDK			
	KSNPCATGGA	LKGYQITNKF	ES ~~L~~ KSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS			
	IADSKKFISS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK			
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS			

Variant	Sequence						SEQ ID NO:
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAQ	ENAILPKNAD	ANGAYNIARK	VLWAIGQFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
K932R, T988F	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	106
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRCINENL	
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTSERGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKV	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQYADADLSV	VEKLEKIIQ	
	KVDEIYKVYG	SSEKLFDAF	VLEKSLKKN	AVVAIMKDLL	DSVKSFEYNI	KAFFGEGKET	
	NRDESFGYDF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK	
	KWMAYYNPSE	DIQKIYKNGT	FKKGD MFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNP KKT T T L S	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVKH	ICELVEKYDA	VIALEDLNSG	F ~~Y~~ NSRVKVEK	QVYQKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMS ~~F~~ Q	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFIS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAQ	ENAILPKNAD	ANGAYNIARK	VLWAIGQFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
K932Y, F983L	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	107
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRCINENL	
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTSERGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKV	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQYADADLSV	VEKLEKIIQ	
	KVDEIYKVYG	SSEKLFDAF	VLEKSLKKN	AVVAIMKDLL	DSVKSFEYNI	KAFFGEGKET	
	NRDESFGYDF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK	
	KWMAYYNPSE	DIQKIYKNGT	FKKGD MFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNP KKT T T L S	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVKH	ICELVEKYDA	VIALEDLNSG	F ~~Y~~ NSRVKVEK	QVYQKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ES ~~L~~ KSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFIS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAQ	ENAILPKNAD	ANGAYNIARK	VLWAIGQFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
K932Y, T988F	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	108
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRCINENL	
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTSERGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKV	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQYADADLSV	VEKLEKIIQ	
	KVDEIYKVYG	SSEKLFDAF	VLEKSLKKN	AVVAIMKDLL	DSVKSFEYNI	KAFFGEGKET	
	NRDESFGYDF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK	
	KWMAYYNPSE	DIQKIYKNGT	FKKGD MFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNP KKT T T L S	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVKH	ICELVEKYDA	VIALEDLNSG	F ~~Y~~ NSRVKVEK	QVYQKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMS ~~F~~ Q	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFIS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAQ	ENAILPKNAD	ANGAYNIARK	VLWAIGQFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
N933L, Q944M	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	109
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRCINENL	
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTSERGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKV	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQYADADLSV	VEKLEKIIQ	
	KVDEIYKVYG	SSEKLFDAF	VLEKSLKKN	AVVAIMKDLL	DSVKSFEYNI	KAFFGEGKET	
	NRDESFGYDF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK	

Variant	Sequence	SEQ ID NO:
	KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKLSRVKVEK QVYMKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYPVEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	
V936G, Q944D	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESE KIKGLNEYIN LYNQKTQKQL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAYDDIH LKKKAVVTEK YEDDRKSFK KIGSFSLEQL QEYADADLSV VEKKEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNIEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRGKVEK QVYDKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYPVEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	110
V936G, S982W	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESE KIKGLNEYIN LYNQKTQKQL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAYDDIH LKKKAVVTEK YEDDRKSFK KIGSFSLEQL QEYADADLSV VEKKEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNIEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRGKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYPVEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	111
V936G, M986G	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK KDIIETILPE FLDDKDEIAL VNSFNFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESE KIKGLNEYIN LYNQKTQKQL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAYDDIH LKKKAVVTEK YEDDRKSFK KIGSFSLEQL QEYADADLSV VEKKEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNIEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRGKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSGSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYPVEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVEC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	112
V936G,	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLLDRYYLS	113

Variant	Sequence	SEQ ID NO:
T988F	FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTLT YDVVKDRRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRQKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	
Q944D, S982W	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTLT YDVVKDRRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRQKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	114
Q944D, F983L	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTLT YDVVKDRRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRQKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEVVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	115
Q944D, T988F	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTLT	116

Variant	Sequence	SEQ ID NO:
	YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRXKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYPVEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	
S982W, F983L	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDIAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIR KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRXKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF EWLKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYPVEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	117
S982W, F988F	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDIAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIR KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRXKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF EWLKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYPVEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	118
F983G, M986G	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDIAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIR KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGN I VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRXKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESCKSGSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYPVEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	119
K932G, N933G	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDIAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI	367

Variant	Sequence	SEQ ID NO:
	IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IIVVVDGKGN I VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FGGSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	
R833L (LbAA9)	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDLGERNLLY IIVVVDGKGN I VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	368
R833K (LbAA19)	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDKGERNLLY IIVVVDGKGN I VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	369
R833M (LbEF1s9)	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVYG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDMGERNLLY IIVVVDGKGN I VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNVMVDK	370

Variant	Sequence	SEQ ID NO:
	KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	
K932E (LbAA23)	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGFEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKIKEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYK SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSIENTIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	371
K932G (LbMS07)	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGFEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKIKEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYK SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSIENTIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	372
K940G (LbAA49)	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGFEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKIKEIIQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYK SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSIENTIKELK AGYISQVVKH ICELVEKYDA VIALEDLNSG FNSRVKVEK QVYQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	373
Q944K (LbAC10)	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGFEFFNFVLT QEGIDVYNAI IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKIKEIIQ	374

Variant	Sequence						SEQ ID NO:
	KVDEIYKVYG	SSEKLFDAF	VLEKSLKND	AVVAIMKDLL	DSVKSFENYI	KAFFGEGKET	
	NRDESFYGF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGPN	KMLPKVFFSK	
	KWMAYYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYKV	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTTLTS	
	YDVVKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVHK	ICELVEKYDA	VIALEDLNSG	FKNSRVKVEK	QVYKKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFISS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIQGFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
K932G, N933G, V936G, S929G (LbMS3n5)	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	375
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYSKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL	
	TRYISNMDIF	EKVDIAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTESGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKN	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QYADADLSV	VEKLKEIIQ	
	KVDEIYKVYG	SSEKLFDAF	VLEKSLKND	AVVAIMKDLL	DSVKSFENYI	KAFFGEGKET	
	NRDESFYGF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGPN	KMLPKVFFSK	
	KWMAYYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYKV	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTTLTS	
	YDVVKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVHK	ICELVEKYDA	VIALEDLNGG	FGGSRGKVEK	QVYQKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFISS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIQGFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
K940G, Q944K (LbTN37)	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	376
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYSKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL	
	TRYISNMDIF	EKVDIAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTESGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKN	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QYADADLSV	VEKLKEIIQ	
	KVDEIYKVYG	SSEKLFDAF	VLEKSLKND	AVVAIMKDLL	DSVKSFENYI	KAFFGEGKET	
	NRDESFYGF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGPN	KMLPKVFFSK	
	KWMAYYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYKV	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTTLTS	
	YDVVKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVHK	ICELVEKYDA	VIALEDLNSG	FKNSRVKVEK	QVYKKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFISS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAO	ENAILPKNAD	ANGAYNIARK	VLWAIQGFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
R836G, Q944K (LbTN39)	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	377
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYSKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRICINENL	
	TRYISNMDIF	EKVDIAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTESGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKN	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QYADADLSV	VEKLKEIIQ	
	KVDEIYKVYG	SSEKLFDAF	VLEKSLKND	AVVAIMKDLL	DSVKSFENYI	KAFFGEGKET	
	NRDESFYGF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGPN	KMLPKVFFSK	
	KWMAYYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYKV	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTTLTS	
	YDVVKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGERNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVVHK	ICELVEKYDA	VIALEDLNSG	FKNSRVKVEK	QVYKKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFISS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVFDWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	

Variant	Sequence						SEQ ID NO:
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAQ	ENAILPKNAD	ANGAYNIARK	VLWAIGQFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
R833M, E835D, Y943T (LbTN2)	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	378
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRCINENL	
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTSERGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKN	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQADADLSV	VEKLKEII IQ	
	KVDEIYKVYG	SSEKLFDAF	VLEKSLKKN	AVVAIMKDLL	DSVKSFEYNI	KAFFGEGKET	
	NRDESFGYDF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK	
	KWMAAYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTITLS	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDMGDRNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVYHK	ICELVEKYDA	VIALEDLNSG	FKNSRVKVEK	QVTQKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFIS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVDFWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAQ	ENAILPKNAD	ANGAYNIARK	VLWAIGQFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
R836G, Q944K, R935G (LbFM14)	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	379
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRCINENL	
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTSERGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKN	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQADADLSV	VEKLKEII IQ	
	KVDEIYKVYG	SSEKLFDAF	VLEKSLKKN	AVVAIMKDLL	DSVKSFEYNI	KAFFGEGKET	
	NRDESFGYDF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK	
	KWMAAYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTITLS	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDRGEGNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVYHK	ICELVEKYDA	VIALEDLNSG	FKNSGVKVEK	QVYKKEFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFIS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVDFWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAQ	ENAILPKNAD	ANGAYNIARK	VLWAIGQFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
R833M, E835D, Y943T, R935G (LbFM17)	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	380
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRCINENL	
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTSERGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKN	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQADADLSV	VEKLKEII IQ	
	KVDEIYKVYG	SSEKLFDAF	VLEKSLKKN	AVVAIMKDLL	DSVKSFEYNI	KAFFGEGKET	
	NRDESFGYDF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK	
	KWMAAYNPSE	DIQKIYKNGT	FKKGDMFNLN	DCHKLIDFFK	DSISRYPKWS	NAYDFNFSET	
	EKYKDIAGFY	REVVEEQGYK	SFESASKKEV	DKLVEEGKLY	MFQIYNKDFS	DKSHGTPNLH	
	TMFYKLLFDE	NNHGQIRLSG	GAELFMRRAS	LKKEELVVHP	ANSP IANKNP	DNPKKTITLS	
	YDVYKDKRFS	EDQYELHIPI	AINKCPKNIF	KINTEVRVLL	KHDDNPYVIG	IDMGDRNLLY	
	IVVVDGKGNI	VEQYSLNEII	NNFNIGIRIKT	DYHSLLDKKE	KERFEARQNW	TSIENIKELK	
	AGYISQVYHK	ICELVEKYDA	VIALEDLNSG	FKNSGVKVEK	QVTQKFEKML	IDKLNVMVDK	
	KSNPCATGGA	LKGYQITNKF	ESFKSMSTQN	GFIFYIPAWL	TSKIDPSTGF	VNLLKTKYTS	
	IADSKKFIS	FDRIMYVPEE	DLFEFALDYK	NFSRTDADYI	KKWKLYSYGN	RIRIFRNPCK	
	NNVDFWEEVC	LTSAYKELFN	KYGINYQQGD	IRALLCEQSD	KAFYSSFMAL	MSLMLQMRNS	
	ITGRTDVDFL	ISPVKNSDGI	FYDSRNYEAQ	ENAILPKNAD	ANGAYNIARK	VLWAIGQFKK	
	AEDEKLDKVK	IAISNKEWLE	YAQTSVKH				
R833M, E835D, Y943T, Q941K (LbFM28)	MSKLEKFTNC	YSLSKTLRFK	AIPVGKTQEN	IDNKRLLED	EKRAEDYKGV	KKLLDRYYLS	381
	FINDVLHSIK	LKNLNYYISL	FRKKTRTEKE	NKELENLEIN	LRKEIAKAFK	GNEGYKSLFK	
	KDIIETILPE	FLDDKDEIAL	VNSFNGFTTA	FTGFFDNREN	MFSEEAKSTS	IAFRCINENL	
	TRYISNMDIF	EKVDAIFDKH	EVQEIKEKIL	NSDYDVEDFF	EGEFFNFVLT	QEGIDVYNAI	
	IGGFVTSERGE	KIKGLNEYIN	LYNQKTKQKL	PKFKPLYKQV	LSDRESLSFY	GEGYTSDEEV	
	LEVFRNTLNK	NSEIFSSIKK	LEKLFKNFDE	YSSAGIFVKN	GPAISTISKD	IFGEWNVIRD	
	KWNAEYDDIH	LKKKAVVTEK	YEDDRRKSFK	KIGSFSLEQL	QEQADADLSV	VEKLKEII IQ	
	KVDEIYKVYG	SSEKLFDAF	VLEKSLKKN	AVVAIMKDLL	DSVKSFEYNI	KAFFGEGKET	
	NRDESFGYDF	VLAYDILLKV	DHIYDAIRNY	VTQKPYSKDK	FKLYFQNPQF	MGGWDKDKET	
	DYRATILRYG	SKYYLAIMDK	KYAKCLQKID	KDDVNGNYEK	INYLKLLPGN	KMLPKVFFSK	

Variant	Sequence	SEQ ID NO:
	KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTITLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDMGDRNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK KVTQKFEMML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	
R833M, E835D, E125A (LbFM44)	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIATILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAYEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNIEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTITLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDMGDRNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEMML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	382
Y943F, Q944K, K932G- N933G, E125A (LbFM51)	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIATILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAYEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNIEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTITLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FGGSRVKVEK QVYQKFEMML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	383
R836G, Q944K, R935G, E125A (LbFM64)	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLLDRYYLS FINDVLHSIK LKNLNYYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSKSLFK KDIATILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAYEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLEII IQ KVDEIYKVG SSEKLFDAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNIEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTITLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGEGNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLLDKKE KERFEARQNW TSINIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSGVKVEK QVYQKFEMML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFIS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KWKLYSYGN RIRIFRNPCK NNVDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKWELE YAQTSVKH	384
R833M,	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KKLLDRYYLS	385

Variant	Sequence	SEQ ID NO:
E835D, Y943T, R935G, E125A (LbFM65)	FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIATILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVG SSEKLFADAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDMGDRNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSGVKVEK QVTQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPCK NNVDFWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	
R833M, E835D, Y943T, Q941K, E125A (LbFM67)	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIATILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVG SSEKLFADAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDMGDRNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FKNSRVKVEK KVTQKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPCK NNVDFWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	386
D832A, Y943F, Q944K, K932G- N933G, E125A (LbFM76)	MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLED EKRAEDYKGV KLLDRYYLS FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYSLSFK KDIATILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL TRYISNMDIF EKVDALFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI IGGFVTSERGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKV GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRSFK KIGSFSLEQL QEYADADLSV VEKLEIIIQ KVDEIYKVG SSEKLFADAF VLEKSLKKN AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFGYDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNVEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKTTTSL YDVVKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI VEQYSLNEII NNFNIGIRIKT DYHSLDKKE KERFEARQNW TSINIENIKEL AGYISQVVKH ICELVEKYDA VIALEDLNSG FGGSRVKVEK QVFKFEKML IDKLNVMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPCK NNVDFWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIQGFKK AEDEKLDKVK IAINSKEWLE YAQTSVKH	387

VI. Polynucleotides

The present disclosure provides, in some aspects, polynucleotides encoding the variant Cas12a endonucleases. Nucleic acids comprise a polymer of nucleotides (nucleotide monomers). Thus, nucleic acids are also referred to as polynucleotides. Nucleic acids may be or may include, for example, deoxyribonucleic acid (DNA), ribonucleic acid (RNA), threose

nucleic acid (TNA), glycol nucleic acid (GNA), peptide nucleic acid (PNA), locked nucleic acid (LNA, including LNA having a β -D-ribo configuration, α -LNA having an α -L-ribo configuration (a diastereomer of LNA), 2'-amino-LNA having a 2'-amino functionalization, and 2'-amino- α -LNA having a 2'-amino functionalization), ethylene nucleic acid (ENA),
5 cyclohexenyl nucleic acid (CeNA) and/or chimeras and/or combinations thereof.

In some embodiments, the polynucleotide encoding the variant Cas12a endonuclease is an RNA, such as an mRNA. Messenger RNA (mRNA) is RNA that encodes a (at least one) protein (a naturally-occurring, non-naturally-occurring, or modified polymer of amino acids) and can be translated to produce the encoded protein *in vitro*, *in vivo*, *in situ*, or *ex vivo*. An
10 mRNA provided herein typically comprises an open reading frame (ORF) encoding a variant Cas12a endonuclease. In some embodiments, an mRNA also comprises an ORF encoding a crRNA or multiple crRNAs. In some embodiments, an mRNA further comprises a 5' cap, a 5' untranslated region (UTR), a 3' UTR, and a poly(A) tail.

An ORF is a continuous stretch of DNA or RNA beginning with a start codon (e.g.,
15 methionine (ATG or AUG)) and ending with a stop codon (e.g., TAA, TAG or TGA, or UAA, UAG or UGA). An ORF typically encodes a protein. It will be understood that the sequences disclosed herein may further comprise additional elements, e.g., 5' and/or 3' UTRs, but that those elements, unlike the ORF, need not necessarily be present in an RNA (e.g., mRNA) of the present disclosure.

In some embodiments, an ORF encoding a variant Cas12a endonuclease of the disclosure is codon optimized. Codon optimization methods are known in the art. An open reading frame of any one or more of the variant Cas12a endonucleases provided herein may be codon optimized. Codon optimization, in some embodiments, may be used to match codon frequencies in target and host organisms to ensure proper folding; bias GC content to increase
25 RNA (e.g., mRNA) stability or reduce secondary structures; minimize tandem repeat codons or base runs that may impair gene construction or expression; customize transcriptional and translational control regions; insert or remove protein trafficking sequences; remove/add post translation modification sites in encoded protein (e.g., glycosylation sites); add, remove or shuffle protein domains; insert or delete restriction sites; modify ribosome binding sites and
30 RNA (e.g., mRNA) degradation sites; adjust translational rates to allow the various domains of the protein to fold properly; or reduce or eliminate problem secondary structures within the polynucleotide. Codon optimization tools, algorithms and services are known in the art – non-limiting examples include services from GeneArt (Life Technologies), DNA2.0 (Menlo

Park CA) and/or proprietary methods. In some embodiments, the open reading frame sequence is optimized using optimization algorithms.

5 A “5′ untranslated region” (UTR) refers to a region of an mRNA that is directly upstream (i.e., 5′) from the start codon (i.e., the first codon of an mRNA transcript translated by a ribosome) that does not encode a polypeptide. A “3′ untranslated region” (UTR) refers to a region of an mRNA that is directly downstream (i.e., 3′) from the stop codon (i.e., the codon of an mRNA transcript that signals a termination of translation) that does not encode a polypeptide. When RNA transcripts are being generated, the 5′ UTR may comprise a promoter sequence.

10 In some embodiments, an RNA (e.g., mRNA) comprises a 5′ terminal cap. 5′-capping of polynucleotides may be completed concomitantly during an in vitro transcription reaction using, for example, the following chemical RNA cap analogs to generate the 5′-guanosine cap structure according to manufacturer protocols: 3′-O-Me-m⁷G(5′)ppp(5′) G [the ARCA cap]; G(5′)ppp(5′)A; G(5′)ppp(5′)G; m⁷G(5′)ppp(5′)A; m⁷G(5′)ppp(5′)G (New England BioLabs, Ipswich, MA). 5′-capping of modified RNA (e.g., mRNA) may be completed post-transcriptionally using, for example, a Vaccinia Virus Capping Enzyme to generate the “Cap 0” structure: m⁷G(5′)ppp(5′)G (New England BioLabs, Ipswich, MA). Cap 1 structure may be generated using both Vaccinia Virus Capping Enzyme and a 2′-O methyl-transferase to generate: m⁷G(5′)ppp(5′)G-2′-O-methyl. Cap 2 structure may be generated from the Cap 1
20 structure followed by the 2′-O-methylation of the 5′-antepenultimate nucleotide using a 2′-O methyl-transferase. Cap 3 structure may be generated from the Cap 2 structure followed by the 2′-O-methylation of the 5′-preantepenultimate nucleotide using a 2′-O methyl-transferase. Enzymes may be derived from a recombinant source. Other cap analogs may be used.

25 A “poly(A) tail” is a region of mRNA that is downstream, e.g., directly downstream (i.e., 3′), from the 3′ UTR that contains multiple, consecutive adenosine monophosphates. A poly(A) tail may contain 10 to 300 adenosine monophosphates. It can, in some instances, comprise up to about 400 adenine nucleotides. For example, a poly(A) tail may contain 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290 or 300 adenosine monophosphates. In some
30 embodiments, a poly(A) tail contains 50 to 250 adenosine monophosphates. In a relevant biological setting (e.g., in cells, in vivo) the poly(A) tail functions to protect mRNA from enzymatic degradation, e.g., in the cytoplasm, and aids in transcription termination, and/or export of the mRNA from the nucleus and translation. In some embodiments, the length of the 3′-poly(A) tail may be an essential element with respect to the stability of the individual

mRNA. In some embodiments, a poly(A) tail has a length of about 50, about 100, about 150, about 200, about 250, about 300, about 350, or about 400 nucleotides. In some embodiments, a poly(A) tail has a length of 100 nucleotides.

5 VII. Fusion Proteins

Some aspects relate to fusion proteins comprising any one or more of the variant Cas12a endonucleases provided herein and one or more effector proteins (e.g., a protein, such as an enzyme, that regulates a biological activity). Non-limiting examples include proteins that exhibit deaminase activity (e.g., adenosine deaminase and/or cytidine deaminase),
 10 reverse transcriptase, endonuclease (e.g., FokI), exonuclease activity (e.g., T5 exonuclease), methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, or
 15 demyristoylation activity.

Provided herein, in some aspects, are base editing fusion proteins comprising one or more base editing enzyme(s). A “base editing enzyme” is an enzyme that is capable of converting a target nucleobase or base pair into a different nucleobase or base pair (e.g. conversion from adenine to thymine (A to T), cytosine to thymine (C to T), adenine to
 20 guanine (A to G), cytosine to guanine (C to G)) without requiring the creation and/or repair of double-stranded breaks in a polynucleotide chain. Base editing enzymes can be specific for DNA bases or specific for RNA bases. Any base editing enzyme may be utilized in a fusion protein as described herein.

A base editing enzyme may be capable of converting adenine to guanine; adenine to
 25 thymine; adenine to uracil; adenine to cytosine; guanine to adenine; guanine to thymine; guanine to uracil; guanine to cytosine; thymine to adenine; thymine to guanine; thymine to uracil; thymine to cytosine; uracil to adenine; uracil to guanine; uracil to thymine; uracil to cytosine; cytosine to adenine; cytosine to guanine; cytosine to thymine; or cytosine to uracil. In some embodiments, a base editing enzyme is capable converting a standard nucleobase
 30 (e.g., A, C, G, T, U) into a modified nucleobase (e.g., hypoxanthine, xanthine 7-methylguanine, 5,6-dihydrouracil, 5-methylcytosine, 5-hydroxymethylcytosine). In some embodiments, a base editing enzyme is capable converting a modified nucleobase (e.g., hypoxanthine, xanthine 7-methylguanine, 5,6-dihydrouracil, 5-methylcytosine, 5-hydroxymethylcytosine) into a standard nucleobase (e.g., A, C, G, T, U). In some

embodiments, a base editing enzyme is capable of converting a modified nucleobase into a different modified nucleobase.

In some embodiments, a base editing enzyme converts a target base pair. For example, in some embodiments, a base editing enzyme converts C-G base pairs to T-A base pairs. In some embodiments, a base editing enzyme converts A-T base pairs to G-C base pairs.

In some embodiments, a base editing enzyme is a deaminase (*e.g.*, cytidine deaminase or adenosine deaminase) that is capable of removing an amino group from a molecule. Cytidine deaminases are capable of removing an amino group from a cytidine; adenosine deaminases are capable of removing an amino group from an adenosine. A cytidine deaminase may be an apolipoprotein B mRNA editing enzyme complex (APOBEC1) family protein. In some embodiments, the deaminase is an APOBEC1 polypeptide, an APOBEC2 polypeptide, an APOBEC3 polypeptide, an APOBEC3A polypeptide, an APOBEC3B polypeptide, an APOBEC3C polypeptide, an APOBEC3D polypeptide, an APOBEC3E polypeptide, an APOBEC3F polypeptide, an APOBEC3G deaminase polypeptide, an APOBEC3H polypeptide, an APOBEC4 polypeptide, or an activation-induced deaminase (AID). In some embodiments, an adenosine deaminase is a TadA polypeptide.

In some embodiments, a base editing enzyme is an oxidase (*e.g.*, a guanine oxidase) that is capable of oxidizing a certain nucleobase. For example, a guanine oxidase functions to oxidize a certain guanine nucleobase in a target gene to form 8-oxoguanine (8-oxo-G). 8-oxo-G induces steric rotation of the nucleobase around the glycosidic bond, forcing base pairing in the Hoogsteen orientation of 8-oxo-G. Cellular recognition of the mismatched 8-oxo-G/cytosine pairing leads to natural repair of the cytosine to an adenine. After an additional replication or mismatch repair, the 8-oxo-G is converted to a thymine, thereby producing a guanine-to-thymine conversion. In some embodiments, a guanine oxidase is a wild-type guanine oxidase, a xanthine dehydrogenase (XHD), a cytochrome P450 enzyme (*e.g.*, CYP1A2, CYP2A6 or CYP3A6), a TET-oxidase (*e.g.*, TET1, TET 1 -CD, TET2 or TET3), an alpha-ketoglutarate-dependent hydroxylase (*e.g.*, AlkB), . In some embodiments, a xanthine dehydrogenase is a *Streptomyces cyanogenus* xanthine dehydrogenase, *C. capitata* xanthine dehydrogenase, *N. crassa* xanthine dehydrogenase, *M. hansupus* xanthine dehydrogenase, *E. cloacae* xanthine dehydrogenase, *S. snoursei* xanthine dehydrogenase, *S. albulus* xanthine dehydrogenase, *S. himastatinicus* xanthine dehydrogenase, or a *S. lividans* xanthine dehydrogenase.

In some embodiments, a base editing enzyme is a methyltransferase (*e.g.*, a guanine methyltransferase) that is capable of methylating a certain nucleobase. For example, a guanine methyltransferase functions to methylate a certain guanine nucleobase in a target gene to form N2,N2-dimethyl-guanine or N-methyl-guanine. These methylated bases disrupt the hydrogen bonding interactions with the paired cytosine. Cellular recognition of the mismatched pairing leads to natural repair of the cytosine to an adenine. After an additional replication or mismatch repair, the methylated guanine is converted to a thymine, thereby producing a guanine-to-thymine conversion. In some embodiments, a guanine methyltransferase is a wild-type RlmA, *Escherichia coli* RlmA, human TrmTIOA, *Escherichia coli* TrmD, *M. Jannaschii* Trm5b, *P. Abyssii* Trm5a, a Trm5c from an archaeon, or a *Staphylococcus scirui* Cfr.

In some embodiments, a base editing enzyme (*e.g.*, a deaminase) is a base editing enzyme derived from a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse deaminase. In some embodiments, a base editing enzyme is a human base editing enzyme (*e.g.*, a human deaminase, *e.g.*, hAPOBEC polypeptide). In some embodiments, a base editing enzyme is a rat base editing enzyme (*e.g.*, a rat deaminase, *e.g.*, rAPOBEC1 polypeptide). In some embodiments, a base editing enzyme (*e.g.*, a deaminase) is an evolved variant of a wild-type base editing enzyme. For example, in some embodiments, a base editing enzyme is an evolved APOBEC polypeptide (*e.g.*, evoAPOBEC1 polypeptide), an evolved cytidine deaminase polypeptide (*e.g.*, evoCDA polypeptide), or an evolved FERNY polypeptide (*e.g.*, evoFERNY polypeptide). In some embodiments, a base editing enzyme is as described in Thuronyi, *et. al.* Nat Biotechnol. 2019 Sep; 37(9): 1070–1079, the contents of which are incorporated herein by reference. In some embodiments, a base editing enzyme is as described in US20200172931, the contents of which are incorporated herein by reference.

Exemplary, non-limiting, base editing enzyme sequences are provided in **Table 4**. In some embodiments, a base editing enzyme comprises the amino acid sequence of any one of SEQ ID NOs: 27-69. In some embodiments, a base editing enzyme comprises an amino acid sequence that includes one or more mutation(s) (*e.g.*, amino acid substitution(s)) relative to the amino acid sequence of any one of SEQ ID NOs: 27-69 and has at least 70% (*e.g.*, at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 98%) identity to the amino acid sequence of any one of the base editing enzymes in Table 4 (*e.g.*, SEQ ID NOs: 120-162), an ortholog thereof, or other base editing enzyme sequence.

Table 4. Non-limiting Examples of Base Editing Enzyme Sequences

Name	Sequence	SEQ ID NO:
<i>rAPOBEC1</i>	SSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNF IEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPNRQGL RDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPYPHLLWVRLYVLELYCIIILGLPPCLNIL RRKQFQLTFTTIALQSCHYQRLPPHILWATGLK	120
<i>evoCDA</i>	STDAEYVRIHEKLDIYTFKKQFSNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGI HAEIFSIRKVEEYLRDNPQGFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWVCKLYEK NARNQIGLWNLRDNGVGLNMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMF QVKILHTTKSPAV	121
<i>evoFERNY</i>	SFERNYDPRELRKETYLLYEIKWGKSGKLWRHWCQNNRTQHAENVYFLENIFNARRFNPSTHCSIT WYLSWSPCAECSQKIVDFLKEHPNVNLEIYVARLYPENERNRQGLRDLVNSGVITIRIMDLPDYN YCWKTFVSDQGGDEYWPGHFAPWIKQYSLKL	122
<i>evoAPOBEC1</i>	SSKTGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNF IEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPNVTLFIYIARLYHIANPNRQGL RDLISSGVTIQIMTEQESGYCWHNFVNYSPSNESHWPYPHLLWVRLYVLELYCIIILGLPPCLNIL RRKQSQLTSTTIALQSCHYQRLPPHILWATGLK	123
<i>hAPOBEC3A</i>	EASPASGPRHLMDFHIFTSNFNNGIGRHKTYLCYEVRDLNNGTSVKMDQHRGFLHNQAKNLLCGF YGRHAELRFLDLVPSLQLDPAQIYRVTFWISWSPCSWGCAGEVRAFLQENTHVRIRIFAARIYD YDPLYKEALQMLRDAGAQVSMITYDEFKHCWDTFVDHQGCFQPWDGLDEHSQALSGRLRAILQN QGN	124
<i>TadA</i> <i>Sequence A</i>	SEVEFSHEYWMRHALTAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPATAHAEIMALRQG GLVMQNYRLIDATLYVTTFEPCVMCAGAMIHSRIGRVVFGVRNSKRGAAAGSLMNVNYPGMNHRVE ITEGILADECAALLCDFYRMPRQVFNQAQKKAQSSIN	125
<i>TadA</i> <i>Sequence B</i>	SEVEFSHEYWMRHALTAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPATAHAEIMALRQG GLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVE ITEGILADECAALLSDFFRMRQEIKAQKKAQSSD	126
<i>E. coli TadA</i>	RRAFITGVFFLSEVEFSHEYWMRHALTAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPT AHAEIMALRQGGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDV LHHPGMNHRVEITEGILADECAALLSDFFRMRQEIKAQKKAQSSD	127
<i>Staphylococcus aureus TadA</i>	GSHMTNDIYFMTLATEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERA AKVLGSRWLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLQSSNFNHRAI VDKGVLEKACSTLLTFFKNLRANKKSTN	128
<i>Bacillus subtilis TadA</i>	TQDELYMKEAIKEAKAEKEGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKAL GTWRLEGATLYVTLEPCVMCAGAVLSRVEKVVFAGADDPKGGCSGTLMNLQEEFRNHQAQEVVSG VLEEECGMLSAFFRELRRKKKAARKNLS	129
<i>Salmonella typhimurium TadA</i>	PPAFITGVTSLSDELDEHEYWMRHALTAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPT AHAEIMALRQGGGLVQNRYRLDITLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDV LHHPGMNHRVEIEGVLRDECATLLSDFFRMRQEIKALKKADRAEGAGPAV	130
<i>Shewanella putrefaciens TadA</i>	DEYMQVAMQMAEKAAEAGEVPGAVLVKDGQIATGYNLSISQHDPTAHAEILCLRSAGKLEN YRLDITLYITLEPCVMCAGAMVHSRIARVVYGARDEKTGAAGTVNLLQHPAFNHQVEVTSGLV AEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE	131
<i>Haemophilus influenzae F3031 TadA</i>	DAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEIIA LRNAGNIQNYRLNLTLYVTLEPCVMCAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDYKMN HTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK	132
<i>Caulobacter crescentus TadA</i>	RTDESEDQDHRMRLALDAARAAAEAGETPGAVILDPTSTGEVIATAGNGPIAAHDPTAHAEIAA MRAAAKLGNYRLDITLYVTLEPCVMCAGAISSHARIGRVVFGADDPKGGAVVHGPKFFAQPTCH WRPEVTGGVLADESADLLRGFFRRRKAKI	133
<i>Geobacter sulfurreducens TadA</i>	SSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRHNLRGNSNDPSAHAEMIA IRQAARRSANWRLTGATLYVTLEPCVMCMGAIILARLERVVFGCYDPKGAAGSLYDLSADPRLN HQVRLSPGVCQEECTMLSDFFRDLRRRKAKATPALFIDERKVPPEP	134
<i>S. cyanogenus XDH</i>	MSHLSEPEKPVVGVSMPEHASAVQHVTGAALYTDDLQRTKDVLAHYPVQVMKARGRVTAALRTGA ALAVPGVVRVLTGADVPGVNDAGMKHDEPLFPDEVMFHGHAVAWVLGETLEAARIGAAAVEVDLE ELPSVITLQDAIAADSYHGARPVMTHGDVDAGFADSAHVFTGEFQFSGQEHFYLETHAALAQVDE NGQVFIQSSTQHPSETQEIVSHVLGVPAHEVTVQCLRMGGGFGGKEMQPHGFAAIAALGAKLTGR PVRFRNLNRTQDLTMSGKRHGFHATWKIGFDTEGRIQALDITLADGGWSLDLSEPVLARALCHID NTYWI PNARVAGRIARTNTVNTAFRGFGPGQMLVIEDILGRCAPRLGVDKELRERNFYRPGQ GQTPPYGQPVTPERIAAVWQVQDNGHIDREREIAAFNAAPHKRALAVTGKFGISFNLT FNQGGALVLIYKDGSVLINHGTEMQGLHTKMLQVAATTLGIPLHKVRLAPTRTDKVPNTSATA ASSGADLNGGAVKNACEQLRERLLRVAASQLGTNASDVRIEAGVARSLSGSDQELAWDDLVRTAYF QRVQLSAGGYRTEGLHWDKASFRGSPFKYFAIGAAATEVEVDGFTGAYRIRRVDIVHDVGDLSLS PLIDIGQVEGGFVQAGWLTLEDLRWDTGDPNRRGLLTQAASTYKLPFSFEMPEEFNVTLLENA TEGAVFGSKAVGEPPLMLAFSVREALRQAAAFGPRGTAVELASPATPEAVYWAIESARQGGTA	135

Name	Sequence	SEQ ID NO:
	GDGRTHGAAASDAVAVRTGVEALSGA	
<i>C. capitata XDH</i>	MTTNGNSFIVPVEKESPLIFFVNGKKVIDPTDPECTLLTYLREKLRLCGTKLGCGEGGCGACTV MLSRVDRATNSVKHLAVNACLMPVCAMHGCAVTTIEGIGSTRTRLHPVQERLAKAHGSCQCGFCTP GIVMSMYALLRSMPLPSMKDLEVAFQGNLCRCTGYRPILEGYKFTTKEFSCGMGEKCKCKLQSNNGN DVEKNGDDKLFERSAFLPFDPSQEPIFPPELHLNSQFDAENLLFKGPRSTWYRPLAPLAPAGGMIA ENPHGKIIIVGNTEVGVEMKFKQFLYTVHINPIKVPPELNMQELEDSSILFGSAVTLMDIEEYLRER IAKLPEHETRFRCVAKMLHYFAGKQIRNVASLGGNIMTGSPISDMNPILTAACAKLKVCSLVEG RIETREVCMPGPGFTTGYRKNTIQPHEVLVAIHFPKSKKDQHFVAFKQARRRDDDIAIVNAAVNVT FESNTNIVRQIYMAFGGMAPTVMVPKTSQIMAKQKWNRVLVERVSESLCAELPLAPTAPGGMIA YRRSLVSLFFKAYLAISQELVKSNIIEEADIPEREQSGAAIFHTPILKSAQLFERVCEVQSTCD PIGRPKVHASAFKQATGEAIYCDDIPRHENELYLALVLSSTKAHAKIVSVDESALQAGVHAFFS SKDITEYENKVGVSFVHDEEVFASERVYCCQGVIGAIVADSQVLAQRAARLVHIKYEELTPVITII EQAIAKHKSYFFNPQYIVQGDVATAFEEADHVEYENSCRMGGQEHFYLETNACVATPRDSDEIELF CSTQNPTEVQKLVAHVLSVPCHRVVCRSKRLGGGFGGKESRSIILALPVALASYRLRRPVRCLMD RDEDMMTTGTRHPFLFKYKVGFTKEGLITACDIECYNNAAGCSMDLSFSVLDAMNHFENCYRIPN VKVAGWVCRTNLPNTAFRGFGGPGQGMFAAEHIVRDVARIVGKDYLDIMQMNFTYKTGDYTHYNQK LENFPIEKCFDCLNQSEFHKKRLAIEEFNKKNRWRKRGIALVPTKYGIAFGAMHLNAGALINI YGDGSVLLSHGGVEIGQGLHTKMIQCCARALGIPTELIHIAETATDKVPNTSPTAASVGSINGM AVLDACEKLNQRLKPIREANPKATWQECISKAYFDRISLSASGFYKMPDVGGDPKTNPNARTYNY FTNGVGVSVVEIDCLTGDHQLVSTDIVMDIGSSLNPAIDIGQIEGAFMQGYGLFVLEELIYSPQG ALYSRGPGRMYKLPGFADIPGEFNVSLLTGAPNPRAVYSSKAVGEPPLFIGSTVFFAIKQAIAAA AERGLSITFELDAPATAARIIRMACQDEFTDLIEQSPGTYTPWNVVP	136
<i>N. crassa XDH</i>	MTTNGNSFIVPVEKESPLIFFVNGKKVIDPTDPECTLLTYLREKLRLCGTKLGCGEGGCGACTV MLSRVDRATNSVKHLAVNACLMPVCAMHGCAVTTIEGIGSTRTRLHPVQERLAKAHGSCQCGFCTP GIVMSMYALLRSMPLPSMKDLEVAFQGNLCRCTGYRPILEGYKFTTKEFSCGMGEKCKCKLQSNNGN DVEKNGDDKLFERSAFLPFDPSQEPIFPPELHLNSQFDAENLLFKGPRSTWYRPLAPLAPAGGMIA ENPHGKIIIVGNTEVGVEMKFKQFLYTVHINPIKVPPELNMQELEDSSILFGSAVTLMDIEEYLRER IAKLPEHETRFRCVAKMLHYFAGKQIRNVASLGGNIMTGSPISDMNPILTAACAKLKVCSLVEG RIETREVCMPGPGFTTGYRKNTIQPHEVLVAIHFPKSKKDQHFVAFKQARRRDDDIAIVNAAVNVT FESNTNIVRQIYMAFGGMAPTVMVPKTSQIMAKQKWNRVLVERVSESLCAELPLAPTAPGGMIA YRRSLVSLFFKAYLAISQELVKSNIIEEADIPEREQSGAAIFHTPILKSAQLFERVCEVQSTCD PIGRPKVHASAFKQATGEAIYCDDIPRHENELYLALVLSSTKAHAKIVSVDESALQAGVHAFFS SKDITEYENKVGVSFVHDEEVFASERVYCCQGVIGAIVADSQVLAQRAARLVHIKYEELTPVITII EQAIAKHKSYFFNPQYIVQGDVATAFEEADHVEYENSCRMGGQEHFYLETNACVATPRDSDEIELF CSTQNPTEVQKLVAHVLSVPCHRVVCRSKRLGGGFGGKESRSIILALPVALASYRLRRPVRCLMD RDEDMMTTGTRHPFLFKYKVGFTKEGLITACDIECYNNAAGCSMDLSFSVLDAMNHFENCYRIPN VKVAGWVCRTNLPNTAFRGFGGPGQGMFAAEHIVRDVARIVGKDYLDIMQMNFTYKTGDYTHYNQK LENFPIEKCFDCLNQSEFHKKRLAIEEFNKKNRWRKRGIALVPTKYGIAFGAMHLNAGALINI YGDGSVLLSHGGVEIGQGLHTKMIQCCARALGIPTELIHIAETATDKVPNTSPTAASVGSINGM AVLDACEKLNQRLKPIREANPKATWQECISKAYFDRISLSASGFYKMPDVGGDPKTNPNARTYNY FTNGVGVSVVEIDCLTGDHQLVSTDIVMDIGSSLNPAIDIGQIEGAFMQGYGLFVLEELIYSPQG ALYSRGPGRMYKLPGFADIPGEFNVSLLTGAPNPRAVYSSKAVGEPPLFIGSTVFFAIKQAIAAA AERGLSITFELDAPATAARIIRMACQDEFTDLIEQSPGTYTPWNVVP	137
<i>M. Hansupus XDH</i>	MSNMFEFRLNGATVRVDGVSPNTTLLDLFRNRLGTGKQCAEGDCGACTVALVDRDAQGNRCLR AFNACIALVPMVAGRELVTVEGVSSSEKPHVPVQQAMVKHYGSCQCGFCTPGFIVSMAEGYSRKDVC TPSSVADQLCGNLCRCTGYRPIRDAMMEALAEERDADASPATAIPSAPLGGPAEPLSALHYEATGQ TFLRPTSWKELLDLRARHPEAHLVAGATELGVDITKKARRFPFLISTEGVESLREVRREKDCWYV GGAASLVALEEALGDALPEVTKMLNVFASRQIRQRATLAGNLVTASPIGDMAPVLLALDARLVLG SVRGERTVALSEFFLAYRKTAQADEVVRHIVIPHPAVPERGQRLSDSFKVSKRRELDISIVAAG FRVELDAHGVVSLARLGYGGVAATPVRAVRAEAALTGQFWTRETVDQVLPVLAEEITPISDQRGS AEYRRGLVAGLFEKFFAGTYSPVLDAAFGFEKGDQVPAADAGRALRHESAMGHVTGSARYVDDFA QRQPMFEVWPVCAPHAHARIFKRDPTAARKVPGVVRVMAEDIPGTNDTGP IRHDEPFFADREVF FHGQIVAFVVGESVEACRAGARAVEVEYEPFAIFTVEDAMAQGSYHTEPHVIRRGDVDAAFASS PHRFSGTMAIGGQEHFYFETQAAFAERGGDGDITVVSSTQHPSEVQAIISHVHFPPRSRVVVKSP RMGGGFGGKETQGNSPAFAVAFASWHTGRPTRWMDRDVDMVVTGKRHPFHAAYEVGFDDEGKFF AFRVQFVSNNGWSFDFSESETDRAFFHEDNAYYVPAFTYTRGVAKTHFVSNTAFRGFGGPGQMLV TEEVLAHVARSVGPADVVRERNLYRGTTGETNTTHYQGELEDERIHRVWHEELKRTSDFERRAEV DAFNARSFFIKRGLAITPMKFGISFTATFLNQAGALVHLYRDGSMVSHGGTEMGGGFHTKVQGV AMREFGVEASAVRIAKTATDKVPNTSATAASSGSDENGAAVRFACITFRERFAPVAVRFFADRHG RTVAPEAFFFSEGKVGFRGEPEVSLPFANVVEAAYLARVGLSATGYQTPGIGYDKAKGRGRFFL YFAYGASVCEVEVDGHTGVKRVLRVDLLEDVGDLSNPGVDRGQIEGGFVQGLGWLTEELRWAN GRLLTHSASTYAVPAFSDAPIDFRVRLLERAHQHNTIHGSKAVGEPPLMLAMSAREALRDVAGAF GQAGGGVALASPATHEALFLAIQKRLSRGAREDGREAA	138

Name	Sequence	SEQ ID NO:
<i>E. cloacae</i> XDH	MKFDKPATTNPIDTLRVVGPHTRIDGPRKTTGSAHYAYEWHDIAPNAAAYGH VVGAPIAKGRITAIIDTKAAEAPGVLAVITADNAGPLGKGEKNTATLLGGPEIEHYHQAVALLVVA ETFEQARAAAAALVKVTCRAQGAIDLAAEKASVTEPPEDTPDKNVGDVATAFASAAVKLDIAYTT PDQSHMAMEPHASMAVWEGDNVTWTSNQIMIDWCRTDLALTLPKIPPENVRIVSPYIGGGFGGKLF LRSDALLAALGARAVKRPVKVMLPRPTIPNNTHRPATLQHIRIGTDTEGKIVAIAHDSWSGNLP GGTPETAVQQTELLYAGANRHTGLRLATLDLPEGNAMRAPGEAPGLMALEIAIDEIADKAGVDPV AFRILNDTQVDPANPERRFSTRQLVECLQTGAERFGWQKRHAQPGQVRDGRWLVMGMAAGFRNN LVATSGARVHLNADGSVAVETDMDITGTSYTIIAQTAAEMLGLPLEKVDVRLGDSRFPVSAGSG GQWGANTSTAGVYACVKLRERAIARQLGFDPAEFADETISAQGRSAPLAEAAKSGVLTAEDSI EFGDLDKEYQQSTFAGHFVEVGVDSATGEVRVRMLAVCAAGRILNPI TARSQVIGAMTMGLGAA LMEELAVDTRLGYFVNHDMAAYEVPVHADIPQEVEVIFLEDTPDISSPMKAKGVGELGLCGVSAAI ANAIYNATGVRVRDYPITLTKLIDALPDVA	139
<i>S. snoursei</i> XDH	MSHDPVPHLPAPLPHPLGAPSVRREGREKVTGAARYAAEHTPPGCAYAWPVPATVARGRITEL DTAAALALPGVIAVLTHENAPRLASTGDP TLAVLQEDRVPHRGWYVALAVADTLEAARDAEEAVH VGYATEPHDVRIADHPRLYVPEEVFGGPGARERGFDAFAAATPVDAVYVTPPLHNPMPH AATAQWTDGHLTVHDSSQGATRVCEDLAALFKLGTDEITVSEHVGGGFGAGKTGPRPQVLAAMA ARHTGRPVKLALPRQLPGVVGHRAPTLHRVRIGAGHDGVITALAHEIVTHTSTVTEFVEQAAP ARMMYTSPHSRTVHRLAALDVP TSPWMRAPGEAPGMYALESALDELAVVLDIDPVELRIRNDPAT EPDTGRPFSSRHLVECLRAGAERFGWLPRDPRPAVRRRGDLLGTGVAATYPVQISETAEAAHA AADGGYRIRVNATDIGTGARTVLTQIAAAVLGAPEDRVVDIGSSDLPPAVLAGGSTGTASWGA VHKACTSLARLRAHHGLPAEGIMAELESEWAPMALRAWRIISGLGLPTKYGSTPVALVMRAATE PVAGSGPSVEGVPVSSGLVAMKRAPFSMSRMALVSASKL	140
<i>S. albulus</i> XDH	MTPTPTTRTRAMSHPEEAPFPFGPPHPLGDPLVRREGREKVTGTARYAAEHTPDGCAYAWPVP ATVVRGRITELDTGAALALPGVIAVLTHENAPRLASTGDP TLALLQEDRVPHRGWYVALAVADTL EAARDAEEAVHVS YATEPHDVLTADHPRLYVPAEVFGGPGARERGFDAFAAATPVDAVYVTP PPLHNPMPHHAATLWTHGHLTVHDSSQGATRVREDLALFKLGGDQITVHSEHVGGGFGSGKT PRPQVLAAMAAARHTGRPVKLALPRRHLPAVVGHRAPTLHRVRLGAGPDGVITALAHEIVTHTST VAEFVEQAAMPARIMYTSPHSRTVHRLAALDVP TSPWMRAPGEAPGMYALESADDELAVVLDIDP IDLRLRNEPGTEPDTRGPFSSRHLVDCLRAGAARFGWSSRDPRPAVRRRGDLLGTGVAATYPV QISATDAEAAADGTFRVRVNATDIGTGARTVLAQIAAAALGAPADRVREIGSSDLPPAVLAG GSTGTASWGWAVHKACTVLLARLREHRGLPAEGVTVTEDTRRETEQSPSPSRHAFGAVFAEVQV DTRTGEVRARLLGQYAAAGHILNPTARSQFVGGVMVGLGMALTEDSALDPVYGDFTARDLAAYH VPACADVPAIEAHLDEEDPHLNPMSGKIGIEIGVGTAAIGNAVWHATGVRLRDLPLTPDRIL TARTVPLT	141
<i>S. himastatinicus</i> XDH	MTRVDGLDKVTGAATYAYEFPTPDVGVVWPQATIAARGVTEVDGAPALARPGVLAVLDSGNAPR INTEAQAGPDLFVLQSPVAYHGQIVAAVVATSLAAREGAAAVRVS YEQEPHDVVLRFDDERAQ VAETVTDGSPGFVEHGDAGALAAAPVRTEAMYTTPVEHTSPMEPHATIAAWDEDRLTYNADQG PFMSSQLLAAVFGLDQGAVEVVAEYIGGGFGSGKIPRSPAVLAALAACHLGRPVKIALTRQQMFQ LIPYRAPTIQIRIRLGAERDGRLLTAIDHEVVQORSMAEFADQTGSSTRVMYAAPNIRTTVKTAPL DVLTPAWFRAPGHTPGMFALESAMDELATELEIDPVELRIRNDTGVPDPSGKPFSSRGLVACIRE GAARFDWALRDPKPGIRREGRLVGTGVAASHHPDYVFPSSATARAEDGTFTVRVGAVDIGTGG RTALTQLAADALGIPVERLRLEIGRASLGPAFAGGSLGTASWGWAVDKACRALLAELDTYGGAV PDGGLVLRADTTEDVELRASFSRHSFGAHFAQVRVDDTGEIRVDRMLGVFAAGRIVNPKTARSQ FVGAMTMGLSMALLEIGEVDVPVFGDFANHDFAGYHVAANADVPKLEALWLDEQDDNPNPVRGKGI GELGIVGAAAAVTNAFHATGQVRVRLDIRVERSREALRAARAEEAQKRGFGAAEQGKPVG	142
<i>S. lividans</i> XDH	MSHLSEPEKPVVGVSMPHESAVQHVTTGAALYTDLDLVQRTKDVLAHYVPQVMKARGRVLTALRTGA ALAVPGVVRVLTGADVPGVNDAGMKHDELPFDEVFMFGHAVA VVLGETLEAARI GAAAVEVDLE ELPSVITLQDAIAADSYHGARPVMTHGDVDAGFADSAHVFTGEFQFSGQEHFYLETHAALAQVDE NGQVFIQSSTQHPSETQEI VSHVFGVPAHEVTVCFRMGGGFGGKEMQPHGFAAIAAFGAKFTGR FVRFRFNRTQDFTMSGKRHGFHATWKIGFDTEGRIQAFDATFTADGGWSFDFSEPVFAARAFCHID NTYWI PNARVAGRIARTNTVNTAFRGFGGPGQGMFVIEDIFGRCAPRFGVDAKEFRERNFYRPGQ GQTTYPGQPVTPERIAAVWQVQDNGHIADREIEIAAFNAHPHTKRAFAVTGVKFGISFNFTA FNQGGAFVFIYKDGSVFINHGGTEMGQGFHTKMFQVAATTFGIPFHKVRFAPTRTDKVPNTSATA ASSGADENGGA VKNACEQFRERFFRVAASQFGTNASDVRIE VEGVARSFGSDQEFADWDFVRTAYF QRVQFSAAGYRTEGFWDAKSFSGSPFKYFAIGAAATEVEVDGFTGAYRIRRVDIHVHDVGSFS FFIDIGQVEGGFVQAGWFTFEDFRWDTGDGPNRGRFFTQAASTYKFPFSSEMPPEEFNVTFEENA TEEGAVFGSKAVGEPPFMFAFSVREAFRQAAAAFGPRGTAVEFASPATPEAVYWAIESARQGGTA GDGRTHGAAAADAVRTGVEAESGA	143
Cytochrome P1A2	MLASGMLLVALLVCLTMVLMVSVWQQRKSKGLPPGPTLPFFIGNYLQINTEQMYNSLMKISERY GPVFTIHLGPRRVVLCGHDAVREALVDQAEFSGRGEQATFDWVFKGYGVVFSNGERAKQLRRF SIATLRDFGVGKRGIEERI QEEAGFLIDALRGTTGGANIDPTFFLSRTVSNVSIIVFGDRFDYKD KEFLSLLRMLGIGIFTSTSTGQLYEMFSSVMKHLPGPQQQAFQLLQGLDFIAKKVEHNQRTLD PNSPRDFIDSFLIRMQEEKNPNTFYLKLNLMVMTLNLNFI GGTEVSTTLRYGFLLLMKHPEVEA KVHEEIDRVI GKNRQPKFEDRAKMPYMEAVIHEIQRFQDVIPMSLARRVKKDKTFDFFLPKGTE VFPMLGSLVRDPSFFSNPDQFNQHFLENEKGQFKKSDAFVFPFSIGKRNCFEGELARMELFLFFT	144

Name	Sequence	SEQ ID NO:
	VMQNFRLKSSQSPKIDVSPKHVGFATIPRNYTMSFFPR	
<i>Cytochrome P2A6</i>	MLASGMLLVALLVCLTMVLMVSVWQQRKSKGKLP GPPTPLPFI GNYLQLNTEQMYNSLMKISERY GPVFTIHLGPRRVVVL CGHDVAREALVDQAEFSGRGEQATFDWVFKGYGVVFSNGERAKQLRRF SIATLRDFGVGKRGIEERI QEEAGFLIDALRGTTGGANIDPTFFLSRTVSNVSISSIVGDRFDYKD KEFLSLLRMLGIFQFTSTSTGQLYEMFSSVMKHLPGPQQQAFQLLQGLDFIAKKVEHNQRTLD PNSPRDFIDSFLIRMQEEKNPNTIFYLKNLVMTTNLNFIGGTETVSTTLRYGFLLLMKHPEVEA KVHEEIDRVIGKNRQPKFEDRAKMPYMEAVIHEIQRFQGDVIPMSLARRVKKDTKFRDFFLPKGTE VFPMLGSVLRDP SFFSNPQDFNPQHFLNEKGQFKKSDAFVPPFSIGKRNCFGEGLARMELFFFTT VMQNFRLKSSQSPKIDVSPKHVGFATIPRNYTMSFFPR	145
<i>Cytochrome P3A4</i>	MALIPDLAMETWLLAVSLVLLYLYGTHSHGLFKKLGIPGPPTPLPFLGNILSYHKGF CMFDMECH KKYGVWGFYDGGQPVLAITDPDMIKTVLVKECYSVFTNRRPFGPVGFMKSAISIAEDEEWKRLR SLLSPTFTSGKLKEMVPIIAQYGDVLRNLRREAETGKPVTLKDVFGAYSMVDITSTSGFVNIDS LNNPQDPFVENTKKLLRDFLDPPFLSITVFPFLIPILEVLNICVFPREVTNFRKSVKMKESR LEDTQKHRVDFLQLMIDSQNSKETESHKALSDELVAQSIIFIFAGYETTSSVLSFIMYELATHP DVQQLQEEIDAVLPNKAPPTYDVLQMEYLDMVNETLRLFIAMRLERVCKKDVINGMFI PK GVVMIPSYALHRDPKYWTEPEKFLPERFSKKNKDNIDPIYIYTPFGSGPRNCIGMRFALNMMLKLA LIRVLQNF SFKPCKETQIPKLKSLGGLLQPEKPVVLKVESRDGTVSGA	146
<i>TET1</i>	MSRSRHARPSRLVRKEDVNKKKNSQLRKT TKGANKNVASVKTLSPGKLGKLIQERDVKKKTEPK PPVPVRSLLTRAGAARMNLDRETVLFQNPESITCNGFTMALRSTSLRRLSQPPLVVAKSKKVP SKGLEKQHDCKYKILPALGVKHSENDSVPMQDTQVLPDIETLIGVQNP SLLKGKSETTQFWQR VEDSKINIPHTSGPAAEILPGPLEGTRCGELFSEETLNDTSGSPKMFADTVCAFPQ RQATPKV TSQGNPSIQLEELGSRVESLKLSDSYLDP IKSEHDCYPTSSLNKVIPDLNLRNCLALGDSPTS VIKFLLAGSKQATLGAKPDHQEAFEATANQQEVSDTTSFLGQAFGAIPHQWELPGADPVHGEALG ETPDLPEIPGAIPVQGEVFGTILDQOETLGMSSGVVDPVLPVLPVPPNP IATFNAPSKWPEPQST VSYGLAVQGAIQILPLGSGHTPQSSSNSEKNSLPPVMAISNVENEKQVHISFLPANTQGFPLAPE RGLFHASLGIAQLSQAGPSKSDRGSSQSVSTSTVHVNTTVTMPVPMVSTSSSSYTTLLPLEK KKRKRGCVCPECCQKTNCGECTYCKNRKNSHQICKRCKCEELKKKPSVVVP LEVIKENKRPQREK KPKVLKADFDNKPVNGPKSEMDYSRCGHGEEQKLELNPHTVENVTKNEDSMTGIEVEKWTQNK SQLTDHVKGDFSANVPEAEKSKNSEVDKKRTKSPKLFVQTVRNGIKHVHCLPAETNVSPFKFNIE EFGKTLNNSYKFLKDTANHNKAMSSVATDMSCDHLKGRSNVLVFPQPGFNCSSIPHSSHSIINH HASIHNEGDQPKTPENIPSKPKDGSFPVQPSLLSLMKDRRLTLEQVVAIEALTQLEAPSENSSP SKSEKDEESEQRTASLLNSCKAILYTVRKDLQDPNLQGEPPKLNHCPSLEKQSSCNTVVFNGQTT TLSNSHINSATNQASTKSEHSKVTNSLSLPIPKSNSKIDTNKSIAGGIITLDNCSNDLHQLPP RNNEVEYCNQLDSSKKLSDDLSCQDATHTQIEEDVATQLTQLASI IKINYIKPEDKKVESTPT SLVTCNVQKYNQEKGTIQKFPSSVHNNHSSLTQKKNPTQKKTSPDRDRKKKPTVVSQ NDRQKWEKLSYMYGTICDIWIASKFQNFQGFCHDFPTVFGKISSSTKIWKPLAQTRSIMQPKTV FPPLTQIKLQRYPESAEEKVVEPLDSL SLFHLKTESNGKAFTKAYNSQVQLTVNANQKAPLT QPSSFPNQCANVMAGDDQIRFQQVVEQLMHQRLPTLPGISHETPLPESALTLRNVNVCSSG VSTKSEEEVCSSEFGTSEFSTVD SAQKNFNDYAMNFTNPTKNLVSITDKSELPTCSCLDRVIO KDKGFPYYTHLGAGPSVAAVREIMENRYGQKGNARIEIVVYTGEKGSSSHGCP IAKWVLRSSDE EKVLCCLVRQRTGHHCTAVMVVLI MVWDGIPLMADRLYTELTENLKS YNGHPTDRRCTLNENRT CTCQGIDPETCGASFSGCSWSMYFNGCKFGRSPSPRRFRIDPSSPLHEKNLEDNLQSLATRLAP IYQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACLD FCAHPHRDIHNMNNGSTVVCILTR EDNRS LGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEAKIKSGAIEVLAPRRKKRTCTQPVPRSG KKRAAMTEVLAHKIRAVEKKPIPRIKRKNSTTTNNSKPSLPTLGSNTETVQPEVKSETEPHF ILKSSDNTKTYSLMP SAPHVPKEASPGFSWSPKTASATPAPLKNDATASC GFSESRSTPHCTMPS GRLSGANAAAADGPGISQLGEVAPLPTLSAPVMEPLINSEPTGVTEPLTPHQPNHQPSFLTSPQ DLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYWSDEHIFLDANIGGVAIAPAH GSVLIECARRELHATTPVEHPNRNHPTRL SLVFYQHKNLKNPQHGFELNKIKFEAKEAKNKKMKA SEQKDQAANEGPEQSSEVNELNQIPSHKALTTLTHDNVTVSPYALTHVAGPYNHVW	147
<i>TET1 catalytic domain</i>	MGSLPTCSCLDRVIQKDKGFPYYTHLGAGPSVAAVREIMENRYGQKGNARIEIVVYTGEKGS GCP IAKWVLRSSDEEKVLCVRQRTGHHCTAVMVVLI MVWDGIPLMADRLYTELTENLKS YN GHPTDRCTLNENRTCTCQGIDPETCGASFSGCSWSMYFNGCKFGRSPSPRRFRIDPSSPLHEK NLEDNLQSLATRLAP IYQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACLD FCAHPHRDI HNMNNGSTVVCILTR EDNRS LGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEAKIKSGAIEVLAP RKKRTCTQPVPRSGKKRAAMTEVLAHKIRAVEKKPIPRIKRKNSTTTNNSKPSLPTLGSNT ETVQPEVKSETEPHFILKSSDNTKTYSLMP SAPHVPKEASPGFSWSPKTASATPAPLKNDATASC GFSESRSTPHCTMPSGRLSGANAAAADGPGISQLGEVAPLPTLSAPVMEPLINSEPTGVTEPLT PHQPNHQPSFLTSPQDLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYWSDEHIF LDANIGGVAIAPAHGSVLIECARRELHATTPVEHPNRNHPTRL SLVFYQHKNLKNPQHGFELNK IKFEAKEAKNKKMKA SEQKDQAANEGPEQSSEVNELNQIPSHKALTTLTHDNVTVSPYALTHVAG PYNHVW	148
<i>TET2</i>	MEQDRTNHVEGNRLSPFLIPSPPICQTEPLATKLQNGSPLPERAHPEVNGDTKWHFSKSYGIPC MKGSQNSRVSPDFTQESRGYSKCLQNGGIKRTVSEPSLSGLLQIKKLKQDKANGERRNFVGSQE RNPGESSQPNVSDLSKKESVSSVAQENAVKDFTSFSTHNCSGPENPELQILNEQEGKSANYHDK	149

113

Name	Sequence	SEQ ID NO:
<i>E. coli GMP Synthase</i>	MTENIHKKRILILDFGSQYTLVARRVRELGVYCEWAWDVTEAQIRDFNPSGIIILSGGPESTTE ENSPRAPQYVFEAGVPVFGVCYGMQTMAMQLGGHVEASNEREFGYAQVEVVNDALVRGIEDALT ADGKPLLDVWMSHGDKVTAIPSDFITVASTESCFFAIMANEERKRYGVQFHPEVTHTRQGMRLME RFVRDICQCEALWTPAKIIDDAVARIREQVGDDKVILGLSGGVDSSTVAMLLHRAIGKNLTCTVFV DNGLRLNEAEQVLDMFGDHFGNLIVHVPADRFLSALAGENDPEAKRKIIGRVFVEVFDEEALK LEDVKWLAQGTIYPDVIESAASATGKAHVIKSHHNVGGLPKEMKMGLVEPLKELFKDEVRKIGLE LGLPYDMLYRHPFPGPLGVRVLGEVKKEYCDLLRRADAIFIEELRKADLYDKVSQAFTVFLPVR SVGVMGDGRKYDWVSLRAVETIDFMTAHWAHLPYDFLGRVSNRIINEVNGISRVVYDISGKPPA TIEWE	153
<i>S. scirui Cfr</i>	MNFNNKTKYKGIQEFRLSNNEPDYRIKQITNAIFKQIRISRFEDMKVLPKLLREDLINNFGETVLN IKLLAEQNSEQVTKVLFVSKNERVETVNMKYKAGWESFCISSQCGCNFGCKFCATGDI GLKKNL TVDEITDQVLYFHLLGHQIDSISFMGMGEALANRQVFDALDSFTDPNLFALSPRRLSISTIGIIP SIKKITQEYPPQVNLTFSLHSPYSEERSKLMPIINDRYPIDVMNILDHRLTSRKVYIAYIMLP VNDLSLEHANEVVSLLKSKYSGKLYHVNLI RYNPTISAPEMYGEANEQVEAFYKVLKSAGIHVT IRSQFGIDIDAACGGLYGNYSNSQ	154
<i>A. aeolicus TrmI</i>	MEIVQEGIAKIIIVEIPKTVSSDMPVFYNPRMRVNRDLAVLGLLEYLCKKLGRPVKADPLSASGI RAIRFLLETSCKEAYANDISSKAIEIMKENFKLNNIPEDRYEIHGMEANFRLRKEWGFDFDVID LDPPGTPVPFFIESVALSMKRGILSLTATDTAPLSGTYPKTCMRRYMARPLRNEFKHEVGIRILI KKVIELAAQYDIAMIPIFAYSHLHYFKLFFVKERGVKVDKLEQFGYIQCFCNMCNREVVDLY KFKEKCPHCGSKFHIIGGPLWIGKLWDEEFTNFLYEEAQKREEIEKETKRILKLIKESQLQTVGF YVLSKLAEKVKLPAQPPIRIAVKFFNGVRTHFVGDFRNTNLSFEVMMKMEELKEKQKEFLEKKK QG	155
<i>S. cerevisiae TrmI</i>	MEGFFRIPLKRANLHGMKAAISKIKANFTAYGAPRINIEDFNIVKEGKAEILFPKKETVFYNPI QQFNRLDSVTICKAWDNLYGEECGQKRNNKKSKKKRCAETNDDSSKRQKMGNGSPKEAVGNSNRN EPYINILEALSATGLAIRYAHEIPHVREVIANDLLPEAVESIKRNVYEYNSVENIVKPNLDDANV LMYRNKATNNKFHVIDLDPYGTVTPFVDAAIQSIEEGGLMLVTCTDLSVLAGNGYPEKCFALYGG ANMVSHESTHESALRLVLNLLKQTAAYKKTVEPLLSLSIDFYVRVFKVKTSPIEVKNVMSSTM TTYHCSRCSYHNQPLGRISQREGRNNKTFTKYSVAQGPVDTCKCFCEGTYHLAGPMYAGPLHN KEFIEEVLRLNKEEHRDQDDTYGTRKRIEGLMSLAKNELSDSPFFYSPNHASVILKQVPLPKKV VAGLSGLGFECSLTHAQPSLKTNPWDIAIYWMQKDEKDL SKMNPNTTGKILSAMPGWLS GTVKSEYDSKLSFAPNEQSGNIEKRLKLIKIVRYQENPTKNWGP KARPNTS	156
<i>Human TRM1</i>	MQGSSLWLSLTFRSARVLSRARFFEWQSPGLPNTAAMENGTPGYGEERPREVQETT VTGEGAAKIA FPSANEVFNYPVQEFNRDLTCAVITEFARIQLGAKGIQIKVPGEKDTQKVVDLSQEEEEKVELK ESENLAGDQPRTAAGVEICEEGLHVLEGLAASGLRSIRFALEVPLGRSVVANDASTRAVDLIRR NVQNDVAHLVQPSQADARMLMYQHQRVSERFVIDLDLPYGS PATFLDAAVQAVSEGGLLCVTCT DMAVLAGNSGETCYSKYGAMALKSPACHEMALRIVLHSLDLRANCYQRFVVPPLLSISADFYVRVF VRVFTGQAKVKASASKQALVFCVCGCAFHLQRLGKASGVP SGRAKFSAAACGFPVTPCECHCQR HQLGGPMWAEPIHDLDVFGRVLEAVSANPGRFHTSERIRGVLSVITEELPDVPLYTLDQLSSTI HCNTPSLQLRSALLHADFRVSLSHACKNAVKTADAPASALWDIMRCWECECPVKRRELSETSPAF RILSVEPRLQANFTIREDANPSSRQRLKRFQANPEANWGPRPRARP GGKAADEAMEERRRLQN KRKEPPEDVAQRAARLKTFPCKRKEGTCQRGDQCCYSHSPPTPRVSADAAPDCPETSNTQTPPGP GAAAGPGID	157
<i>E. coli RlmA</i>	MSFSCPLCHQPLSREKNSYICPQRHQFDMAKEGYVNLFPVQHKRSRDPGDSAEMMQARRAFLDAG HYQPLRDAIVAQLRERLDDKATAVLDIGCGEGYYTHAFADALPEITTFGLDVSKVAIKAAAKRYP QVTFCVASSHRLPFSDTSMDAIRIYAPCKAEELARVVKPGGWITATPGPRHLMELKGLIYNEV HLHAPHAEQLEGFTLQQSAELCYPMRLRGDEAVALLQMTPF AWRKPEVWQTLAAKEVDFCQTD NIHLWQRSY	158
<i>E. coli TrmD</i>	MWIGTISLFPEMFRAITDYGVTRAVKNGLLSIQSWSPRDFTHDRHRTVDDRYPYGGPGMLMMVQ PLRDAIHAAKAAAGGAKVIYLS PQGRKLDQAGVSELATNQKLILVCGRYEGIDERVIQTEIDEE WSIGDYVLSGGELPAMTLIDSVSRFIPGVLGHEASATEDSFAEGLDCPHYTRPEVLEGMVEPPV LLSGNHAEIRWRLLKQSLGRTWLRPELLENLALTEEQARLLAEFKTEHAQQQHKHDGMA	159
<i>Human TRMT10A</i>	MSSEMLPAFIETSNVDKKQGINEDQEEQKPRLGEGCEPISKRMKKLIKQKQWEEQRELKQKR KEKRKKKLERQCQMEPNSDGHRKRVRRDVHSTLRLIIDCSFDHMLVLKDIKKLHKQIQRCYA ENRRALHPVQFYLTSHGGQLKKNMDENDKGWVNWKDIHKEPEHYSELIKKEDLIYLTSDSPNLIK ELDESKAYVIGGLVDHNNHGLTYKQASDYGINHAQLPLGNFVKMNSRKVLAVNVHVFIEILEYLE TRDWEAEFTTILPQRKGAVPTDRACESASHDNQSVRMEEGGSDSDSSEEEYSRNELDSPHEEKQD KENHTESTVNSLPH	160
<i>M. Jannaschii Trm5b</i>	MPLCLKINKKHGEQTRRILNIENLLNKDYKITSEGNLYLPKIDVDEDILKSILNIEFELVDKEL EEKKIIKKPSFREIIISKRYKEIDEGLISLSYDVVGDVLVILQISDEVDEKIRKEIGELAYKLIP KGVFRKSEVKGEFRVRELEHLAGENRILTIIHKENGYRLWVDIAKYVFSRPLGGERARIMKKVSL NDVVDMFAGVGFPSIACKNAKKIY AIDINPHAIELLKKNIKLNKLEHKIIPILSDVREVDVGN RVIMNLPKFAHKFIDKALDIVEEGGVIHYTTIGKDFDKAIKLFKKKDCDEVLEKRIKVSYPREY ILALDFKINKK	161

Name	Sequence	SEQ ID NO:
<i>P. Abyssis Trm5a</i>	MTLAVKVP LKEGEIVRRRLIELGALDNTYKIKREGNLLIPVKFPVKGFVVEAELEQVSRPNS YREIVNVPQELRRFLPTSFDIIGNIAITEIPEELKGYAKEIGRAIVEVHKNVKAVYMGSKIEGE YRTRELIHIAGENITETIHTRENGIRLKLDAKVYFSPRLATERMRVFKMAQEGEVVDFMFAGVGP FSILLAKKAELVFACDINPWAIKYLEENIKLNKVNVPVILGDSREIEVKADRIIMNLPKYAHEF LEHAISCINDGGVIHYGFGPEGDPYGWHLERIRELANKFVKVEVLGKRVIRNYAPRQYNIAID FRVSF	162

In some embodiments, the fusion protein comprises one or more variant Cas12a endonuclease, one or more base editing enzymes, and one or more additional proteins (e.g., a protein, such as an enzyme, that regulates a biological activity). Non-limiting examples of additional protein elements include a polypeptide having uracil glycosylase inhibitor (UGI) activity, a polypeptide having uracil DNA glycosylase activity, a DNA binding domain (e.g., a Rad51 DNA binding domain), reverse transcriptase, endonuclease (e.g., FokI), a polypeptide having exonuclease activity (e.g., T5 exonuclease), a polypeptide having methyltransferase activity, a polypeptide having demethylase activity, a polypeptide having acetyltransferase activity, a polypeptide having deacetylase activity, a polypeptide having kinase activity, a polypeptide having phosphatase activity, a polypeptide having ubiquitin ligase activity, a polypeptide having deubiquitinating activity, a polypeptide having adenylation activity, a polypeptide having deadenylation activity, a polypeptide having SUMOylating activity, a polypeptide having deSUMOylating activity, a polypeptide having ribosylation activity, a polypeptide having deribosylation activity, a polypeptide having myristoylation activity, or a polypeptide having demyristoylation activity.

A fusion protein may comprise any one of the variant Cas12a endonucleases described herein (e.g., any one of SEQ ID NOs: 48-119 and 367-387, preferably any one of SEQ ID NOs: 367-387). In some embodiments, a fusion protein comprises any one of the base editing enzymes described herein.

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position R833 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is R833L, R833K, or R833M. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position R833 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position R833 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ

ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position K932 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is K932E or K932G. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K932 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K932 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position Q944 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is Q944K. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises a mutation at an amino acid position corresponding to position K940 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutation is K940G. In some embodiments, a variant LbCas12a endonuclease comprises a mutation at position K940 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant

LbCas12a endonuclease comprises a mutation at position K940 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

- 5 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions K932, N933, V936, and S929 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are K932G, N933G, V936G, and S929G. In some embodiments, a variant LbCas12a endonuclease comprises
- 10 mutations at positions K932, N933, V936, and S929 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions K932, N933, V936, and S929 with
- 15 reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

- In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions K940
- 20 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are K940G and Q944K. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions K940 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least
- 25 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions K940 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID
- 30 NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions R836 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are R836G and Q944K. In some

embodiments, a variant LbCas12a endonuclease comprises mutations at positions R836 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

5 In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R836 and Q944 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

10 In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions R833, E835, and Y943 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are R833M, E835D, and Y943T. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions
15 R833, E835, and Y943 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R833, E835, and Y943 with reference to amino acid
20 position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions R836, Q944, and R935 with reference to amino acid position numbering of LbCas12a ND2006
25 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are R836G, Q944K, and R935G. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R836, Q944, and R935 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%,
30 at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R836, Q944, and R935 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no

more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions R833, E835, Y943, and R935 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are R833M, E835D, Y943T, and R935G. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R833, E835, Y943, and R935 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R833, E835, Y943, and R935 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions R833, E835, Y943, and Q941 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are R833M, E835D, Y943T, and Q941K. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R833, E835, Y943, and Q941 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R833, E835, Y943, and Q941 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions R833, E835, and E125 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are R833M, E835D, and E125A. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R833, E835, and E125 with reference to amino acid position numbering of LbCas12a

ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R833, E835, and E125 with reference to amino acid position
5 numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions Y943,
10 Q944, K932, N933, and E125 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are Y943F, Q944K, K932G, N933G, and E125A. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions Y943, Q944, K932, N933, and E125 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1)
15 and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions Y943, Q944, K932, N933, and E125 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional
20 substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions R836, Q944, R935, and E125 with reference to amino acid position numbering of LbCas12a
25 ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are R836G, Q944K, R935G, and E125A. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R836, Q944, R935, and E125 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference
30 LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R836, Q944, R935, and E125 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions R833, E835, Y943, R935, and E125 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are R833M, E835D, Y943T, R935G, and E125A. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R833, E835, Y943, R935, and E125 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R833, E835, Y943, R935, and E125 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions R833, E835, Y943, Q941, and E125 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are R833M, E835D, Y943T, Q941K, and E125A. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R833, E835, Y943, Q941, and E125 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions R833, E835, Y943, Q941, and E125 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, the variant Cas12a endonuclease comprises a polypeptide sequence that comprises mutations at amino acid positions corresponding to positions D832, Y943, Q944, K932, N933, and E125 with reference to amino acid position numbering of LbCas12a ND2006 (e.g., SEQ ID NO: 1). In some embodiments, the mutations are D832A, Y943F, Q944K, K932G, N933G, and E125A. In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions D832, Y943, Q944, K932, N933, and E125

with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% identity to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1). In some embodiments, a variant LbCas12a endonuclease comprises mutations at positions

5 D832, Y943, Q944, K932, N933, and E125 with reference to amino acid position numbering of LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1) and has no more than 1, 2, 3, 4, or 5 additional substitutions relative to a wild-type reference LbCas12a ND2006 endonuclease (e.g., SEQ ID NO: 1).

In some embodiments, a fusion protein comprises a variant Cas12a endonuclease that

10 is located at or near the N-terminal end of the protein. In some embodiments, a fusion protein comprises a variant Cas12a endonuclease that is located at or near the C-terminal end of the protein. In some embodiments, a fusion protein comprises a base editing enzyme that is located at or near the N-terminal end of the protein. In some embodiments, a fusion protein comprises a base editing enzyme that is located at or near the C-terminal end of the protein.

15 In some embodiments, a fusion protein comprises an N-terminal variant Cas12a endonuclease and a C-terminal base editing enzyme (i.e., variant Cas12a endonuclease is located closer to the N-terminus of the protein than the base editing enzyme). In some embodiments, a fusion protein comprises an N-terminal base editing enzyme and a C-terminal variant Cas12a endonuclease (i.e., base editing enzyme is located closer to the N-terminus of the protein than

20 the variant Cas12a endonuclease).

A fusion protein may comprise one or more nuclear localization signals (NLSs). In some embodiments, a fusion protein comprises 1, 2, 3, 4, 5, or more NLSs. An NLS is an amino acid sequence that directs the fusion protein for import into the cell nucleus by nuclear transport. In some embodiments, an NLS is a positively charged amino acid sequence that

25 comprises several lysine and/or arginine amino acids. In some embodiments, a fusion protein comprises an NLS that is located at or near the N-terminal end of the protein. In some embodiments, a fusion protein comprises an NLS that is located at or near the C-terminal end of the protein. In some embodiments, a fusion protein comprises an NLS that is located at or near the N-terminal end of the protein and an NLS that is located at or near the C-terminal

30 end of the protein.

Non-limiting examples of an NLS include an SV40 NLS, a nucleoprotein (NP) NLS, and a bipartite (BP) NLS. In some embodiments, an SV40 NLS comprises the amino acid sequence of PKKKRKV (SEQ ID NO: 193). In some embodiments, a nucleoprotein NLS comprises the amino acid sequence of KRPAATKKAGQAKKKK (SEQ ID NO: 194). In

some embodiments, a bipartite NLS comprises the amino acid sequence of KRTADGSEFESPKKKRKV (SEQ ID NO: 195). In some embodiments, a fusion protein comprises an SV40 NLS that is located at or near the N-terminal end of the protein and/or an SV40 NLS that is located at or near the C-terminal end of the protein. In some embodiments, a fusion protein comprises an SV40 NLS that is located at or near the N-terminal end of the protein, an SV40 NLS that is located at or near the C-terminal end of the protein, and an NP NLS that is located at or near the C-terminal end of the protein. In some embodiments, a fusion protein comprises an NP NLS that is located at or near the N-terminal end of the protein, an NP NLS that is located at or near the C-terminal end of the protein, and an SV40 NLS that is located at or near the C-terminal end of the protein. In some embodiments, a fusion protein comprises a BP NLS that is located at or near the N-terminal end of the protein, an SV40 NLS that is located at or near the C-terminal end of the protein, and an NP NLS that is located at or near the C-terminal end of the protein. In some embodiments, a fusion protein comprises a BP NLS that is located at or near the N-terminal end of the protein and a BP NLS that is located at or near the C-terminal end of the protein. In some embodiments, a fusion protein comprises a BP NLS that is located at or near the N-terminal end of the protein and an NP NLS that is located at or near the C-terminal end of the protein.

A fusion protein may comprise one or more linkers. A linker for use in a fusion protein of the disclosure is generally an amino acid linker. In some embodiments, a linker functions to provide separation between different protein elements of the fusion protein (e.g., variant Cas12a endonuclease and base editing enzyme). In some embodiments, the presence of a linker between two protein elements of the fusion protein provides flexibility between the two elements of the fusion protein (e.g., to allow for each protein to fold and perform its function, e.g., enzymatic function). In some embodiments, a fusion protein comprises a linker between a variant Cas12a endonuclease and a base editing enzyme.

In some embodiments, a linker is a flexible linker. In some embodiments, a linker is a flexible linker comprising serine and/or glycine amino acids. In some embodiments, a linker is an amino acid sequence, wherein the majority of the amino acids of the linker are serine and/or glycine amino acids. In some embodiments, a linker comprises the amino acid sequence of (GS)_n (SEQ ID NO: 196), (GGS)_n (SEQ ID NO: 197), (GSS)_n (SEQ ID NO: 198), (GGSS)_n (SEQ ID NO: 199), (SGGGS)_n (SEQ ID NO: 200) or (SGGS)_n (SEQ ID NO: 201) wherein n is 1-10. In some embodiments, a linker comprises the amino acid sequence of GSSGGSGGSGSGS (SEQ ID NO: 202). In some embodiments, a linker comprises the amino acid sequence of SGSETPGTSESATPES (SEQ ID NO: 203). In other

embodiments, a linker comprises the amino acid sequence of SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 204). In some embodiments, a linker comprises the amino acid sequence of SGGSSGGSSGGS (SEQ ID NO: 205). In some embodiments, a linker comprises the amino acid sequence of GGGGGGS (SEQ ID NO: 206);

5 GSSGGSSGGSSGGS (SEQ ID NO: 207); or SGGS (SEQ ID NO: 208).

In various embodiments, the base editor fusion protein further comprises an inhibitor of base excision repair ("iBER") that covalently or non-covalently binds to a mutated nucleobase to prevent its excision during subsequent mismatch repair. Use of an iBER in the base editor fusion protein may increase base editing efficiency for the

10 deamination-oxidation and other strategies. In certain embodiments, the iBER is an 8-oxo-guanine glycosylase (OGG or OGGI) inhibitor ("OGG inhibitor"), a thymine-DNA glycosylase (TDG) inhibitor, a uracil-DNA glycosylase (UDG) inhibitor, a Methyl-CpG Binding Domain 4 (MBD4) inhibitor. In certain embodiments, the iBER comprises a catalytically inactive OGG that binds 8-oxo-inosine to prevent its excision during subsequent

15 mismatch repair.

A fusion protein may comprise one or more uracil glycosylase inhibitors (UGI). In some embodiments, a fusion protein comprises 1, 2, 3, 4, or 5 UGI polypeptides. In some embodiments, a UGI is a polypeptide that is capable of inhibiting a uracil-DNA glycosylase base-excision repair enzyme (e.g., from the uracil base excision repair (UBER) pathway). In

20 some embodiments, a UGI is capable of inhibiting repair machinery such that the UGI polypeptide increases the efficiency of a C to T conversion. In some embodiments, a UGI polypeptide reduces the rate of conversions that are not the C to T conversion. A UGI polypeptide may be a wild-type UGI polypeptide or a variant UGI polypeptide. In some embodiments, a UGI is a UGI from *Bacillus subtilis* bacteriophage PBS1. In some

25 embodiments, a UGI polypeptide comprises the amino acid sequence of
 TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVML
 LTSDAPEYKPWALVIQDSNGENKIKML (SEQ ID NO: 209). In some embodiments, a UGI polypeptide comprises at least 70%, 75%, 80%, 85%, 90%, 95%, or 97% identity to SEQ ID NO: 209.

30 A fusion protein may comprise one or more DNA glycosylases. In some embodiments, a fusion protein comprises 1, 2, 3, 4, or 5 DNA glycosylases. In some embodiments, a fusion protein comprises a uracil DNA glycosylase (UNG). In some embodiments, a fusion protein comprises a N-methyl purine glycosylase (MPG).

In some embodiments, an N-methyl purine glycosylase (MPG) functions in recognition and repair of base pairs comprising deoxyinosine. Hypoxanthine (the nucleobase of deoxyinosine) is recognized and excised by MPG, which results in the generation of an abasic site (AP site). The abasic site is then processed by the base excision repair pathway.

- 5 Accordingly, MPG is useful, in some embodiments, for A-to-C or A-to-T conversions, particularly when in combination with an adenosine deaminase (e.g., TadA deaminase).

In some embodiments, a fusion protein comprises an MPG and an adenosine deaminase (e.g., TadA deaminase). In such embodiments, the adenosine deaminase converts adenine to inosine; and the MPG subsequently removes the inosine to produce an abasic site.

- 10 The abasic site may subsequently be processed (e.g., to place a cytosine at that position). In some embodiments, a fusion protein comprises an MPG and a cytidine deaminase (e.g., APOBEC1 deaminase).

In some embodiments, an MPG polypeptide comprises the amino acid sequence of SEQ ID NO: 210. In some embodiments, an MPG polypeptide comprises at least 70%, 75%,
15 80%, 85%, 90%, 95%, or 97% identity to SEQ ID NO: 210.

- VTPALQMKKPKQFCRRMGQKKQRPARAGQPHSSSDAAQAPAEQPHSSSDAAQAPCPRERCLG
PPTTPGPYRSIYFSSPKGHLTRLGLEFFDQPAVPLARAFLGQVLVRRLPNGTELGRIVETE
AYLGPEDEAAHSRGGRTPRNRGMFMKPGTLYVYIIYGMYFCMNISSQGDGACVLLRALEPL
EGLETMRQLRSTLRKGTASRVLKDRELCSGPSKLCQALAINKSFDQRDLAQDEAVWLERGPL
20 EPSEPAVVAAARVGVGHAGEWARKPLRFYVRGSPWVSVDRAEQDTQA (SEQ ID NO: 210)

In some embodiments, a UNG is a polypeptide that is capable of recognition and excision of uracil from DNA strands. A UNG polypeptide is able to remove unwanted uracil bases from DNA molecules by cleaving the N-glycosidic bond and initiating the base-excision repair (BER) pathway. In some embodiments, a UNG is capable of increasing the efficiency of a C to G conversion. A UNG polypeptide may be a human UNG (hUNG) or an *Escherichia coli* UNG (eUNG). In some embodiments, a hUNG polypeptide is a mitochondrial UNG1 or the nuclear UNG2A. A UNG polypeptide may be a wild-type UNG polypeptide or a variant UNG polypeptide. In some embodiments, an UNG polypeptide comprises the amino acid sequence of ANELTWHDVLAEEKQQPYFLNTLQTVASERQS

- 30 GVTIYPPQKDVFNFRFTELGDVKVVILGQDPYHGPGQAHGLAFSVRPGIAIPPSLLNMYKE
LENTIPGFTRPNHGYLESWARQGVLLNTVLTVRAGQAHSHASLGWETFTDKVISLINQHRE
GVVFLWGSQAQKKGAIIDKQRHHVLKAPHPSPLSAHRGFFGCNHFVLANQWLEQRGETPID
WMPVLPASE (SEQ ID NO: 211). In some embodiments, a UNG polypeptide comprises the

amino acid sequence of IGQKTLYSFFSPSPARKRHAPSPEPA

VQGTGVAGVPPEESGDAAAIPAKKAPAGQEEPSTPPSSPLSAEQLDRIQRNKAALLRLAARN
VPVGFGESWKKHLSGEFGKPYFIKLMGFVAEERKHYTVYPPPHQVFTWTQMCDIKDVKVIL
GQDPYHGNQAHGLCFVQRPVPPPSLENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLLN
5 AVLTVRAHQANSHKERGWEQFTDAVVSQNSNGLVFLWGSYAQKKGSAIDRKRHHVLQT
AHPSPLSVYRGFFGCRHFSKTNELLQKSGKKPIDWKEL (SEQ ID NO: 212). In some
embodiments, an UNG polypeptide comprises at least 70%, 75%, 80%, 85%, 90%, 95%, or
97% identity to SEQ ID NO: 211 or 212.

A fusion protein may comprise one or more DNA binding domains (DBD). In some
10 embodiments, a fusion protein comprises 1, 2, 3, 4, or 5 DBDs. In some embodiments, a
DBD is a DBD that recognizes a sequence-specific single-stranded DNA molecule. In some
embodiments, a DBD is a DBD that recognizes a non-sequence-specific single-stranded DNA
molecule. In some embodiments, a DBD is a DBD that recognizes a sequence-specific
double-stranded DNA molecule. In some embodiments, a DBD is a DBD that recognizes a
15 non-sequence-specific double-stranded DNA molecule. In some embodiments, a DBD
comprises a Rad51 DNA binding domain (DBD). A DBD may be a wild-type DBD
polypeptide or a variant DBD polypeptide. In some embodiments, a DBD polypeptide
comprises the amino acid sequence of MAMQMQLANADTSVEEESFGPQPISRLEQ
CGINANDVKKLEEAGFHTVEAVAYAPKKELINIKGISEAKADKILAEAAKLVPMTTATEF
20 HQRRSEIIQITTGSKELDKLLQ (SEQ ID NO: 213). In some embodiments, a DBD
polypeptide comprises at least 70%, 75%, 80%, 85%, 90%, 95%, or 97% identity to SEQ ID
NO: 213.

A fusion protein can be designed to be specific for a certain type of enzymatic
conversion. For example, certain fusion proteins are designed for C to T conversion, A to G
25 conversion, or C to G conversion.

In some embodiments, a fusion protein comprising a variant Cas12a endonuclease, a
cytidine deaminase (e.g., APOBEC1), and a UGI is a C to T base editor (i.e., enzymatically
converts C to T). In some embodiments, a fusion protein comprising a variant Cas12a
endonuclease, a cytidine deaminase (e.g., APOBEC1), a UGI, and a Rad51 DBD is a C to T
30 base editor (i.e., enzymatically converts C to T). In some embodiments, a fusion protein
comprising a variant Cas12a endonuclease and a base editing enzyme (e.g., TadA) is an A to
G base editor (i.e., enzymatically converts A to G). In some embodiments, a fusion protein
comprising a variant Cas12a endonuclease, an adenosine deaminase (e.g., TadA), and a

Rad51 DBD is an A to G base editor (i.e., enzymatically converts A to G). In some embodiments, a fusion protein comprising a variant Cas12a endonuclease and a cytidine deaminase (e.g., APOBEC1) is a C to G base editor (i.e., enzymatically converts C to G). In some embodiments, a fusion protein comprising a variant Cas12a endonuclease, a cytidine deaminase (e.g., APOBEC1), and a UNG is a C to G base editor (i.e., enzymatically converts C to G).

Exemplary, non-limiting, fusion protein sequences are provided in **Table 5**.

Table 5. Non-limiting Examples of Fusion Proteins Comprising Variant Cas12a

Endonucleases and Base Editing Enzyme

Name*	Sequence	SEQ ID NO:
<i>LbBEv2</i> (<i>rAPOBEC1</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40 NLSs</i> , <i>NP NLS</i> , <i>BP NLS</i>) <i>C to T base editor</i>	MGSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLYEINWGRHSIWRHTSQNTNKHVEV NFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHTLFIYIARLYHADPRNRQ GLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPYPHPLWVRLYVLELYCIILGLPPCLN ILRRKQPLTFFTTIALQSCHYQRLPPLHILWATGLKSGGSSGGSSGSETPGTSESATPESGGSSG GSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLEVEDEKRAEDYKGVKKLLDRYYLSFIND VLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDIETILP EFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAPRCINENLTRYISNMDIFEKVD AIFDKHEVQEIKEKILNSDYDVEDFFEGEFTNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYI NLYNQTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKL PKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRKSF KKIGSFSLEQLQEYADADLSVVEKLKEIIQKVDEIYKVYGSSEKLFDAFVLEKSLKKNDAVVA IMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAIDILLKVDHIYDAIRNYVTQKPYSKD KFKLYFQNPQFMGGWDDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNVEKINYK LLPGPNKMLPKVFFSKKWMAYNPSIEDIQKIYKNGTFKKGDMFNLDCHKLIDFFKDSISRYPKW SNAYDNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVVEGKLYMFQIYNKDFSDKSH GTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVHPANSP IANKNPDNPKTTTL SYDVYKDKRFSQYELHIP IAINKCPKNIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVV DGKGNIVEQYSLNEIINNFGIRIKTDYHSLDKKEKERFEARQNWTSIENIKELKAGYISQVHH KICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKLNMYVDKKSNNPCATGALKGY QITNKFESFKMSMTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISFSFDRIMYVPE EDLEFALDYKNFRTDADYIKKWKLYSYGNRIRIFRNPKNVDFDWEVCLTSAYKELFNKYGI NYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLISPVKNSDGFIFYSRNVEA QENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAISNKEWLEYAQTSVKKRPAATK KAGQAKKKKSGSGSGSGSGSTNLSDIIEKETGKQLVQESILMLPEEVEEVIGNKPESDILVHT AYDESTDENVMLLTSDAPEYKFWALVIQDSNGENKIKMLSGGSPKKRRKV	163
<i>LbBEv3</i> (<i>rAPOBEC1</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40 NLSs</i> , <i>NP NLS</i> , <i>BP NLS</i>) <i>C to T base editor</i>	MPKKRKRKVGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLYEINWGRHSIWRHTSQN TNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHTLFIYIARLYHH ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPYPHPLWVRLYVLELYCIIL GLPPCLNILRRKQPLTFFTTIALQSCHYQRLPPLHILWATGLKSGGSSGGSSGSETPGTSESATPE SSGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLEVEDEKRAEDYKGVKKLLDRY YLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFKKD IETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAPRCINENLTRYISNMD DIFEKVD AIFDKHEVQEIKEKILNSDYDVEDFFEGEFTNFVLTQEGIDVYNAIIGGFVTESGEKI KGLNEYINLYNQTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSS IKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYE DDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIQKVDEIYKVYGSSEKLFDAFVLEKSLK KNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAIDILLKVDHIYDAIRNYVT QKPYSKDKFKLYFQNPQFMGGWDDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGN YEKINYKLLPGPNKMLPKVFFSKKWMAYNPSIEDIQKIYKNGTFKKGDMFNLDCHKLIDFFKDS ISRYPKWSNAYDNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVVEGKLYMFQIYNK DFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVHPANSP IANKNPDN PKTTTLSDYDVYKDKRFSQYELHIP IAINKCPKNIFKINTEVRVLLKHDDNPYVIGIARGERN LLYIVVVDGKGNIVEQYSLNEIINNFGIRIKTDYHSLDKKEKERFEARQNWTSIENIKELKAG YISQVHHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKLNMYVDKKSNNPCAT GALKGYQITNKFESFKMSMTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISFSF RIMYVPEEDLEFALDYKNFRTDADYIKKWKLYSYGNRIRIFRNPKNVDFDWEVCLTSAYKE LFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLISPVKNSDGFIFY DSRNVEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAISNKEWLEYAQTSVK KRPAATKAGQAKKKKSGSGSGSGSGSTNLSDIIEKETGKQLVQESILMLPEEVEEVIGNKPE SDILVHTAYDESTDENVMLLTSDAPEYKFWALVIQDSNGENKIKMLSGGSPKKRRKV	164

Name*	Sequence	SEQ ID NO:
<i>LbBEv4</i> (<i>rAPOBEC1</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40 NLSs</i> , <i>NP NLS</i> , <i>BP NLS</i>) <i>C to T base editor</i>	MKRPAATKKAGQAKKKKGSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLLYEINWGGRH SIWRHTSQNTNKHVEVNFIEKFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTILF IYIARLYHHADPRNRQGLRDLISSGVTTIQIMTEQESGYCWRNFVNYSNEAHWPYPHLLVRLY VLELYCIIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSGSGSGSGSETP GTSESATPESGGSGSGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYK GVKKLLDRYYLSFINDVLHSIKLKNLNYYISLFRKKTRTEKENKELENLEINLRKEIAKAFKNE GYKSLFKKDI IETILPEFLDDKDEIALVNSFNGFTTFTAGFFDNRENMFSEEAKSTSI AFR CINE NLT RYISNM DIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFFNFVLTQEGIDVYNAIIGG FVTESGEKIKGLNEYINLYNQTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVFNTL NKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKK KAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLEKII IQKVDEIYKVYGSSEKLFDA DFVLEKSLKKNDVAVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGYGFVLAYDILLKV VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYLPGPNKMLPKVFFSKKWMAYNPNSEDIQKIYKNGTFFKKGDMFN NDCHKLIDFFKDSISRYPKWSNAYDNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELV VHPANSP IANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKH DHPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEIINNPNNGIRIKTDYHSLDDKKEKERFEARQ NWTISIENIKELKAGYISQVHHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKL NYMVDKSNPCATGGALGKYQITNKFESFKSMSTQNGFIYIPAWLTSKIDPSTGFVNLLKTKYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNV FDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRD VDVFLISPVKNSDGI FYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAIS NKEWLEYAQTSVKKRPAATKKAGQAKKKKGS SGGSGSGSGSTNLSDI IEKETGKQLV IQESILMLPEEV EEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPFKKRRK V	165
<i>LbBEv5</i> (<i>rAPOBEC1</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40 NLSs</i> , <i>NP NLS</i> , <i>BP NLS</i>) <i>C to T base editor</i>	MGSKRTADGSEFESPKKKRKVGSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLLYEINW GGRHSIWRHTSQNTNKHVEVNFIEKFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPH VTILFIYIARLYHHADPRNRQGLRDLISSGVTTIQIMTEQESGYCWRNFVNYSNEAHWPYPHLLV RLYVLELYCIIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSGSGSGSG SETPGTSESATPESGGSGSGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNYYISLFRKKTRTEKENKELENLEINLRKEIAKAF KNEGYKSLFKKDI IETILPEFLDDKDEIALVNSFNGFTTFTAGFFDNRENMFSEEAKSTSI AFR CINE NLT RYISNM DIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFFNFVLTQEGIDVYNAIIGG FVTESGEKIKGLNEYINLYNQTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVFNTL NKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKK KAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLEKII IQKVDEIYKVYGSSEKLFDA DFVLEKSLKKNDVAVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGYGFVLAYDILLKV VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYLPGPNKMLPKVFFSKKWMAYNPNSEDIQKIYKNGTFFKKGDMFN NDCHKLIDFFKDSISRYPKWSNAYDNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELV VHPANSP IANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKH DHPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEIINNPNNGIRIKTDYHSLDDKKEKERFEARQ NWTISIENIKELKAGYISQVHHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKL NYMVDKSNPCATGGALGKYQITNKFESFKSMSTQNGFIYIPAWLTSKIDPSTGFVNLLKTKYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNV FDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRD VDVFLISPVKNSDGI FYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAIS NKEWLEYAQTSVKKRPAATKKAGQAKKKKGS SGGSGSGSGSTNLSDI IEKETGKQLV IQESILMLPEEV EEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPFKKRRK V	166
<i>LbBEv6</i> (<i>rAPOBEC1</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40 NLSs</i> , <i>NP NLS</i> , <i>BP NLS</i>) <i>C to T base editor</i>	MGSKRTADGSEFESPKKKRKVGSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLLYEINW GGRHSIWRHTSQNTNKHVEVNFIEKFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPH VTILFIYIARLYHHADPRNRQGLRDLISSGVTTIQIMTEQESGYCWRNFVNYSNEAHWPYPHLLV RLYVLELYCIIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSGSGSGSG SETPGTSESATPESGGSGSGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNYYISLFRKKTRTEKENKELENLEINLRKEIAKAF KNEGYKSLFKKDI IETILPEFLDDKDEIALVNSFNGFTTFTAGFFDNRENMFSEEAKSTSI AFR CINE NLT RYISNM DIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFFNFVLTQEGIDVYNAIIGG FVTESGEKIKGLNEYINLYNQTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVFNTL NKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKK KAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLEKII IQKVDEIYKVYGSSEKLFDA DFVLEKSLKKNDVAVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGYGFVLAYDILLKV VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYLPGPNKMLPKVFFSKKWMAYNPNSEDIQKIYKNGTFFKKGDMFN NDCHKLIDFFKDSISRYPKWSNAYDNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELV VHPANSP IANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKH DHPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEIINNPNNGIRIKTDYHSLDDKKEKERFEARQ NWTISIENIKELKAGYISQVHHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKL NYMVDKSNPCATGGALGKYQITNKFESFKSMSTQNGFIYIPAWLTSKIDPSTGFVNLLKTKYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNV FDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRD VDVFLISPVKNSDGI FYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAIS NKEWLEYAQTSVKKRPAATKKAGQAKKKKGS SGGSGSGSGSTNLSDI IEKETGKQLV IQESILMLPEEV EEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPFKKRRK V	167

Name*	Sequence	SEQ ID NO:
	LISPVKNSDGIIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAIS NKEWLEYAQTSVKKRPAATKKAGQAKKKKSGSGSGSGSGSTNLSDIEKETGKQLVQIESILML PEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGSGSG SGGSTNLSDIEKETGKQLVQIESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDA PEYKPWALVIQDSNGENKIKMLSGSGSPKKRRKV	
<i>LbBEv16</i> (<i>rAPOBEC1</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40 NLSs</i> , <i>NP</i> <i>NLS</i> , <i>BP NLS</i>) <i>C to T base editor</i>	MGSKRTADGSEFESPKKKRKVSGSGTNLSDIEKETGKQLVQIESILMLPEEVEEVIGNKPESDIL VHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGSGSGSGSGSGSGSETGPVAVD PTLRRRIEPHEFEVFFDPRRLRKETCLLYEINWGGRHISWRHTSQNTNKHVEVNFIEKFTTTERYF CPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTI QIMTEQESGYCWRNFVNYSPSNEAHWPYPHVLWVRLYVLELYCIIILGLPPCLNILRRKQPQLTFF TIALQSCHYQRLPPHILWATGLKSGSGSGSGSGSETPGTSESATPESSGGSGSGSGSKLEKFTNCTY SLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKLLDRYYLSFINVVLHSIKLNNLN YISLFRKKTREKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDIETILPEFLDDKDEIALV NSFNNGFTTAFGTGFNDRENMFSEEAKSTSIAPRCINENLTRYISNMDIFEKVDAIFDKHEVQEI EKILNSDYDVEDFFEGEFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQTKQKLP KFKPLYKQVLSDRRESLSFYGEGYTSDEEVLEVFRNTLNKNSIEFSSIKKLEKLFKNFDEYSSAGI FVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFLEQLQ EYADADLSVVEKLKEIIQKVDEIYKVYGSSEKLFDAFVLEKSLKKNDAVVAIMKDLDSVKFSF ENYIKAFFGEGKETNRDESFGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFM GGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKV FFSKWMAYNPSEDIQKIYKNGTFKKGDMFNLDCHKLIDFFKDSISRYPKWSNAYDNFSETE KYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKL LFDENNHGQIRLSGGAELFMRRASLKKEELVHPANSP IANKNPDNPKKTTTSLSYDVYKDKRFSE DQYELHIP IAINKCPKNIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVDGKGNIEQYSL NEIINNFGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVKICELVEKYDAV IAEDLNSGFKNSRVKVEKQVYQFKEKMLIDKLNVMVDKSNPCATGGALKGYQITNKFESEFKSM STQNGFIFYIPAWLTISKIDPSTGFVNLLKTKYTSIADSKKFISSEFDRIMYVPEEDLFEFALDYKN FSRTDADYIKKWKLYSYGNRIRIFRNPKNVDFDWEVCLTSAYKELFNKYGINYQGGDIRALLC EQSDKAFYSSFMALMSMLQMRNSITGRTDVDFLISPVKNSDGIIFYDSRNYEAQENAILPKNADA NGAYNIARKVLWAIQGFKADEKLDKVKIAISNKEWLEYAQTSVKKRPAATKKAGQAKKKKSGS GGSGSGSGSPKKRRKV	168
<i>LbBEv17</i> (<i>rAPOBEC1</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40 NLSs</i> , <i>NP</i> <i>NLS</i> , <i>BP NLS</i>) <i>C to T base editor</i>	MGSKRTADGSEFESPKKKRKVSGSETGPVAVDPTLRRRIEPHEFEVFFDPRRLRKETCLLYEINW GGRHSIWRHTSQNTNKHVEVNFIEKFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPH VTLFIYIARLYHHADPRNRQGLRDLISSGVTI QIMTEQESGYCWRNFVNYSPSNEAHWPYPHVLW VRLYVLELYCIIILGLPPCLNILRRKQPQLTFTTIALQSCHYQRLPPHILWATGLKSGSGSGSGS SETPGTSESATPESSGGSGSGSTNLSDIEKETGKQLVQIESILMLPEEVEEVIGNKPESDILVH TAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGSGSGSGSGSGSGSKLEKFTNCTY SLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKLLDRYYLSFINVVLHSIKLNNLN YISLFRKKTREKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDIETILPEFLDDKDEIALVNSF NGFTTAFGTGFNDRENMFSEEAKSTSIAPRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKI LNSDYDVEDFFEGEFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQTKQKLPKFK PLYKQVLSDRRESLSFYGEGYTSDEEVLEVFRNTLNKNSIEFSSIKKLEKLFKNFDEYSSAGIFVK NGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFLEQLQYEA DADLSVVEKLKEIIQKVDEIYKVYGSSEKLFDAFVLEKSLKKNDAVVAIMKDLDSVKFSFENY IKAFFGEGKETNRDESFGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGW DKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFS KKWMAYNPSEDIQKIYKNGTFKKGDMFNLDCHKLIDFFKDSISRYPKWSNAYDNFSETEKYK DIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFD ENNHGQIRLSGGAELFMRRASLKKEELVHPANSP IANKNPDNPKKTTTSLSYDVYKDKRFSE DQYELHIP IAINKCPKNIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVDGKGNIEQYSL NEIINNFGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVKICELVEKYDAVIAL EDLNSGFKNSRVKVEKQVYQFKEKMLIDKLNVMVDKSNPCATGGALKGYQITNKFESEFKSMSTQ NGFIFYIPAWLTISKIDPSTGFVNLLKTKYTSIADSKKFISSEFDRIMYVPEEDLFEFALDYKNFSR TDADYIKKWKLYSYGNRIRIFRNPKNVDFDWEVCLTSAYKELFNKYGINYQGGDIRALLCEQS DKAFYSSFMALMSMLQMRNSITGRTDVDFLISPVKNSDGIIFYDSRNYEAQENAILPKNADANGA YNIARKVLWAIQGFKADEKLDKVKIAISNKEWLEYAQTSVKKRPAATKKAGQAKKKKSGSGS GGSGSGSPKKRRKV	169
<i>LbBEv7</i> (<i>rAPOBEC1</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40 NLSs</i> , <i>NP</i> <i>NLS</i> , <i>BP NLS</i> , <i>Rad51 DBD</i>) <i>C to T base editor</i>	MGSKRTADGSEFESPKKKRKVSGSETGPVAVDPTLRRRIEPHEFEVFFDPRRLRKETCLLYEINW GGRHSIWRHTSQNTNKHVEVNFIEKFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPH VTLFIYIARLYHHADPRNRQGLRDLISSGVTI QIMTEQESGYCWRNFVNYSPSNEAHWPYPHVLW VRLYVLELYCIIILGLPPCLNILRRKQPQLTFTTIALQSCHYQRLPPHILWATGLKSGSGSGSGS SETPGTSESATPESSGGSGSGSMAMQMLEANADTSVEESFGQPQISRLEQCGINANVKKLEE AGFHTVEAVAYAPKKELINIKGISEAKADKILAEAAKLVPMGFTTATEFHRQRSEIIQITTSKE LDKLLQSGSGSGSGSGSETPGTSESATPESSGGSGSGSGSKLEKFTNCTYSLSKTLRFKAIPVGKTQ ENIDNKRLLVEDEKRAEDYKGVKLLDRYYLSFINVVLHSIKLNNLN YISLFRKKTREKENKE LENLEINLRKEIAKAFKGNEGYKSLFKKDIETILPEFLDDKDEIALVNSFNNGFTTAFGTGFNDNR ENMFSEEAKSTSIAPRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKI LNSDYDVEDFFEGE FFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQTKQKLPKFKPLYKQVLSDRRESLS FYGEGYTSDEEVLEVFRNTLNKNSIEFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFG EWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFLEQLQYEAADADLSVVEKLKEII IQKVDEIYKVYGSSEKLFDAFVLEKSLKKNDAVVAIMKDLDSVKFSFENYIKAFFGEGKETNRD ESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILR YGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKWMAYNPSEDIQ KIYKNGTFKKGDMFNLDCHKLIDFFKDSISRYPKWSNAYDNFSETEKYKDIAGFYREVEEQGY	170

Name*	Sequence	SEQ ID NO:
	KVSFESASKKEVDKLVVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHHGQIRLSGGAE LFMRASLKKEELVVHPANSP IANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKN IFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEI INNFNGIRIKTDYH SLLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKV EKQVYQKFEKMLIDKLNMYMDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSK IDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSY GNRIRIFRNPKNVDFWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSL MLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGF KKADEKLDKVKIAISNKEWLEYAQTSVKKRPAATKKAGQAKKKKGSSGGSGGSTNLSDII KETGKQLVQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQD SNGENKIKMLSGGSPKKKRKV	
<i>LbBEv8</i> (<i>rAPOBEC1</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40 NLSs</i> , <i>NP</i> <i>NLS</i> , <i>BP NLS</i> , <i>Rad51 DBD</i>) <i>C to T base editor</i>	MGSKRTADGSEFESPKKKRKVGSGSETGPVAVDPTLRRRIEPHEFEVFFDPRRLRKETCLLYEINW GGRHSIWRHTSQNTNKHVEVNFIEKFTTERRYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPH VTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPYRPHLW VRLYVLELYCIIILGLPPCLNILLRRKQPLTFTTIALQSCHYQRLPPHILWATGLKSGGSGSGSSG SETPGTSESATPESGGSGSGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKT RTEKENKELENLEINLRKEIAKAF KGNIEGKSLFKKDIETILPEFLDDKDEIALVNSFNFTTAF TGFFDNRENMFSEAKSTSI AFR CINENLTRYISNMDFEKFVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAI IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYTSDEEVLEVF RNTLNKNSIEFSSIKKLEKLFKNFDEYSSAGIFVKNPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRKSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKYVGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVFLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYLKLPKPNKMLPKVFFSKKWMAYNPNSED IQKIYKNGTFFKKGDMFN NDCHKLIDFFKDSISRYPKWSNAYDNFSETEKYKD IAGFYREVEEQGYKVSFESASKKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHHGQIRLSGGAE LFMRRASLKKEELVVH PANSP IANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEI INNFNGIRIKTDYHSLLDKKEKERFEARQ WTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKL NYMVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNVDF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFQKKADEKLDKVKIAIS NKEWLEYAQTSVKKRPAATKKAGQAKKKKGSSGGSGGSTNLSDII EKETGKQLVQESILML PEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPK KKRKVSGGSMAMQMQLANADTSVEEESFGQPISRLEQCGGINANDVKKLEAGFHTVEAVAYAP KKELINIKGISEAKADKILAEAAKLVPMGFTTATEFHQRSEIIQITTGSKELDKLLQ	171
<i>LbBEv9</i> (<i>rAPOBEC1</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40 NLSs</i> , <i>NP</i> <i>NLS</i> , <i>BP NLS</i> , <i>Rad51 DBD</i>) <i>C to T base editor</i>	MGSKRTADGSEFESPKKKRKVGSGAMQMQLANADTSVEEESFGQPISRLEQCGGINANDVKKLE EAGFHTVEAVAYAPKKELINIKGISEAKADKILAEAAKLVPMGFTTATEFHQRSEIIQITTGSK ELDKLLQSGGSGSGSGSETPGTSESATPESGGSGSGSGSETGPVAVDPTLRRRIEPHEFEVFF DPRRLRKETCLLYEINWGGGRHSIWRHTSQNTNKHVEVNFIEKFTTERRYFCPNTRCSITWFLSWSP CGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFV NYSPSNEAHWPYRPHLWRLYVLELYCIIILGLPPCLNILLRRKQPLTFTTIALQSCHYQRLPPH LWATGLKSGGSGSGSGSETPGTSESATPESGGSGSGSGSKLEKFTNCYSLSKTLRFKAIPVGKT QENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKT RTEKEN ELENLEINLRKEIAKAFKGNIEGKSLFKKDIETILPEFLDDKDEIALVNSFNFTTAF TGFFDN RENMFSEAKSTSI AFR CINENLTRYISNMDFEKFVDAIFDKHEVQEIKEKILNSDYDVEDFFEG EFFNFVLTQEGIDVYNAI IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRES LSFYGEGYTSDEEVLEVFRTNLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNPAISTISKDIF GEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRKSFKKIGSFSLEQLQEYADADLSVVEKLKEI IIQKVDEIYKYVGSSEKLFDAADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNR DESFGDFVFLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATIL RYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYLKLPKPNKMLPKVFFSKKWMAYNPNSED I QKIYKNGTFFKKGDMFNNDCHKLIDFFKDSISRYPKWSNAYDNFSETEKYKD IAGFYREVEEQG YKVSFESASKKEVDKLVVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHHGQIRLSGG ELFMRRASLKKEELVVHPANSP IANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPK NIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEI INNFNGIRIKTDY HSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVK VEKQVYQKFEKMLIDKLNMYMDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTS KIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYS YGNRIRIFRNPKNVDFWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMS LMLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGF FKKADEKLDKVKIAISNKEWLEYAQTSVKKRPAATKKAGQAKKKKGSSGGSGGSTNLSDII EKETGKQLVQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQ DSNGENKIKMLSGGSPKKKRKV	172
<i>LbBEv10</i> (<i>evoCDA</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40</i> <i>NLSs</i> , <i>NP NLS</i> , <i>BP</i> <i>NLS</i>)	MGSKRTADGSEFESPKKKRKVGSGSTDAEYVRIHEKLDIYTFKKQFSNNKKS VSHRCYVLFELKR RGERRACFWGAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPQGQTINWYSSWSPPCADCAEKILE WYNQELRNGHTLKIWVCKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQL NENRWLEKTLKRAEKRRSELSIMFQVKILHTTKSPAVSGSGSGSGSGSETPGTSESATPESGGSG SGGSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLSFI NDVLHSIKLKNLNNYISLFRKKT RTEKENKELENLEINLRKEIAKAFKGNIEGKSLFKKDIETI LPEFLDDKDEIALVNSFNFTTAF TGFFDNRENMFSEAKSTSI AFR CINENLTRYISNMDFEKF VDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAI IIGGFVTESGEKIKGLNE	173

Name*	Sequence	SEQ ID NO:
<i>C to T base editor</i>	YINLYNQKTKQKLPKFKPLYKQVLS DRESLSFYGEGYTSDEEVLEVFNRNTLNKNS EIFFSSIKKLE KLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRK SFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEKLFDAADFVLEKSLKKNDV VAIMKDLLDSVKS FENYIKAFFGEGKETNRDES FYGDFVLAYDILLKVDHIYDAIRNYVTQKPYS KDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKIN YKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLDCHKLIDFFKDSISRYP KWSNAYDFNFS ETEKYKD IAGFYREV EEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFS DK SHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVVHPANSP IANKNPDNPKKTT TLSYDVYKDKRFS EDOYELHIP IAINKCPKNIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIV VVDGKGNIVEQYSLNEIINNFGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELKAGYISQV VHKICELVEKYDAVIALEDLNSGFKNRVRKVEKQVYQKFEKMLIDKLYMVDKKS NPNCATGGALK GYQITNKFSFESKMS TQNGFI FYIPAWLT SKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYV PEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNNVDFWEEVCLTSAYKELFNKY GINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNY EAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAISNKEWLEYAQT SVKKRPAA TKKAGQAKKKKSGSGSGSGSGSTNLSDIIEKETGKQLV IQESILMLPEEVEEVIGNKPESDILV HTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKRRKV	
<i>LbBEv11</i> (<i>evoFERNY</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40</i> <i>NLSs</i> , <i>NP NLS</i> , <i>BP</i> <i>NLS</i>) <i>C to T base editor</i>	MGSKRTADGSEFESPKKKRKVGSGGSFERNYDPRELRKETYLLEYIKWKGSGKLWRHWCQNNRTQH AEVYFLENIFNARRFNPSTHCSITWYLSWSPCAECSQKIVDFLKEHPNVNLEYIVARLYYPENER NRQGLRDLVNSGV TIRIMLDPDYNICWKTFVSDQGGDEYWP GHFAPWIKQYSLKLSGGSSGGSS GSETPGTSESATPESSGSGSGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNRLLVEDEKR AEDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKA FKGNEGYKSLFKKDI IETILPEFLDDKDEIALVNSFNFTTAFTGFFDNRENMFSEEA KSTSI A RCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDDFEGEFNFVLTQEGIDVYN AIIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLS DRESLSFYGEGYTSDEEVLE VFRNTLNKNS EIFFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDD IHLKKKAVVTEKYEDDRKSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSE KLFDAADFVLEKSLKKNDVVAIMKDLLDSVKS FENYIKAFFGEGKETNRDES FYGDFVLAYDILL KVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYA KCLQKIDKDDVNGNYEKINYLKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFN LNDCHKLIDFFKDSISRYPKWSNAYDFNFS ETEKYKD IAGFYREV EEQGYKVSFESASKEVDKLV VEEGKLYMFQIYNKDFS DKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVV HPANSP IANKNPDNPKKTTTSLSYDVYKDKRFS EDOYELHIP IAINKCPKNIFKINTEVRVLLKH DNPYVIGIARGERNLLYIVVDGKGNIVEQYSLNEIINNFGIRIKTDYHSLLDKKEKERFEARQ NWTSIENIKELKAGYISQV VHKICELVEKYDAVIALEDLNSGFKNRVRKVEKQVYQKFEKMLIDK LNYMVDKKS NPNCATGGALKGYQITNKFSFESKMS TQNGFI FYIPAWLT SKIDPSTGFVNLLKTKY TSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNN VDFWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDV FLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAI SNKEWLEYAQT SVKKRPAA TKKAGQAKKKKSGSGSGSGSGSTNLSDIIEKETGKQLV IQESILM LPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSP KKRRKV	174
<i>LbBEv12</i> (<i>evoAPOBEC1</i> , <i>Linker</i> , <i>Cas12a</i> , <i>UGI</i> , <i>sv40 NLSs</i> , <i>NP</i> <i>NLS</i> , <i>BP NLS</i>) <i>C to T base editor</i>	MGSKRTADGSEFESPKKKRKVGSGSSKTPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEI NWGGRHSIWRHTSQNTKNHVEVNFIEKFTTERYFCNTRCSITWFLSWSPCGECSRAITEFLSRY PNVTLFYIYARLYHLANPRNRQGLRDLISSGV TIQIMTEQESGYCWHNFVNYSNESHWPYPH LWVRLYVLELYCIILGLPPCLNILRRKQSQLTSTFTIALQSCHYQRLPPHILWATGLKSGSGSGSS GSETPGTSESATPESSGSGSGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNRLLVEDEK RAEDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAK AFKNEGYKSLFKKDI IETILPEFLDDKDEIALVNSFNFTTAFTGFFDNRENMFSEEA KSTSI A RCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDDFEGEFNFVLTQEGIDVYN NAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLS DRESLSFYGEGYTSDEEVLE VFRNTLNKNS EIFFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYD DIHLKKKAVVTEKYEDDRKSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSE EKLFDADFVLEKSLKKNDVVAIMKDLLDSVKS FENYIKAFFGEGKETNRDES FYGDFVLAYDIL LKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKY AKCLQKIDKDDVNGNYEKINYLKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMF NLNDCHKLIDFFKDSISRYPKWSNAYDFNFS ETEKYKD IAGFYREV EEQGYKVSFESASKEVDK LVEEGKLYMFQIYNKDFS DKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELV VHPANSP IANKNPDNPKKTTTSLSYDVYKDKRFS EDOYELHIP IAINKCPKNIFKINTEVRVLLKH DDNPYVIGIARGERNLLYIVVDGKGNIVEQYSLNEIINNFGIRIKTDYHSLLDKKEKERFEAR QNWTSIENIKELKAGYISQV VHKICELVEKYDAVIALEDLNSGFKNRVRKVEKQVYQKFEKMLID KLYMVDKKS NPNCATGGALKGYQITNKFSFESKMS TQNGFI FYIPAWLT SKIDPSTGFVNLLKTK YTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNN VDFWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDV DFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAI SNKEWLEYAQT SVKKRPAA TKKAGQAKKKKSGSGSGSGSGSTNLSDIIEKETGKQLV IQESIL MLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGS PKRRKV	175
<i>LbBEv13</i> (<i>hAPOBEC3A</i> , <i>Linker</i> , <i>Cas12a</i> ,	MGSKRTADGSEFESPKKKRKVGSGEASPGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDN GTSVKMDQHRGFLHNQAKNLLCGFYGRHAELRFLDLPVSLQDPAQIYRVTWFIWSWPCFSWGCA GEVRAFLQENTHVRIRIFAARIYDYDPLYKEALQMLRDAGAQVSIIMTYDEFKHCWDFTFVDHQGCP FQPWDGLDEHSQALSGRLRAILQNGSGSGSGSGSGSETPGTSESATPESSGSGSGSGSSKLEK TNCYSLSKTLRFKAIPVGKTQENIDNRLLVEDEKRAEDYKGVKKLLDRYYLSFINDVLHSIKLK NLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKNEGYKSLFKKDI IETILPEFLDDKDE	176

Name*	Sequence	SEQ ID NO:
<i>UGI, sv40 NLSs, NP NLS, BP NLS)</i> <i>C to T base editor</i>	IALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSI AFR C INENLTRYISNMDFE KVD AIFDKHEV QEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESEGEKIKGLNEYINLYNQKTK QKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVFRTNLNKNSIEFSSIKKLEKLFKNFDEYS SAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRKRSFKKIGSFS EQLQEQYADADLSVVEKLEKIIQKVDEIYKVYGSSEKLFDAFVLEKSLKKNDAVVAIMKDLLDS VKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQN PQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYLKLLPGPNKM LPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLDCHKLIDFFKDSISRYPKWSNAYDFNF SETEKYKDIAFGYREVEEQGYKVSFESASKKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTM YFKLLFDENNHGQIRLSGGAE LFMRRASLKK EELV VHPANSP IANKNPDPNPKTTTSLSYDVYKDK RFSEdqYELHIP IAINKCPKNIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVDGKGNIVE QYSLNEIINNFGIRIKTDYHSLDDKKEKERFEARQNWTSIENIKELKAGYISQVVKICELVEK YDAVIALEDLNSGFKNRVRVVEKQVYQKFEKMLIDKLNMYMDKSNPCATGGALKGYQITNKFES FKSMSTQNGFIFYIPAWLT SKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALD DYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNNVFDWEEVCLTSAYKELFNKYGINYQQGD IRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILP NADANGAYNIARKVLAIGQFKKAEDEKLDKVKIAISNKEWLEYAQTSVKKRPAAATFKKGAQKKK KGS SGGSGGSGGSGTNSDIIEKETGKQLV IQESILMLPEEVEEVIGNKPESSDILVHTAYDESTDE NVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKRKV	
<i>LbABE8e</i> <i>(TadA, Linker, Cas12a, BP NLS)</i> <i>A to G base editor</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSGGSSG SETPGTSESATPESGGSGGSGSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNYYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSI AFR C INENLTRYISNMDFE KVD AIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAI IIGGFVTESEGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVF RNTLNKNSIEFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRKRSFKKIGSFSLEQLQEQYADADLSVVEKLEKIIQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYLKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFN LDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAFGYREVEEQGYKVSFESASKKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAE LFMRRASLKK EELV V HPANSP IANKNPDPNPKTTTSLSYDVYKDKRFSEdqYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEIINNFGIRIKTDYHSLDDKKEKERFEARQ NWTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNRVRVVEKQVYQKFEKMLIDK LNMYMDKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLT SKIDPSTGFVNLLKTKYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNNV FDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLAIGQFKKAEDEKLDKVKIAIS NKEWLEYAQTSVKSGGSKRTADGSEFEPKKRKV	177
<i>LbABE8ev11</i> <i>(TadA, Linker, Cas12a, BPNLS, Rad51 DBD)</i> <i>A to G base editor</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSGGSSG SETPGTSESATPESGGSGGSGSMAMQMQLANADTSVEEESFGPQISRLEQCGINANDVKKLEE AGFHTVEAVAYAPKKELINIKGISEAKADKILAEAAKLVPMGFTTATEFHRQRSEIQTITNKFES IDKLLQSGGSGGSGGSETPGTSESATPESGGSGGSGSKLEKFTNCYSLSKTLRFKAIPVGKTQ ENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLSFINDVLHSIKLKNLNYYISLFRKKTRTEKENKE LENLEINLRKEIAKAFKGNEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDN RENMFSEEAKSTSI AFR C INENLTRYISNMDFE KVD AIFDKHEVQEIKEKILNSDYDVEDFFEGE FFNFVLTQEGIDVYNAIIGGFVTESEGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLS FYGEGYTSDEEVLEVFRTNLNKNSIEFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFG EWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRKRSFKKIGSFSLEQLQEQYADADLSVVEKLEKII IQKVDEIYKVYGSSEKLFDAFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRD ESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILR YGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYLKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQ KIYKNGTFKKGDMFNLDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAFGYREVEEQGY KVSFESASKKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAE LFMRRASLKK EELV VHPANSP IANKNPDPNPKTTTSLSYDVYKDKRFSEdqYELHIP IAINKCPKN IFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEIINNFGIRIKTDYH SLDDKKEKERFEARQNWTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNRVRV EKQVYQKFEKMLIDKLNMYMDKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLT SK IDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYS YGNRIRIFRNPKNKNNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSL MLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLAIGQF KKADEKLDKVKIAISNKEWLEYAQTSVKSGGSKRTADGSEFEPKKRKVEFGSG	178
<i>LbABE8ev12</i> <i>(TadA, Linker, Cas12a, BP NLS, Rad51 DBD)</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPTAHAEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSGGSSG SETPGTSESATPESGGSGGSGSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNYYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSI AFR C INENLTRYISNMDFE KVD AIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAI	179

Name*	Sequence	SEQ ID NO:
<i>A to G base editor</i>	IIGGFVTESGEKIKGLNEYINLYNQTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVF RNTLNKNSIEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDVAVAIMKDLLDSVKSFENYKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPCKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEI INNFNGIRIKTDYHSLDKKEKERFEARQN WTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLT SKIDPSTGFVNLLKTKYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNV DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAI GQFKKADEKLDKVKIAIS NKEWLEYAQT SVKSGGSKRTADGSEFEPKKRKYSGGSMAMQMQLANADTSVEEESFQGPQPSR LEQCGINANDVKKLEEA GFHTVEAVAYAPKKELINIKGISEAKADKILAEAAKLVPMGFTTATEF HQRSEIIQITGSKELDKLLQ	180
<i>LbABE8ev13</i> <i>(TadA, Linker, Cas12a, BP NLS, Rad51 DBD)</i> <i>A to G base editor</i>	MKRTADGSEFESPPKKRKYSGGAMQMQLANADTSVEEESFQGPQPSRLEQCGINANDVKKLEEA GFHTVEAVAYAPKKELINIKGISEAKADKILAEAAKLVPMGFTTATEFHQRSEIIQITGSKEL DKLLQSGSSGGSSGSETPGTSESATPESGSSGSSGSEVEFSHEYWMRHALTAKRARDEREVP VGAVLVNNRVI GEGWNRAIGLHDP TAAHEIMALRQGGGLVMQNYRLIDATLVYTFEPCVMCAGAM IHSRIGRVVFGVRNSKRGAAGSLMNVNLYPGMNRHVEITEGILADECAALLCDFYRMPRQVFNAQ KKAQSSINSGSSGGSSGSETPGTSESATPESGSSGSSGSKLEKFTNCYLSKTLRFKAIPVGK TQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLSFINVLSHISIKLKNLNNYISLFRKKTRTEKEN KELENLEINLRKEIAKAFKGNIEGYSKLFKKDIETILPEFLDDKDEIALVNSFNNGFTTAFTEGFFD NRENMFSEEA KSTSI AFR CINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFE GEFFNFVLTQEGIDVYNAI IIGGFVTESGEKIKGLNEYINLYNQTKQKLPKFKPLYKQVLSDRS LSFYGEGYTSDEEVLEVFRTLNKNSIEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDI FGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKE II IQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDVAVAIMKDLLDSVKSFENYKAFFGEGKETN RDESFGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATI LRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSE DIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQ GYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGG AELFMRRASLKKEELVVHPANSPIANKNPDNPCKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCP KNIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEI INNFNGIRIKTD YHSLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRV KVEKQVYQKFEKMLIDKLNMYMVDKKSNNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLT SKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLY SYGNRIRIFRNPKNKNV DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALM SLMLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAI G QFKKADEKLDKVKIAISNKEWLEYAQT SVKSGGSKRTADGSEFEPKKRKY	181
<i>LbCGBEv0</i> <i>(rAPOBEC1, Linker, LbCas12a, sv40 NLSs, NP NLS, BP NLS)</i> <i>C to G base editor</i>	MGSRTADGSEFESPPKKRKYSGSSETGPVAVDPTLRRIIEPHEFEVFFDPRELKTECLLYEINW GGRHSIWRHTSQNTNKHVEVNFIEKFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPH VTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSNEAHWPRYPHLW VRLYVLELYCIIILGLPPCLNILRRKQPLTFTTIALQSCHYQRLPHILWATGLKSGSSGGSSG SETPGTSESATPESGSSGSSGSKLEKFTNCYLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINVLSHISIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYSKLFKKDIETILPEFLDDKDEIALVNSFNNGFTTAFTEGFFDNRENMFSEEA KSTSI AFR CINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVF RNTLNKNSIEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDVAVAIMKDLLDSVKSFENYKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPCKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEI INNFNGIRIKTDYHSLDKKEKERFEARQN WTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLT SKIDPSTGFVNLLKTKYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNV DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAI GQFKKADEKLDKVKIAIS NKEWLEYAQT SVKKRPAATKKAGQAKKKGSGSGSGSGSPKKRKY	

Name*	Sequence	SEQ ID NO:
LbCGBEv1 <i>(rAPOBEC1, Linker, LbCas12a, sv40 NLSs, NP NLS, BP NLS, eUNG)</i> <i>C to G base editor</i>	MGSKRTADGSEFESPKKKRKVGSSETGPVAVDPTLRRRIEPHEFEVFFDPRRLRKETCLLYEINW GGRHSIWRHTSQNTNKHVEVNFIEKFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPH VTLFIYIARLYHHADPRNRQGLRDLISSGVTTIQIMTEQESGYCWRNFVNYSPSNEAHWPYPHPLW VRLYVLELYCIIILGLPPCLNILLRRKQPQLTFTTIALQSCHYQRLPPHILWATGLKSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNYYISLFRKKTRTEKENKELEINLRKEIAKAF KGNEGYKSLFKKDI IETILPEFLDDKDEIALVNSFNGFTTFTAGFFDNRENMFSEEAKSTSI AFR CINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLQEQYADADLSVVEKLKEII IQKVDEIYKVGSSSEK LFDADFVLEKSLKKNDAVVAIMKDLDSVKSFENYIKAFFGEGKETNRDESFYGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMPNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKHT NPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEI INNFNGIRIKTDYHSLDKKEKERFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNV DWEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRDVF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKI AIS NKEWLEYAQTSVKKRPAATKKAGQAKKKKSGSGSGSGGSANELTWHVDVLAEEKQPPYFLNTLQ TVASERQSGVTIYPPQKDVFNARFTELGDVKVILGQDPYHGGPQAAGLAFSVRPGIAIPPSSLNMYKELENT IPGFTRPNGHYLESWARQGVLLNTVLTVRAGQAHSASLGWETFTDKVISLINQHREGVFLWGS HAQKKGAIIDKQRHHVLKAPHSPLSAHRGFFGCNHFVLANQWLEQORGETPIDWMPVLP AESESSESGSPKKKKRV	182
LbCGBEv2 <i>(rAPOBEC1, Linker, LbCas12a, sv40 NLSs, NP NLS, BP NLS, eUNG)</i> <i>C to G base editor</i>	MGSKRTADGSEFESPKKKRKVGSANELTWHVDVLAEEKQPPYFLNTLQTVASERQSGVTIYPPQK DVFNARFTELGDVKVILGQDPYHGGPQAAGLAFSVRPGIAIPPSSLNMYKELENTIPGFTRP NGHYLESWARQGVLLNTVLTVRAGQAHSASLGWETFTDKVISLINQHREGVFLWGS HAQKKGAIIDKQRHHVLKAPHSPLSAHRGFFGCNHFVLANQWLEQORGETPIDWMPVLP AESESSESGSGSGSETPGTSESATPESGGSSGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRRLRKETCLLYE INWGGRHSIWRHTSQNTNKHVEVNFIEKFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSR YPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTTIQIMTEQESGYCWRNFVNYSPSNEAHWPYP HLWVRLYVLELYCIIILGLPPCLNILLRRKQPQLTFTTIALQSCHYQRLPPHILWATGLKSGGSSGG SSGSETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDE KRAEDYKGVKKLLDRYYLSFINDVLHSIKLKNLNYYISLFRKKTRTEKENKELEINLRKEIA KAFKGNEGYKSLFKKDI IETILPEFLDDKDEIALVNSFNGFTTFTAGFFDNRENMFSEEAKSTSI AFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDV YNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEYTSDEEVLE VFRTNLKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEY DDIHLKKKAVVTEKYEDDRRSFKKIGSFSLQEQYADADLSVVEKLKEII IQKVDEIYKVGSSSEK SEKLFADADFVLEKSLKKNDAVVAIMKDLDSVKSFENYIKAFFGEGKETNRDESFYGDFVLAYD I LLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKK YAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGD MNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVD KLVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEEL VVHPANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLK HDDNPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEI INNFNGIRIKTDYHSLDKKEKERFEA RQNWTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLI DKLNYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLK TYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKN KNVFDWEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRD VDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKI AISNKEWLEYAQTSVKKRPAATKKAGQAKKKKSGSGSGSGSGSPKKKKRV	183

Name*	Sequence	SEQ ID NO:
LbCGBEv3 <i>(rAPOBEC1, Linker, LbCas12a, sv40 NLSs, NP NLS, BP NLS, hUNG)</i> <i>C to G base editor</i>	MGSKRTADGSEFESPKKKRKVGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLLYEINW GGRHSIWRHTSQNTNKHVEVNFIEKFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPH VTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPYPHPLW VRLYVLELYCIIILGLPPCLNILLRRKQPQLTFTTIALQSCHYQRLPPHILWATGLKSGGSSGGSSG SETPGTSESATPESGGSSGGSSGSKLEKFTNCYSLSKTLRFKAIPVGKQTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNYYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDI IETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSI AFR CINENLTRYISNMDIFEKVDAIFDKHEVQEIKEIKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQTKQKLPKFKPLYKQVLSDRSLSFYEGGYTSDEEVLEVF RNTLNKNSIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDVAVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVPH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIPAIANKCPKNIFKINTEVRVLLKHDD NPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEIINNFGIRIKTDYHSLDDKKEKERFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKL NYMVDKSNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWLYSYGNRIRFRNPKNNVF DWEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRDVF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKAEDEKLDKVKIAIS NKEWLEYAQTSVKKRPAATKKAGQAKKKKSGSSGGSSGGSSIGQKTLYSFFSPSPARKRHAPSPE PAVQGTGVAGVPEESGDAAPAKKAPAGQEEPPTPSSPLSAEQLDRIQRNKAALLRLAARNV PVGFGESWKKHLSGEFGKPYFIKLMGFVAEERKHYTVYPPPHQVFTWTQMDIKDKVKVILGQDP YHGPNAHQGLCFSVQRPVPPPPSLENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLNNAVLTVRA HQANSHKERGWQFTDAVVSWLNQNSGLVFLWGSYAQKKGSAIDRRKHHVLTQTAHPSPLSVYR GFFGCRHFSKTNELLQKSGKKPIDWKELSGGSPKKRKV	184
LbCGBEv4 <i>(rAPOBEC1, Linker, LbCas12a, sv40 NLSs, NP NLS, BP NLS, hUNG)</i> <i>C to G base editor</i>	MGSKRTADGSEFESPKKKRKVGSGIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGD AAAIPAKKAPAGQEEPPTPSSPLSAEQLDRIQRNKAALLRLAARNVVPVGFESWKKHLSGEFG KPYFIKLMGFVAEERKHYTVYPPPHQVFTWTQMDIKDKVKVILGQDPYHGPNAHQGLCFSVQRP VPPPPSLENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLNNAVLTVRAHQANSHKERGWQFTDA VVSWLNQNSGLVFLWGSYAQKKGSAIDRRKHHVLTQTAHPSPLSVYRGFFGCRHFSKTNELLQK SGKKPIDWKELSGGSSGGSSGSETPGTSESATPESGGSSGGSSGSETGPVAVDPTLRRRIEPHE FEVFFDPRELKRETCCLLYEINWGRHSIWRHTSQNTNKHVEVNFIEKFTTTERYFCPNTRCSITW LSWSPCGECSRAITEFLSRYPHTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYC WRNFVNYSNPSNEAHWPYPHPLWVRLYVLELYCIIILGLPPCLNILLRRKQPQLTFTTIALQSCHYQ LPPHILWATGLKSGGSSGGSSGSETPGTSESATPESGGSSGGSSGSKLEKFTNCYSLSKTLRFKAIP VGKQTQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLSFINDVLHSIKLKNLNYYISLFRKKTRT EKENKELENLEINLRKEIAKAFKGNIEGYKSLFKKDI IETILPEFLDDKDEIALVNSFNGFTTAFT GFFDNRENMFSEEAKSTSI AFR CINENLTRYISNMDIFEKVDAIFDKHEVQEIKEIKILNSDYDVE DFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQTKQKLPKFKPLYKQVLS DRESLSFYEGGYTSDEEVLEVFRTNLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISK DIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVE KLKEII IQKVDEIYKVYGSSEKLFDAFVLEKSLKKNDVAVAIMKDLLDSVKSFENYIKAFFGEG KETNRDESFGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFLYFQNPQFMGGWDKDKETDY RATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYY PSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYRE VEEQGYKVSFESASKEVDKLVVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIR LSGGAEFLMRRASLKKEELVPHANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIPAI NKCPKNIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEIINNFGIR IKTDYHSLDDKKEKERFEARQNWTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFK NSRVKVEKQVYQKFEKMLIDKLNMYMVDKSNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIP AWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKK WKLYSYGNRIRIFRNPKNVDFDWEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSF MALMSLMLQMRNSITGRDVF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVL WAIQGFKKAEDEKLDKVKIAISNKEWLEYAQTSVKKRPAATKKAGQAKKKKSGSSGGSSGGSSGPK KKRKV	185

Name*	Sequence	SEQ ID NO:
<i>TBN04 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPATAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEEAKSTSI AFR CINENLTRYISNMDFEKFVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDESLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNIEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFGSRVKVEKQVYQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQTSVKGSGSKRTADGSEFEPKKKRKV	388
<i>LbAA9 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPATAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEEAKSTSI AFR CINENLTRYISNMDFEKFVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDESLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNIEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFGKNSRVKVEKQVYQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQTSVKGSGSKRTADGSEFEPKKKRKV	389
<i>LbAA19 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPATAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEEAKSTSI AFR CINENLTRYISNMDFEKFVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDESLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNIEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFGKNSRVKVEKQVYQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQTSVKGSGSKRTADGSEFEPKKKRKV	390

Name*	Sequence	SEQ ID NO:
<i>LbEF1s9 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFFPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITTEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEAKSTSI AFR CINENLTRYISNMDFEKFVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDESLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLQEQYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNRSRVKVEKQVYQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNV DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQT SVKSGGSKRTADGSEFEPKKRKV	391
<i>LbAA23 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFFPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITTEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEAKSTSI AFR CINENLTRYISNMDFEKFVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDESLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLQEQYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFENRSRVKVEKQVYQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNV DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQT SVKSGGSKRTADGSEFEPKKRKV	392
<i>LbMS07 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFFPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITTEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEAKSTSI AFR CINENLTRYISNMDFEKFVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDESLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLQEQYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFENRSRVKVEKQVYQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNV DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQT SVKSGGSKRTADGSEFEPKKRKV	393

Name*	Sequence	SEQ ID NO:
<i>LbAA49 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITTEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTISIAFR CINENLTRYISNMDFIEKVDAlFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDESLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLQEQYADADLSVVEKLKEIIIQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKERFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNSRVKVEGQVYQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSEFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNV DWEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAIS NKEWLEYAQTSVKS GSGSKRTADGSEFEPKKRKV	394
<i>LbAC10 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITTEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTISIAFR CINENLTRYISNMDFIEKVDAlFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDESLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLQEQYADADLSVVEKLKEIIIQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKERFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSEFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNV DWEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAIS NKEWLEYAQTSVKS GSGSKRTADGSEFEPKKRKV	395
<i>LbMS3n5 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITTEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTISIAFR CINENLTRYISNMDFIEKVDAlFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDESLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLQEQYADADLSVVEKLKEIIIQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKERFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNGGFGGSRGKVEKQVYQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSEFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNV DWEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAIS NKEWLEYAQTSVKS GSGSKRTADGSEFEPKKRKV	396

Name*	Sequence	SEQ ID NO:
<i>LbTN37 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEAKSTISIAFR CINENLTRYISNMDFIEKVDIAFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDESLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLQEQYADADLSVVEKLKEIIQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNIEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKERFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNRSRVKVEGQVYKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSEFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQTSVKS GSGSKRTADGSEFEPKKRKV	397
<i>LbTN39 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEAKSTISIAFR CINENLTRYISNMDFIEKVDIAFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDESLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLQEQYADADLSVVEKLKEIIQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNIEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKERFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNRSRVKVEKQVYKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSEFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQTSVKS GSGSKRTADGSEFEPKKRKV	398
<i>LbTN2 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEAKSTISIAFR CINENLTRYISNMDFIEKVDIAFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDESLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLQEQYADADLSVVEKLKEIIQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNIEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDMGRNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKERFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNRSRVKVEKQVTFEKLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSEFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQTSVKS GSGSKRTADGSEFEPKKRKV	399

Name*	Sequence	SEQ ID NO:
<i>LbFM14 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFFPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITTEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEAKSTSI AFR CINENLTRYISNMDFEKFVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPLNHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGEGNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNSGVKVEKQVYKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQTSVKS GSGSKRTADGSEFEPKKRKV	400
<i>LbFM17 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFFPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITTEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEAKSTSI AFR CINENLTRYISNMDFEKFVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPLNHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDMGDRNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNSGVKVEKQVYKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQTSVKS GSGSKRTADGSEFEPKKRKV	401
<i>LbFM28 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFFPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITTEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEAKSTSI AFR CINENLTRYISNMDFEKFVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPLNHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDMGDRNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNSGVKVEKQVYKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQTSVKS GSGSKRTADGSEFEPKKRKV	402

Name*	Sequence	SEQ ID NO:
<i>LbFM44 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFFPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNLYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIATILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEAKSTSI AFR CINENLTRYISNMDFIEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNIEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDMGDRNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNRSRVKVEKQVYKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSEFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQT SVKSGGSKRTADGSEFEPKKKRKV	403
<i>LbFM51 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFFPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNLYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIATILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEAKSTSI AFR CINENLTRYISNMDFIEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNIEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFGGSRVKVEKQVYKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSEFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQT SVKSGGSKRTADGSEFEPKKKRKV	404
<i>LbFM64 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFFPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSIKLKNLNLYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIATILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEAKSTSI AFR CINENLTRYISNMDFIEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNIEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDRGEGNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNRSRVKVEKQVYKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSEFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQT SVKSGGSKRTADGSEFEPKKKRKV	405

Name*	Sequence	SEQ ID NO:
<i>LbFM65 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFFPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIATILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEEAKSTSI AFR CINENLTRYISNMDFIEKVD AIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNIEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDMGDRNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNKSGVKVEKQVTQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQTSVKGSGSKRTADGSEFEPKKRKV	406
<i>LbFM67 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFFPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIATILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEEAKSTSI AFR CINENLTRYISNMDFIEKVD AIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNIEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIDMGDRNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFKNKSRVKVEKQVTQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQTSVKGSGSKRTADGSEFEPKKRKV	407
<i>LbFM76 ABE8e</i>	MKRTADGSEFESPKKKRKVSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNR AIGLHDPHTAAHEIMALRQGGGLVMQNYRLIDATLYVTFFPCVMCAGAMIHSRIGRVVFGVRNSKRG AAGSLMNVNLNYPGMNHRVEITEGILADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSG SETPGTSESATPESGGSSGGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRA EDYKGVKKLLDRYYLSFINDVLHSLKLNKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAF KGNIEGYKSLFKKDIATILPEFLDDKDEIALVNSFNGFTTAF TGFFDNRENMFSEEAKSTSI AFR CINENLTRYISNMDFIEKVD AIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNA IIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVF RNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDI HLKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVYGSSEK LFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAK CLQKIDKDDVNGNIEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNL NDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKD IAGFYREVEEQGYKVSFESASKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAEFLMRRASLKKEELVVH PANSPIANKNPDNPKKTTTSLSYDVYKDKRFSEDQYELHIP IAINKCPKNIFKINTEVRVLLKHDD NPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEIINNENFNGIRIKTDYHSLDDKKEKFEARQN WTSIENIKELKAGYISQVVKICELVEKYDAVIALEDLNSGFGRVKVEKQVTQKFEKMLIDKL NYMVDKKSNNPCATGGALKGYQITNKFESEFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLTKTYT SIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKNKNVF DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDF LISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIQGFKKADEKLDKVKIAIS NKEWLEYAQTSVKGSGSKRTADGSEFEPKKRKV	408

*Protein Elements ordered from N-terminus to C-terminus

In some embodiments, a fusion protein comprises the amino acid sequence of any one of SEQ ID NO: 163-185. In some embodiments, a fusion protein comprises an amino acid sequence that includes any one or more mutation(s) (*e.g.*, amino acid substitution(s)) described herein and has at least 80% (*e.g.*, at least 80%, at least 85%, at least 90%, at least 95%, or at least 98%) identity to the amino acid sequence of any one of the fusion protein in Table 5 (*e.g.*, SEQ ID NOs: 163-185).

VIII. Cells

Aspects of the present disclosure relate to cells comprising any more of the variant Cas12a endonucleases described herein. Further, as described below, the variant Cas12a endonucleases described herein may be used to modify cells.

The cells may be eukaryotic cells or prokaryotic cells. Non-limiting examples of eukaryotic cells include animal cells, plant cells, and fungal cells. In some embodiments, the cells are mammalian cells. In some embodiments, the cells are human cells (*e.g.*, human primary cells or human immortalized cells). In some embodiments, the cells are stem cells, such as adult stem cells or induced pluripotent stem cells (iPSCs). Non-human cells are also provided herein. For example, the cells may be selected from non-human primate cells, porcine cells, bovine cells, canine cells, feline cells, or rodent cells (*e.g.*, rat or mouse cells).

Various human cell types are contemplated herein, including without limitation, immune cells (*e.g.*, T cells (*e.g.*, NKT cells, CD4+ T cells, CD8+ T cells, regulatory T cells, engineered T cells, *e.g.*, CAR-T, TCR)), B cells, NK cells, tumor-infiltrating lymphocytes, etc.), neural cells, cardiovascular cells, epidermal cells, and metabolic cells. The cells may be cancerous or non-cancerous. In some embodiments, the cells are tumor cells. Other cell types are contemplated herein.

25

IX. Methods of Use

Cas12a endonucleases provided herein have numerous uses, many of which are known in the art. The variant Cas12a endonucleases of the present disclosure may be used, in some instances, to improve those uses. Several non-limiting examples of such uses include genome editing, bioengineering, diagnostics, and agricultural advancement.

30

A. Genome Editing

In some embodiments, the variant Cas12a endonuclease are used for genome editing. Genome editing is a type of genetic engineering where a DNA is inserted, deleted or replaced

in the genome of a living organism. Application of CRISPR-Cas systems as molecular tools for genome editing exploits their ability to produce a double strand break (DSB) at a specific genomic locus and depends entirely on the host cell DNA repair machinery to fix the lesion produced by these systems. The repair mechanisms can be either of the following processes:

5 homology-directed repair (HDR) or non-homologous end joining (NHEJ). HDR utilizes a template DNA that is homologous to the break site (an unbroken sister chromatid or a homologous chromosome) to repair the DSB, whereas NHEJ is based on direct joining of broken ends of the DSB, making NHEJ the more error prone mechanism of the two. HDR can thus be used to supply exogenous template DNA to implement a user defined change in

10 the host genome. NHEJ can be applied for gene disruption whereas HDR allows for the scope of introducing new genetic information or direct correction of the sequence at a specific locus.

At the center of CRISPR mediated genome engineering today is Cas9, with applications including, but not limited to, gene knockout and precise genome editing. Despite

15 the rapid advances in genome editing by Cas9, it still presents challenges owing to the possibility of off-target effects and difficulty of delivering the ribonucleoprotein particle. Cas12a, owing to its substantial differences with Cas9, presents an alternate molecular genome editing tool. The use of Cas12a in genome editing for various cell types has been probed in several studies up to date. Comparative studies of gene repression by catalytically

20 dead Cas9 from *S. pyogenes* (SpdCas9) and catalytically dead Cas12a from *Eubacterium eligens* (EedCas12a) revealed that the latter displays a higher gene repression in the template strand of the target DNA than SpdCas9. It was also shown that the pre-crRNA processing activity of Cas12a makes it an attractive candidate for multiplex gene regulation, which is cumbersome when attempted with Cas9. This auto-processing of its own crRNA has been

25 used to modify multiple genetic elements simultaneously generating constitutive, conditional, inducible, orthogonal and multiplexed genome engineering of endogenous targets using multiple CRISPR RNAs delivered on a single plasmid.

The viability of this approach has been further established by other studies, in which multiplex gene regulation by Cas12a was successfully observed in bacteria, plants, as well as

30 in mammalian cells. Cas12a can also serve as a solution in cell types where use of Cas9 is toxic, such as in some industrial strains of *Streptomyces*.

Targeted mutagenesis in plants can also be achieved through co-expression of Cas12a and its cognate crRNA *in vivo*, as was shown in rice. Additionally, it was also shown that the mutagenesis was more efficient through the use of pre-crRNAs with full-length direct repeat

sequences than with mature crRNAs. Efficient mutagenesis through delivery of the pre-assembled ribonucleoprotein (RNP) particle was also observed in soybean and wild tobacco. The RNP was assembled from recombinantly expressed Cas12a and *in vitro* transcribed or chemically synthesized crRNAs.

- 5 Successful gene editing of mammalian cells using Cas12a include correction of mutations causing Duchenne muscular dystrophy (DMD) in patient derived induced pluripotent stem cells (iPSCs) and in *mdx* mice, a popular model for studying DMD. Dystrophin expression was reinstated in iPSCs after Cas12a-mediated gene editing, while in the *mdx* mice, corrections in the pathophysiological hallmarks of muscular dystrophy were
- 10 observed. Delivery of the adenovirus vector with an AsCas12a expression cassette yielded successful mutations in primary human hepatocytes from humanized mice with chimeric liver. Cas12a-mediated genome editing was also used to engineer rat models that mimic human atherosclerosis and this system may have potential applications in understanding early stage atherosclerosis.
- 15 *See Paul, B. & Montoya, G. et al. Biomedical Journal 2020; 43(1): 8-17, incorporated herein in its entirety.*

B. Bioengineering

- In some embodiments, the variant Cas12a endonuclease are used for bioengineering.
- 20 Currently, a vast effort is ongoing to redesign all these tools for biomedical and biotechnological applications. However, recent studies have envisioned the possibility of using CRISPR-Cas nucleases in bioengineering of smart materials, for example hydrogels. These water-filled polymers are encapsulated by DNA. Cas12a has been used to specifically degrade the DNA scaffold of DNA hydrogels, thus opening the possibility that this smart
- 25 cutter can be turned into a programmable device to deliver the cargo of DNA encaged hydrogels in a determined location at a certain time. The cleavage properties of Cas12a make it an ideal candidate to promote controlled delivery of the cargo. *See Paul, B. & Montoya, G. et al. Biomedical Journal 2020; 43(1): 8-17, incorporated herein in its entirety.*

30 C. Nucleic Acid Detection and Quantification (e.g., Diagnostics)

 In some embodiments, the variant Cas12a endonucleases are used for detecting and/or quantifying nucleic acids. For example, the variants may be used as *in vitro* diagnostic tools for pathogenic (e.g., bacterial or viral) nucleic acids, or for identification of biomarkers indicative of disease, such as cancer (e.g., for identification and quantification of single CpG

methylation sites – see, for example, van Dongen, JE *et al. Biosensors and Bioelectronics* 2021; 194(15): 113624).

In some embodiments, the variant Cas12a endonucleases are used with the Specific Enhancer for Detection of PCR-amplified Nucleic Acids (SENA) method, which combines
5 the transcleavage activity of Cas12a with the sensitivity offered by real-time PCR. See, e.g., Huang W, *et al. EBioMedicine* 2020; 61: 103036.

In some embodiments, the variant Cas12a endonucleases are used with the DNA Endonuclease-Targeted CRISPR Trans Reporter (DETECTR) technology, which performs simultaneous reverse transcription and isothermal amplification using loop-mediated
10 amplification (RT–LAMP) for RNA extracted from a biological sample. See, e.g., Broughton JP, *et al. Nature Biotechnology* 2020; 38: 870-874.

In some embodiments, the variant Cas12a endonucleases are used with the one-hour low-cost multipurpose highly efficient system (HOLMES) technology. See, e.g., Li, L. *et al. ACS Synth Biol.* 2019; 8(10): 2228-2237.

15 Other nucleic acid detection methods are contemplated herein.

D. Agriculture

In some embodiments, the variant Cas12a endonuclease are used for applications relating to agricultural advancement. Cas12a editing has been widely utilized in many crops
20 including rice, wheat, maize, soybean, cotton, tomato, citrus, tobacco, and the model plant Arabidopsis. At present, three Cas12a genome editing systems AsCas12a, FnCas12a, and LbCas12a have been demonstrated in plants with varied efficiency.

Rice is one of the most well-studied crops due to its agricultural importance, small genome size, ease of transformation and available genetic resources making it an ideal
25 flagship genome for the grasses. These factors have also made it an ideal testing ground for developing genome editing technologies. Codon optimized FnCas12a binary vectors were utilized for targeted mutagenesis in rice (OsDL, OsALS, OsNCED1, OsAO1) and tobacco (NtPDS and NtSTF1) with average targeted mutation frequencies of 47.2% and 28.2%, respectively. Utilizing the LbCas12a nuclease two endogenous rice genes OsPDS and OsBEL
30 were targeted with mutation frequencies of 21.4 and 41.2%, respectively. An independent study that targeted the disruption of OsPDS by LbCas12a resulted in a similarly high editing frequency of 32.3%. It was also demonstrated that pre-crRNAs were more efficient in generating mutants than mature crRNAs in rice. However, the opposite was observed in HEK293T cells. In addition to these proof-of-concept experiments, LbCas12a was also used

to create loss-of-function alleles of OsEPFL9 which regulates stomatal density. These lines increased water use efficiency eight-fold in T2 generation plants. *See, Bandyopadhyay, A. et al. Front. Plant Sci. 2020.*

5 Additional Embodiments

Additional embodiments are described in the following numbered paragraphs:

1. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising a mutation at an amino acid position corresponding to position E95, E125, V245, N260, Y277, R747, H759, I765, F810, N813, T814, I831, T870, G902, K960, S982, K984, or
10 T988 with reference to amino acid position numbering of LbCas12a ND2006.
2. The engineered variant Cas12a endonuclease of paragraph 1, wherein the variant Cas12a endonuclease exhibits hyperactivity, low single indiscriminate strand deoxyribonuclease (DNase) activity, target nickase activity (or a preference for cleaving one strand over the other of a dsDNA), or protospacer adjacent motif (PAM) nickase activity.
- 15 3. The engineered variant Cas12a endonuclease of paragraph 1 or 2, wherein the mutation is an amino acid substitution.
4. The engineered variant Cas12a endonuclease of paragraph 3, wherein the amino acid substitution is an amino acid having an equivalent charge, polarity, and/or chemical class.
5. The engineered variant Cas12a endonuclease of any one of paragraphs 1-4, wherein
20 (a) the polypeptide sequence comprises a mutation at an amino acid position corresponding to position E95, E125, V245, Y277, R747, H759, I765, F810, or T814 with reference to amino acid position numbering of LbCas12a ND2006 and (b) the variant Cas12a endonuclease exhibits hyperactivity.
6. The engineered variant Cas12a endonuclease of paragraph 5, wherein the polypeptide
25 sequence comprises a mutation at an amino acid position corresponding to position E95 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
7. The engineered variant Cas12a endonuclease of paragraph 6, wherein the mutation is
30 a substitution of a polar, positively charged, and/or basic amino acid at position E95 with reference to amino acid position numbering of LbCas12a ND2006, preferably E95R or E95H, more preferably E95R.

8. The engineered variant Cas12a endonuclease of paragraph 6, wherein the mutation is a substitution of a polar, uncharged, and/or aromatic amino acid at position E95 with reference to amino acid position numbering of LbCas12a ND2006, preferably E95Y.
9. The engineered variant Cas12a endonuclease of paragraph 5, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position E125 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
10. The engineered variant Cas12a endonuclease of paragraph 9, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position E125 with reference to amino acid position numbering of LbCas12a ND2006, preferably E125A, E125G, E125I, E125L, E125P, E125V, more preferably E125A.
14. The engineered variant Cas12a endonuclease of paragraph 5, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position V245 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
15. The engineered variant Cas12a endonuclease of paragraph 14, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position V245 with reference to amino acid position numbering of LbCas12a ND2006, preferably V245I, V245A, V245G, V245L, or V245P, more preferably V245I.
16. The engineered variant Cas12a endonuclease of paragraph 14, wherein the mutation is a substitution of a polar, uncharged, and/or aromatic amino acid at position V245 with reference to amino acid position numbering of LbCas12a ND2006, preferably V245Y.
21. The engineered variant Cas12a endonuclease of paragraph 5, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position Y277 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
22. The engineered variant Cas12a endonuclease of paragraph 21, wherein the mutation is a substitution of a polar, uncharged, and/or hydroxyl amino acid at position Y277 with reference to amino acid position numbering of LbCas12a ND2006, preferably Y277S or Y277T, more preferably Y277S.

29. The engineered variant Cas12a endonuclease of paragraph 5, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position R747 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
30. The engineered variant Cas12a endonuclease of paragraph 29, wherein the mutation is a substitution of a polar, uncharged, and/or aromatic amino acid at position R747 with reference to amino acid position numbering of LbCas12a ND2006, preferably R747Y.
31. The engineered variant Cas12a endonuclease of paragraph 5, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position H759 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
32. The engineered variant Cas12a endonuclease of paragraph 31, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position H759 with reference to amino acid position numbering of LbCas12a ND2006, preferably H759V, H759A, H759G, H759I, H759L, or H759P, more preferably H759V.
33. The engineered variant Cas12a endonuclease of paragraph 31, wherein the mutation is a substitution of a polar, negatively charged, and/or acidic amino acid at position H759 with reference to amino acid position numbering of LbCas12a ND2006, preferably H759D.
34. The engineered variant Cas12a endonuclease of paragraph 5, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position I765 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
35. The engineered variant Cas12a endonuclease of paragraph 34, wherein the mutation is a substitution of a polar, uncharged, and/or aromatic amino acid at position I765 with reference to amino acid position numbering of LbCas12a ND2006, preferably I765Y.
36. The engineered variant Cas12a endonuclease of paragraph 5, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position F810 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

37. The engineered variant Cas12a endonuclease of paragraph 36, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position F810 with reference to amino acid position numbering of LbCas12a ND2006, preferably F810W.
38. The engineered variant Cas12a endonuclease of paragraph 36, wherein the mutation is
5 a substitution of a polar, charged, and/or acidic amino acid at position F810 with reference to amino acid position numbering of LbCas12a ND2006, preferably F810Q.
39. The engineered variant Cas12a endonuclease of paragraph 5, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position T814 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein
10 the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
40. The engineered variant Cas12a endonuclease of paragraph 39, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position T814 with reference to amino acid position numbering of LbCas12a ND2006, preferably T814E.
41. The engineered variant Cas12a endonuclease of paragraph 39, wherein the mutation is
15 a substitution of a polar, positively charged, and/or basic amino acid at position T814 with reference to amino acid position numbering of LbCas12a ND2006, preferably T814R, T814H, or T814K, more preferably T814R.
42. The engineered variant Cas12a endonuclease of any one of paragraphs 1-4, wherein
20 (a) the polypeptide sequence comprises a mutation at an amino acid position corresponding to position N813, I831, T870, G902, S982, K984, or T988 with reference to amino acid position numbering of LbCas12a ND2006 and (b) the variant Cas12a endonuclease exhibits low indiscriminate single strand deoxyribonuclease (ssDNase) activity.
43. The engineered variant Cas12a endonuclease of paragraph 42, wherein the
25 polypeptide sequence comprises a mutation at an amino acid position corresponding to position N813 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
44. The engineered variant Cas12a endonuclease of paragraph 43, wherein the mutation is
30 a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position N813 with reference to amino acid position numbering of LbCas12a ND2006, preferably N813W.
45. The engineered variant Cas12a endonuclease of paragraph 43, wherein the mutation is a substitution of a polar, positively charged, and/or basic amino acid at position N813 with

reference to amino acid position numbering of LbCas12a ND2006, preferably N813R, N813H, or N813K, more preferably N813R or N813H.

46. The engineered variant Cas12a endonuclease of paragraph 42, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position I831 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

47. The engineered variant Cas12a endonuclease of paragraph 46, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position I831 with reference to amino acid position numbering of LbCas12a ND2006, preferably I831A, I831G, I831L, I831P, or I831V, more preferably I831A.

48. The engineered variant Cas12a endonuclease of paragraph 46, wherein the mutation is a substitution of a polar, uncharged, and/or aromatic amino acid at position I831 with reference to amino acid position numbering of LbCas12a ND2006, preferably I831Y.

49. The engineered variant Cas12a endonuclease of paragraph 42, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position T870 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

50. The engineered variant Cas12a endonuclease of paragraph 47, wherein the mutation is a substitution of a polar, uncharged, and/or aromatic amino acid at position T870 with reference to amino acid position numbering of LbCas12a ND2006, preferably T870Y.

51. The engineered variant Cas12a endonuclease of paragraph 47, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position T870 with reference to amino acid position numbering of LbCas12a ND2006, preferably T870F or T870W, more preferably T870F.

52. The engineered variant Cas12a endonuclease of paragraph 42, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position G902 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

53. The engineered variant Cas12a endonuclease of paragraph 52, wherein the mutation is a substitution of a polar, positively charged, and/or basic amino acid at position G902 with reference to amino acid position numbering of LbCas12a ND2006, preferably G902R, G902H, or G902K, more preferably G902R.
- 5 54. The engineered variant Cas12a endonuclease of paragraph 52, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position G902 with reference to amino acid position numbering of LbCas12a ND2006, preferably G902W or G902F, more preferably G902W.
55. The engineered variant Cas12a endonuclease of paragraph 42, wherein the
10 polypeptide sequence comprises a mutation at an amino acid position corresponding to position S982 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
- 15 56. The engineered variant Cas12a endonuclease of paragraph 55, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position S982 with reference to amino acid position numbering of LbCas12a ND2006, preferably S982F or S982W, more preferably S982W.
57. The engineered variant Cas12a endonuclease of paragraph 42, wherein the
20 polypeptide sequence comprises a mutation at an amino acid position corresponding to position K984 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
- 25 58. The engineered variant Cas12a endonuclease of paragraph 57, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position K984 with reference to amino acid position numbering of LbCas12a ND2006, preferably K984F or K984W, more preferably K984F.
59. The engineered variant Cas12a endonuclease of paragraph 57, wherein the mutation is
30 a substitution of a polar, positively charged, and/or basic amino acid at position K984 with reference to amino acid position numbering of LbCas12a ND2006, preferably K984R, K984H, or K984K, more preferably K984R.
60. The engineered variant Cas12a endonuclease of paragraph 42, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to

position T988 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

5 61. The engineered variant Cas12a endonuclease of paragraph 60, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position T988 with reference to amino acid position numbering of LbCas12a ND2006, preferably T988F or T988W, more preferably T988F.

62. The engineered variant Cas12a endonuclease of any one of paragraphs 1-4, wherein
10 (a) the polypeptide sequence comprises a mutation at an amino acid position corresponding to position N260 or G902 with reference to amino acid position numbering of LbCas12a ND2006 and (b) the variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).

63. The engineered variant Cas12a endonuclease of paragraph 62, wherein the
15 polypeptide sequence comprises a mutation at an amino acid position corresponding to position N260 with reference to amino acid position numbering of LbCas12a, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

64. The engineered variant Cas12a endonuclease of paragraph 63, wherein the mutation is
20 a substitution of a polar, positively charged, and/or basic amino acid at position N260 with reference to amino acid position numbering of LbCas12a ND2006, preferably N260R, N260H, or N260K, more preferably N260R.

65. The engineered variant Cas12a endonuclease of paragraph 62, wherein the
25 polypeptide sequence comprises a mutation at an amino acid position corresponding to position G902 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

66. The engineered variant Cas12a endonuclease of paragraph 65, wherein the mutation is
30 a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position G902 with reference to amino acid position numbering of LbCas12a ND2006, preferably G902W or G902F, more preferably G902W.

67. The engineered variant Cas12a endonuclease of any one of paragraphs 1-4, wherein (a) polypeptide sequence comprises a mutation at an amino acid position corresponding to

position K960 with reference to amino acid position numbering of LbCas12a ND2006 and (b) the variant Cas12a endonuclease exhibits PAM nickase activity (or a preference for cleaving one strand over the other of a dsDNA).

- 5 68. The engineered variant Cas12a endonuclease of paragraph 67, wherein the mutation is a substitution of a polar, negatively charged, and/or amide amino acid amino acid at position K960 with reference to amino acid position numbering of LbCas12a ND2006, preferably K960E, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
- 10 69. The engineered variant Cas12a endonuclease of any one of the preceding paragraphs comprising an amino acid sequence having at least 85%, at least 90%, or least 95%, but less than 100% identity with the amino acid sequence of a wild-type Cas12a endonuclease selected from *Acidaminococcus* sp., *Lachnospiraceae* sp., and *Francisella* sp.
- 15 70. The engineered variant Cas12a endonuclease of any one of the preceding paragraphs further comprising no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 additional amino acid substitutions relative to a wild-type reference Cas12a endonuclease.
71. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising the amino acid sequence of any one of SEQ ID NOs: 48-119.
- 20 72. A polynucleotide encoding the variant Cas12a endonuclease of any one of the preceding paragraphs.
73. A cell comprising (a) the variant Cas12a endonuclease of any one of the preceding paragraphs or the polynucleotide of paragraph 72 and (b) a guide RNA or a polynucleotide encoding a guide RNA.
- 25 74. A method comprising introducing into a cell (a) the variant Cas12a endonuclease of any one of the preceding paragraphs or the polynucleotide of paragraph 72 and optionally (b) a guide RNA or a polynucleotide encoding a guide RNA.
75. Use of the variant Cas12a endonuclease of any one of the preceding paragraphs for cleaving a nucleic acid.
- 30 76. A method for introducing a double strand break in a target nucleic acid, comprising introducing into a cell comprising a target nucleic acid (a) the variant Cas12a endonuclease of any one of paragraphs 5-41 and (b) a guide RNA; and incubating the cell to produce a double strand break in the target nucleic acid.

77. A method for introducing a double strand break in a target nucleic acid, comprising introducing into a cell comprising a target nucleic acid (a) the variant Cas12a endonuclease of any one of paragraphs 42-61 and (b) a guide RNA; and incubating the cell to produce a double strand break in the target nucleic acid.
- 5 78. The method of paragraph 77, wherein off-target single strand nucleic acid cleavage in the cell is reduced relative to off-target single strand nucleic acid cleavage in a control cell comprising a wild-type Cas12a endonuclease and a guide RNA.
79. A method for introducing a single strand break in a target nucleic acid, comprising introducing into a cell comprising a target nucleic acid (a) the variant Cas12a endonuclease of any one of paragraphs 62-68 and (b) a guide RNA; and
- 10 incubating the cell to produce a single strand break in the target nucleic acid.

Further Embodiments

Further embodiments are described in the following numbered paragraphs:

- 15 1. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising a mutation at an amino acid position corresponding to position R833, E835, R836, S929, F931, K932, N933, S934, R935, V936, K937, V938, K940, Q941, Y943, Q944, F983, or M986 with reference to amino acid position numbering of LbCas12a ND2006.
2. The engineered variant Cas12a endonuclease of paragraph 1, wherein the engineered
- 20 variant Cas12a endonuclease exhibits hyperactivity, low indiscriminate single strand deoxyribonuclease (DNase) activity, or target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).
3. The engineered variant Cas12a endonuclease of paragraph 1 or 2, wherein the mutation is an amino acid substitution.
- 25 4. The engineered variant Cas12a endonuclease of paragraph 3, wherein the amino acid substitution is an amino acid having an equivalent charge, polarity, and/or chemical class.
5. The engineered variant Cas12a endonuclease of any one of paragraphs 1-4, wherein (a) the polypeptide sequence comprises a mutation at an amino acid position corresponding to position K932, N933, or V936, with reference to amino acid position numbering of LbCas12a
- 30 ND2006 and (b) the engineered variant Cas12a endonuclease exhibits hyperactivity.
6. The engineered variant Cas12a endonuclease of paragraph 5, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position K932 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein

the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

7. The engineered variant Cas12a endonuclease of paragraph 6, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position K932 with reference to amino acid position numbering of LbCas12a ND2006, preferably K932I, K932L, K932V, K932A, or K932P, more preferably K932I, K932L, or K932V.

8. The engineered variant Cas12a endonuclease of paragraph 5, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position N933 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

9. The engineered variant Cas12a endonuclease of paragraph 8, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position N933 with reference to amino acid position numbering of LbCas12a, preferably N933L, N933A, N933G, N933I, or N933P, more preferably N933L.

10. The engineered variant Cas12a endonuclease of paragraph 5, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position V936 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

11. The engineered variant Cas12a endonuclease of paragraph 10, wherein the mutation is a substitution of a nonpolar, uncharged, and/or sulfuric amino acid at position V936 with reference to amino acid position numbering of LbCas12a, preferably V936M or V936C, more preferably V936M.

12. The engineered variant Cas12a endonuclease of any one of paragraphs 1-4, wherein (a) the polypeptide sequence comprises a mutation at an amino acid position corresponding to position S929, K932, N933, S934, V936, K937, Q944, F983, or M986 with reference to amino acid position numbering of LbCas12a ND2006 and (b) the variant Cas12a endonuclease exhibits low indiscriminate single strand deoxyribonuclease (ssDNase) activity.

13. The engineered variant Cas12a endonuclease of paragraph 12, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position S929 with reference to amino acid position numbering of LbCas12a, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

14. The engineered variant Cas12a endonuclease of paragraph 13, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position S929 with reference to amino acid position numbering of LbCas12a, preferably S929L, S929A, S929I, S929P, or S929V, more preferably S929L.
- 5 15. The engineered variant Cas12a endonuclease of paragraph 12, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position K932 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of
- 10 Table 1.
16. The engineered variant Cas12a endonuclease of paragraph 15, wherein the mutation is a substitution of a polar, positively charged, and/or basic amino acid at position K932 with reference to amino acid position numbering of LbCas12a, preferably K932R.
17. The engineered variant Cas12a endonuclease of paragraph 15, wherein the mutation is
- 15 a substitution of a polar, uncharged, and/or hydroxyl amino acid at position K932 with reference to amino acid position numbering of LbCas12a, preferably K932T.
18. The engineered variant Cas12a endonuclease of paragraph 15, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position K932 with reference to amino acid position numbering of LbCas12a, preferably K932F or K932W.
- 20 19. The engineered variant Cas12a endonuclease of paragraph 15, wherein the mutation is a substitution of a polar, uncharged, and/or aromatic amino acid at position K932 with reference to amino acid position numbering of LbCas12a, preferably K932Y.
20. The engineered variant Cas12a endonuclease of paragraph 12, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to
- 25 position N933 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
21. The engineered variant Cas12a endonuclease of paragraph 20, wherein the mutation is
- 30 a substitution of a polar, negatively charged, and/or amide amino acid at position N933 with reference to amino acid position numbering of LbCas12a, preferably N933E.
22. The engineered variant Cas12a endonuclease of paragraph 20, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position N933 with

reference to amino acid position numbering of LbCas12a, preferably N933V, N933A, N933G, N933I, or N933P, more preferably N933V.

23. The engineered variant Cas12a endonuclease of paragraph 12, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position S934 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
24. The engineered variant Cas12a endonuclease of paragraph 23, wherein the mutation is a substitution of a polar, uncharged, and/or acidic amino acid at position S934 with reference to amino acid position numbering of LbCas12a, preferably S934Q.
25. The engineered variant Cas12a endonuclease of paragraph 12, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position V936 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
26. The engineered variant Cas12a endonuclease of paragraph 25, wherein the mutation is a substitution of a polar, negatively charged, and/or amide amino acid at position V936 with reference to amino acid position numbering of LbCas12a, preferably V936E.
27. The engineered variant Cas12a endonuclease of paragraph 25, wherein the mutation is a substitution of a polar, positively charged, and/or basic amino acid at position V936 with reference to amino acid position numbering of LbCas12a, preferably V936K, V936R, or V936H, more preferably V936K.
28. The engineered variant Cas12a endonuclease of paragraph 12, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position K937 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
29. The engineered variant Cas12a endonuclease of paragraph 28, wherein the mutation is a substitution of a polar, uncharged, and/or aromatic amino acid at position K937 with reference to amino acid position numbering of LbCas12a, preferably K937Y.

30. The engineered variant Cas12a endonuclease of paragraph 12, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position Q944 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
31. The engineered variant Cas12a endonuclease of paragraph 30, wherein the mutation is a substitution of a polar, negatively charged, and/or acidic amino acid at position Q944 with reference to amino acid position numbering of LbCas12a, preferably Q944D.
32. The engineered variant Cas12a endonuclease of paragraph 30, wherein the mutation is a substitution of a nonpolar, uncharged, and/or sulfur amino acid at position Q944 with reference to amino acid position numbering of LbCas12a, preferably Q944M or Q944C, more preferably Q944M.
33. The engineered variant Cas12a endonuclease of paragraph 30, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position Q944 with reference to amino acid position numbering of LbCas12a, preferably Q944G, Q944I, Q944A, Q944L, Q944P, or Q944V, more preferably Q944G.
34. The engineered variant Cas12a endonuclease of paragraph 12, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position F983 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
35. The engineered variant Cas12a endonuclease of paragraph 34, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position F983 with reference to amino acid position numbering of LbCas12a, preferably F983L, F983A, F983G, F983I, F983P, or F983V, more preferably F983L.
36. The engineered variant Cas12a endonuclease of paragraph 12, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position M986 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

37. The engineered variant Cas12a endonuclease of paragraph 36, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position M986 with reference to amino acid position numbering of LbCas12a, preferably M986F or M986W, more preferably M986F.
- 5 38. The engineered variant Cas12a endonuclease of paragraph 36, wherein the mutation is a substitution of a polar, uncharged, and/or hydroxyl amino acid at position M986 with reference to amino acid position numbering of LbCas12a, preferably M986S or M986T, more preferably M986S.
39. The engineered variant Cas12a endonuclease of any one of paragraphs 1-4, wherein
- 10 (a) the polypeptide sequence comprises a mutation at an amino acid position corresponding to position R833, E835, R836, F931, K932, R935, V936, V938, K940, Q941, Y943, Q944, or M986 with reference to amino acid position numbering of LbCas12a ND2006 and (b) the variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).
- 15 40. The engineered variant Cas12a endonuclease of paragraph 39, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position R833 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of
- 20 Table 1.
41. The engineered variant Cas12a endonuclease of paragraph 40, wherein the mutation is a substitution of a polar, positively charged, and/or basic amino acid at position R833 with reference to amino acid position numbering of LbCas12a, preferably R833K or R833H, more preferably R833K optionally wherein the engineered variant Cas12a endonuclease also
- 25 exhibits low indiscriminate ssDNase activity.
42. The engineered variant Cas12a endonuclease of paragraph 40 wherein the mutation is a substitution of a nonpolar, uncharged, and/or sulfuric amino acid at position R833 with reference to amino acid position numbering of LbCas12a, preferably R833M or R833C, more preferably R833M, optionally wherein the engineered variant Cas12a endonuclease also
- 30 exhibits low indiscriminate ssDNase activity.
43. The engineered variant Cas12a endonuclease of paragraph 40, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position R833 with reference to amino acid position numbering of LbCas12a, preferably R833L, R833A, R833I,

R833P, or R833V, more preferably R833L, optionally wherein the engineered variant Cas12a endonuclease also exhibits low indiscriminate ssDNase activity.

44. The engineered variant Cas12a endonuclease of paragraph 39, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position E835 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

45. The engineered variant Cas12a endonuclease of paragraph 44, wherein the mutation is a substitution of a polar, negatively charged, and/or acidic amino acid at position E835 with reference to amino acid position numbering of LbCas12a, preferably E835D, optionally wherein the engineered variant Cas12a endonuclease exhibits hypoactivity.

46. The engineered variant Cas12a endonuclease of paragraph 44, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position E835 with reference to amino acid position numbering of LbCas12a, preferably E835G, E835A, E835I, E835L, E835P, or E835V, more preferably E835G or E835A.

47. The engineered variant Cas12a endonuclease of paragraph 39, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position R836 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

48. The engineered variant Cas12a endonuclease of paragraph 47, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position R836 with reference to amino acid position numbering of LbCas12a, preferably R836G, R836A, R836I, R836L, R836P, or R836V, more preferably R836G or R836A.

49. The engineered variant Cas12a endonuclease of paragraph 39, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position F931 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

50. The engineered variant Cas12a endonuclease of paragraph 49, wherein the mutation is a substitution of a polar, positively charged, and/or basic amino acid at position F931 with

reference to amino acid position numbering of LbCas12a, preferably F931H or F931K, more preferably F931H, optionally wherein the engineered variant Cas12a endonuclease exhibits hypoactivity.

51. The engineered variant Cas12a endonuclease of paragraph 49, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position F931 with reference to amino acid position numbering of LbCas12a, preferably F931L, F931A, F931G, F931I, F931P, or F931V, more preferably F931L.

52. The engineered variant Cas12a endonuclease of paragraph 39, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position K932 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

53. The engineered variant Cas12a endonuclease of paragraph 52, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position K932 with reference to amino acid position numbering of LbCas12a, preferably K932G or K932P, more preferably K932G, optionally wherein the engineered variant Cas12a endonuclease exhibits hypoactivity.

54. The engineered variant Cas12a endonuclease of paragraph 52, wherein the mutation is a substitution of a polar, negatively charged, and/or amide amino acid at position K932 with reference to amino acid position numbering of LbCas12a, preferably K932E, optionally wherein the engineered variant Cas12a endonuclease exhibits hypoactivity.

55. The engineered variant Cas12a endonuclease of paragraph 52, wherein the mutation is a substitution of a polar, positively charged, and/or basic amino acid at position K932 with reference to amino acid position numbering of LbCas12a, preferably K932H or K932R, more preferably K932H.

56. The engineered variant Cas12a endonuclease of paragraph 52, wherein the mutation is a substitution of a nonpolar, uncharged, and/or sulfur amino acid at position K932 with reference to amino acid position numbering of LbCas12a, preferably K932M or K932C, more preferably K932M.

57. The engineered variant Cas12a endonuclease of paragraph 52, wherein the mutation is a substitution of a polar, uncharged, and/or amide amino acid at position K932 with reference to amino acid position numbering of LbCas12a, preferably K932N.

58. The engineered variant Cas12a endonuclease of paragraph 52, wherein the mutation is a substitution of a polar, uncharged, and/or acidic amino acid at position K932 with reference to amino acid position numbering of LbCas12a, preferably K932Q.
59. The engineered variant Cas12a endonuclease of paragraph 52, wherein the mutation is
5 a substitution of a polar, uncharged, and/or hydroxyl amino acid at position K932 with reference to amino acid position numbering of LbCas12a, preferably K932S.
60. The engineered variant Cas12a endonuclease of paragraph 39, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position R935 with reference to amino acid position numbering of LbCas12a ND2006,
10 optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
61. The engineered variant Cas12a endonuclease of paragraph 60, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position R935 with
15 reference to amino acid position numbering of LbCas12a, preferably R935L, R935G, R935I, or R935P, more preferably R935L, R935G, or R935I, optionally wherein the engineered variant Cas12a endonuclease having the R935I substitution exhibits hypoactivity.
62. The engineered variant Cas12a endonuclease of paragraph 60, wherein the mutation is a substitution of a polar, positively charged, and/or basic amino acid at position R935 with
20 reference to amino acid position numbering of LbCas12a, preferably R935H or R935K.
63. The engineered variant Cas12a endonuclease of paragraph 60, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position R935 with reference to amino acid position numbering of LbCas12a, preferably R935F or R935W,
25 optionally wherein the engineered variant Cas12a endonuclease having the R935W substitution exhibits hypoactivity.
64. The engineered variant Cas12a endonuclease of paragraph 60, wherein the mutation is a substitution of a nonpolar, uncharged, and/or sulfur amino acid at position R935 with reference to amino acid position numbering of LbCas12a, preferably R935M or R935C, more preferably R935M.
- 30 65. The engineered variant Cas12a endonuclease of paragraph 60, wherein the mutation is a substitution of a polar, uncharged, and/or amide amino acid at position R935 with reference to amino acid position numbering of LbCas12a, preferably R935N.

66. The engineered variant Cas12a endonuclease of paragraph 60, wherein the mutation is a substitution of a polar, uncharged, and/or hydroxyl amino acid at position R935 with reference to amino acid position numbering of LbCas12a, preferably R935S or R935T.
67. The engineered variant Cas12a endonuclease of paragraph 39, wherein the
5 polypeptide sequence comprises a mutation at an amino acid position corresponding to position V936 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
- 10 68. The engineered variant Cas12a endonuclease of paragraph 67, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position V936 with reference to amino acid position numbering of LbCas12a, preferably V936G, V936I, V936L, or V936P, more preferably V936G, optionally wherein the engineered variant Cas12a endonuclease also exhibits low indiscriminate ssDNase activity.
- 15 69. The engineered variant Cas12a endonuclease of paragraph 39, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position V938 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of
20 Table 1.
70. The engineered variant Cas12a endonuclease of paragraph 69, wherein the mutation is a substitution of a polar, negatively charged, and/or amide amino acid at position V938 with reference to amino acid position numbering of LbCas12a, preferably V938E, optionally wherein the engineered variant Cas12a endonuclease also exhibits low indiscriminate
25 ssDNase activity.
71. The engineered variant Cas12a endonuclease of paragraph 39, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position K940 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type
30 reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.
72. The engineered variant Cas12a endonuclease of paragraph 71, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position K940 with

reference to amino acid position numbering of LbCas12a, preferably K940G, K940A, K940I, K940L, K940P, or K940V, more preferably K940G.

73. The engineered variant Cas12a endonuclease of paragraph 39, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position Q941 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

74. The engineered variant Cas12a endonuclease of paragraph 73, wherein the mutation is a substitution of a polar, positively charged, and/or basic amino acid at position Q941 with reference to amino acid position numbering of LbCas12a, preferably Q941K, Q941R, or Q941H.

75. The engineered variant Cas12a endonuclease of paragraph 73, wherein the mutation is a substitution of a polar, uncharged, and/or aromatic amino acid at position Q941 with reference to amino acid position numbering of LbCas12a, preferably Q941Y.

76. The engineered variant Cas12a endonuclease of paragraph 39, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position Y943 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

77. The engineered variant Cas12a endonuclease of paragraph 76, wherein the mutation is a substitution of a polar, uncharged, and/or hydroxyl amino acid at position Y943 with reference to amino acid position numbering of LbCas12a, preferably Y943T or Y943S, more preferably Y943T.

78. The engineered variant Cas12a endonuclease of paragraph 76, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aromatic amino acid at position Y943 with reference to amino acid position numbering of LbCas12a, preferably Y943F or Y943W, more preferably Y943F.

79. The engineered variant Cas12a endonuclease of paragraph 39, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position Q944 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type

reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

80. The engineered variant Cas12a endonuclease of paragraph 79, wherein the mutation is a substitution of a polar, positively charged, and/or basic amino acid at position Q944 with reference to amino acid position numbering of LbCas12a, preferably Q944K, Q944R, or Q944H, more preferably Q944K.

81. The engineered variant Cas12a endonuclease of paragraph 79, wherein the mutation is a substitution of a polar, negatively charged, and/or amide amino acid at position Q944 with reference to amino acid position numbering of LbCas12a, preferably Q944E.

82. The engineered variant Cas12a endonuclease of paragraph 39, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position M986 with reference to amino acid position numbering of LbCas12a ND2006, optionally wherein the polypeptide sequence has at least 90% identity to a wild-type reference Cas12a endonuclease, optionally to a wild-type reference Cas12a endonuclease of Table 1.

83. The engineered variant Cas12a endonuclease of paragraph 82, wherein the mutation is a substitution of a nonpolar, uncharged, and/or aliphatic amino acid at position M986 with reference to amino acid position numbering of LbCas12a, preferably M986G, M986A, M986I, M986P, or M986V, more preferably M986G, optionally wherein the engineered variant Cas12a endonuclease also exhibits low indiscriminate ssDNase activity.

84. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising amino acid substitutions corresponding to the following amino acid substitutions: R833L, S929L, K932M, Q944F, and E947M, with reference to amino acid position numbering of LbCas12a ND20006, preferably wherein the engineered variant Cas12a endonuclease exhibits low indiscriminate single strand DNase activity, optionally wherein the engineered variant Cas12a endonuclease also exhibits hypoactivity.

85. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising amino acid substitutions corresponding to the following amino acid substitutions: N933L and Q944M, with reference to amino acid position numbering of LbCas12a ND20006, preferably wherein the engineered variant Cas12a endonuclease exhibits low indiscriminate single strand DNase activity, optionally wherein the engineered variant Cas12a endonuclease also exhibits hypoactivity.

86. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising amino acid substitutions corresponding to the following amino acid substitutions:

R833K and E947D, with reference to amino acid position numbering of LbCas12a ND20006, preferably wherein the engineered variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).

87. An engineered variant Cas12a endonuclease comprising a polypeptide sequence
5 comprising amino acid substitutions corresponding to the following amino acid substitutions: E835G and E880G, with reference to amino acid position numbering of LbCas12a ND20006, preferably wherein the engineered variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).

88. An engineered variant Cas12a endonuclease comprising a polypeptide sequence
10 comprising amino acid substitutions corresponding to the following amino acid substitutions: S929G, K932G, and N933G, with reference to amino acid position numbering of LbCas12a ND20006, preferably wherein the engineered variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).

89. An engineered variant Cas12a endonuclease comprising a polypeptide sequence
15 comprising amino acid substitutions corresponding to the following amino acid substitutions: S929G, K932G, N933G, and V936G, with reference to amino acid position numbering of LbCas12a ND20006, preferably wherein the engineered variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).

20 90. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising amino acid substitutions corresponding to the following amino acid substitutions: S929G, K932G, N933G, and V936F, with reference to amino acid position numbering of LbCas12a ND20006, preferably wherein the engineered variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a
25 dsDNA).

91. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising amino acid substitutions corresponding to the following amino acid substitutions: S929G, V936G, F983G, and M986G, with reference to amino acid position numbering of LbCas12a ND20006, preferably wherein the engineered variant Cas12a endonuclease
30 exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).

92. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising amino acid substitutions corresponding to the following amino acid substitutions: G930A and F931L, with reference to amino acid position numbering of LbCas12a ND20006,

preferably wherein the engineered variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA) and low indiscriminate single strand DNase activity.

93. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising amino acid substitutions corresponding to the following amino acid substitutions: G930A, F931L, and S934Q with reference to amino acid position numbering of LbCas12a, preferably wherein the engineered variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).

94. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising amino acid substitutions corresponding to the following amino acid substitutions: K932G and N933G, with reference to amino acid position numbering of LbCas12a ND20006, preferably wherein the engineered variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).

95. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising amino acid substitutions corresponding to the following amino acid substitutions: K932G, N933G, and V936F, with reference to amino acid position numbering of LbCas12a ND20006, preferably wherein the engineered variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).

96. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising amino acid substitutions corresponding to the following amino acid substitutions: V936G, F983G, and M986G, with reference to amino acid position numbering of LbCas12a ND20006, preferably wherein the engineered variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).

97. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising amino acid substitutions corresponding to the following amino acid substitutions: F983G and M986G, with reference to amino acid position numbering of LbCas12a ND20006, preferably wherein the engineered variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA) and low indiscriminate single strand DNase activity.

98. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising a Lid-hub domain, wherein the polypeptide comprises a mutation at an amino acid position in the Lid-hub domain or in the vicinity of the Lid-Hub domain.

99. The engineered variant Cas12a endonuclease of paragraph 98, wherein the mutation is a substitution at a position corresponding to the following amino acid positions: R833, E835,

E880, R836, S929, G930, F931, K932, N933, S934, R935, V936, K937, V938, K940, Q941, Y943, Q944, E947, F983, or M986, with reference to amino acid position numbering of LbCas12a ND20006.

100. The engineered variant Cas12a endonuclease of paragraph 98 or 99, wherein the
5 engineered variant Cas12a endonuclease exhibits hyperactivity.

101. The engineered variant Cas12a endonuclease of paragraph 100, wherein the mutation is a substitution corresponding to any one of the following amino acid substitutions: K932I, K932L, K932V, N933L, or V936M.

102. The engineered variant Cas12a endonuclease of paragraph 98 or 99, wherein the
10 engineered variant Cas12a endonuclease exhibits low indiscriminate single strand deoxyribonuclease (DNase) activity.

103. The engineered variant Cas12a endonuclease of paragraph 102, wherein the mutation is a substitution corresponding to any one of the following amino acid substitutions: S929L, K932F, K932R, K932T, K932Y, K932W, N933E, N933V, S934Q, V936E, V936K, K937Y,
15 Q944D, Q944G, Q944M, F983L, M986F, or M986S.

104. The engineered variant Cas12a endonuclease of paragraph 98 or 99, wherein the engineered variant Cas12a endonuclease exhibits target nickase activity (or a preference for cleaving one strand over the other of a dsDNA).

105. The engineered variant Cas12a endonuclease of paragraph 104, wherein the mutation
20 is a substitution corresponding to any one of the following amino acid substitutions: R833K, R833L, R833M, E835A, E835D, E835G, R836A, R836G, F931H, F931L, K932E, K932G, K932H, K932M, K932N, K932Q, K932S, R935F, R935G, R935H, R935I, R935K, R935L, R935M, R935N, R935S, R935T, R935W, V936G, V938E, K940G, Q941H, Q941K, Q941R, Q941Y, Y943F, Y943T, Q944E, Q944K, or M986G.

25 106. The engineered variant Cas12a endonuclease of any one of the preceding paragraphs comprising an amino acid sequence having at least 85%, at least 90%, or least 95%, but less than 100% identity with the amino acid sequence of a wild-type Cas12a endonuclease selected from *Acidaminococcus* sp., *Lachnospiraceae* sp., and *Francisella* sp.

107. The engineered variant Cas12a endonuclease of any one of the preceding paragraphs
30 further comprising no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 additional amino acid substitutions relative to a wild-type reference Cas12a endonuclease.

108. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising the amino acid sequence of any one of SEQ ID NOs: 48-119.

109. A polynucleotide encoding the variant Cas12a endonuclease of any one of the preceding paragraphs.
110. A cell comprising (a) the variant Cas12a endonuclease of any one of the preceding paragraphs or the polynucleotide of paragraph 109 and (b) a guide RNA or a polynucleotide
5 encoding a guide RNA.
111. A method comprising introducing into a cell (a) the variant Cas12a endonuclease of any one of the preceding paragraphs or the polynucleotide of paragraph 109 and optionally (b) a guide RNA or a polynucleotide encoding a guide RNA.
112. Use of the variant Cas12a endonuclease of any one of the preceding paragraphs for
10 cleaving a nucleic acid.
113. A method for introducing a double strand break in a target nucleic acid, comprising introducing into a cell comprising a target nucleic acid (a) the variant Cas12a endonuclease of any one of paragraphs 5-11 and (b) a guide RNA; and incubating the cell to produce a double strand break in the target nucleic acid.
114. A method for introducing a double strand break in a target nucleic acid, comprising
15 introducing into a cell comprising a target nucleic acid (a) the variant Cas12a endonuclease of any one of paragraphs 12-38, 84, or 85 and (b) a guide RNA; and incubating the cell to produce a double strand break in the target nucleic acid.
115. The method of paragraph 114, wherein off-target single strand nucleic acid cleavage
20 in the cell is reduced relative to off-target single strand nucleic acid cleavage in a control cell comprising a wild-type Cas12a endonuclease and a guide RNA.
116. A method for introducing a single strand break in a target nucleic acid, comprising introducing into a cell comprising a target nucleic acid (a) the variant Cas12a endonuclease of any one of paragraphs 39-83 or 86-97 and (b) a guide RNA; and
25 incubating the cell to produce a single strand break in the target nucleic acid.

EXAMPLES

Example 1. Cas12a Endonuclease Variants

- Purified Cas12a variants of Table 3 in complex with crRNA were tested against (1)
30 dsDNA containing a quencher on one site of the cleavage site and a fluorophore on the other site of the cleavage site, on separate DNA strands, and (2) a combination dsDNA containing no quencher or fluorophore but together with a ssDNA containing both a quencher and fluorophore. The retained activity on dsDNA was confirmed by (1) where the quencher and fluorophore were separated upon cleavage, and the emission signal increases over time.

Hyperactive Cas12a endonuclease variants were identified as having a higher reaction speed or by initiating the reaction faster than the naturally-occurring (i.e., wildtype (WT)) Cas12a. Within the subset of Cas12a endonuclease variants that had hyperactivity, several variants were also identified as having low ssDNase activity. Low ssDNase activity is identified by a lowered ability to cleave the ssDNA and thereby separate the quencher and fluorophore when activated by specific dsDNA, thus by lacking an emission signal in comparable magnitude to that of the wildtype Cas12a in (2). See FIGs. 2A-2D and Table 6.

Hypoactive Cas12a endonuclease variants were identified as having a lower reaction speed or by initiating the reaction slower than the wildtype (WT) Cas12a. Within the subset of Cas12a endonuclease variants that had hypoactivity, several variants were also identified as having low ssDNase activity. Low ssDNase activity is identified by a lowered ability to cleave the ssDNA and thereby separate the quencher and fluorophore when activated by specific dsDNA, thus by lacking an emission signal in comparable magnitude to that of the wildtype Cas12a in (2). See FIGs. 3A-3E, 4A-4C, and Table 6.

Three Cas12a endonuclease variants were identified as having similar activity profile as wildtype Cas12a while demonstrating low ssDNase activity. See FIG. 5 and Table 6.

Table 6. Variant Cas12a Endonucleases of Example 1

Variant	dsDNA Phenotype	Reduced/Lowered Activity towards ssDNA?
E95R	Hyperactive	
E95Y	Hyperactive	
E125A	Hyperactive	
E125W	Hyperactive	
N256A	Hyperactive	
N256K	Hypoactive	
R747Y	Hyperactive	
H759V	Hyperactive	
H759D	Hyperactive	
N813R	Hyperactive	Yes
N813W	Wildtype	Yes
N813H	Hyperactive	Yes
I831A	Hypoactive	Yes
I831Y	Hypoactive	Yes
K932L	Hyperactive	
K932I	Wildtype	
K932V	Wildtype	
K932M	Hypoactive	Yes
K932F	Hypoactive	Yes
K932R	Hypoactive	Yes

K932A	Hypoactive	Yes
K932H	Hypoactive	Yes
K932N	Hypoactive	Yes
K932Q	Hypoactive	Yes
K932S	Hypoactive	Yes
K932T	Hypoactive	Yes
K932Y	Hypoactive	Yes
K932W	Hypoactive	Yes
N933E	Hyperactive	Yes
N933V	Hyperactive	
N933L	Hypoactive	Yes
S934Q	Hyperactive	Yes
S934K	Wildtype	Yes
S934W	Hypoactive	
V936E	Hyperactive	Yes
V936M	Hyperactive	
V936K	Hyperactive	
V936G	Hypoactive	Yes
Q944D	Hypoactive	Yes
Q944E	Hypoactive	Yes
Q944K	Hypoactive	Yes
Q944M	Hypoactive	
S982W	Hypoactive	Yes
S982T	Hypoactive	
S982A	Wildtype	
S982N	Hyperactive	
F983L	Hypoactive	Yes
F983G	Hypoactive	Yes
K984R	Hyperactive	
K984F	Hypoactive	Yes
M986G	Hypoactive	Yes
M986F	Wildtype	Yes
M986L	Hypoactive	
M986S	Hypoactive	
T988F	Hypoactive	Yes
K932F, F983L	Hypoactive	Yes
K932F, T988F	Hypoactive	Yes
K932R, Q944D	Hypoactive	Yes
K932R, F983L	Hypoactive	Yes
K932R, T988F	Hypoactive	Yes
K932Y, F983L	Hypoactive	Yes
K932Y, T988F	Hypoactive	Yes
N933L, Q944M	Hypoactive	Yes
V936G, Q944D	Hypoactive	Yes
V936G, S982W	Hypoactive	Yes
V936G, M986G	Hypoactive	Yes
V936G, T988F	Hypoactive	Yes

Q944D, S982W	Hypoactive	Yes
Q944D, F983L	Hypoactive	Yes
Q944D, T988F	Hypoactive	Yes
S982W, F983L	Hypoactive	Yes
S982W, T988F	Hypoactive	Yes
F983G, M986G	Hypoactive	Yes

Example 2. Enzyme-to-Enzyme Linker Assessment

The activity of different C-to-T base editors (fusion proteins comprising a Cas12a variant and cytidine deaminase) containing linkers of varying sizes and sequences between different domains was tested in U2OS cells. U2OS cells were transfected using a plasmid expressing the base editor together with a plasmid expressing a gRNA (AGCCTCAC₈C₉C₁₀CTC₁₃TAGCCCT (SEQ ID NO: 186)). Transfected cells were sorted after 3 days and re-cultured for another 3 days after which the cells were lysed and genomic DNA was extracted. Genomic region around the gRNA was PCR amplified and editing was analyzed by next generation sequencing. In these experiments, LbCas12a was either catalytic inactive (inactive Cas12a comprising a D832A mutation) LbCas12a or a LbCas12a variant (TBN04).

LbBEv2 (comprising Linker4 between rAPOBEC1 and LbCas12a and Linker5 between LbCas12a and UGI) showed efficient base editing (FIG. 6A-6C). **FIG. 6A:** % of total reads showing base editing at individual positions. This includes the reads that show base editing without indels plus the reads that harbor base editing as well as the indels. **FIG. 6B:** % of total reads showing only base editing at individual positions. This only includes the reads that show base editing but no indels. **FIG. 6C:** % of total reads showing base editing at only one specific position. This excludes all the reads with base editing at multiple positions as well as the reads harboring indels. Inactive Cas12a was used as the control.

The improvement in the base editing efficiency with LbBEv2 was more significant when using a base editor containing LbCas12a variant (TBN04). These data suggest that inclusion of a longer linker between the Cas12a variant and the base editing enzyme in the fusion proteins described herein is useful in improving base editing efficiency (likely by decreasing undesired interactions, *e.g.*, steric interactions, between the Cas12a variant and the base editing enzyme).

Structure of LbBEv2 is as follows:

Linker3-rAPOBEC1-Linker4-LbCas12a-NP NLS-Linker5-UGI-Linker2-SV40 NLS; wherein Linker3 sequence is the amino acid sequence of GS, Linker4 sequence is the amino acid

sequence of SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 204) and Linker5 sequence is the amino acid sequence of GSSGGSGGSGGS (SEQ ID NO: 207).

Example 3. Nuclear Localization Signal Assessment

5 The activity of C-to-T base editors (fusion proteins comprising a LbCas12a Cas12a variant (TBN04) and cytidine deaminase) with different NLSs at the N-terminus (LbVEv3, LbBEv4 and LbBEv5) was tested in U2OS cells and compared with LbBEv2, which lacks an NLS at the N-terminus. U2OS cells were transfected using a plasmid expressing the base editor together with a plasmid expressing the gRNA. Transfected cells were sorted after 3
10 days and re-cultured for another 3 days after which the cells were lysed, and genomic DNA was extracted. Genomic region around the gRNA was PCR amplified and editing was analyzed by next generation sequencing. Two different gRNAs were tested- TTCTCCCC₈TC₁₀TGCTGGATAC (SEQ ID NO:187) (FIG. 7A-7D) and CTGATGGTC₉C₁₀ATGTC₁₅TGTTA (SEQ ID NO: 191) (FIG. 8A-8D).

15 **FIG. 7A:** % of total reads showing base editing at individual positions. This includes the reads that show base editing without indels plus the reads that harbor base editing as well as the indels. **FIG. 7B:** % of total reads showing only base editing at individual positions. This only includes the reads that show base editing but no indels. **FIG. 7C:** % of total reads showing base editing at only one specific position. This excludes all the reads with base
20 editing at multiple positions as well as the reads harboring indels. **FIG. 7D:** % of total reads showing indels. Inactive Cas12a was used as the control.

 In these experiments, LbCas12a can be catalytic inactive (inactive LbCas12a comprising a D832A mutation) or a variant LbCas12a (TBN04). Relative to LbBEv2 (*i.e.*, lacking an NLS at the N-terminus), C-to-T base editors with N-terminus NLS (LbBEv3,
25 LbBEv4 and LbBEv5) showed improved base editing efficiency in a gRNA-specific manner (FIG. 8A-8D). **FIG. 8A:** % of total reads showing base editing at individual positions. This includes the reads that show base editing without indels plus the reads that harbor base editing as well as the indels. **FIG. 8B:** % of total reads showing only base editing at individual positions. This only includes the reads that show base editing but no indels. **FIG.**
30 **8C:** % of total reads showing base editing at only one specific position. This excludes all the reads with base editing at multiple positions as well as the reads harboring indels. **FIG. 8D:** % of total reads showing indels. Inactive Cas12a was used as the control.

These data demonstrate that inclusion of an NLS located at or near the N-terminal of the fusion proteins described herein is useful in improving base editing efficiency (likely by increasing the amount of fusion protein located in the nucleus).

Structure of LbBEv3 is as follows:

5 SV40 NLS-Linker3-rAPOBEC1-Linker4-LbCas12a-NP NLS-Linker5-UGI-Linker2-SV40 NLS

Structure of LbBEv4 is as follows:

NP NLS-Linker3-rAPOBEC1-Linker4-LbCas12a-NP NLS-Linker5-UGI-Linker2-SV40 NLS

Structure of LbBEv5 is as follows:

10 Linker3-BP NLS-Linker3-rAPOBEC1-Linker4-LbCas12a-NP NLS-Linker5-UGI-Linker2-SV40 NLS

Example 4. Efficiency and Specificity of C-to-T Base Editors

The base editing efficiency and specificity of C-to-T base editors (fusion proteins comprising a Cas12a variant and cytidine deaminase) was determined when comprising either inactive LbCas12a (Cas12a comprising a D832A mutation) or TBN04 (LbCas12a variant). U2OS cells were transfected using a plasmid expressing the base editor together with a plasmid expressing the gRNA. Transfected cells were sorted after 3 days and re-cultured for another 3 days after which the cells were lysed and genomic DNA was extracted. Genomic region around the gRNA was PCR amplified and editing was analyzed by next generation sequencing. Two different gRNAs were tested - AGCCTC₆AC₈C₉C₁₀C₁₁TC₁₃TAGCCCT (SEQ ID NO:192) (FIG. 9A-9F) and TTCTCCCC₈TC₁₀TGCTGGATAC (SEQ ID NO: 187) (FIG. 10A-10F).

C-to-T base editor containing the TBN04 shows overall higher base editing efficiency than base editors containing inactive Cas12a. Moreover, TBN04 base editor shows remarkable base selectivity where the majority of edited alleles has editing only at one specific position whereas C-to-T base editor containing inactive LbCas12a lacks base selectivity and edits multiple Cs in the editing window.

For example, for gRNA AGCCTC₆AC₈C₉C₁₀C₁₁TC₁₃TAGCCCT (SEQ ID NO: 192), base editor containing TBN04 shows editing primarily at C₁₃ whereas base editor with inactive LbCas12a edits all the Cs between position 6 and 13 with comparable efficiency (FIG. 9A-9F).

FIGs. 9A-9D provide graphs of data comparing percent (%) of total reads having a C-to-T nucleotide edit at genomic positions corresponding to positions C6, C8, C9, C10, C11,

and C13 of the gRNA using the LbBEv5 base editor. **FIG. 9A:** % of total reads showing base editing at individual positions. This includes the reads that show base editing without indels plus the reads that harbor base editing as well as the indels. **FIG. 9B:** % of total reads showing only base editing at individual positions. This only includes the reads that show base editing but no indels. **FIG. 9C:** % of total reads showing base editing at only one specific position. This excludes all the reads with base editing at multiple positions as well as the reads harboring indels. **FIG. 9D:** % of total reads showing indels. **FIGs. 9E-9F** show allele frequency tables around the gRNA. **Fig. 9E:** frequency of alleles harboring different edits by LbBEv5 containing inactive LbCas12a and gRNA AGCCTC₆AC₈C₉C₁₀C₁₁TC₁₃TAGCCCT (SEQ ID NO: 192) in U2OS cells. **FIG. 9F:** frequency of alleles harboring different edits by LbBEv5 containing TBN04 (LbCas12a) and gRNA AGCCTC₆AC₈C₉C₁₀C₁₁TC₁₃TAGCCCT (SEQ ID NO: 192) in U2OS cells.

Likewise, for gRNA TTCTCCCC₈TC₁₀TGCTGGATAC (SEQ ID NO: 187), base editor containing TBN04 shows editing primarily at C₁₀ whereas base editor with inactive LbCas12a edits both C₈ and C₁₀ with comparable efficiency (FIG. 10A-10F).

FIGs. 10A-10D provide graphs of data comparing percent (%) of total reads having a C-to-T nucleotide edit at genomic positions corresponding to positions C8 and C10 of the gRNA using the LbBEv5 base editor. **FIG. 10A:** % of total reads showing base editing at individual positions. This includes the reads that show base editing without indels plus the reads that harbor base editing as well as the indels. **FIG. 10B:** % of total reads showing only base editing at individual positions. This only includes the reads that show base editing but no indels. **FIG. 10C:** % of total reads showing base editing at only one specific position. This excludes all the reads with base editing at multiple positions as well as the reads harboring indels. **FIG. 10D:** % of total reads showing indels. **FIGs. 10E-10F** show allele frequency tables around the gRNA. **FIG. 10E:** frequency of alleles harboring different edits by LbBEv5 containing inactive LbCas12a and gRNA TTCTCCCC₈TC₁₀TGCTGGATAC (SEQ ID NO: 187) in U2OS cells. **FIG. 10F:** frequency of alleles harboring different edits by LbBEv5 containing TBN04 (LbCas12a) and gRNA TTCTCCCC₈TC₁₀TGCTGGATAC (SEQ ID NO: 187) in U2OS cells.

Example 5. Efficiency and Specificity of A-to-G Base Editors

The base editing efficiency and specificity of A-to-G base editor (fusion protein comprising a Cas12a variant and adenosine deaminase) containing inactive LbCas12a (Cas12a comprising a D832A mutation) or TBN04 (LbCas12a variant) was determined.

U2OS cells were transfected using a plasmid expressing the base editor together with a plasmid expressing the gRNA. Transfected cells were sorted after 3 days and re-cultured for another 3 days after which the cells were lysed and genomic DNA was extracted. Genomic region around the gRNA was PCR amplified and editing was analyzed by next generation sequencing.

The A-to-G base editor containing the TBN04 shows overall higher base editing efficiency than base editors containing inactive Cas12a. Moreover, TBN04 base editor shows significant base selectivity where the majority of edited alleles has editing only at one specific position. This is in contrast to the A-to-G base editor containing inactive LbCas12a which lacked base selectivity and edited multiple As in the editing window. For example, for gRNA TGCTGCA₇A₈GTA₁₁A₁₂GCA₁₅TGCATTTG (SEQ ID NO: 188), base editor containing TBN04 shows editing primarily at A₁₁ whereas base editor with inactive LbCas12a edits all the As between position 7 and 15 with comparable efficiency (FIG. 11A-11F).

FIGs. 11A-11D provide graphs of data comparing percent (%) of total reads having a C-to-T nucleotide edit at genomic positions corresponding to positions A7, A8, A11, A12, and A15 of the gRNA using the LbABE8e base editor. **FIG. 11A:** % of total reads showing A-to-G base editing at individual positions. This includes the reads that show base editing without indels plus the reads that harbor base editing as well as the indels. **FIG. 11B:** % of total reads showing only base editing at individual positions. This only includes the reads that show base editing but no indels. **FIG. 11C:** % of total reads showing base editing at only one specific position. This excludes all the reads with base editing at multiple positions as well as the reads harboring indels. **FIG. 11D:** % of total reads showing indels. **FIGs. 11E-11F** show allele frequency tables around the gRNA. **Fig. 11E:** frequency of alleles harboring different edits by LbABE8e containing inactive LbCas12a and gRNA TGCTGCA₇A₈GTA₁₁A₁₂GCA₁₅TGCATTTG (SEQ ID NO: 188) in U2OS cells. **FIG. 11F:** frequency of alleles harboring different edits by LbABE8e containing TBN04 (LbCas12a) and gRNA TGCTGCA₇A₈GTA₁₁A₁₂GCA₁₅TGCATTTG (SEQ ID NO: 188) in U2OS cells.

The structure of LbABE8e is as follows:

BP NLS-TadA-Linker4-LbCas12a-Linker2-BP NLS; wherein LbCas12a can be inactive LbCas12a or TBN04 (LbCas12a variant).

A selection of fusion proteins comprising a Cas12a variants as provided in Table 7 were assayed for their ability to produce indels and perform an A-to-G base conversion. The Cas12a variants were placed into the LbABE8e structural framework. The data in Table 7

and FIGs. 12A-12C demonstrate that the A-to-G base editors were selective for the A₁₁ in the gRNA TGCTGCA₇A₈GTA₁₁A₁₂GCA₁₅TGCATTTG (SEQ ID NO: 188). These data demonstrate that single and combinatorial mutations in Cas12a endonuclease provide higher base editing efficiency and higher base editing selectivity relative to inactive Cas12 and TBN04.

FIGs. 12A-12C provide graphs of data comparing percent (%) of total reads having a A-to-G nucleotide edit at genomic positions corresponding to positions A11 and A12 of the gRNA using the base editors. **FIG. 12A:** % of total reads showing A-to-G base editing at individual positions. This includes the reads that show base editing without indels plus the reads that harbor base editing as well as the indels. **FIG. 12B:** % of total reads showing only base editing at individual positions. This only includes the reads that show base editing but no indels. **FIG. 12C:** % of total reads showing base editing at only one specific position (left) and % of total reads showing indels (right). This excludes all the reads with base editing at multiple positions as well as the reads harboring indels.

Table 7. Cas12a variant for use in A-to-G base editors*

Variant name	Mutation	Indels (Normalized to TBN04)	Total base editing (Normalized to TBN04)	Only base editing without indels (Normalized to TBN04)	Unique base editing at single position without indels (Normalized to TBN04)
Inactive Cas12a	D832A	2.9	65.6	109.2	28.2
TBN04	K932G, N933G	100	100	100	100
LbAA9	R833L	101.7	60.1	35.2	34.7
LbAA19	R833K	105.4	94	94.4	92.3
LbEF1s9	R833M	111.6	69	40.5	38.7
LbAA23	K932E	88.6	82.6	84.8	78.8
LbMS07	Q944K	108.2	90.4	83.9	87.1
LbAA49	K940G	130.4	86.8	62.5	58
LbAC10	Q944K	189.5	119.4	98.2	84.5
LbMS3n5	K932G, N933G, V936G, S929G	106.3	92.5	90.8	87.7
LbTN37	K940G, Q944K	37.7	75.7	100.9	83.8
LbTN39	R836G, Q944K	21,3	26,3	83,3	61,393,7
LbTN2	R833M,	52,9	105,4	132,3	77,8

Variant name	Mutation	Indels (Normalized to TBN04)	Total base editing (Normalized to TBN04)	Only base editing without indels (Normalized to TBN04)	Unique base editing at single position without indels (Normalized to TBN04)
	E835D, Y943T				
LbFM14	R836G, Q944K, R935G	0,7	65,4	112	34,3
LbFM17	R833M, E835D, Y943T, R935G	1,2	66,7	113,8	24,8
LbFM28	R833M, E835D, Y943T, Q941K	1	93,7	160,2	27,5
LbFM44	R833M, E835D, E125A	23,2	94,5	146,8	71,7
LbFM51	Y943F, Q944K, K932G, N933G, E125A	9,3	113,8	199,9	86,7
LbFM64	R836G, Q944K, R935G, E125A	0,6	145,8	262,0	85,4
LbFM65	R833M, E835D, Y943T, R935G, E125A	0,4	131,2	235,8	43,1
LbFM67	R833M, E835D, Y943T, Q941K, E125A	0,4	134,5	241,4	42,9
LbFM76	D832A, Y943F, Q944K, K932G, N933G, E125A	1,0	153,8	276,1	65,8

*based on Base Editor LbABE8e SEQ ID NO: 177 and gRNA SEQ ID NO: 189 in HEK 293T cells. Editing efficiencies are corresponding to A-to-G base editing at position A11.

Example 6. Efficiency and Specificity of A-to-C Base Editors

The N-methyl purine glycosylase (MPG) coding sequence was inserted at the N-terminus (LbABE8e_nMPG) or at the C-terminus (LbABE8e_cMPG) of the LbABE8e construct (A-to-G base editor). The TadA domain functions to convert adenine to inosine. In the absence of MPG, the inosine is then replaced by guanine during DNA replication and DNA repair. However, in the presence of MPG, inosine is instead removed by MPG as part of the base excision (BER) repair pathway to form an AP site, which is further processed to result in an A-to-C conversion.

The base editing efficiency and specificity of these A-to-C base editors (containing inactive LbCas12a or TBN04 (LbCas12a variant)) were determined. HEK 293T cells were transfected using a plasmid expressing the base editor together with a plasmid expressing the gRNA. Transfected cells were sorted after 3 days and re-cultured for another 3 days after which the cells were lysed and genomic DNA was extracted. Genomic region around the gRNA was PCR amplified and editing was analyzed by next generation sequencing.

The A-to-C base editor containing the TBN04 and MPG (TBN04_nMPG and TBN04_cMPG) showed higher A-to-C base editing efficiency than base editors containing inactive Cas12a. The A-to-C base editors demonstrated high selectivity for specific sites. For gRNA TGCTGCA₇A₈GTA₁₁A₁₂GCA₁₅TGCATTTG (SEQ ID NO: 189), the base editors showed selective editing at A₁₁ (FIG. 13A). For gRNA GTTTA₅A₆A₇CA₉CA₁₁CCGGGTTA₁₉A₂₀TA₂₂A₂₃ (SEQ ID NO: 190), the base editors showed selective editing at A₉ (FIG. 13B).

FIG. 13A: % of total reads showing A-to-C, A-to-G, or A-to-T base editing at position A₁₁ of TGCTGCA₇A₈GTA₁₁A₁₂GCA₁₅TGCATTTG (SEQ ID NO: 189). **FIG. 13B:** % of total reads showing A-to-C, A-to-G, or A-to-T base editing at position A₉ of GTTTA₅A₆A₇CA₉CA₁₁CCGGGTTA₁₉A₂₀TA₂₂A₂₃ (SEQ ID NO: 190).

Construct structures

The structure of LbABE8e_nMPG is as follows:

MPG-Linker5-BP NLS-TadA-Linker4-LbCas12a-Linker2-BP NLS; wherein LbCas12a can be inactive LbCas12a or TBN04 (LbCas12a variant).

The structure of LbABE8e_cMPG is as follows:

BP NLS-TadA-Linker4-LbCas12a-Linker5-MPG-Linker2-BP NLS; wherein LbCas12a can be inactive LbCas12a or TBN04 (LbCas12a variant).

All references, patents and patent applications disclosed herein are incorporated by reference with respect to the subject matter for which each is cited, which in some cases may encompass the entirety of the document.

5 The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

10 In the claims, as well as in the specification above, all transitional phrases such as “comprising,” “including,” “carrying,” “having,” “containing,” “involving,” “holding,” “composed of,” and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases “consisting of” and “consisting essentially of” shall be closed or semi-closed transitional phrases, respectively, as set forth in the United
15 States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

The terms “about” and “substantially” preceding a numerical value mean $\pm 10\%$ of the recited numerical value.

Where a range of values is provided, each value between and including the upper and lower ends of the range are specifically contemplated and described herein.

20

What is claimed is:

CLAIMS

1. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising a mutation at an amino acid position corresponding to position E95, E125, N256,
5 R747, H759, N813, K932, N933, S934, V936, S982, or K984 with reference to amino acid position numbering of LbCas12a ND2006.
2. The engineered variant Cas12a endonuclease of claim 1, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position E95R, E95Y, E125A, E125W, N256A, R747Y, H759V, H759D, N813R, N813H, K932L, N933E,
10 N933V, S934Q, V936E, V936M, V936K, S982N, or K984R.
3. The engineered variant Cas12a endonuclease of claim 1 or 2, wherein the endonuclease exhibits hyperactivity.
4. The engineered variant Cas12a endonuclease of any one of the preceding claims comprising a mutation at an amino acid position corresponding to position E95, optionally
15 wherein the mutation is E95R or E95Y, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 48 or 49.
5. The engineered variant Cas12a endonuclease of any one of the preceding claims comprising a mutation at an amino acid position corresponding to position E125, optionally
20 wherein the mutation is E125A or E125W, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 50 or 51.
6. The engineered variant Cas12a endonuclease of any one of the preceding claims comprising a mutation at an amino acid position corresponding to position N256, optionally
25 wherein the mutation is N256A, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 52 or 53.
7. The engineered variant Cas12a endonuclease of any one of the preceding claims comprising a mutation at an amino acid position corresponding to position R747, optionally
30 wherein the mutation is R747Y, further optionally wherein the polypeptide sequence has at

least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 54.

8. The engineered variant Cas12a endonuclease of any one of the preceding claims comprising a mutation at an amino acid position corresponding to position H759, optionally
5 wherein the mutation is H759V or H759D, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 55 or 56.

9. The engineered variant Cas12a endonuclease of any one of the preceding claims comprising a mutation at an amino acid position corresponding to position N813, optionally
10 wherein the mutation is N813R or N813H, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 57, 58, or 59.

10. The engineered variant Cas12a endonuclease of any one of the preceding claims comprising a mutation at an amino acid position corresponding to position K932, optionally
15 wherein the mutation is K932L, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 62.

11. The engineered variant Cas12a endonuclease of any one of the preceding claims comprising a mutation at an amino acid position corresponding to position N933, optionally
20 wherein the mutation is N933E or N933V, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 76 or 77.

12. The engineered variant Cas12a endonuclease of any one of the preceding claims comprising a mutation at an amino acid position corresponding to position S934, optionally
25 wherein the mutation is S934Q, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 79.

13. The engineered variant Cas12a endonuclease of any one of the preceding claims comprising a mutation at an amino acid position corresponding to position V936, optionally
30 wherein the mutation is V936E, V936M, or V936K, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 82, 83, or 84.

14. The engineered variant Cas12a endonuclease of any one of the preceding claims comprising a mutation at an amino acid position corresponding to position S982, optionally wherein the mutation is S982N, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 92.
15. The engineered variant Cas12a endonuclease of any one of the preceding claims comprising a mutation at an amino acid position corresponding to position K984, optionally wherein the mutation is K984R, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 96.
16. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising a mutation at an amino acid position corresponding to position N256, I831, K932, N933, S934, V936, Q944, S982, F983, K984, M986, or T988 with reference to amino acid position numbering of LbCas12a ND2006.
17. The engineered variant Cas12a endonuclease of claim 16, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position N256K, I831A, I831Y, K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, K932Y, N933L, S934W, V936G, Q944D, Q944E, Q944K, Q944M, S982T, S982W, F983G, F983L, K984F, M986G, M986L, M986S, or T988F.
18. The engineered variant Cas12a endonuclease of claim 16 or 17, wherein the endonuclease exhibits hypoactivity.
19. The engineered variant Cas12a endonuclease of any one of claims 16-18 comprising a mutation at an amino acid position corresponding to position N813, optionally wherein the mutation is N813H, N813R, or N813W, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 57, 58, or 59.
20. The engineered variant Cas12a endonuclease of any one of claims 16-19 comprising a mutation at an amino acid position corresponding to position I831, optionally wherein the mutation is I831A or I831Y, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 60 or 61.

21. The engineered variant Cas12a endonuclease of any one of claims 16-20 comprising a mutation at an amino acid position corresponding to position K932, optionally wherein the mutation is K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 68, 66, 69, 65, 70, 71, 67, 72, 73, 75, or 74.
22. The engineered variant Cas12a endonuclease of any one of claims 16-21 comprising a mutation at an amino acid position corresponding to position N933, optionally wherein the mutation is N933L, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 78.
23. The engineered variant Cas12a endonuclease of any one of claims 16-22 comprising a mutation at an amino acid position corresponding to position S934, optionally wherein the mutation is S934W, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 81.
24. The engineered variant Cas12a endonuclease of any one of claims 16-23 comprising a mutation at an amino acid position corresponding to position V936, optionally wherein the mutation is V936G, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 85.
25. The engineered variant Cas12a endonuclease of any one of claims 16-24 comprising a mutation at an amino acid position corresponding to position Q944, optionally wherein the mutation is Q944D, Q944E, Q944K, Q944M, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 86, 87, 88, or 89.
26. The engineered variant Cas12a endonuclease of any one of claims 16-25 comprising a mutation at an amino acid position corresponding to position S982, optionally wherein the mutation is S982T or S982W, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 90 or 90.

27. The engineered variant Cas12a endonuclease of any one of claims 16-26 comprising a mutation at an amino acid position corresponding to position F983, optionally wherein the mutation is F983G or F983L, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 94 or 93.
28. The engineered variant Cas12a endonuclease of any one of claims 16-27 comprising a mutation at an amino acid position corresponding to position K984, optionally wherein the mutation is K984F, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 95.
29. The engineered variant Cas12a endonuclease of any one of claims 16-28 comprising a mutation at an amino acid position corresponding to position M986, optionally wherein the mutation is M986G, M986L, or M986S, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 97, 99, or 100.
30. The engineered variant Cas12a endonuclease of any one of claims 16-29 comprising a mutation at an amino acid position corresponding to position T988, optionally wherein the mutation is T988F, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 101.
31. The engineered variant Cas12a endonuclease of any one of claims 16-18, wherein the polypeptide sequence comprises mutations at an amino acid positions corresponding to positions: K932F and F983L; K932F and T988F; K932R and Q944D; K932R and F983L; K932R and T988F; K932Y and F983L; K932Y and T988F; N933L and Q944M; V936G and Q944D; V936G and S982W; V936G and M986G; V936G and T988F; Q944D and S982W; Q944D and F983L; Q944D and T988F; S982W and F983L; S982W and T988F; or F983G and M986G.
32. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations K932F and F983L, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 102.

33. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations K932F and T988F, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 103.
- 5 34. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations K932R and Q944D, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 104.
- 10 35. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations K932R and F983L, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 105.
- 15 36. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations K932R and T988F, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 106.
- 20 37. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations K932Y and F983L, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 107.
38. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations K932Y and T988F, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 108.
- 25 39. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations N933L and Q944M, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 109.
- 30 40. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations V936G and Q944D, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 110.

41. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations V936G and S982W, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 111.
- 5 42. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations V936G and M986G, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 112.
43. The engineered variant Cas12a endonuclease of claim 31 comprising the following
10 mutations V936G and T988F, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 113.
44. The engineered variant Cas12a endonuclease of claim 31 comprising the following
15 mutations Q944D and S982W, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 114.
45. The engineered variant Cas12a endonuclease of claim 31 comprising the following
20 mutations Q944D and F983L, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 115.
46. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations Q944D and T988F, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 116.
- 25 47. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations S982W and F983L, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 117.
48. The engineered variant Cas12a endonuclease of claim 31 comprising the following
30 mutations S982W and T988F, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 118.

49. The engineered variant Cas12a endonuclease of claim 31 comprising the following mutations F983G and M986G, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 119.
- 5 50. An engineered variant Cas12a endonuclease comprising a polypeptide sequence comprising a mutation at an amino acid position corresponding to position N813, I831, K932, N933, S934, V936, Q944, S982, F983, K984, M986, or T988 with reference to amino acid position numbering of LbCas12a ND2006.
- 10 51. The engineered variant Cas12a endonuclease of claim 50, wherein the polypeptide sequence comprises a mutation at an amino acid position corresponding to position N813H, N813R, N813W, I831A, I831Y, K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, K932Y, N933E, N933L, S934K, S934Q, V936E, V936G, Q944D, Q944E, Q944K, S982W, F983G, F983L, K984F, M986F, M986G, or T988F.
- 15 52. The engineered variant Cas12a endonuclease of claim 50 or 51, wherein the endonuclease exhibits low indiscriminate ssDNase activity.
- 20 53. The engineered variant Cas12a endonuclease of any one of claims 50-52 comprising a mutation at an amino acid position corresponding to position N813, optionally wherein the mutation is N813H, N813R, or N813W, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 59, 58, or 57.
- 25 54. The engineered variant Cas12a endonuclease of any one of claims 50-53 comprising a mutation at an amino acid position corresponding to position I831, optionally wherein the mutation is I831A or I831Y, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 60 or 61.
- 30 55. The engineered variant Cas12a endonuclease of any one of claims 50-54 comprising a mutation at an amino acid position corresponding to position K932, optionally wherein the mutation is K932A, K932F, K932H, K932M, K932N, K932Q, K932R, K932S, K932T, K932W, or K932Y, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 68, 66, 69, 65, 70, 71, 67, 72, 73, 75, or 74.

56. The engineered variant Cas12a endonuclease of any one of claims 50-55 comprising a mutation at an amino acid position corresponding to position N933, optionally wherein the mutation is N933E or N933L, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 76 or 78.

57. The engineered variant Cas12a endonuclease of any one of claims 50-56 comprising a mutation at an amino acid position corresponding to position S934, optionally wherein the mutation is S934K or S934Q, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 80 or 79.

58. The engineered variant Cas12a endonuclease of any one of claims 50-57 comprising a mutation at an amino acid position corresponding to position V936, optionally wherein the mutation is V936E or V936G, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 82 or 85.

59. The engineered variant Cas12a endonuclease of any one of claims 50-58 comprising a mutation at an amino acid position corresponding to position Q944, optionally wherein the mutation is Q944D, Q944E, or Q944K, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 86, 87, or 88.

60. The engineered variant Cas12a endonuclease of any one of claims 50-59 comprising a mutation at an amino acid position corresponding to position S982, optionally wherein the mutation is S982W, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 90.

61. The engineered variant Cas12a endonuclease of any one of claims 50-60 comprising a mutation at an amino acid position corresponding to position F983, optionally wherein the mutation is F983G or F983L, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 94 or 93.

62. The engineered variant Cas12a endonuclease of any one of claims 50-61 comprising a mutation at an amino acid position corresponding to position K984, optionally wherein the

mutation is K984F, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 95.

63. The engineered variant Cas12a endonuclease of any one of claims 50-62 comprising a
5 mutation at an amino acid position corresponding to position M986, optionally wherein the mutation is M986F or M986G, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 98 or 97.

64. The engineered variant Cas12a endonuclease of any one of claims 50-63 comprising a
10 mutation at an amino acid position corresponding to position T988, optionally wherein the mutation is T988F, further optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 101.

65. The engineered variant Cas12a endonuclease of claim 50, wherein the polypeptide
15 sequence comprises mutations at an amino acid positions corresponding to positions: N933L and Q944M; or F983G and M986G.

66. The engineered variant Cas12a endonuclease of claim 65 comprising the following
mutations N933L and Q944M, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid
20 sequence of SEQ ID NO: 109.

67. The engineered variant Cas12a endonuclease of claim 65 comprising the following
mutations F983G and M986G, optionally wherein the polypeptide sequence has at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or 100% identity to the amino acid sequence of SEQ ID NO: 119.

25 68. A polynucleotide encoding an engineered variant Cas12a endonuclease of any one of the preceding claims.

69. A cell comprising (a) an engineered variant Cas12a endonuclease of any one of the preceding claims or a polynucleotide endonuclease of any one of the preceding claims and (b) a guide RNA or a polynucleotide encoding a guide RNA.

70. A method comprising introducing into a cell (a) an engineered variant Cas12a endonuclease of the present disclosure or a polynucleotide of any one of the preceding claims and optionally (b) a guide RNA or a polynucleotide encoding a guide RNA.
71. Use of an engineered variant Cas12a endonuclease of any one of the preceding claims
5 for cleaving a nucleic acid.
72. A method for introducing a double strand break in a target nucleic acid comprising introducing into a cell comprising a target nucleic acid (a) an engineered variant Cas12a endonuclease of any one of the preceding claims and (b) a guide RNA and incubating the cell to produce a double strand break in the target nucleic acid.
- 10 73. A method for introducing a double strand break in a target nucleic acid comprises introducing into a cell comprising a target nucleic acid (a) an engineered variant Cas12a endonuclease of any one of the preceding claims and (b) a guide RNA and incubating the cell to produce a double strand break in the target nucleic acid.
74. The method of claim 73, wherein the off-target single strand nucleic acid cleavage in
15 the cell is reduced relative to off-target single strand nucleic acid cleavage in a control cell comprising a wild-type Cas12a endonuclease and a guide RNA.
75. A method for introducing a single strand break in a target nucleic acid, comprises introducing into a cell comprising a target nucleic acid (a) an engineered variant Cas12a endonuclease of any one of the preceding claims (b) a guide RNA, and incubating the cell to
20 produce a single strand break in the target nucleic acid.
76. A fusion protein comprising an engineered variant Cas12a endonuclease of any one of the preceding claims and a base editing enzyme.
77. A fusion protein comprising the amino acid sequence of any one of SEQ ID NOs: 163-185 and 388-408 or a variant thereof having at least 80%, at least 85%, at least 90%, at
25 least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the amino acid sequence of any one of SEQ ID NOs: 163-185 and 388-408.
78. A fusion protein comprising an engineered variant Cas12a endonuclease an amino acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the amino acid sequence of any one
30 of SEQ ID NOs: 367-387.

79. The fusion protein of any one of the preceding claims, wherein the base editing enzyme is capable of converting a purine into a different purine or a pyrimidine into a different pyrimidine.
80. The fusion protein of claim 79, wherein the base editing enzyme comprises a
5 deaminase, a guanine oxidase, or a guanine methyltransferase.
81. The fusion protein of claim 79, wherein the deaminase is a cytidine deaminase or an adenosine deaminase.
82. The fusion protein of claim 79, wherein the deaminase comprises a rAPOBEC1 polypeptide, an evoAPOBEC1 polypeptide, a hAPOBEC3A polypeptide, an evoCDA
10 polypeptide, an evoFERNY polypeptide, or a TadA polypeptide.
83. The fusion protein of any one of the preceding claims, further comprising a uracil glycosylase inhibitor (UGI).
84. The fusion protein of any one of the preceding claims, further comprising one or more nuclear localization signal (NLS), optionally selected from an SV40 NLS, a nucleoprotein
15 (NP) NLS, and a bipartite (BP) NLS.
85. The fusion protein of any one of the preceding claims, further comprising a uracil DNA glycosylase (UNG), optionally a human UNG (hUNG) or an Escherichia coli UNG (eUNG).
86. The fusion protein of any one of the preceding claims, further comprising a N-methyl
20 purine glycosylase (MPG), optionally wherein the MPG is positioned at or near the N-terminal or C-terminal ends of the fusion protein.
87. The fusion protein of any one of the preceding claims, further comprising one or more linker.
88. The fusion protein of any one of the preceding claims, wherein the linker comprises
25 the sequence of SGSETPGTSESATPES (SEQ ID NO: 203).
89. The fusion protein of any one of claims 1-10, wherein the linker comprises the sequence of SGGSSGGSSGSETPGTSESATPESGGSSGGS (SEQ ID NO: 204)
90. The fusion protein of any one of the preceding claims, further comprising a DNA binding domain (DBD).
- 30 91. The fusion protein of claim 90, wherein the DBD is a Rad51 DBD.

92. A polynucleotide encoding the fusion protein of any one of the preceding claims.
93. A cell comprising:
a target nucleic acid comprising a target strand and a non-target strand;
a guide RNA (gRNA) or a nucleic acid encoding a gRNA that binds to the target
5 strand; and
the fusion protein of any one of the preceding claims or the polynucleotide of any one
of the preceding claims.
94. The cell of claim 93, wherein the cell is a human cell.
95. The cell of claim 93, wherein the human cell is from a human subject, wherein the
10 human subject has a disease, disorder or condition associated with the target nucleic acid.
96. The cell of claim 93, wherein the cell is a plant cell.
97. A method comprising introducing into a cell (a) the fusion protein of any one of the
preceding claims or the polynucleotide of claim 92 and optionally (b) a guide RNA or a
polynucleotide encoding a guide RNA.
- 15 98. A method of gene editing comprising
(i) contacting a target nucleic acid sequence with the fusion protein of any one of the
preceding claims and a guide RNA, wherein the target nucleic acid comprises a target
nucleobase; and
(ii) modifying the target nucleobase.
- 20 99. The method of claim 98, wherein the target nucleic acid is a target double-stranded
DNA nucleic acid.
100. The method of claim 98 or 99, wherein the guide RNA directs the fusion protein to
bind to a specific segment of the target nucleic acid and in proximity to the target nucleobase.
101. The method of any one of claims 97-100, wherein the fusion protein cleaves the target
25 nucleic acid.
102. The method of any one of claims 97-101, wherein the fusion protein comprises a
cytidine deaminase and the target nucleobase is a cytidine.
103. The method of any one of claims 97-101, wherein the fusion protein comprises a
guanine oxidase and the target nucleobase is a guanosine.

104. The method of any one of claims 97-101, wherein the fusion protein comprises a guanine methyltransferase and the target nucleobase is a guanosine.
105. The method of any one of claims 97-104, wherein the method is performed in a cell,
5 optionally a human cell or a plant cell.
106. The method of any one of claims 97-105, wherein the method is performed in vitro or ex vivo.
107. The method of any one of claims 97-105, wherein the method is performed in vivo.
108. The method of any one of claims 97-107, wherein the target nucleic acid is a gene
10 comprising a nucleobase mutation relative to a wild-type gene.
109. The method of claim 108, wherein the gene comprising a nucleobase mutation is associated with a disease or disorder.