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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
03 June 2021 (03.06.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/108324 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 15/113* (2010.01)
C12N 15/10 (2006.01) *C12N 15/90* (2006.01)

(21) International Application Number:
PCT/US2020/061850

(22) International Filing Date:
23 November 2020 (23.11.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/941,392 27 November 2019 (27.11.2019) US

(71) Applicant: **TECHNICAL UNIVERSITY OF DENMARK** [DK/DK]; Anker Engelunds Vej 1, Building 101A, 2800 Kgs. Lyngby (DK).

(72) Inventors; and
(71) Applicants: **GILL, Ryan, T.** [US/US]; 2025 Ash Street, Denver, CO 80207 (US). **WARNECKE, Tanya** [US/US]; 460 Oneida Street, Boulder, CO 80303 (US). **JEDRZEJCZYK, Dominika, Joanna** [PL/DK]; Robert Jacobsens Vej 40 A, 2.1, 2300 Copenhagen S. (DK).

(74) Agent: **HANSON, Roberta, J.** et al.; 1401 Lawrence Street, Suite 2300, Denver, CO 80202 (US).

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

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

(54) **Title:** CONSTRUCTS, COMPOSITIONS AND METHODS THEREOF HAVING IMPROVED GENOME EDITING EFFICIENCY AND SPECIFICITY

(57) **Abstract:** Embodiments disclosed herein include novel nucleic acid-guided nucleases, novel guide nucleic acids, and novel targetable nuclease systems, and methods of use. In some embodiments, engineered non-naturally occurring nucleic acid-guided nucleases, can be used with known guide nucleic acids in a targetable nuclease system. In certain embodiments, targetable nuclease systems can be used to edit targeted genomes of humans and other species. In some embodiments, methods include, but are not limited to, recursive genetic engineering and trackable genetic engineering methods.

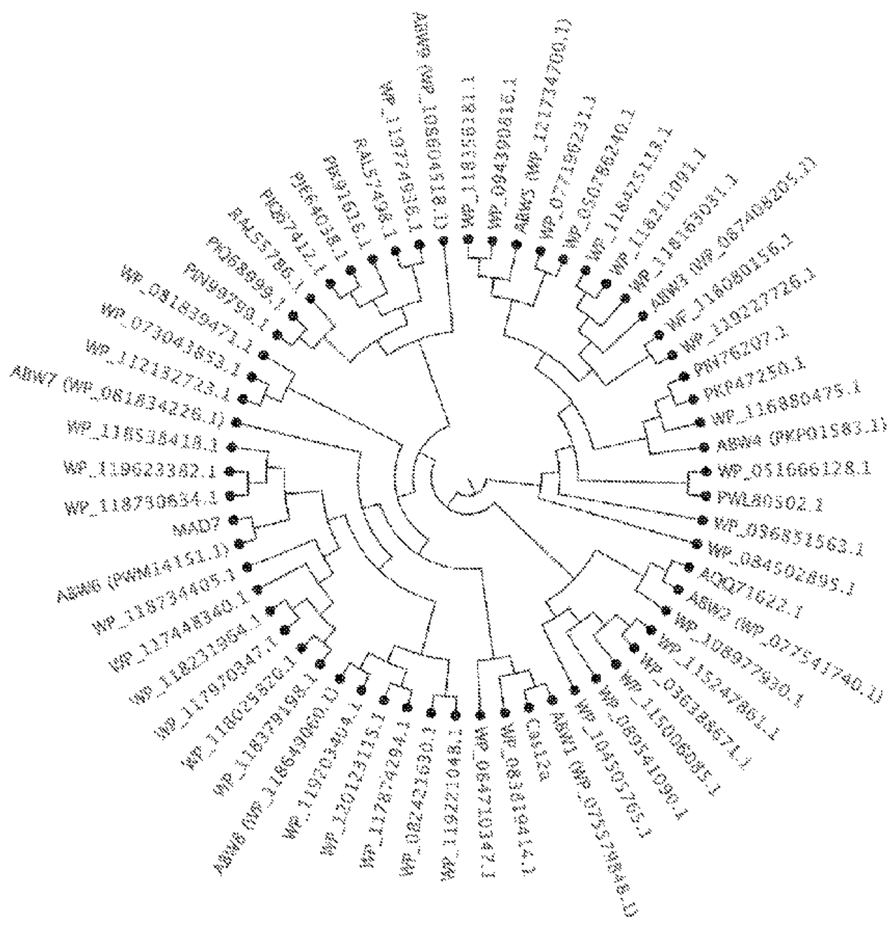


FIG. 1

[Continued on next page]

WO 2021/108324 A1

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

CONSTRUCTS, COMPOSITIONS AND METHODS THEREOF HAVING IMPROVED GENOME EDITING EFFICIENCY AND SPECIFICITY

PRIORITY

[001] This application claims priority to U.S. Provisional Application No. 62/941,392 filed November 27, 2019. This provisional application is incorporated herein by reference in its entirety for all purposes.

STATEMENT REGARDING SEQUENCE LISTING

[002] The instant application contains a Sequence Listing which has been submitted via ASCII copy created on November 21, 2020 referred to as '20201121 105529.640658 Sequence Listing.txt' having 150 sequences.

FIELD

[003] Some embodiments disclosed herein concern novel nucleic acid-guided nucleases, guide nucleic acids (*e.g.* gRNAs), and targetable nuclease systems, and methods of use. In other embodiments, methods for making and using engineered non-naturally occurring nucleic acid-guided nucleases, guide nucleic acids, and targetable nuclease systems are disclosed. In some embodiments, targetable nuclease systems can be used to edit mammalian such as human genomes or genomes of other species.

BACKGROUND

[004] CRISPR is an abbreviation of Clustered Regularly Interspaced Short Palindromic Repeats. In a palindromic repeat, the sequence of nucleotides is the same in both directions. Each of these palindromic repetitions is followed by short segments of spacer DNA. Small clusters of Cas (CRISPR-associated system) genes are located next to CRISPR sequences. The CRISPR/Cas system is a prokaryotic immune system that can confer resistance to foreign genetic elements such as those present within plasmids and phages providing the prokaryote a form of acquired immunity. RNA harboring a spacer sequence assists Cas (CRISPR-associated) proteins to recognize and cut exogenous DNA. CRISPR sequences, found in approximately 50% of bacterial genomes and nearly 90% of sequenced archaea, select for efficient and robust metabolic and regulatory networks that prevent unnecessary metabolite biosynthesis and optimally distribute resources to maximize overall cellular fitness. The complexity of these networks with limited approaches to understand their structure and function, and the ability to re-program cellular networks to modify these systems for a diverse range of applications have complicated

advances in this space. Certain approaches to re-program cellular networks are directed to modifying single genes of complex pathways but as a consequence of modifying single genes, unwanted modifications to the genes or other genes can result, getting in the way of identifying changes necessary to achieve a particular endpoint as well as complicating the endpoint sought by the modification.

[005] CRISPR-Cas driven genome editing and engineering has dramatically impacted biology and biotechnology in general. CRISPR-Cas editing systems require a polynucleotide guided nuclease, a guide polynucleotide (*e.g.* a guide RNA (gRNA)) that directs by homology the nuclease to cut a specific region of the genome, and, optionally, a donor DNA cassette that can be used to repair the cut dsDNA and thereby incorporate programmable edits at the site of interest. The earliest demonstrations and applications of CRISPR-Cas editing used Cas9 nucleases and associated gRNA. These systems have been used for gene editing in a broad range of species encompassing bacteria, plants, to higher order mammalian systems such as animals and in certain cases, humans. It is well established, however, that key editing parameters such as protospacer adjacent motif (PAM) specificity, editing efficiency, and off-target rates, among others, are species, loci, and nuclease dependent. There is increasing interest in identifying and rapidly characterizing novel nuclease systems that can be exploited to broaden and improve overall editing capabilities.

[006] One version of the CRISPR/Cas system, CRISPR/Cas9, has been modified to provide useful tools for editing targeted genomes. By delivering the Cas9 nuclease complexed with a synthetic guide RNA (gRNA) into a cell, the cell's genome can be cut/edited at a predetermined location, allowing existing genes to be removed and/or new ones added. These systems are useful but have some important limitations regarding efficiency and accuracy of targeted editing, imprecise editing complications, as well as, impediments when used for commercially relevant situations such as gene replacement. Therefore, a need exists for improved nucleic acid guided nuclease systems for directed and accurate editing with improved efficiency.

SUMMARY

[007] Some embodiments disclosed herein concern novel and improved nucleic acid-guided nucleases and guide nucleic acids (*e.g.* gRNAs) of use to target genomes such as mammalian genomes for improved genome editing and reduced off-targeting. In certain embodiments, eukaryotic or prokaryotic genomes can be edited using targeted systems disclosed herein. In

other embodiments, systems for using these novel nucleic acid-guided nucleases with known gRNAs or with novel gRNAs disclosed herein are contemplated. In addition, it is contemplated that known nucleic acid-guided nucleases can be used in systems for genome editing that include novel guide nucleic acids (*e.g.* gRNAs) disclosed in the instant application.

[008] In other embodiments, methods for making and using engineered non-naturally occurring nucleic acid-guided nucleases, guide nucleic acids, and targetable nuclease systems are disclosed. In some embodiments, targetable nuclease systems can be used to edit human genomes or genomes of other species. In some embodiments, nucleic acid-guided nucleases of use in compositions, methods and systems disclosed herein can be represented by the amino acid sequence represented by one or more of SEQ ID NO: 94 (ABW8), 29 (ABW3), 81 (ABW7), 107 (ABW9), 3 (ABW1), 16 (ABW2), 42 (ABW4), 55 (ABW5), and 68 (ABW6). In other embodiments, nucleic acid-guided nucleases can be represented by the polynucleotides encoding polypeptides represented by one or more of SEQ ID NO: 95-104 (ABW8 variants 1-10), 30-39 (ABW3 variants 1-10), 82-91 (ABW7 variants 1-10), 108-117 (ABW9 variants 1-10), 4-13 (ABW1 variants 1-10), 17-26 (ABW2 variants 1-10), 43-52 (ABW4 variants 1-10), 56-65 (ABW5 variants 1-10), and 69-78 (ABW6 variants 1-10). In other embodiments, gRNAs of use in compositions and methods disclosed herein can be represented by gRNAs represented by SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128 and can be a split gRNA of use as a synthetic tracrRNA and crRNA.

[009] In some embodiments, a nucleic acid-guided nuclease system can include, but is not limited to, an engineered nucleic acid-guided nuclease; and an engineered guide polynucleotide (gRNA) for complexing with the nucleic acid-guided nuclease, wherein the engineered guide polynucleotide has an amino acid sequence selected from SEQ ID NO: 3, 16, 29, 42, 55, 68, 81, 94 and 107. In certain methods, the target region is eukaryotic genome. In other embodiments, the target region a mammalian genome. In other embodiments, a nucleic acid-guided nuclease system can include an engineered nucleic acid-guided nuclease, wherein the engineered nucleic acid-guided nuclease has a nucleic acid sequence represented by SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and 108-117 and an engineered guide polynucleotide for complexing with the nucleic acid-guided nuclease. In certain embodiments, the target region is a eukaryotic genome. In other embodiments, the target region a mammalian genome. In certain embodiments, the targeted genome is a prokaryotic genome. In some embodiments, mammalian

genomes can include, pets, livestock or other animals. In certain embodiments, mammalian genomes contemplated to be edited by systems disclosed herein can include human genomes for example, adult, children, infant and/or fetal genomes.

[0010] In other embodiments, a nucleic acid-guided nuclease system disclosed herein can include, but is not limited to, an engineered nucleic acid-guided nuclease, wherein the engineered nucleic acid-guided nuclease has an amino acid sequence represented by SEQ ID NO: 3, 16, 29, 42, 55, 68, 81, 94 and 107; and an engineered guide polynucleotide for complexing with the nucleic acid-guided nuclease, wherein the engineered guide polynucleotide includes a nucleic acid sequence represented by SEQ ID NO:118 to SEQ ID NO:126 or SEQ ID NO:128. In other embodiments, the engineered polynucleotide (gRNA) represented by SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128 and can be a split gRNA of use as a synthetic tracrRNA and cfRNA. In certain methods, the target region is a eukaryotic genome. In other embodiments, the target region is a mammalian genome (*e.g.* animal or human genome). In certain embodiments, the targeted genome is a prokaryotic genome.

[0011] In other embodiments, methods for modifying a genome are disclosed. In accordance with these embodiments, methods can include, but are not limited to, contacting a targeted genome with an engineered nucleic acid-guided nuclease; and an engineered guide polynucleotide for complexing with the nucleic acid-guided nuclease, wherein the engineered guide polynucleotide includes a nucleic acid sequence represented by SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128; and allowing the nuclease and gRNA to modify the targeted genome. In some embodiments, the engineered polynucleotide (gRNA) disclosed herein can be split into fragments encompassing a synthetic tracrRNA and crRNA of use in methods for targeting a genome. In other embodiments, methods can further include contacting the targeted genome with a novel engineered nucleic acid-guided nuclease wherein the engineered nucleic acid-guided nuclease has an amino acid sequence represented by SEQ ID NO: 3, 16, 29, 42, 55, 68, 81, 94 and 107 or wherein the engineered nucleic acid-guided nuclease has nucleic acid sequence represented by SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and 108-117. In other embodiments, the engineered guide nucleic acid and an editing sequence are provided as a single nucleic acid. In other embodiments, the editing sequence further includes a protospacer adjacent motif (PAM) site or a mutation in a protospacer adjacent motif (PAM) site.

[0012] In other embodiments, kits are contemplated. In some embodiments, the kit can include an engineered nucleic acid-guided nuclease and a gRNA having a nucleic acid sequence represented by SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128 and a container. In other embodiments, a kit can include an engineered nucleic acid-guided nuclease having a polypeptide sequence represented by SEQ ID NO: 3, 16, 29, 42, 55, 68, 81, 94 and 107 or an engineered nucleic acid-guided nuclease having a nucleic acid sequence represented by SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and 108-117; and a container.

[0013] Other embodiments include methods of modifying a target region in the genome of a cell, the method includes, but is not limited to, contacting a cell with: a non-naturally occurring nucleic-acid-guided nuclease encoded by a nucleic acid having at least 80% identity to one or more of SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and/or 108-117; an engineered guide nucleic acid capable of complexing with the nucleic acid-guided nuclease; and an editing sequence encoding a nucleic acid complementary to said target region having a change in sequence relative to the target region; and permitting the nuclease, guide nucleic acid, and editing sequence to create an edited region in a targeted region of the genome of the cell. In other embodiments, a non-naturally occurring nucleic-acid-guided nuclease encodes an amino acid sequence represented by at least 80% identity to SEQ ID NO: 3, 16, 29, 42, 55, 68, 81, 94 and/or 107. In some embodiments, an engineered guide nucleic acid (*e.g.* gRNA) and the editing sequence are provided as a single nucleic acid construct. In some embodiments, the engineered polynucleotide (gRNA) disclosed herein can be split into fragments encompassing a synthetic tracrRNA and crRNA of use in methods for targeting a genome. In other embodiments, the single nucleic acid construct can include a protospacer adjacent motif (PAM) site and/or a mutation in a protospacer adjacent motif (PAM) site. In some aspects, the nucleic acid-guided nuclease is encoded by a nucleic acid with at least 85% identity to one or more of SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and/or 108-117. In some embodiments, the nucleic acid-guided nuclease is encoded by a nucleic acid having at least 85% identity to SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128.

[0014] In yet other embodiments, nucleic acid-guided nuclease systems are disclosed that include, but are not limited to, a non-naturally occurring nuclease encoded by a nucleic acid having at least 80% identity to one or more of SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65,

69-78, 82-91, 95-104 and/or 108-117; a known engineered guide nucleic acid capable of complexing with the nucleic acid-guided nuclease or a novel engineered guide nucleic acid capable of complexing with the nucleic acid-guided nuclease having at least 85% identity to SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 or 128 and an editing sequence, wherein the system can edit a targeted genome in the target region of the genome of the cell facilitated by the nuclease, the engineered guide nucleic acid, and the editing sequence. In some aspects, the nucleic acid-guided nuclease is encoded by a nucleic acid with at least 85% identity to one or more of SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and/or 108-117. In some embodiments, the nucleic acid-guided nuclease can be codon optimized for the cell to be edited. In other aspects, the engineered guide nucleic acid and the editing sequence are provided as a single nucleic acid. In some aspects, the single nucleic acid further comprises a wild type or mutated proto-spacer adjacent motif (PAM) site.

[0015] In other embodiments, compositions disclosed herein can include a non-naturally occurring nuclease encoded by a nucleic acid having at least 75% identity to one or more of SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and/or 108-117. In some aspects, the nucleic acid has at least 80% identity to one or more of SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and/or 108-117. In some embodiments, the nucleic acid has at least 90% identity to one or more of SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and/or 108-117. In certain embodiments, the nuclease can be codon optimized for use in cells from a particular organism. In certain embodiments, the nuclease is codon optimized for a human genome. In other embodiments, the nuclease is codon optimized for a mammalian genome such as a pet, livestock or other mammal. In certain embodiments, a nuclease disclosed herein can be codon optimized for a bird or fish. In other embodiments, a nuclease disclosed herein can be codon optimized for a plant. In other embodiments, a nuclease disclosed herein can be codon optimized for a prokaryotic genome.

BRIEF DESCRIPTION OF THE FIGURES

[0016] The following drawings form part of the present specification and are included to further demonstrate certain embodiments of the present disclosure. Certain embodiments can be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[0017] **FIG. 1** is an exemplary image illustrating a circular phylogram representing some evolutionary relationships among novel engineered nucleases of some embodiments disclosed herein.

[0018] **FIG. 2** is an exemplary image illustrating novel guide polynucleotide sequences (*e.g.* guide RNA (gRNAs)) used in a DNMT1 amplicon *in vitro* cleavage assay to assess the efficiently of ABW nucleases of some embodiments disclosed herein.

[0019] **FIG. 3** is an exemplary image illustrating an *in vitro* cleavage assay to assess the efficiently of ABW nucleases and cognate gRNAs of some embodiments disclosed herein.

[0020] **FIG. 4** is an exemplary image illustrating an *in vitro* cleavage assay to assess the efficiently of ABW nucleases and Cas12a gRNA of some embodiments disclosed herein.

[0021] **FIG. 5** is an exemplary image illustrating an *in vitro* cleavage assay to assess the efficiently of ABW nucleases and STAR gRNA of some embodiments disclosed herein.

[0022] **FIGS. 6A and 6B** are exemplary images illustrating *in vitro* cleavage assays to assess the efficiently of Cas12a Ultra, LbaCas12a, MAD7, ABW1, ABW5, M21, M44 (**FIG. 6A**) or a Cas12a Ultra, LbaCas12a, MAD7, ABW1, ABW5, ABW8 (**FIG. 6B**) and Cas12a gRNA of some embodiments disclosed herein.

[0023] **FIGS. 7A-7C** are exemplary graphs illustrating Next Generation Sequencing (NGS) data of cleaved TRAC (**FIG. 7A and FIG. 7B**) and DNMT1 (**FIG. 7C**) target sequences resulting from an activity and editing efficiency test performed in Jurkat cells.

[0024] **FIG. 8** is an exemplary image illustrating a T7 endonuclease assay to assess the efficiently of ABW nuclease editing of the DNMT1 gene in Jurkat cells of some embodiments disclosed herein.

[0025] **FIG. 9** is an exemplary image illustrating a T7 endonuclease assay to assess the efficiently of ABW nuclease editing of the TRAC gene in Jurkat cells of some embodiments disclosed herein.

[0026] **FIG. 10** is an exemplary graph illustrating a depletion assay to assess cutting efficiency of an ABW1 nucleic acid-guided nuclease of some embodiments disclosed herein.

[0027] **FIG. 11** is an exemplary graph illustrating a depletion assay to assess cutting efficiency of an ABW4 nucleic acid-guided nuclease of some embodiments disclosed herein.

[0028] **FIG. 12** is an exemplary graph illustrating a depletion assay to assess cutting efficiency of an ABW7 nucleic acid-guided nuclease of some embodiments disclosed herein.

[0029] **FIG. 13** is an exemplary graph illustrating a depletion assay to assess cutting efficiency of an ABW2 nucleic acid-guided nuclease of some embodiments disclosed herein.

[0030] **FIG. 14** is an exemplary graph illustrating a depletion assay to assess cutting efficiency of an ABW5 nucleic acid-guided nuclease of some embodiments disclosed herein.

[0031] **FIG. 15** is an exemplary graph illustrating a depletion assay to assess cutting efficiency of an ABW8 nucleic acid-guided nuclease of some embodiments disclosed herein.

[0032] **FIG. 16** is an exemplary graph illustrating a depletion assay to assess cutting efficiency of an ABW3 nucleic acid-guided nuclease of some embodiments disclosed herein.

[0033] **FIG. 17** is an exemplary graph illustrating a depletion assay to assess cutting efficiency of an ABW6 nucleic acid-guided nuclease of some embodiments disclosed herein.

[0034] **FIG. 18** is an exemplary graph illustrating a depletion assay to assess cutting efficiency of an ABW9 nucleic acid-guided nuclease of some embodiments disclosed herein.

DETAILED DESCRIPTION

[0035] In the following sections, various exemplary compositions and methods are described in order to detail various embodiments of the disclosure. It will be obvious to one of skill in the relevant art that practicing the various embodiments does not require the employment of all or even some of the details outlined herein, but rather that concentrations, times and other details can be modified through routine experimentation. In some cases, well-known methods or components have not been included in the description.

[0036] As used herein, the term "modulating" and "manipulating" of genome editing can mean an increase, a decrease, upregulation, downregulation, induction, a change in editing activity, a change in binding, a change cleavage or the like, of one or more of targeted genes or gene clusters of certain embodiments disclosed herein.

[0037] In certain embodiments of the present disclosure, there can be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature and understood by those of skill in the art.

[0038] In other embodiments, primers used herein for preparation per conventional techniques can include sequencing primers and amplification primers. In some embodiments, plasmids and oligomers used in conventional techniques can include synthesized oligomers and oligomer cassettes.

[0039] In some embodiments disclosed herein, nucleic acid-guided nuclease systems and methods of use are provided. A nuclease system can include transcripts and other elements involved in the expression of an engineered nuclease disclosed herein, which can include sequences encoding a novel engineered nucleic acid-guided nuclease protein and a guide sequence (gRNA) or a novel gRNA as disclosed herein. In some embodiments, nucleic acid-guided nuclease systems can include at least one CRISPR-associated nucleic acid guided nuclease construct, the disclosure of which are provided herein. In other embodiments, nucleic acid-guided nuclease systems can include at least one known guide sequence (gRNA) or at least one novel gRNA. In some embodiments, an engineered nucleic acid-guided nuclease of the instant invention can be used in systems for editing a gene of interest in humans or other species.

[0040] Bacterial and archaeal targetable nuclease systems have emerged as powerful tools for precision genome editing. However, naturally occurring nucleases have some limitations including expression and delivery challenges due to the nucleic acid sequence and protein size. In certain embodiments, novel engineered nucleic acid-guided nuclease constructs disclosed herein can be created for altered targeting of a targeted gene and/or increased efficiency and/or accuracy of targeted gene editing in a subject.

[0041] In accordance with these embodiments, it is known that Cas12a is a single RNA-guided CRISPR/Cas endonuclease capable of genome editing having differing features when compared to Cas9. Compared to other known Cas nucleases, Cas12a nucleases can process gRNAs from a transcribed CRISPR array lacking accessory factors (e.g. tracrRNA), recognize T-rich PAMs located 5′ of the displaced strand of target DNA, utilize a RuvC endonucleolytic domain to nick both strands of target DNA, and/or can non-specifically cleave single-stranded DNA upon target recognition. In certain embodiments, a Cas12a-based system disclosed herein can allow for fast and reliable introduction of donor DNA into a genome. In some embodiments, a Cas12a-based system disclosed herein can broaden genome editing. CRISPR/Cas12a genome editing has been evaluated in human cells as well as other organisms including plants.

[0042] It is known that a Cas12a nuclease recognizes T-rich protospacer adjacent motif (PAM) sequences (*e.g.* 5'-TTTN-3' (AsCas12a, LbCas12a) and 5'-TTN-3' (FnCas12a); whereas, the comparable sequence for SpCas9 is NGG. The PAM sequence of Cas12a is located at the 5' end of the target DNA sequence, where it is at the 3' end for Cas9. In addition, Cas12a is capable of cleaving DNA distal to its PAM around the +18/+23 position of the protospacer.

This cleavage creates a staggered DNA overhang (*e.g.* sticky ends), whereas Cas9 cleaves close to its PAM after the 3' position of the protospacer at both strands and creates blunt ends. In certain methods, creating altered recognition of nucleases can provide an improvement over Cas9 or Cas12a to improve accuracy. Further, Cas12a is guided by a single crRNA and does not require a tracrRNA, resulting in a shorter gRNA sequence than the gRNA used by Cas9.

[0043] It is also known that Cas12a displays additional ribonuclease activity that functions in crRNA processing. Cas12a is used as an editing tool for different species (*e.g. S. cerevisiae*), allowing the use of an alternative PAM sequence compared with the one recognized by CRISPR/Cas9. Novel nucleases disclosed herein can further recognize the same or alternative PAM sequences. These novel nucleases can provide an alternative system for multiplex genome editing as compared with known multiplex approaches and can be used as an improved system in mammalian gene editing.

[0044] Well-known Cas12a protein–RNA complexes recognize a T-rich PAM and cleavage leads to a staggered DNA double-stranded break. Cas12a-type nuclease interacts with the pseudoknot structure formed by the 5′-handle of crRNA. A guide RNA segment, composed of a seed region and the 3′ terminus, possesses complementary binding sequences with the target DNA sequences. Cas12a type nucleases characterized to date have been demonstrated to work with a single gRNA and to process gRNA arrays. While Cas12a-type and Cas9 nuclease systems have proven highly impactful, neither system has been demonstrated to function as predictably as is desired to enable the full range of applications envisioned for gene-editing technologies.

[0045] In the current state, a range of efforts have attempted to engineer improved CRISPR editing systems having increased efficiency and accuracy, which have included engineering of the PAM specificity, stability, and sequence of the gRNA and-or the nuclease. For example, chemical modifications of CRISPR/Cas9 gRNA expected to increase gRNA stability was found to lead to a 3.8-fold higher indel frequencies in human cells. In addition, other studies included structure-guided mutagenesis of Cas12a and screened to identify variants with an increased range of recognized PAM sequences. These engineered AsCas12a recognized TYCV and TATV PAMs in addition to the established TTTV sequence, with enhanced activities *in vitro* and in tested human cells.

[0046] In other embodiments, Cas12a-like nucleases and engineered gRNAs disclosed herein are contemplated of use in bacteria, yeast, Archaea, and other prokaryotes. In other

embodiments, engineered designer nucleases are contemplated of use in eukaryotes such as mammals as well as of use in birds and fish. In other embodiments, engineered designer nucleases are contemplated of use in plants. In accordance with these embodiments, these constructs are created in order to alter certain features of the wild-type gRNA sequences while preserving other desirable features compared to the control the gRNAs are derived from.

[0047] In certain embodiments, engineered gRNA constructs of embodiments disclosed herein can be created from Cas12as gRNAs known in the art or not yet discovered and can include, but are not limited to, *Acidaminococcus massiliensis sp.* (*e.g.* AM _Cas12a strain Marseille-P2828), *Sedimentisphaera cyanobacteriorum sp.* (SC_Cas12a, strain L21-RPul-D3), *Barnesiella sp.* An22 (B_Cas12a; An22 An22), *Bacteroidetes bacterium HGW-Bacteroidetes-6 sp*. XS5, (BB_Cas12a, 08E140C01), *Parabacteroides distasonis sp.* (PD_Cas12a, strain 8-P5) *Collinsella tanakaei sp.* (CT_Cas12a, isolate CIM:MAG 294), *Lachnospiraceae bacterium* MC2017 *sp.* (LB_Cas12a, T350), *Coprococcus sp*. AF16-5 (Co_Cas12a, AF16-5 AF16-5.Scaf1), or *Catenovulum sp.* CCB-QB4 (Ca_Cas12a, species CCB-QB4) Eubacterium rectale,(a positive control is a derivative of this Cas12a), Flavobacterium branchiophilum (FB_Cas12a), and/or a synthetic construct (SC_Cas12a) or similar. In certain embodiments, constructs can include 60% or less identity to a known Cas12a to create a novel nuclease. In certain embodiments, novel Cas12a derived constructs can include constructs with reduced off-targeting rates and/or improved editing functions compared to a control or wild-type Cas12a nuclease.

[0048] In some embodiments, off-targeting rates for nuclease constructs disclosed herein can be reduced compared to a control for improved editing. For example, off-targeting rates can be readily tested. In accordance with these embodiments, a wild-type gRNA plasmid can be used to assess baseline off-target editing compared to experimentally designed gRNAs to assess accuracy of novel nucleases compared to control Cas12a nucleases or other nucleases known in the art as a positive control (*e.g.* MAD7). In certain methods, spacer mutations can be introduced to a plasmid to test when a substitution gRNA sequence is created or a deletion or insertion mutant. Each of these plasmid constructs can be used to test genome editing accuracy and efficiency, for example, with deletions, substitutions or insertions.

[0049] In certain embodiments, spacer mutations can be introduced to a plasmid to test when a substitution gRNA sequence is created or a deletion or insertion mutant is created. Each of these plasmid constructs can be used to test genome editing accuracy and efficiency, for

example, having a deletion, substitution or insertion. Alternatively, in some embodiments, nuclease constructs created by compositions and methods disclosed herein can be tested for optimal genome editing time on a select target by observing editing efficiencies over pre-determined time periods. In accordance with these embodiments, nuclease constructs created by compositions and methods disclosed herein can be tested for optimal genome editing windows to optimize editing efficiency and accuracy.

[0050] In some embodiments, nuclease constructs created by compositions and methods disclosed herein having optimal genome editing efficiency and accuracy are an improvement over control nuclease constructs. In some embodiments, nuclease constructs created by compositions and methods disclosed herein can have at least a 10% increase, a 15% increase, a 20% increase or more in enzymatic activity, efficiency and/or accuracy compared to control nucleases. In other embodiments, nuclease constructs created by compositions and methods disclosed herein can have about 10% to about 99.5% or more increase in enzymatic activity and/or editing efficiency and/or editing accuracy compared to nucleases having a native sequence compared to nucleases disclosed herein. In some embodiments, nuclease constructs disclosed herein having increased enzymatic activity and/or editing efficiency compared to control nuclease sequences can have a polypeptide sequence having at least 85% homology to the polypeptide represented by SEQ ID NO: 94 (ABW8), 29 (ABW3), 81 (ABW7), 107 (ABW9), 3 (AWBW1), 16 (AWBW2), 42 (AWBW4), 55 (AWBW5), and/or 68 (AWBW6). In some embodiments, nuclease constructs herein having increased enzymatic activity and/or editing efficiency and/or accuracy compared to control nuclease sequences can have a polynucleotide sequence at least 85% homologous to the polynucleotide encoding the polypeptide having a polynucleotide represented by SEQ ID NO: 95-104 (ABW8 variants 1-10), 30-39 (ABW3 variants 1-10), 82-91 (ABW7 variants 1-10), 108-117 (ABW9 variants 1-10), 4-13 (ABW1 variants 1-10), 17-26 (ABW2 variants 1-10), 43-52 (ABW4 variants 1-10), 56-65 (ABW5 variants 1-10), and/or 69-78 (ABW6 variants 1-10).

[0051] In some embodiments, nuclease constructs herein having a polypeptide of at least 85% homology to the polypeptide represented SEQ ID NO: 94 (ABW8) can have increased activity and/or editing accuracy compared to other nuclease constructs. In some embodiments, nuclease constructs herein having a polypeptide of at least 85% homology to the polypeptide represented by SEQ ID NO: 94 (ABW8), 29 (ABW3), 81 (ABW7) and/or 107 (ABW9) can have increased

enzymatic activity and/or editing efficiency and/or accuracy compared to other nuclease constructs such as control nuclease constructs or native sequence-containing nucleases.

[0052] In some embodiments, nuclease constructs disclosed herein having a polynucleotide encoding a polypeptide having a polynucleotide of at least 85% homology to a polynucleotide represented by SEQ ID NO: 95-104 (ABW8 variants 1-10) can have increased enzymatic activity and/or editing efficiency and/or accuracy compared to control nuclease constructs or nuclease constructs having native sequences. In some embodiments, nuclease constructs disclosed herein having a polynucleotide encoding a polypeptide of at least 85% homology to a polynucleotide represented by SEQ ID NO: 95-104 (ABW8 variants 1-10), 30-39 (ABW3 variants 1-10) or 82-91 (ABW7 variants 1-10) can have increased activity (*e.g.* editing and/or efficiency) compared to control nuclease constructs or other nuclease constructs.

[0053] Examples of target polynucleotides for use with engineered nucleic acid guided nucleases disclosed herein can include a sequence/gene or gene segment associated with a signaling biochemical pathway, *e.g.*, a signaling biochemical pathway-associated gene or polynucleotide. Other embodiments contemplated herein concern examples of target polynucleotides related to a disease-associated genes or polynucleotides.

[0054] A "disease-associated" or “disorder-associated” gene or polynucleotide can refer to any gene or polynucleotide which results in a transcription or translation product at an abnormal level compared to a control or results in an abnormal form in cells derived from disease-affected tissues compared with tissues or cells of a non-disease control. It can be a gene that becomes expressed at an abnormally high level; it can be a gene that becomes expressed at an abnormally low level, or where the gene contains one or more mutations and where altered expression or expression directly correlates with the occurrence and/or progression of a health condition or disorder. A disease or disorder-associated gene can refer to a gene possessing mutation(s) or genetic variation that are directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the cause or progression of a disease or disorder. The transcribed or translated products can be known or unknown, and can be at a normal or abnormal level.

[0055] It is understood by one of skill in the relevant art that examples of disease-associated genes and polynucleotides are available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web.

[0056] **Genetic Disorders** contemplated herein can include, but are not limited to:

[0057] **Neoplasia**: Genes linked to this disorder: PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; Notchl; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF; HIFI a; HIF3a; Met; HRG; Bc12; PPAR alpha; PPAR gamma; WT1 (Wilms Tumor); FGF Receptor Family members (5 members: 1, 2, 3, 4, 5); CDKN2a; APC; RB (retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR (Androgen Receptor); TSG101; IGF; IGF Receptor; Igfl (4 variants); Igf2 (3 variants); Igf 1 Receptor; Igf 2 Receptor; Bax; Bc12; caspases family (9 members:1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc;

[0058] **Age-related Macular Degeneration**: Genes linked to these disorders Abcr; Cc12; Cc2; cp (cemloplasmin); Timp3; cathepsinD; VIdlr; Ccr2;

[0059] **Schizophrenia Disorders:** Genes linked to this disorder: Neuregulinl (Nrgl); Erb4 (receptor for Neuregulin); Complexinl (Cp1x1); Tphl Tryptophan hydroxylase; Tph2 Tryptophan hydroxylase 2; Neurexin 1; GSK3; GSK3a; GSK3b;

[0060] **Trinucleotide Repeat Disorders:** Genes linked to this disorder: 5 HTT (Huntington's Dx); SBMA/SMAX1/AR (Kennedy's Dx); FXN/X25 (Friedrich's Ataxia); ATX3 (Machado-Joseph's Dx); ATXN1 and ATXN2 (spinocerebellar ataxias); DMPK (myotonic dystrophy); Atrophin-1 and Atnl (DRPLA Dx); CBP (Creb-BP - global instability); VLDLR (Alzheimer's); Atxn7; Atxn10;

[0061] **Fragile X Syndrome:** Genes linked to this disorder: FMR2; FXR1; FXR2; mGLURS;

[0062] **Secretase Related Disorders:** Genes linked to this disorder: APH-1 (alpha and beta); Presenil n (Psenl); nicastrin (Ncstn); PEN-2;

[0063] **Others:** Genes linked to this disorder: Nosl ; Paipl; Nati; Nat2;

[0064] **Prion - related disorders:** Gene linked to this disorder: Prp;

[0065] **ALS:** Genes linked to this disorder: SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a; VEGF-b; VEGF-c);

[0066] **Drug addiction:** Genes linked to this disorder: Prkce (alcohol); Drd2; Drd4; ABAT (alcohol); GRIA2; GrmS; Grinl; Htrlb; Grin2a; Drd3; Pdyn; Grial (alcohol);

[0067] **Autism:** Genes linked to this disorder: Mecp2; BZRAP1; MDGA2; SemaSA; Neurexin 1; Fragile X (FMR2 (AFF2); FXR1; FXR2; MglurS);

[0068] **Alzheimer's Disease** Genes linked to this disorder: El; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterin; PS1; SORL1; CR1; VIdlr; Ubal; Uba3; CHIP28 (Aqp1, Aquaporin 1); Uchll; Uch13; APP;

[0069] **Inflammation and Immune-related disorders** Genes linked to this disorder: IL-10; IL-1 (IL-la; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL-17b; IL-17c; IL-17d; IL-17f); 11-23; Cx3crl; ptpn22; TNFa; NOD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b); CTLA4; Cx3c11, AAT deficiency/mutations, AIDS (KIR3DL1, NKAT3, NKB1, ANIB11, KIR3DS1, IFNG, CXCL12, SDF1); Autoimmune lymphoproliferative syndrome (TNFRSF6, APT1, FAS, CD95, ALPS1A); Combined immunodeficiency, (IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228), HIV susceptibility or infection (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCKR5 (CCR5)); Immunodeficiencies (CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD4OLG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI); Inflammation (IL-10, IL-1 (IL-la, IL-1b), IL-13, IL-17 (IL-17a (CTLA8), IL-17b, IL-17c, IL-17d, IL-17f), 11-23, Cx3crl, ptpn22, TNFa, NOD2/CARD15 for IBD, IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3c11); Severe combined immunodeficiencies (SCIDs)(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4);

[0070] **Parkinson's,** Genes linked to this disorder: x-Synuclein; DJ-1; LRRK2; Parkin; PINK1;

[0071] **Blood and coagulation disorders:** Genes linked to these disorders: Anemia (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH I, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH I, ASB, ABCB7, ABC7, ASAT); Bare lymphocyte syndrome (TAPBP, TPSN, TAP2, ABCB3, PSF2, RINGI 1, MHC2TA, C2TA, RFX5, RFXAP, RFX5), Bleeding disorders (TBXA2R, P2RX I, P2X I); Factor H and factor H-like 1 (HF1, CFH, HUS); Factor V and factor VIII (MCFD2); Factor VII deficiency (F7); Factor X deficiency (F10); Factor XI deficiency (F11); Factor XII deficiency (F12, HAF); Factor XIIIA deficiency (F13A1, F13A); Factor XIIIB deficiency (F13B); Fanconi anemia (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCJ, PHF9, FANCL, FANCM, ICIAA1596); Hemophagocytic lymphohistiocytosis disorders (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemophilia A (F8, F8C, HEMA); Hemophilia

B (F9, HEMB), Hemorrhagic disorders (PI, ATT, F5); Leukocyde deficiencies and disorders (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); Sickle cell anemia (HBB); Thalassemia (HBA2, HBB, HBD, LCRB, HBA1);

[0072] Cell dysregulation and oncology disorders: Genes linked to these disorders: B-cell non-Hodgkin lymphoma (BCL7A, BCL7); Leukemia (TALI TCL5, SCL, TAL2, FLT3, NBS 1 , NBS, ZNFNIAI, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AFIO, ARHGEFI2, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX 1 , CBFA2, AML1, WHSC 1 LI , NSD3, FLT3, AF1Q, NPM 1 , NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AFI 0, CALM, CLTH, ARLI 1, ARLTS1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NFI, VRNF, WSS, NFNS, PTPNI 1, PTP2C, SHP2, NS 1 , BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP2I4, D9S46E, CAN, CAIN);

[0073] Metabolic, liver, kidney disorders: Genes linked to these disorders: Amyloid neuropathy (TTR, PALS); Amyloidosis (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, UR, PALS); Cirrhosis (KATI 8, KRT8, CaHlA, NAIC, TEX292, KIAA1988); Cystic fibrosis (CFTR, ABCC7, CF, MRP7); Glycogen storage diseases (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPS, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Hepatic adenoma, 142330 (TCF1, HNF1A, MODY3), Hepatic failure, early onset, and neurologic disorder (SCOD1, SCO1), Hepatic lipase deficiency (LIPC), Hepatoblastoma, cancer and carcinomas (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; Medullary cystic kidney disease (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); Phenylketonuria (PAH, PKU1, QDPR, DHPR, PTS); Polycystic kidney and hepatic disease (FCYT, PKHD1, ARPKD, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63);

[0074] Muscular/Skeletal Disorders: Genes linked to these disorders: Becker muscular dystrophy (DMD, BMD, MYF6), Duchenne Muscular Dystrophy (DMD, BMD); Emery-Dreifuss muscular dystrophy (LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); Facioscapulohumeral muscular dystrophy (FSHMD1A, FSHD1A); Muscular dystrophy (FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B,

SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3, LGMD1C, SEPN1, SELN, RSMD1, PLEC1, PLTN, EBS1); Osteopetrosis (LAPS, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, 0C116, OPTB1); Muscular atrophy (VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1);

[0075] Neurological and Neuronal disorders: Genes linked to these disorders: ALS (SOD1, ALS2, STEX, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c); Alzheimer disease (APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLAU, URK, ACE, DCPI, ACEI, MPO, PACIP1, PAXIPIL, PTIP, A2M, BLMH, BMH, PSEN1, AD3); Autism (Mecp2, BZRAP I, MDGA2, Sema5A, Neurex 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); Fragile X Syndrome (FMR2, FXR1, FXR2, mGLUR5); Huntington's disease and disease like disorders (HD, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); Parkinson disease (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARKS, PINK1, PARK6, UCHL1, PARKS, SNCA, NACP, PARK1, PARK4, PRKN, PARK-2, PDJ, DBH, NDUFV2); Rett syndrome (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Synuclein, DJ-1); Schizophrenia (Neuregulinl (Nrgl), Erb4 (receptor for Neuregulin), Complexinl (Cp1x1), Tphl Tryptophan hydroxylase, Tph2, Tryptophan hydroxylase 2, Neurexin 1, GSK3, GSK3a, GSK3b, 5-HTT (S1c6a4), COMT, DRD (Drd la), SLC6A3, DAOA, DTNBP1, Dao (Daol)); Secretase Related Disorders (APH-1 (alpha and beta), Preseni I in (Psenl), nicastrin, (Ncstn), PEN-2, Nosl, Parpl, Natl, Nat2); Trinucleotide Repeat Disorders (HTT (Huntington's Dx), SBMA/SMAX1/AR (Kennedy's Dx), FXN/X25 (Friedrich's Ataxia), ATX3 (Machado- Joseph's Dx), ATXN1 and ATXN2 (spinocerebellar ataxias), DMPK (myotonic dystrophy), Atrophin-1 and Atnl (DRPLA Dx), CBP (Creb-BP - global instability), VLDLR (Alzheimer's), Atxn7, Atxn10);

[0076] Occular-related disorders: Genes linked to these disorders: Age-related macular degeneration (Aber, Cc12, Cc2, cp (ceruloplasmin), Timp3, cathepsinD, Vld1r, Ccr2); Cataract (CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4,

BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQPO, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); Corneal clouding and dystrophy (APOA1, TGFBI, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD); Cornea plana congenital (KERA, CNA2); Glaucoma (MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPAL, NTG, NPG, CYP1B1, GLC3A); Leber congenital amaurosis (CRB1, RP12, CRX, CORD2, CRD, RPGRIP1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3); Macular dystrophy (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2);

[0077] P13K/AKT Cellular Signaling disorders: Genes linked to these disorders: PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2; PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1; AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2; PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2; ITGAl; KRAS; EIF4EBP1; RELA; PRKCD; NOS3; PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7; YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A; CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1; PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2; TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SOK; HS P9OAA1; RP S 6KB1;

[0078] ERK/MAPK Cellular Signaling disorders: Genes linked to these disorders: PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1; PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A; PIK3C3; MAPK8; MAPK3; ITGAl; ETS1; KRAS; MYCN; EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ; PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1; PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1; CRKL; BRAE; ATF4; PRKCA; SRF; STAT1; SGK;

[0079] Glucocorticoid Receptor Cellular Signaling disorders: Genes linked to these disorders: RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1; MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I; PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2; MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1; MAPK3; TSC22D3; MAPK10; NRIP1;

KRAS; MAPK13; RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1; PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3; MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP; CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1; ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP 1; STAT1; IL6; HSP9OAA1;

[0080] Axonal Guidance Cellular Signaling disorders: Genes linked to these disorders: PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12; IGF1; RAC1; RAP1A; El F4E; PRKCZ; NRP1; NTRK2; ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2; CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11; PRKD1; GNB2L1; ABL1 ; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1; GUI; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B; AKT3; PRKCA;

[0081] Ephrin Recptor Cellular Signaling disorders: Genes linked to these disorders: PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2; DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14; CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4, AKT1; JAK2; STAT3; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4; AKT3; SGK;

[0082] Actin Cytoskeleton Cellular Signaling disorders: Genes linked to these disorders: ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1; PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6; ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8; F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7; PPP1CC; PXN; VIL2; RAF1; GSN; DYRK1A; ITGB1; MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3; ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL; BRAF; VAV3; SGK;

[0083] Huntington's Disease Cellular Signaling disorders: Genes linked to these disorders: PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2; MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2; PIK3CA; HDAC5; CREB1; PRKC1; HS PA5 ; REST; GNAQ;

PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1; GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2; HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A; HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1; PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX; ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3;

[0084] Apoptosis Cellular Signaling disorders: Genes linked to these disorders: PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1; BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB; CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8; BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF; RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2; CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2; BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK; CASP3: BTRC3: PARPI;

[0085] B Cell Receptor Cellular Signaling disorders: Genes linked to these disorders: RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11; AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A; MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1; MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9; EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1; NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN; GSK3B; ATF4; AKT3; VAV3; RPS6KB1;

[0086] Leukocyte Extravasation Cellular Signaling disorders: Genes linked to these disorders: ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA; RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11; MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12; PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB; MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK; MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2; CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; FUR; ITK; CRKL; VAV3; CTTN; PRKCA; MMPl; MMP9;

[0087] Integrin Cellular Signaling disorders: Genes linked to these disorders: ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A; TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2; CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8; CAV1; CAPN1; ABL1; MAPK3; ITGAl; KRAS; RHOA; SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1; TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2; CRKL; BRAF; GSK3B; AKT3;

[0088] Acute Phase Response Cellular Signaling disorders: Genes linked to these disorders: IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11; AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14; PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1; TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN; AKT3; IL1R1; IL6;

[0089] PTEN Cellular Signaling disorders: Genes linked to these disorders: ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11; MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA; CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1; MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR; RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2; AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1; NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2; GSK3B; AKT3; FOXO1; CASP3;

[0090] p53 Cellular Signaling disorders: Genes linked to these disorders: RPS6KB1 PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A; BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2; PIK3CB; PIK3C3; MAPK8; THBS 1; ATR; BCL2L1; E2F1; PMAIP1; CHEK2; TNFASF10B; TP73; RB1; HDAC9; CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A; HIPK2; AKT1; PIK3R1; RAM2B; APAF1; CTNNB1; SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN; SNAI2; GSK3B; BAX; AKT3;

[0091] Aryl Hydrocarbon Receptor Cellular Signaling disorders: Genes linked to these disorders: HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1; NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1; SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1; MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1; SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF; CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1; CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1; HSP9OAA1;

[0092] Xenobiotic Metabolism Cellular Signaling disorders: Genes linked to these disorders: PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1; NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A; PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1; ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD; GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL; NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1; CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1; NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1; HSP9OAA1;

[0093] SAPL/JNK Cellular Signaling disorders: Genes linked to these disorders: PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1; GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA; FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1; GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2; PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1; CRKL; BRAF; SGK;

[0094] PPAr/RXR Cellular Signaling disorders: Genes linked to these disorders: PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN; RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2; ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8; IASI; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A; NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1; TGFBAl; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP9OAA1; ADIPOO;

[0095] NF-KB Cellular Signaling disorders: Genes linked to these disorders: IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ: TRAF6; TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2; MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2; KRAS; RELA; PIK3C2A; TRAF2; TLR4: PDGFRB; TNF; INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1; PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10; GSK3B; AKT3; TNFAIP3; IL1R1;

[0096] Neuregulin Cellular Signaling disorders: Genes linked to these disorders: ERBB4; PRKCE; ITGAM; ITGA5: PTEN; PRKCZ; ELK1; MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI; CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS; PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2; ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3; EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL; AKT3; PRKCA; HS P9OAA1; RPS6KB1;

[0097] Wnt and Beta catenin Cellular Signaling disorders: Genes linked to these disorders: CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO; AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A; WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2: ILK; LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1; PPP2R5C; WNT5A; LAPS; CTNNB1; TGFBR1; CCND1; GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B; AKT3; SOX2;

[0098] Insulin Receptor Signaling disorders: Genes linked to these disorders: PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1; PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3; MAPK8; IASI; MAPK3; TSC2; KRAS; EIF4EBP1; SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1; GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK; RPS6KB1;

[0099] IL-6 Cellular Signaling disorders: Genes linked to these disorders: HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11; IKBKB; FOS; NFKB2: MAP3K14; MAPK8; MAPK3; MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3; MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6;

[00100] Hepatic Cholestasis Cellular Signaling disorders: Genes linked to these disorders: PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA; RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8; PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1; TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8; CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4; JUN; IL1R1; PRKCA; IL6;

[00101] IGF-1 Cellular Signaling disorders: Genes linked to these disorders: IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2; PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8; IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A; YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3; FOXO1; SRF; CTGF; RPS6KB1;

[00102] NRF2-mediated Oxidative Stress Response Signaling disorders: Genes linked to these disorders: PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1; NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL; NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP; MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1; GSK3B; ATF4; PRKCA; EIF2AK3; HSP9OAA1;

[00103] Hepatic Fibrosis/Hepatic Stellate Cell Activation Signaling disorders: Genes linked to these disorders: EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF; SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9; IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8; PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX; IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9;

[00104] PPAR Signaling disorders: Genes linked to these disorders: EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB; NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3; NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2; PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA; MAP2K1; NFKB1; JUN; IL1R1; HSP9OAA1;

[00105] Fc Epsilon RI Signaling disorders: Genes linked to these disorders: PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD; MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3; VAV3; PRKCA;

[00106] G-Protein Coupled Receptor Signaling disorders: Genes linked to these disorders: PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB; PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB; PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1; IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK; PDPK1; S TAT3 ; MAP2K1; NFKB1; BRAF; ATF4; AKT3; PRKCA;

[00107] Inositol Phosphate Metabolism Signaling disorders: Genes linked to these disorders: PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6; MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3; MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1; MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK;

[00108] PDGF Signaling disorders: Genes linked to these disorders: EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB; P IK3 C3 ; MAPK8; CAV1; ABL1 ; MAPK3; KRAS; SRC; PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2; PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC; JUN; CRKL; PRKCA; SRF; STAT1; SPHK2 **VEGF Signaling disorders:** Genes linked to these disorders: ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF; AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3; BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN; RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN; VEGFA; AKT3; FOXO1; PRKCA;

[00109] Natural Killer Cell Signaling disorders: Genes linked to these disorders: PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11; KIR2DL3; AKT2; PIK3CA; SYK; PRKCI;

PIK3CB; PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6; PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1; PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA;

[00110] Cell Cycle: Gl/S Checkpoint Regulation Signaling disorders: Genes linked to these disorders: HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC; ATR; ABL1 ; E2F1; HDAC2; HDAC7A; RB1; HDAC11 ; HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1; E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1; GSK3B; RBL1; HDAC6;

[00111] T Cell Receptor Signaling disorders: Genes linked to these disorders: RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS; NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; RELA, PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB, FYN; MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10; JUN; VAV3;

[00112] Death Receptor disorders: Genes linked to these disorders: CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD; FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8; DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB; CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3; BIRC3;

[00113] FGF Cell Signaling disorders: Genes linked to these disorders: RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11; AKT2;PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8; MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1; AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4; AKT3; PRKCA; HGF;

[00114] GM-CSF Cell Signaling disorders: Genes linked to these disorders: LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A; STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3; ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3; STAT1;

[00115] Amyotrophic Lateral Sclerosis Cell Signaling disorders: Genes linked to these disorders: BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2; PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1; PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1; APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3 PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B; PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A; PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3; STAT1;

[00116] JAK/Stat Cell Signaling disorders: Genes linked to these disorders: PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B; PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1;

STAT5A; PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3; STAT1;

[00117] Nicotinate and Nicotinamide Metabolism Cell Signaling disorders: Genes linked to these disorders: PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1; PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1; PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2; MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK;

[00118] Chemokine Cell Signaling disorders: Genes linked to these disorders: CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ; CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13; RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1; MAP2K2; MAP2K1; JUN; CCL2; PRKCA;

[00119] IL-2 Cell Signaling disorders: Genes linked to these disorders: ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS; STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2; JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3;

[00120] Synaptic Long Term Depression Signaling disorders: Genes linked to these disorders: PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS; PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3; KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA; YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA;

[00121] Estrogen Receptor Cell Signaling disorders: Genes linked to these disorders: TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2; SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1; HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP; MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2;

[00122] Protein Ubiquitination Pathway Cell Signaling disorders: Genes linked to these disorders: TRAF6; SMURF1; BIRC4; BRCAl; UCHL1; NEDD4; CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7; USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8; USP1; VHL; HSP9OAA1; BIRC3;

[00123] IL-10 Cell Signaling disorders: Genes linked to these disorders: TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF; IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1; JUN; IL1R1; IL6;

[00124] VDR/RXR Activation Signaling disorders: Genes linked to these disorders: PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1; NCOR2; SP1; PRKCI; CDKN1B;

PRKD1; PRKCD; RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1; LAPS; CEBPB; FOXO1; PRKCA;

[00125] TGF-beta Cell Signaling disorders: Genes linked to these disorders: EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1; FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2; SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2; MAP2K1; TGFBR1; SMAD4; JUN; SMAD5;

[00126] Toll-like Receptor Cell Signaling disorders: Genes linked to these disorders: IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK; NFKB1; TLR2; JUN;

[00127] p38 MAPK Cell Signaling disorders: Genes linked to these disorders: HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS; CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2; MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1; SRF; STAT1; and

[00128] Neurolrophin/TRK Cell Signaling disorders: Genes linked to these disorders: NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A; RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; CDC42; JUN; ATF4.

[00129] Other cellular dysfunction disorders linked to a genetic modification are contemplated herein for example, FXR/RXR Activation, Synaptic Long Term Potentiation, Calcium Signaling EGF Signaling, Hypoxia Signaling in the Cardiovascular System, LPS/IL-1 Mediated Inhibition of RXR Function LXR/RXR Activation, Amyloid Processing, IL-4 Signaling, Cell Cycle: G2/M DNA Damage Checkpoint Regulation, Nitric Oxide Signaling in the Cardiovascular System Purine Metabolism, cAMP-mediated Signaling, Mitochondrial Dysfunction Notch Signaling Endoplasmic Reticulum Stress Pathway Pyrimidine Metabolism, Parkinson's Signaling Cardiac & Beta Adrenergic Signaling Glycolysis/ Gluconeogenesis Interferon Signaling Sonic Hedgehog Signaling Glycerophospholipid Metabolism, Phospholipid Degradation, Tryptophan Metabolism Lysine Degradation Nucleotide Excision Repair Pathway, Starch and Sucrose Metabolism, Aminosugars Metabolism Arachidonic Acid Metabolism, Circadian Rhythm Signaling, Coagulation System Dopamine Receptor Signaling, Glutathione Metabolism Glycerolipid Metabolism Linoleic Acid Metabolism Methionine Metabolism Pyruvate Metabolism Arginine and Praline Metabolism, Eicosanoid Signaling Fructose and Mannose Metabolism, Galactose

Metabolism Stilbene, Coumarine and Lignin Biosynthesis Antigen Presentation Pathway, Biosynthesis of Steroids Butanoate Metabolism Citrate Cycle Fatty Acid Metabolism Glycerophosphol ipid Metabolism, Histidine Metabolism Inositol Metabolism Metabolism of Xenobiotics by Cytochrome p450, Methane Metabolism, Phenylalanine Metabolism, Propanoate Metabolism Selenoamino Acid Metabolism Sphingolipid Metabolism Aminophosphonate Metabolism, Androgen and Estrogen Metabolism Ascorbate and Aldarate Metabolism, Bile Acid Biosynthesis Cysteine Metabolism Fatty Acid Biosynthesis Glutamate Receptor Signaling, NRF2-mediated, Oxidative Stress Response Pentose Phosphate Pathway, Pentose and Glucuronate Interconversions, Retinol Metabolism Riboflavin Metabolism Tyrosine Metabolism Ubiquinone Biosynthesis Valine, Leucine and Isoleucine Degradation Glycine, Serine and Threonine Metabolism Lysine Degradation Pain/Taste, or Mitochondrial Function Developmental Neurology or combinations thereof.

[00130] Nucleic acid-guided nucleases disclosed herein can encompass a native sequence, an engineered sequence, or engineered nucleotide sequences of synthetized variants. Non-limiting examples of types of engineering that can be done to obtain a non-naturally occurring nuclease system are as follows. Engineering can include codon optimization to facilitate expression or improve expression in a host cell, such as a heterologous host cell. Engineering can reduce the size or molecular weight of the nuclease in order to facilitate expression or delivery. Engineering can alter PAM selection in order to change PAM specificity or to broaden the range of recognized PAMs. Engineering can alter, increase, or decrease stability, processivity, specificity, or efficiency of a targetable nuclease system. Engineering can alter, increase, or decrease protein stability. Engineering can alter, increase, or decrease processivity of nucleic acid scanning. Engineering can alter, increase, or decrease target sequence specificity. Engineering can alter, increase, or decrease nuclease activity. Engineering can alter, increase, or decrease editing efficiency. Engineering can alter, increase, or decrease transformation efficiency. Engineering can alter, increase, or decrease nuclease or guide nucleic acid expression. As used herein, a non-naturally occurring nucleic acid sequence can be an engineered sequence or engineered nucleotide sequences of synthetized variants. Such non-naturally occurring nucleic acid sequences can be amplified, cloned, assembled, synthesized, generated from synthesized oligonucleotides or dNTPs, or otherwise obtained using methods known by those skilled in the art. Examples of non-naturally occurring nucleic acid sequences which are disclosed herein

include those for nucleic acid-guided nucleases with engineered sequences (*e.g.*, SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and 108-117) and those for nucleic acid-guided nucleases with engineered nucleotide sequences of synthetized variants (*e.g.*, SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128).

[00131] Disclosed herein are nucleic acid-guided nucleases. Subject nucleases are functional *in vitro*, or in prokaryotic, archaeal, or eukaryotic cells for *in vitro*, *in vivo*, or *ex vivo* applications. Suitable nucleic acid-guided nucleases can be from an organism from a genus which includes but is not limited to Thiomicrospira, Succinivibrio, Candidatus, Porphyromonas, Acidaminococcus, Acidomonococcus, Barnesiella, Prevotella, Smithella, Moraxella, Synergistes, Francisella, Leptospira, Catenibacterium, Kandleria, Clostridium, Dorea, Coprococcus, Enterococcus, Fructobacillus, Weissella, Pediococcus, Collinsella, Corynebacter, Sutterella, Legionella, Treponema, Roseburia, Filifactor, Lachnospiraceae, Eubacterium, Sedimentisphaera, Streptococcus, Lactobacillus, Mycoplasma, Bacteroides, Flaviivola, Flavobacterium, Sphaerochaeta, Azospirillum, Gluconacetobacter, Neisseria, Roseburia, Parvibaculum, Parabacteroides, Staphylococcus, Nitratifractor, Mycoplasma, Alicyclobacillus, Brevibacilus, Bacillus, Bacteroidetes, Brevibacilus, Carnobacterium, Clostridiaridium, Clostridium, Desulfonatronum, Desulfovibrio, Helcococcus, Leptotrichia, Listeria, Methanomethyophilus, Methylobacterium, Opitutaceae, Paludibacter, Rhodobacter, Sphaerochaeta, Tuberibacillus, Oleiphilus, Omnitrophica, Parcubacteria, and Campylobacter. Species of organism of such a genus can be as otherwise herein discussed. Suitable gRNAs can be from an organism from a genus or unclassified genus within a kingdom which includes but is not limited to Firmicute, Actinobacteria, Bacteroidetes, Proteobacteria, Spirochates, and Tenericutes. Suitable gRNAs can be from an organism from a genus or unclassified genus within a phylum which includes but is not limited to Erysipelotrichia, Clostridia, Bacilli, Actinobacteria, Bacteroidetes, Catenovulum, Coprococcus, Flavobacteria, Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, Epsilonproteobacteria, Spirochaetes, and Mollicutes. Suitable gRNAs can be from an organism from a genus or unclassified genus within an order which includes but is not limited to Clostridiales, Lactobacillales, Actinomycetales, Bacteroidales, Flavobacteriales, Rhizobiales, Rhodospirillales, Burkholderiales, Neisseriales, Legionellales, Nautiliales, Campylobacterales, Spirochaetales, Mycoplasmatales, and Thiotrichales. Suitable gRNAs can be from an organism from a genus or unclassified genus within a family which includes but is

not limited to, Lachnospiraceae, Enterococcaceae, Leuconostocaceae, Lactobacillaceae, Streptococcaceae, Peptostreptococcaceae, Staphylococcaceae, Eubacteriaceae, Corynebacterineae, Bacteroidaceae, Flavobacterium, Cryomoorphaceae, Rhodobiaceae, Rhodospirillaceae, Acetobacteraceae, Sutterellaceae, Neisseriaceae, Legionellaceae, Nautiliaceae, Campylobacteraceae, Spirochaetaceae, Mycoplasmataceae, Pisciririckettsiaceae, and Francisellaceae. In some embodiments, suitable gRNAs can be from an organism from a genus or unclassified genus within a family which includes Acidaminococcus, Sedimentisphaera, Barnesiella sp., Bacteroidetes, Parabacteroides, Lachnospiraceae, Coprococcus sp., Catenovulum sp., and Collinsella. Other nucleic acid-guided nucleases have been described in US Patent Application Publication No. US20160208243 filed Dec. 18, 2015, US Application Publication No. US20140068797 filed Mar. 15, 2013, U.S. Pat. No. 8,697,359 filed Oct. 15, 2013, and Zetsche et al., Cell 2015 Oct. 22; 163(3):759-71, each of which are incorporated herein by reference in their entirety.

[00132] Some nucleic acid-guided nucleases suitable for use in the methods, systems, and compositions of the present disclosure can include, but are not limited to, those derived from an organism such as, but not limited to, Thiomicrospira sp. XS5, Eubacterium rectale, Succinivibrio dextrinosolvens, Candidatus Methanoplasma termitum, Candidatus Methanomethylophilus alvus, Porphyromonas crevioricanis, Flavobacterium branchiophilum, Acidaminococcus Sp., Acidomonococcus sp., Lachnospiraceae bacterium COE1, Prevotella brevis ATCC 19188, Smithella sp. SCADC, Moraxella bovoculi, Synergistes jonesii, Bacteroidetes oral taxon 274, Francisella tularensis, Leptospira inadai serovar Lyme str. 10, Acidomonococcus sp. crystal structure (5B43) S. mutans, S. agalactiae, S. equisimilis, S. sanguinis, S. pneumonia; C. jejuni, C. coli; N. salsuginis, N. tergarcus; S. auricularis, S. carnosus; N. meningitides, N. gonorrhoeae; L. monocytogenes, L. ivanovii; C. botulinum, C. difficile, C. tetani, C. sordellii; Francisella tularensis 1, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Butyrivibrio proteoclasticus B316, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, Porphyromonas macacae, Catenibacterium sp. CAG:290, Kandleria vitulina, Clostridiales

bacterium KA00274, Lachnospiraceae bacterium 3-2, Dorea longicatena, Coprococcus catus GD/7, Enterococcus columbae DSM 7374, Fructobacillus sp. EFB-N1, Weissella halotolerans, Pediococcus acidilactici, Lactobacillus curvatus, Streptococcus pyogenes, Lactobacillus versmoldensis, Filifactor alocis ATCC 35896, Alicyclobacillus acidoterrestris, Alicyclobacillus acidoterrestris ATCC 49025, Desulfovibrio inopinatus, Desulfovibrio inopinatus DSM 10711, Oleiphilus sp. Oleiphilus sp. HI0009, Candidtus kefeldibacteria, Parcubacteria CasY.4, Omnitrophica WOR 2 bacterium GWF2, Bacillus sp. NSP2.1, Bacillus thermoamylovorans, Catenovulum sp. CCB-QB4, Coprococcus sp. AF16-5, Lachnospiraceae bacterium MC2017, Collinsella tanakaei, Parabacteroides distasonis, Bacteroidetes bacterium HGW-Bacteroidetes-6, Barnesiella sp. An22, Sedimentisphaera cyanobacteriorum, and Acidaminococcus massiliensis.

[00133] In some embodiments, a nucleic acid-guided nuclease disclosed herein includes an amino acid sequence having at least 50% amino acid identity to any one of SEQ ID NO: 3, 16, 29, 42, 55, 68, 81, 94 and/or 107. In some embodiments, a nucleic acid-guided nuclease disclosed herein includes a polypeptide having an amino acid sequence of about 60%, about 65%, or about 75%, or about 85%, or about 95%, or about 99% or about 99.5% identity to about 100% to amino acid sequences of one or more of SEQ ID NO: 3, 16, 29, 42, 55, 68, 81, 94 and/or 107. In some embodiments, a nucleic acid-guided nuclease disclosed herein includes an amino acid sequence having about 85%, about 90%, or about 95%, or about 99%, or about 99.5% or about 100%, amino acid identity to any one of SEQ ID NO: 3, 16, 29, 42, 55, and/or 94.

[00134] In some embodiments, a guide RNA (gRNA) disclosed herein includes a nucleic acid sequence of at least 50% amino acid identity to any one of SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128. In some embodiments, a gRNA disclosed herein includes a nucleic acid sequence of at least 10%, 20%, 30%, 40%, 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, greater than 95%, or 100% nucleic acid identity to any one of SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128. In some embodiments, a gRNA disclosed herein includes a nucleic acid sequence of at least 50%, or about 60%, about 65%, or about 75%, or about 85%, or about 95%, or about 99% or about 99.5% identity to about 100% to, nucleic acid identity to any one of SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128. In some embodiments, the engineered polynucleotide (gRNA) can be split into fragments encompassing a synthetic tracrRNA and crRNA. In some embodiments, a crRNA disclosed

herein can include a nucleic acid sequence of at least 50%, or about 60%, about 65%, or about 75%, or about 85%, or about 95%, or about 99% or about 99.5% identity to about 100% to, nucleic acid identity to any one of SEQ ID NO: 129-139. In some embodiments, a crRNA disclosed herein can include a nucleic acid sequence of at least 50%, or about 60%, or about 65%, or about 75%, or about 85%, or about 95%, or about 99% or about 99.5% identity to about 100% to, nucleic acid identity to any one of SEQ ID NO: 129-137.

[00135] In some embodiments, gRNA disclosed herein can include a nucleic acid sequence of at least 50% nucleic acid identity to SEQ ID NO: 127. In other embodiments, a gRNA disclosed herein can include a nucleic acid sequence of about 10%, 20%, 30%, 40%, 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, greater than 95%, or 100% nucleic acid identity to SEQ ID NO: 127. In some embodiments, a gRNA disclosed herein includes a nucleic acid sequence of at least 50%, or about 60%, about 65%, or about 75%, or about 85%, or about 95%, or about 99% or about 99.5% identity to about 100%, nucleic acid identity to SEQ ID NO: 127.

[00136] In some embodiments, a nucleic acid-guided nuclease disclosed herein includes a nucleic acid sequence of at least 50% nucleic acid sequence identity to any one of SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and/or 108-117. In some embodiments, a nucleic acid-guided nuclease disclosed herein includes a nucleic acid sequence of about 60%, or about 65%, or about 70%, or about 75%, or about 80%, or about 85%, or about 90%, or about 95%, greater than 95%, or 100% amino acid identity to any one of SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and/or 108-117. In some embodiments, a nucleic acid-guided nuclease disclosed herein includes a nucleic acid sequence of at about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, greater than 95%, nucleic acid identity to any one of SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, and/or 95-104.

[00137] In some instances, a nucleic acid-guided nuclease disclosed herein is encoded from a nucleic acid sequence. Such a nucleic acid can be codon optimized for expression in a desired host cell. Suitable host cells can include, as non-limiting examples, prokaryotic cells such as *E. coli, P. aeruginosa, B. subtilus, and V. natriegens,* and *S. cerevisiae*, eukaryotic cells, plant cells, insect cells, nematode cells, amphibian cells, fish cells, or mammalian cells, including human cells.

[00138] A nucleic acid sequence encoding a nucleic acid-guided nuclease can be operably linked to a promoter. Such nucleic acid sequences can be linear or circular. The nucleic acid

sequences can be encompassed on a larger linear or circular nucleic acid sequence that comprises additional elements such as an origin of replication, selectable or screenable marker, terminator, other components of a targetable nuclease system, such as a guide nucleic acid, or an editing or recorder cassette as disclosed herein. In some aspects, nucleic acid sequences can include a at least one glycine, at least one 6X histidine tag, and/or at least one 3x nuclear localization signal tag. Larger nucleic acid sequences can be recombinant expression vectors, as are described in more detail later.

gRNAs

[00139] In general, a guide polynucleotide can complex with a compatible nucleic acid-guided nuclease and can hybridize with a target sequence, thereby directing the nuclease to the target sequence. A subject nucleic acid-guided nuclease capable of complexing with a guide polynucleotide can be referred to as a nucleic acid-guided nuclease that is compatible with the guide polynucleotide. In addition, a guide polynucleotide capable of complexing with a nucleic acid-guided nuclease can be referred to as a guide polynucleotide or a guide nucleic acid that is compatible with the nucleic acid-guided nucleases. In some embodiments, an engineered polynucleotide (gRNA) disclosed herein can be split into fragments encompassing a synthetic tracrRNA and crRNA. Examples of gRNA can include, but are not limited to, gRNAs represented in **Table 1**.

Table 1. Exemplary gRNAs

gRNA SEQ. ID NO.	gRNA Nucleotide Sequence	compatible nucleic acid-guided nuclease
118	GUCUAAAAGACCAUAUGAAUUUCUACUU UCGUAGAUNNNNNNNNNNNNNNNNNNNN	ABW1
119	GUCUAAAGGCCUUAUAAAAUUUCUACUG UCGUAGAUNNNNNNNNNNNNNNNNNNNN	ABW2
120	GUCUAUACAGACACUUUAAUUUCUACUA UUGUAGAUNNNNNNNNNNNNNNNNNNNN	ABW3
121	GUCUGAAAGACAAGUAUAAUUUCUACUA UUGUAGAUNNNNNNNNNNNNNNNNNNNN	ABW4
122	GGCUAUAAGCCUUGUAUAAUUUCUACUA UUGUAGAUNNNNNNNNNNNNNNNNNNNN	ABW5
123	GUUGAAACUGUAAGCGGAAUGUCUACUU GGGUAGAUNNNNNNNNNNNNNNNNNNNN	ABW6
124	GCAUGAGAACCAUGCAUUUCUAAGGUAC UCCAAAACNNNNNNNNNNNNNNNNNNNN	ABW7
125	GUUGAGUAACCUUAAAUAAUUUCUACUG UUGUAGAUNNNNNNNNNNNNNNNNNNNN	ABW8

126	AUCUACAACAGUAGAAAUUUAAGCUAAG GCUUAGACNNNNNNNNNNNNNNNNNNNN	ABW9
127	UAAUUUCUACUCUUGUAGAUNNNNNNNNN NNNNNNNNNNNN	Cas12A
128	UAAUUUCUACUC - UUGUAGAUNNNNNNNNNNNNNNNNNNNN	STAR

[00140] A guide polynucleotide can be DNA. A guide polynucleotide can be RNA. A guide polynucleotide can include both DNA and RNA. A guide polynucleotide can include modified or non-naturally occurring nucleotides. In cases where the guide polynucleotide comprises RNA, the RNA guide polynucleotide can be encoded by a DNA sequence on a polynucleotide molecule such as a plasmid, linear construct, or editing cassette as disclosed herein.

[00141] A guide polynucleotide can comprise a guide sequence. A guide sequence is a polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a complexed nucleic acid-guided nuclease to the target sequence. The degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment can be determined with the use of any suitable algorithm for aligning sequences. In some embodiments, a guide sequence can be about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In other embodiments, a guide sequence can be less than about 75, 50, 45, 40, 35, 30, 25, 20 nucleotides in length. Preferably the guide sequence is 10-30 nucleotides long. The guide sequence can be 15-20 nucleotides in length. The guide sequence can be 15 nucleotides in length. The guide sequence can be 16 nucleotides in length. The guide sequence can be 17 nucleotides in length. The guide sequence can be 18 nucleotides in length. The guide sequence can be 19 nucleotides in length. The guide sequence can be 20 nucleotides in length.

[00142] A guide polynucleotide can include a scaffold sequence. In general, a "scaffold sequence" can include any sequence that has sufficient sequence to promote formation of a targetable nuclease complex, wherein the targetable nuclease complex includes, but is not limited to, a nucleic acid-guided nuclease and a guide polynucleotide can include a scaffold sequence and a guide sequence. Sufficient sequence within the scaffold sequence to promote formation of a targetable nuclease complex can include a degree of complementarity along the length of two

sequence regions within the scaffold sequence, such as one or two sequence regions involved in forming a secondary structure. In some cases, the one or two sequence regions are included or encoded on the same polynucleotide. In some cases, the one or two sequence regions are included or encoded on separate polynucleotides. Optimal alignment can be determined by any suitable alignment algorithm, and can further account for secondary structures, such as self-complementarity within either the one or two sequence regions. In some embodiments, the degree of complementarity between the one or two sequence regions along the length of the shorter of the two when optimally aligned can be about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, at least one of the two sequence regions can be about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length.

[00143] A scaffold sequence of a subject guide polynucleotide can comprise a secondary structure. A secondary structure can comprise a pseudoknot region. In some cases, binding kinetics of a guide polynucleotide to a nucleic acid-guided nuclease is determined in part by secondary structures within the scaffold sequence. In some cases, binding kinetics of a guide polynucleotide to a nucleic acid-guided nuclease is determined in part by nucleic acid sequence with the scaffold sequence. In some aspects, the invention provides a nuclease that binds to a guide polynucleotide can include a conserved scaffold sequence. For example, the nucleic acid-guided nucleases for use in the present disclosure can bind to a conserved pseudoknot region.

[00144] An engineered guide polynucleotide, or engineered gRNA, can be the sequence of any one of SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128 or another suitable known gRNA. In some embodiments, the engineered polynucleotide (gRNA) can be split into fragments encompassing a synthetic tracrRNA and crRNA. In some examples, any one of SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128 can be split into fragments encompassing a synthetic tracrRNA and crRNA.

[00145] As used herein, “guide nucleic acid” or “guide polynucleotide” can refer to one or more polynucleotides and can include 1) a guide sequence capable of hybridizing to a target sequence and 2) a scaffold sequence capable of interacting with or complexing with a nucleic acid-guided nuclease as described herein. A guide nucleic acid can be provided as one or more nucleic acids. In some embodiments, the guide sequence and the scaffold sequence are provided

as a single polynucleotide. In other aspects, guide nucleic acid can include at least one amplicon targeting fragments.

[00146] A guide nucleic acid can be compatible with a nucleic acid-guided nuclease when the two elements can form a functional targetable nuclease complex capable of cleaving a target sequence. In certain methods, a compatible scaffold sequence for a compatible guide nucleic acid can be found by scanning sequences adjacent to a native nucleic acid-guided nuclease loci. For example, native nucleic acid-guided nucleases can be encoded on a genome within proximity to a corresponding compatible guide nucleic acid or scaffold sequence.

[00147] Nucleic acid-guided nucleases can be compatible with guide nucleic acids that are not found within the nucleases endogenous host. Such orthogonal guide nucleic acids can be determined by empirical testing. Orthogonal guide nucleic acids can come from different bacterial species or be synthetic or otherwise engineered to be non-naturally occurring.

[00148] Orthogonal guide nucleic acids that are compatible with a common nucleic acid-guided nuclease can comprise one or more common features. Common features can include sequence outside a pseudoknot region. Common features can include a pseudoknot region. Common features can include a primary sequence or secondary structure.

[00149] A guide nucleic acid can be engineered to target a desired target sequence by altering the guide sequence such that the guide sequence is complementary to the target sequence, thereby allowing hybridization between the guide sequence and the target sequence. A guide nucleic acid with an engineered guide sequence can be referred to as an engineered guide nucleic acid. Engineered guide nucleic acids are often non-naturally occurring and are not found in nature.

[00150] Engineered guide nucleic acids can be formed using a Synthetic Tracr RNA (STAR) system. STAR, when combined with a Cas12a protein, can form at least one ribonucleoprotein (RNP) complex that targets a specific genomic locus. STAR takes advantage of the natural properties of the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) where the CRISPR system functions much like an immune system against invading viruses and plasmid DNA. Short DNA sequences (spacers) from invading viruses are incorporated at CRISPR loci within the bacterial genome and serve as "memory" of previous infections. Reinfection triggers complementary mature CRISPR RNA (crRNA) to find a matching viral sequence. Together, the crRNA and trans-activating crRNA (tracrRNA) guide CRISPR-associated (Cas) nuclease to

cleave double-strand breaks in “foreign” DNA sequences. The prokaryotic CRISPR “immune system” has been engineered to function as an RNA-guided, mammalian genome editing tool that is simple, easy and quick to implement. STAR (which includes synthetic crRNA and tracrRNA) when combined with Cas12a protein can form ribonucleoprotein (RNP) complexes that target a specific genomic locus. Engineered guide nucleic acids formed with the RNA (STAR) system can result in a split gRNA. An example of a split gRNA for use as disclosed herein can include the sequence represented by SEQ ID NO: 128.

[00151] In some embodiments, a ribonucleoprotein (RNP) complex of use herein can include at least one nuclease disclosed herein. In some aspects, a RNP complex can include at least one nuclease having an amino acid sequence of about 75%, about 85%, about 95%, about 99%, or is identical to one or more sequences of SEQ ID NOs: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and 108-117. In some embodiments, an RNP complex including a nuclease disclosed herein can further include at least one STAR gRNA. In another embodiment, an RNP complex including a nuclease disclosed herein can further include at least one non-STAR gRNA. In other embodiments, an RNP complex including a nuclease disclosed herein can further include at least one polynucleotide. In certain embodiments, a polynucleotide included in an RNP complex disclosed herein can be greater than about 50 nucleotides in length. In other embodiments, a polynucleotide included in a RNP complex disclosed herein can be about 50, to about 100, to about 150, to about 200, to about 250, to about 300, to about 350, to about 400, to about 450 to about 500, to about 750, to about 1000 nucleotides, or greater than 1000 nucleotides in length. In some embodiments, more than one nuclease can be included in an RNP complex contemplated herein in order to affect overall editing efficiency of the complex on a targeted genome. In certain embodiments, more than one gRNA can be added to the RNP complex to allow for multiplexed editing of more than one site in a single transfection. In certain embodiments, more than one DNA template can be added to an RNP complex to allow for multiplexed editing at one or more sites based on a desired repair outcome of a targeted genome.

Nuclease Systems

[00152] Other embodiments disclosed herein concern targetable nuclease systems. In certain embodiments, a targetable nuclease system can include a nucleic acid-guided nuclease and a compatible guide nucleic acid (also referred to interchangeably herein as “guide polynucleotide” and “gRNA”). A targetable nuclease system herein can include a novel nucleic acid-guided

nuclease or a polynucleotide sequence encoding the novel nucleic acid-guided nuclease disclosed herein. In other embodiments, a targetable nuclease system can include a guide nucleic acid or a polynucleotide sequence encoding the guide nucleic acid and a known or novel gRNA.

[00153] In accordance with these embodiments, a targetable nuclease system as disclosed herein can be characterized by elements that promote the formation of a targetable nuclease complex at the site of a target sequence (*e.g.* eukaryotic genome sequence for editing), where the targetable nuclease complex includes at least a nucleic acid-guided nuclease and a guide nucleic acid. A guide nucleic acid (gRNA) together with a nucleic acid-guided nuclease forms a targetable nuclease complex capable of binding to a target sequence within a target polynucleotide, as determined by the guide sequence of the guide nucleic acid.

[00154] In certain embodiments, to generate a double stranded break in the target sequence, a targetable nuclease complex can bind to a target sequence as determined by the guide nucleic acid (gRNA), and the nuclease recognizes a protospacer adjacent motif (PAM) sequence adjacent to the target sequence in order to cut the target sequence. In some embodiments, a targetable nuclease complex can include a nucleic acid-guided nuclease encoded by one or more of SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and 108-117 and a compatible guide nucleic acid. In other embodiments, a targetable nuclease complex can include a nucleic acid-guided nuclease encoded by one or more of a nuclease represented by SEQ ID NO: 3, 16, 29, 42, 55, 68, 81, 94 and 107 and a compatible guide nucleic acid. In yet other embodiments, a targetable nuclease complex can include a nucleic acid-guided nuclease and a compatible guide nucleic acid represented by SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128. In other embodiments, a targetable nuclease complex can include a nucleic acid-guided nuclease according to one or more of SEQ ID NO: 4-13, 17-26, 30-39, 43-52, 56-65, 69-78, 82-91, 95-104 and 108-117 and a compatible guide nucleic acid represented by SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 and 128. In accordance with these embodiments, the guide nucleic acid can include a scaffold sequence compatible with the nucleic acid-guided nuclease. In other embodiments, the guide sequence can be engineered to be complementary to any desired target sequence for efficient editing of the target sequence. In other embodiments, the guide sequence can be engineered to hybridize to any desired target sequence. In some embodiments, the target nucleic acid sequence has 20 nucleotides in length. In some embodiments, the target nucleic acid has less than 20 nucleotides in length. In some

embodiments, the target nucleic acid has more than 20 nucleotides in length. In some embodiments, the target nucleic acid has at least: 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 or more nucleotides in length. In some embodiments, the target nucleic acid has at most: 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30 or more nucleotides in length.

[00155] In some embodiments, a target sequence of a targetable nuclease complex can be any polynucleotide endogenous or exogenous to a prokaryotic or eukaryotic cell, or in an i*n vitro* system for verification or otherwise. In other embodiments, a target sequence can be a polynucleotide residing in the nucleus of the eukaryotic cell. A target sequence can be a sequence coding a gene product (*e.g.*, a protein) or a non-coding sequence (*e.g.*, a regulatory polynucleotide or a junk DNA). It is contemplated herein that the target sequence should be associated with a PAM; that is, a short sequence recognized by a targetable nuclease complex. In some embodiments, sequence and length requirements for a PAM differ depending on the nucleic acid-guided nuclease selected. In certain embodiments, PAM sequences can be about 2-5 base pair sequences adjacent the target sequence or longer, depending on the PAM desired. Examples of PAM sequences are given in the Examples section below, and the skilled person will be able to identify further PAM sequences for use with a given nucleic acid-guided nuclease as these are not intended to limit this aspect of the inventions. Further, engineering of a PAM Interacting (PI) domain can allow programming of PAM specificity, improve target site recognition fidelity, and increase the versatility of a nucleic acid-guided nuclease genome engineering platform. Nucleic acid-guided nucleases can be engineered to alter their PAM specificity, for example as previously described.

[00156] In some embodiments, at least one PAM site can be a nucleotide sequence in close proximity to a target sequence. In accordance with these embodiments, a nucleic acid-guided nuclease can only cleave a target sequence if at least one corresponding PAM is present as selected herein. In certain embodiments, PAM sites can be nucleic acid-guided nuclease-specific and can be different between two different nucleic acid-guided nucleases. In accordance with these embodiments, a PAM can be positioned or located 5′ of a target sequence, 3′ of a target sequence, consecutively or combination. A PAM can be upstream of a target sequence, downstream of a target sequence, repeated or a combination. A PAM can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotides in length. In some embodiments, a PAM is between 2-6 nucleotides in length. In some embodiments, a PAM sequence for use herein can be 5′-TTN-3′. In other

embodiments, a PAM sequence for use herein can be 5′-TTTN-3′. In certain embodiments, a PAM sequence for use herein can be different than the 5′-TTN-3′ or 5′-TTTN-3′ sequence described above. In some embodiments, a PAM sequence for use herein can depend on (or for example, correspond to) one or more of nucleases disclosed herein (*e.g.* matching or pairing for efficient editing). In some embodiments, various methods (*e.g.*, in silico and/or wet lab methods) for identification of an appropriate PAM sequence are known in the art and can be used herein.

[00157] In some embodiments disclosed herein, a PAM can be provided on a separate oligonucleotide. In accordance with these embodiments, providing PAM on an adjacent or separate oligonucleotide allows cleavage of a neighboring target sequence that otherwise would not be able to be cleaved or edited because no adjacent PAM is present on the targeted sequence itself.

[00158] Polynucleotide sequences encoding a component of a targetable nuclease system can include one or more vectors. The term "vector" as used herein can refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (*e.g.* circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (*e.g.* retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Other vectors (e.g., non-episomal mammalian vectors) can be integrated into the genome of a host cell upon introduction into the host cell. Recombinant expression vectors can include a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, can mean that the recombinant expression vectors include one or more regulatory elements, which can be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed.

[00159] In some embodiments, a regulatory element can be operably linked to one or more elements of a targetable nuclease system so as to drive expression of the one or more components of the targetable nuclease system.

[00160] In some embodiments, a vector can include a regulatory element operably linked to a polynucleotide sequence encoding a nucleic acid-guided nuclease. The polynucleotide sequence encoding the nucleic acid-guided nuclease can be codon optimized for expression in particular cells, such as prokaryotic or eukaryotic cells. Eukaryotic cells can be yeast, fungi, algae, plant, animal, or human cells. Eukaryotic cells can be those derived from a particular organism, such as a mammal, including but not limited to human, mouse, rat, rabbit, dog, or non-human mammal including non-human primate. Plant cells can include, without limitation, cells from seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen and microspores.

[00161] As used herein, 'codon optimization' can refer to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon or more of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. As contemplated herein, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the "Codon Usage Database."

[00162] In some embodiments, a nucleic acid-guided nuclease and one or more guide nucleic acids can be delivered either as DNA or RNA. Delivery of a nucleic acid-guided nuclease and guide nucleic acid both as RNA (unmodified or containing base or backbone modifications) molecules can be used to reduce the amount of time that the nucleic acid-guided nuclease persist in the cell (*e.g.* reduced half-life). This can reduce the level of off-target cleavage activity in the target cell. Since delivery of a nucleic acid-guided nuclease as mRNA takes time to be translated into protein, an aspect herein can include delivering a guide nucleic acid several hours following the delivery of the nucleic acid-guided nuclease mRNA, to maximize the level of guide nucleic acid available for interaction with the nucleic acid-guided nuclease protein. In other cases, the nucleic acid-guided nuclease mRNA and guide nucleic acid can be delivered concomitantly. In other examples, the guide nucleic acid can be delivered sequentially, such as 0.5, 1, 2, 3, 4, or more hours after the nucleic acid-guided nuclease mRNA.

[00163] In some embodiments, guide nucleic acid in the form of RNA or encoded on a DNA expression cassette can be introduced into a host cell that includes a nucleic acid-guided nuclease encoded on a vector or chromosome. The guide nucleic acid can be provided in the cassette

having one or more polynucleotides, which can be contiguous or non-contiguous in the cassette. In some embodiments, the guide nucleic acid can be provided in the cassette as a single contiguous polynucleotide. In other embodiments, a tracking agent can be added to the guide nucleic acid in order to track distribution and activity.

[00164] In other embodiments, a variety of delivery systems can be used to introduce a nucleic acid-guided nuclease (e.g. DNA or RNA or other nucleic acid construct) and guide nucleic acid (e.g. DNA or RNA or other nucleic acid construct) into a host cell. In accordance with these embodiments, systems of use for embodiments disclosed herein can include, but are not limited to, yeast systems, lipofection systems, microinjection systems, biolistic systems, virosomes, liposomes, immunoliposomes, polycations, lipid:nucleic acid conjugates, virions, artificial virions, viral vectors, electroporation, cell permeable peptides, nanoparticles, nanowires, exosomes. Molecular trojan horse liposomes or similar can be used to deliver an engineered nuclease and guide nuclease for example, across the blood brain barrier.

[00165] In some embodiments, an editing template can also be provided. In accordance with these embodiments, an editing template can be a component of a vector as described herein, contained in a separate vector, or provided as a separate polynucleotide, such as an oligonucleotide, linear polynucleotide, or synthetic polynucleotide. In some embodiments, an editing template is on the same polynucleotide as a guide nucleic acid. In other embodiments, an editing template is designed to serve as a template in homologous recombination, such as within or near a target sequence nicked or cleaved by a nucleic acid-guided nuclease as a part of a complex for editing as disclosed herein. An editing template polynucleotide can be of any suitable length, such as about or less or more than about 5, 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, or more nucleotides in length. In some embodiments, an editing template polynucleotide can be complementary to a portion of a polynucleotide that can include the target sequence or be adjacent or in close proximity to a target sequence for editing. In accordance with these embodiments, when optimally aligned, an editing template polynucleotide can overlap with one or more nucleotides of a target sequences (*e.g.* about or more than about 1, 5, 10, 15, 20, 25, 30, 35, 40, or more nucleotides). In some embodiments, when optimally aligned, an editing template sequence and a polynucleotide can include a target sequence optimally aligned, where the nearest nucleotide of the template polynucleotide is within about 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, or more nucleotides from the target sequence.

[00166] In some embodiments, methods are provided for delivering one or more polynucleotides, such as or one or more vectors or linear polynucleotides as described herein, one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell. In some aspects, the invention further provides cells produced by such methods, and organisms can include or produced from such cells. In some embodiments, an engineered nuclease in combination with (and optionally complexed with) a guide nucleic acid is delivered to a cell.

[00167] In certain embodiments, conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in cells, such as prokaryotic cells, eukaryotic cells, plant cells, mammalian cells, or target tissues. Such methods can be used to administer nucleic acids encoding components of an engineered nucleic acid-guided nuclease system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (*e.g.* a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. Any gene therapy method known in the art is contemplated of use herein. Methods of non-viral delivery of nucleic acids include are contemplated herein. Adeno-associated virus ("AAV") vectors can also be used to transduce cells with target nucleic acids, *e.g.*, in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures.

[00168] In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors, linear polynucleotides, polypeptides, nucleic acid-protein complexes, or any combination thereof as described herein. In some embodiments, a cell can be transfected in vitro, in culture, or ex vivo. In some embodiments, a cell can be transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected can be taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line.

[00169] In some embodiments, a cell transfected with one or more vectors, linear polynucleotides, polypeptides, nucleic acid-protein complexes, or any combination thereof as described herein is used to establish a new cell line can include one or more transfection-derived sequences. In some embodiments, a cell transiently transfected with the components of an engineered nucleic acid-guided nuclease system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity

of an engineered nuclease complex, is used to establish a new cell line can include cells containing the modification but lacking any other exogenous sequence.

[00170] In some embodiments, one or more vectors described herein are used to produce a non-human transgenic cell, organism, animal, or plant. In some embodiments, the transgenic animal is a mammal, such as a mouse, rat, or rabbit. Methods for producing transgenic cells, organisms, plants, and animals are known in the art, and generally begin with a method of cell transformation or transfection, such as described herein.

[00171] In certain embodiments, an engineered nuclease complex, "target sequence" can refer to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of an engineered nuclease complex. A target sequence can include any polynucleotide, such as DNA, RNA, or a DNA-RNA hybrid. A target sequence can be located in the nucleus or cytoplasm of a cell. A target sequence can be located in vitro or in a cell-free environment. A target sequence can be eukaryotic or prokaryotic target sequence.

[00172] In some embodiments, formation of an engineered nuclease complex can include a guide nucleic acid hybridized to a target sequence and complexed with one or more engineered nucleases as disclosed herein leading to cleavage of one or both strands in or near (*e.g.* within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) of the target sequence. In certain embodiments, cleavage can occur within a target sequence, 5′ of the target sequence, upstream of a target sequence, 3′ of the target sequence, or downstream of a target sequence.

[00173] In some embodiments, one or more vectors driving expression of one or more components of a targetable nuclease system can be introduced into a host cell or used *in vitro* such formation of a targetable nuclease complex at one or more target sites. In some embodiments, a nucleic acid-guided nuclease and a guide nucleic acid could each be operably linked to separate regulatory elements on separate vectors. In other embodiments, two or more of the elements expressed from the same or different regulatory elements, can be combined in a single vector, with one or more additional vectors providing any components of the targetable nuclease system not included in the first vector. Targetable nuclease system elements that are combined in a single vector can be arranged in any suitable orientation, such as one element located 5′ with respect to ("upstream" of) or 3′ with respect to ("downstream" of) a second element. In some embodiments, the coding sequence of one element can be located on the same

or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In other embodiments, a single promoter drives expression of a transcript encoding a nucleic acid-guided nuclease and one or more guide nucleic acids. In certain embodiments, a nucleic acid-guided nuclease and one or more guide nucleic acids are operably linked to and expressed from the same promoter. In other embodiments, one or more guide nucleic acids or polynucleotides encoding the one or more guide nucleic acids are introduced into a cell or in vitro environment already can include a nucleic acid-guided nuclease or polynucleotide sequence encoding the nucleic acid-guided nuclease.

[00174] In certain methods, when multiple different guide sequences are used, a single expression construct can be used to target nuclease activity to multiple different, corresponding target sequences (to the selected guide sequences etc.) within a cell or cells within a tissue, *ex vivo* or *in vitro*. For example, a single vector can include about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide sequences. In some embodiments, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-sequence-containing vectors can be provided, and optionally delivered to a cell or *in vitro*.

[00175] In other embodiments, methods and compositions disclosed herein can include more than one guide nucleic acid, such that each guide nucleic acid has a different guide sequence, thereby targeting a different target sequence. In accordance with these embodiments, multiple guide nucleic acids can be using in multiplexing, wherein multiple targets are targeted simultaneously. Additionally, or alternatively, multiple guide nucleic acids can be introduced into a population of cells or cells within a tissue, such that each cell in a population of cells receives a different or random guide nucleic acid, thereby targeting multiple different target sequences across a population of cells for optimal editing outcomes in some embodiments disclosed herein. In certain embodiments, the collection of subsequently altered cells can be referred to as a library.

[00176] In other embodiments, methods and compositions disclosed herein can include multiple different nucleic acid-guided nucleases, each with one or more different corresponding guide nucleic acids, thereby allowing targeting of different target sequences by different nucleic acid-guided nucleases. In some embodiments, each nucleic acid-guided nuclease can correspond to a distinct plurality of guide nucleic acids, allowing two or more non-overlapping, partially overlapping, or completely overlapping multiplexing events to occur.

[00177] In some embodiments, nucleic acid-guided nucleases herein can have DNA cleavage activity or RNA cleavage activity. In some embodiments, the nucleic acid-guided nuclease directs cleavage of one or both strands at the location of a target sequence, such as within the target sequence and/or within the complement of the target sequence. In certain embodiments, the nucleic acid-guided nuclease directs cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence.

[00178] In some embodiments, methods of modifying a target sequence *in vitro*, or in a prokaryotic or eukaryotic cell, which can be *in vivo, ex vivo,* or *in vitro* are disclosed. In some embodiments, the method includes sampling a cell or population of cells such as prokaryotic cells, or those from a human or non-human animal or plant (including micro-algae), and modifying the cell or cells. Culturing can occur at any stage in vitro or ex vivo. The cell or cells can be re-introduced into the host, such as a non-human animal or plant (including micro-algae). In some embodiments, compositions and methods disclosed herein can be used to improve resistance in a plant to microbes or changes in climate. In some embodiments, for re-introduced cells, they can include stem cells or other progenitor cells.

[00179] In some embodiments, methods can include allowing a targetable nuclease complex to bind to the target sequence to effect cleavage of the target sequence, thereby modifying the target sequence, wherein the targetable nuclease complex includes a nucleic acid-guided nuclease complexed with a guide nucleic acid wherein the guide sequence of the guide nucleic acid is hybridized to a target sequence within a target polynucleotide. In other embodiments, methods of modifying expression of a target polynucleotide in *in vitro* or in a prokaryotic or eukaryotic cell are provided. In some embodiments, methods herein can include allowing a targetable nuclease complex to bind to a target sequence with the target polynucleotide such that the binding results in increased or decreased expression of the target polynucleotide. In accordance with these embodiments, the targetable nuclease complex can include a nucleic acid-guided nuclease complexed with a guide nucleic acid, where the guide sequence of the guide nucleic acid is hybridized to a target sequence within the target polynucleotide.

[00180] In some embodiments, kits are provided containing one or more of the elements disclosed in the above methods and compositions and at least one container. Elements can be provided individually or in combinations, and can be provided in any suitable container, such as

a vial, a bottle, or a tube. In some embodiments, kits can include instructions in one or more languages, for example in more than one language. In some embodiments, kits can include components of novel nucleases and/or gRNAs disclosed herein or compositions for making these components. In other embodiments, kits contemplated herein can include all components and containers needed for performing an efficient editing of a target genome.

[00181] In some embodiments, a kit contemplated herein includes one or more reagents for use in a process utilizing one or more of the elements described herein. Reagents can be provided in any suitable container. For example, a kit can provide one or more reaction or storage buffers. Reagents can be provided in a form that is usable in a particular assay, or in a form that requires addition of one or more other components before use (*e.g.* in concentrate or lyophilized form). A buffer can be any buffer, including but not limited to a sodium carbonate buffer, a sodium bicarbonate buffer, a borate buffer, a Tris buffer, a MOPS buffer, a HEPES buffer, and combinations thereof. In some embodiments, the buffer is alkaline. In other embodiments, the buffer has a pH from about 7 to about 10. In other embodiments, the kit includes one or more oligonucleotides corresponding to a guide sequence for insertion into a vector so as to operably link the guide sequence and a regulatory element. In some embodiments, the kit includes an editing template.

[00182] In some embodiments, a targetable nuclease complex has a wide variety of utility including modifying (*e.g.*, deleting, inserting, translocating, inactivating, activating) a target sequence in a multiplicity of cell types. In some embodiments, a targetable nuclease complex can have a broad spectrum of applications in, *e.g.*, biochemical pathway optimization, genome-wide studies, genome engineering, gene therapy, drug screening, disease diagnosis, and prognosis. An exemplary targetable nuclease complex includes a nucleic acid-guided nuclease as disclosed herein complexed with a guide nucleic acid, wherein the guide sequence of the guide nucleic acid can hybridize to a target sequence within the target polynucleotide. A guide nucleic acid can include a guide sequence linked to a scaffold sequence. A scaffold sequence can include one or more sequence regions with a degree of complementarity such that together they form a secondary structure.

[00183] In some embodiments, an editing template polynucleotide can include a sequence to be integrated (*e.g.*, a mutated gene). In accordance with these embodiments, a sequence for integration can be a sequence endogenous or exogenous to the cell. Examples of a sequence to be

integrated include polynucleotides encoding a protein or a non-coding RNA (*e.g.*, a microRNA). In certain embodiments, the sequence for integration can be operably linked to an appropriate control sequence or sequences. Alternatively, the sequence to be integrated can provide a regulatory function. In certain embodiments, sequences to be integrated can be a mutated or variant of an endogenous wild-type sequence. In other embodiments, sequences to be integrated can be a wild-type version of an endogenous mutated sequence. Additionally, or alternatively, sequences to be integrated can be a variant or mutated form of an endogenous mutated or variant sequence.

[00184] An upstream or downstream sequence can encompass from about 20 bp to about 2500 bp, for example, about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, or 2500 bp. In some methods, the exemplary upstream or downstream sequence has about 15 bp to about 50 bp, about 30 bp to about 100 bp, about 200 bp to about 2000 bp, about 600 bp to about 1000 bp, or more particularly about 700 bp to about 1000 bp.

[00185] In some methods, the editing template polynucleotide can further include a marker. In accordance with these embodiments, a marker can make it easy to screen for targeted integrations in order to assess efficiency and accuracy. Examples of suitable markers include, but are not limited to, restriction sites, fluorescent proteins, or selectable markers. In some embodiments, exogenous polynucleotide templates disclosed herein can be constructed using recombinant techniques.

[00186] In some embodiments, methods for modifying a target polynucleotide by integrating an editing template polynucleotide, can be by introducing a double-stranded break into the genome sequence by an engineered nuclease complex, the break can be repaired via homologous recombination using an editing template such that a desired template is integrated into the target polynucleotide. The presence of a double-stranded break can increase the efficiency of integration of the editing template for directed outcome.

[00187] In other embodiments, methods are disclosed for modifying expression of a polynucleotide in a cell. In accordance with these embodiments, some methods can include increasing or decreasing expression of a target polynucleotide by using a targetable nuclease complex that binds to the target polynucleotide.

[00188] Detection of the gene expression level can be conducted in real time in an amplification assay. In one aspect, the amplified products can be directly visualized with fluorescent DNA-binding agents including but not limited to DNA intercalators and DNA groove binders. Because the amount of the intercalators incorporated into the double-stranded DNA molecules is typically proportional to the amount of the amplified DNA products, an amount of the amplified products can be determined by quantifying the fluorescence of the intercalated dye using conventional optical systems in the art. DNA-binding dye suitable for this application include, but are not limited to, SYBR green, SYBR blue, DAPI, propidium iodine, Hoeste, SYBR gold, ethidium bromide, acridines, proflavine, acridine orange, acriflavine, fluorcoumanin, ellipticine, daunomycin, chloroquine, distamycin D, chromomycin, homidium, mithramycin, ruthenium polypyridyls, anthramycin, and others known by one of skill in the art.

[00189] In some embodiments, other fluorescent labels such as sequence specific traceable probes can be employed in the amplification reaction to facilitate the detection and quantification of the amplified products. Probe-based quantitative amplification relies on the sequence-specific detection of a desired amplified product. In these methods, fluorescent, target-specific probes (*e.g.*, TaqMan™ probes) can be used resulting in increased specificity and sensitivity of detection and quantitative analysis. Methods for performing probe-based quantitative amplification are well known in the art and contemplated of use herein.

[00190] In certain embodiments, an agent-induced change in expression of sequences associated with a signaling biochemical pathway can also be determined by examining the corresponding gene products. Determining protein levels can involve (a) contacting the protein contained in a biological sample with an agent that specifically binds to a protein associated with a signaling biochemical pathway; and (b) identifying an agent:polypeptide complex so formed. In one aspect of this embodiment, the agent that specifically binds a protein associated with a signaling biochemical pathway can be an antibody, such as a monoclonal antibody.

[00191] In some embodiments, the amount of agent:polypeptide complexes formed during the binding reaction can be quantified by standard quantitative assays. As disclosed above, the formation of agent:polypeptide complex can be measured directly by the amount of label remained at the site of binding. In an alternative, the protein associated with a signaling biochemical pathway is tested for its ability to compete with a labeled analog for binding sites on the specific agent. In this competitive assay, the amount of label captured is inversely

proportional to the amount of protein sequences associated with a signaling biochemical pathway present in a test sample.

[00192] In other embodiments, a number of techniques for protein analysis based on the general principles outlined above are available in the art. They include, but are not limited to, radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunofluorescent assays, and SDS-PAGE.

[00193] In some embodiments, methods herein can be used to discern the expression pattern of a protein associated with a signaling biochemical pathway in different bodily tissue, in different cell types, and/or in different subcellular structures. These studies can be performed with the use of tissue-specific, cell-specific or subcellular structure specific antibodies capable of binding to protein markers that are preferentially expressed in certain tissues, cell types, or subcellular structures.

[00194] In some embodiments, an altered expression of a gene associated with a signaling biochemical pathway can also be determined by examining a change in activity of the gene product relative to a control cell. The assay for an agent-induced change in the activity of a protein associated with a signaling biochemical pathway will dependent on the biological activity and/or the signal transduction pathway that is under investigation. For example (but not limited to), where the protein is a kinase, a change in its ability to phosphorylate the downstream substrate(s) can be determined by a variety of assays known in the art. Representative assays include but are not limited to immunoblotting and immunoprecipitation with antibodies such as anti-phosphotyrosine antibodies that recognize phosphorylated proteins. In addition, kinase activity can be detected by high throughput chemiluminescent assays.

[00195] In certain embodiments, where the protein associated with a signaling biochemical pathway is part of a signaling cascade leading to a fluctuation of intracellular pH condition, pH sensitive molecules such as fluorescent pH dyes can be used as the reporter molecules. In another example, where the protein associated with a signaling biochemical pathway is an ion channel, fluctuations in membrane potential and/or intracellular ion concentration can be monitored. A number of commercial kits and high-throughput devices are particularly suited for a rapid and robust screening for modulators of ion channels. Representative instruments include FLIPR™

(Molecular Devices, Inc.) and VIPR (Aurora Biosciences). These instruments are capable of detecting reactions in over 1000 sample wells of a microplate simultaneously, and providing real-time measurement and functional data within a second or even a millisecond.

[00196] In practicing any of the methods disclosed herein, a suitable vector can be introduced to a cell, tissue, organism, or an embryo via one or more methods known in the art, including without limitation, microinjection, electroporation, sonoporation, biolistics, calcium phosphate-mediated transfection, cationic transfection, liposome transfection, dendrimer transfection, heat shock transfection, nucleofection transfection, magnetofection, lipofection, impalefection, optical transfection, proprietary agent-enhanced uptake of nucleic acids, and delivery via liposomes, immunoliposomes, virosomes, or artificial virions. In some methods, the vector can be introduced into an embryo by microinjection. The vector or vectors disclosed herein can be microinjected into the nucleus or the cytoplasm of the embryo. In some methods, the vector or vectors can be introduced into a cell by nucleofection.

[00197] In some embodiments, a target polynucleotide of a targetable nuclease complex can be any polynucleotide endogenous or exogenous to the host cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell, the genome of a prokaryotic cell, or an extrachromosomal vector of a host cell. The target polynucleotide can be a sequence coding a gene product (*e.g.*, a protein) or a non-coding sequence (*e.g.*, a regulatory polynucleotide or a junk DNA).

[00198] Some embodiments disclosed herein relate to use of an engineered nucleic acid guided nuclease system disclosed herein; for example, in order to target and knock out genes, amplify genes and/or repair particular mutations associated with DNA repeat instability and a medical disorder. This nuclease system can be used to harness and to correct these defects of genomic instability. In other embodiments, engineered nucleic acid guided nuclease systems disclosed herein can be used for correcting defects in the genes associated with Lafora disease. Lafora disease is an autosomal recessive condition which is characterized by progressive myoclonus epilepsy which can start as epileptic seizures in adolescence. This condition causes seizures, muscle spasms, difficulty walking, dementia, and eventually death.

[00199] In yet another aspect of the invention, the engineered/novel nucleic acid guided nuclease system disclosed herein can be used to correct genetic-eye disorders that arise from several genetic mutations.

[00200] In other embodiments, methods herein can be used to correct defects associated with a wide range of genetic diseases which are described, but not limited to those on the website of the National Institutes of Health under the topic subsection Genetic Disorders. Certain genetic disorders of the brain can include, but are not limited to, Adrenoleukodystrophy, Agenesis of the Corpus Callosum, Aicardi Syndrome, Alpers' Disease, glioblastoma, Alzheimer's, Barth Syndrome, Batten Disease, CADASIL, Cerebellar Degeneration, Fabry's Disease, Gerstmann-Straussler-Scheinker Disease, Huntington's Disease and other Triplet Repeat Disorders, Leigh's Disease, Lesch-Nyhan Syndrome, Menkes Disease, Mitochondrial Myopathies and NINDS Colpocephaly or other brain disorder scontributed to by genetically-linked causation.

[00201] In some embodiments, a genetically-linked disorder can be a neoplasia. In some embodiments, where the condition is neoplasia, targeted genes can include one or more genes listed above. In some embodiments, a health condition contemplated herein can be Age-related Macular Degeneration or a Schizophrenic-related Disorder. In other embodiments, the condition can be a Trinucleotide Repeat disorder or Fragile X Syndrome. In other embodiments, the condition can be a Secretase-related disorder. In some embodiments, the condition can be a Prion-related disorder. In some embodiments, the condition can be ALS. In some embodiments, the condition can be a drug addiction related to prescription or illegal substances. In accordance with these embodiments, addiction-related proteins can include ABAT for example.

[00202] In some embodiments, the condition can be Autism. In some embodiments, the health condition can be an inflammatory-related condition, for example, over-expression of a pro-inflammatory cytokine. Other inflammatory condition-related proteins can include one or more of monocyte chemoattractant protein-1 (MCP1) encoded by the Ccr2 gene, the C C chemokine receptor type 5 (CCR5) encoded by the Ccr5 gene, the IgG receptor IIB (FCGR2b, also termed CD32) encoded by the Fcgr2b gene, or the Fc epsilon Rlg (FCER1g) protein encoded by the Fcerlg gene, or other protein having a genetic-link to these conditions.

[00203] In some embodiments, the condition can be Parkinson's Disease. In accordance with these embodiments, proteins associated with Parkinson's disease can include, but are not limited to, a-synuclein, DJ-1, LRRK2, PINK1, Parkin, UCHL1, Synphilin-1, and NURR1.

[00204] Cardiovascular- associated proteins that contribute to a cardiac disorder, can include, but are not limited to, IL1b (interleukin 1-beta), XDH (xanthine dehy-drogenase), TP53 (tumor protein p53), PTGIS (prostaglandin 12 (prostacyclin) synthase), MB (myoglobin), IL4 (interleu-

kin 4), ANGPT1 (angiopoietin 1), ABCG8 (ATP-binding cas¬sette, sub-family G (WHITE), member 8), or CTSK (cathepsin K), or other known contributors to these conditions.

[00205] In certain embodiments, the condition can be Alzheimer's disease. In accordance with these embodiments, Alzheimer's disease associated proteins can include very low density lipoprotein receptor protein (VLDLR) encoded by the VLDLR gene, ubiquitin-like modifier activating enzyme 1 (UBA1) encoded by the UBA1 gene, or for example, NEDD8-activating enzyme El catalytic subunit protein (UBE1C) encoded by the UBA3 gene or other genetically-related contributor.

[00206] In other embodiments, the condition can be an Autism Spectrum Disorder. In accordance with these embodiments, proteins associated Autism Spectrum Disorders can include the benzodiazapine receptor (peripheral) associated protein 1 (BZRAP1) encoded by the BZRAP1 gene, the AF4/FMR2 family member 2 protein (AFF2) encoded by the AFF2 gene (also termed MFR2), the fragile X mental retardation autosomal homolog 1 protein (FXR1) encoded by the FXR1 gene, or the fragile X mental retardation autosomal homolog 2 protein (FXR2) encoded by the FXR2 gene, or other genetically-related contributor.

[00207] In some embodiments, the condition can be Macular Degeneration. In accordance with these embodiments, proteins associated with Macular Degeneration can include, but are not limited to, the ATP-binding cassette, sub-family A (ABC1) member 4 protein (ABCA4) encoded by the ABCR gene, the apolipoprotein E protein (APOE) encoded by the APOE gene, or the chemokine (CC motif) Llg and 2 protein (CCL2) encoded by the CCL2 gene, or other genetically-related contributor.

[00208] In certain embodiments, the condition can be Schizophrenia. In accordance with these embodiments, proteins associated with Schizophrenia In accordance with these embodiments, proteins associated with Schizophrenia y include NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, DISCI, GSK3B, and combinations thereof.

[00209] In other embodiments, the condition can be tumor suppression. In accordance with these embodiments, proteins associated with tumor suppression can include ATM (ataxia telangiectasia mutated), ATR (ataxia telangiectasia and Rad3 related), EGFR (epidermal growth factor receptor), ERBB2 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 2), ERBB3 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 3), ERBB4 (v-erb-b2 erythroblastic

leukemia viral oncogene homolog 4), Notch 1, Notch2, Notch 3, or Notch 4 or other genetically-related contributor.

[00210] In yet other embodiments, the condition can be a secretase disorder. In accordance with these embodiments, proteins associated with a secretase disorder can include PSENEN (presenilin enhancer 2 homolog (C. elegans)), CTSB (cathepsin B), PSEN1 (presenilin 1), APP (amyloid beta (A4) precursor protein), APH1B (anterior pharynx defective 1 homolog B (C. elegans)), PSEN2 (presenilin 2 (Alzheimer disease 4)), or BACE1 (beta-site APP-cleaving enzyme 1), or other genetically-related contributor.

[00211] In certain embodiments, the condition can be Amyotrophic Lateral Sclerosis. In accordance with these embodiments, proteins associated with can include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof or other genetically-related contributor.

[00212] In some embodiments, the condition can be a prion disease disorder. In accordance with these embodiments, proteins associated with a prion diseases disorder can include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof or other genetically-related contributor. Examples of proteins related to neurodegenerative conditions in prion disorders can include A2M (Alpha-2-Macro-globulin), AATF (Apoptosis antagonizing transcription factor), ACPP (Acid phosphatase prostate), ACTA2 (Actin alpha 2 smooth muscle aorta), ADAM22 (ADAM metallopeptidase domain), ADORA3 (Adenosine A3 receptor), or ADRA1D (Alpha-1D adrenergic receptor for Alpha-1D adrenoreceptor), or other genetically-related contributor.

[00213] In some embodiments, the condition can be an immunodeficiency disorder. In accordance with these embodiments, proteins associated with an immunodeficiency disorder can include A2M [alpha-2-macroglobulin]; AANAT [aryla-lkylamine N-acetyltransferase]; ABCA1 [ATP-binding cassette, sub-family A (ABC1), member 1]; ABCA2 [ATP-binding cassette, sub-family A (ABC1), member 2]; or ABCA3 [ATP-binding cassette, sub-family A (ABC 1), member 3]; or other genetically-related contributor.

[00214] In certain embodiments, the condition can be an immunodeficiency disorder. In accordance with these embodiments, proteins associated with an immunodeficiency disorder can include Trinucleotide Repeat Disorders include AR (androgen receptor), FMR1 (fragile X mental retardation 1), HTT (huntingtin), or DMPK (dystro-phia myotonica-protein kinase), FXN (frataxin), ATXN2 (ataxin 2), or other genetically-related contributor.

[00215] In some embodiments, the condition can be a Neurotransmission Disorders. In accordance with these embodiments, proteins associated with a Neurotransmission Disorders can include SST (somatostatin), NOS1 (nitric oxide synthase 1 (neuronal)), ADRA2A (adrenergic, alpha-2A-, receptor), ADRA2C (adrenergic, alpha-2C-, receptor), TACR1 (tachykinin receptor 1), or HTR2c (5-hydrox-ytryptamine (serotonin) receptor 2C), or other genetically-related contributor. In other embodiments, neurodevelopmental-associated sequences can include, but are not limited to, A2BP1 [ataxin 2-binding protein 1], AADAT [aminoadipate aminotransferase], AANAT [arylalkylamine N-acetyltransferase], ABAT [4-aminobutyrate aminotrans- ABCA1 [ATP-binding cassette, sub-family A (ABC1), member 1], or ABCA13 [ATP-binding cassette, sub-family A (ABC1), member 13], or other genetically-related contributor.

[00216] In yet other embodiments, genetic health conditions targeted for genome editing to treat a condition in a subject can include, but are not limited to Aicardi-Goutieres Syndrome; Alexander Disease; Allan-Herndon-Dudley Syndrome; POLG-Related Disorders; Alpha-Mannosidosis (Type II and III); Alstrom Syndrome; Angelman; Syndrome; Ataxia-Telangiectasia; Neuronal Ceroid-Lipofuscinoses; Beta-Thalassemia; Bilateral Optic Atrophy and (Infantile) 3 Optic Atrophy Type 1; Retinoblastoma (bilateral); Canavan Disease; Cerebrooculofacioskeletal Syndrome 1 [COFS1]; Cerebrotendinous Xanthomatosis; Cornelia de Lange Syndrome; MAPT-Related Disorders; Genetic Prion Diseases; Dravet Syndrome; Early-Onset Familial Alzheimer Disease; 4 Friedreich Ataxia [FRDA]; Fryns Syndrome; Fucosidosis; Fukuyama Congenital Muscular Dystrophy; Galactosialido-sis; Gaucher Disease; Organic Acidemias; Hemophagocytic Lymphohistiocytosis; Hutchinson-Gilford Progeria Syndrome; Mucolipidosis II; Infantile Free Sialic Acid Storage 4 Disease; PLA2G6-Associated Neurodegeneration; Jervell and Lange-Nielsen Syndrome; Junctional Epidermolysis Bullosa; Huntington Disease; Krabbe Disease (Infantile); Mitochondrial DNA-Associated Leigh Syndrome and NARP; Lesch-Nyhan Syndrome; LIST-Associated Lissen- 5 cephaly; Lowe

Syndrome; Maple Syrup Urine Disease; MECP2 Duplication Syndrome; ATP7A-Related Copper Transport Disorders; LAMA2-Related Muscular Dystrophy; Arylsulfatase A Deficiency; Mucopolysaccharidosis Types I, II or III; Peroxisome Biogenesis Disorders, Zellweger Syndrome Spectrum; Neurodegeneration with Brain Iron Accu¬mulation Disorders; Acid Sphingomyelinase Deficiency; Niemann-Pick Disease Type C; Glycine Encephalopathy; ARX-Related Disorders; Urea Cycle Disorders; COL1A1/2-Related Osteogenesis Imperfecta; Mitochondrial DNA Deletion Syndromes; PLP1-Related Disorders; Perry Syndrome; Phelan-McDermid Syndrome; Glycogen Storage Disease Type II (Pompe Disease) (Infantile); MAPT-Related Disorders; MECP2-Related Disorders; Rhizomelic Chondrodys-plasia Punctata Type 1; Roberts Syndrome; Sandhoff Disease; Schindler Disease Type 1; Adenosine Deaminase Deficiency; Smith-Lemli-Opitz Syndrome; Spinal Muscular Atrophy; Infantile-Onset Spinocerebellar Ataxia; Hex-osaminidase A Deficiency; Thanatophoric Dysplasia Type 1; Collagen Type VI-Related Disorders; Usher Syndrome Type I; Congenital Muscular Dystrophy; Wolf-Hirschhorn Syndrome; Lysosomal Acid Lipase Deficiency; and Xeroderma Pigmentosum.

[00217] In other embodiments, genetic disorders in animals targeted by editing systems disclosed herein can include, but are not limited to, Hip Dysplasia, Urinary Bladder conditions, epilepsy, cardiac disorders, Degenerative Myelopathy, Brachycephalic Syndrome, Glycogen Branching Enzyme Deficiency (GBED), Hereditary Equine Regional Dermal Asthenia (HERDA), Hyperkalemic Periodic Paralysis Disease (HYPP), Malignant Hyperthermia (MH), Polysaccharide Storage Myopathy - Type 1 (PSSM1), junctional epdiermolysis bullosa, cerebellar abiotrophy, lavender foal syndrome, fatal familial insomnia, or other animal-related genetic disorder.

[00218] In some embodiments, nuclease and/or gRNA sequences of use in compositions and methods disclosed herein can include sequences having homologous substitution (for example, substitution and replacement are both used herein to mean the interchange of an existing amino acid residue or nucleotide, with an alternative residue or nucleotide) that can occur in the case of amino acids such as basic for basic, acidic for acidic, polar for polar, etc. Non-homologous substitutions are also contemplated; for example, from one class of residue to another or alternatively involving the inclusion of non-naturally occurring amino acids such as ornithine (hereinafter referred to as Z), diamin-obutyric acid ornithine (hereinafter referred to as B), nor-

leucine ornithine (hereinafter referred to as 0), pyridylala-nine, thienylalanine, naphthylalanine and phenylglycine.

[00219] In certain embodiments disclosed herein, engineered nucleic acid guided nuclease constructs can recognize a protospacer adjacent motif (PAM) sequence other than TTTN or in addition to TTTN. In other embodiments, engineered nucleic acid guided nuclease constructs disclosed herein can be further mutated to improve targeting efficiency or can be selected from a library for particular targeted features. Other embodiments disclosed herein concern vectors including constructs disclosed herein of use for further analysis and to select for improved genome editing features.

[00220] Other embodiments include kits for packaging and transporting nucleic acid guided nuclease constructs and/or novel gRNAs disclosed herein or known gRNAs disclosed herein and further include at least one container.

[00221] As will be apparent, it is envisaged that the present system can be used to target any polynucleotide sequence of interest. Some examples of conditions or diseases that might be use fully treated using the present system are included in the figures and tables herein and examples of genes currently associated with those conditions are also provided there. However, the genes exemplified are not exhaustive. Additional objects, advantages, and novel features of this disclosure will become apparent to those skilled in the art upon review of the following examples in light of this disclosure. The following examples are not intended to be limiting.

EXAMPLES

[00222] The following examples are included to demonstrate preferred embodiments of the disclosure. It should be appreciated by those of skill in the art that the techniques disclosed in the examples that follow represent techniques discovered by the inventors to function well in the practice of the present disclosure, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the present disclosure.

Example 1

[00223] In one exemplary method, selection criteria used was set to identify sequences with < 60% AA sequence similarity to Cas12a, < 60% AA sequence similarity to MAD7 (positive

control nuclease), and > 80% query cover. After some screening rounds, nine nucleases were identified and referenced herein as ABW 1-9 for further study.

[00224] In one exemplary method, ABW (as referred to herein) nucleic acid guided nuclease constructs were compared to native amino acid sequences of Cas12a nucleases from different organisms for homology. Exemplary results are provided in **Tables 1-2** below:

Table 1: ***Percent identity between amino acid sequences of ABW nucleases and native Cas12a nucleases.***

	Percent Identity Between Amino Acid Sequences		
	AsCpf1 (WP_021736722.1)	**FnCpf1** (WP_003040289.1)	**EeCpf1** (WP_055225123.1)
ABW1 (WP_075579848.1) SEQ ID NO: 1	48.81	34.75	32.22
ABW2 (WP_077541740.1) SEQ ID NO: 14	34.14	37.25	30.23
ABW3 (WP_087408205.1) SEQ ID NO: 27	33.75	42.64	35.66
ABW4 (PKP01583.1) SEQ ID NO: 40	34.96	41.17	33.65
ABW5 (WP_121734700.1) SEQ ID NO: 53	33.02	42.80	35.05
ABW6 (PWM14151.1) SEQ ID NO: 66	32.64	33.28	52.45
ABW7 (WP_081834226.1) SEQ ID NO: 79	23.28	22.80	26.39
ABW8 (WP_118649060.1) SEQ ID NO: 92	31.65	35.39	48.69
ABW9 (WP_108604518.1) SEQ ID NO: 105	30.67	32.36	34.29

Table 2: ***Percent identity between amino acid sequences of ABW nucleases and native Cas12a nucleases.*** NCBI references are provided for each sequence.

	Percent Identity Amino Acid Between Sequences											
	AsCpf1	**FnCpf1**	**EeCpf1**	**ABW1**	**ABW2**	**ABW3**	**ABW4**	**ABW5**	**ABW6**	**ABW7**	**ABW8**	**ABW9**

AsCpf1 (WP_021736722.1)	100.0											
FnCpf1 (WP_003040289.1)	31.33	100.0										
EeCpf1 (WP_055225123.1)	28.98	32.74	100.0									
ABW1 (WP_075579848.1) SEQ ID NO: 1	48.81	31.80	28.82	100.0								
ABW2 (WP_077541740.1) SEQ ID NO: 14	30.55	32.68	26.24	30.65	100.0							
ABW3 (WP_087408205.1) SEQ ID NO: 27	30.30	41.33	31.74	30.29	33.07	100.0						
ABW4 (PKP01583.1) SEQ ID NO: 40	31.10	39.52	29.23	31.56	33.54	46.88	100.0					
ABW5 (WP_121734700.1) SEQ ID NO: 53	29.67	42.18	30.70	29.44	31.58	53.86	45.85	100.0				
ABW6 (PWM14151.1) SEQ ID NO: 66	28.18	29.60	51.21	28.12	25.15	29.03	27.48	27.71	100.0			
ABW7 (WP_081834226.1) SEQ ID NO: 79	16.23	15.69	16.63	17.32	16.81	16.10	15.69	15.57	16.40	100.0		
ABW8 (WP_118649060.1) SEQ ID NO: 92	27.69	30.66	48.10	28.31	26.13	30.51	28.46	30.69	43.48	15.59	100.0	
ABW9 (WP_108604518.1) SEQ ID NO: 105	23.71	27.11	23.60	24.42	27.71	26.83	29.72	26.78	22.90	19.59	23.59	100.0

[00225] The nucleotide sequences of the ABW nucleases were compared to Cas12a nucleotide sequences from different organism. The exemplary results are provided in **Table 3** below:

Table 3: ***Percent identity between nucleotide sequences of native ABW nucleases and native Cas12a nucleases.*** NCBI references are provided for each sequence.

	Percent Identity Between Nucleotide Sequences											
	AsCpf1	**FnCpf1**	**EeCpf1**	**ABW1**	**ABW2**	**ABW3**	**ABW4**	**ABW5**	**ABW6**	**ABW7**	**ABW8**	**ABW9**
AsCpf1 (NZ_AWUR01000016.1:24220-28143)	100.0											
FnCpf1 (NC_008601.1:c1477344-1473442)	38.63	100.0										

EeCpf1 (NZ_CYYW01000037.1:2537-6328)	40.14	42.32	100.0									
ABW1 (NZ_LT608315.1:1257961-1261869) SEQ ID NO: 2	55.81	39.71	39.41	100.0								
ABW2 (NZ_CP019633.1:c2931404-2927472) SEQ ID NO: 15	37.89	49.37	37.06	38.69	100.0							
ABW3 (NZ_NFJR01000003.1:227868-231620) SEQ ID NO: 28	37.92	42.91	38.30	36.79	40.46	100.0						
ABW4 (PHDB01000067.1:9586-13488)	38.39	46.28	39.66	38.57	41.85	53.33	100.0					
ABW5 (NZ_RAYI01000001.1:346670-350428) SEQ ID NO: 41	36.37	45.11	40.19	36.86	39.39	56.35	51.47	100.0				
ABW6 (QALK01000061.1:4314-8129) SEQ ID NO: 67	38.86	39.54	59.26	38.16	36.16	37.35	37.34	37.95	100.0			
ABW7 (NZ_KL370807.1:41505-45212) SEQ ID NO: 80	31.57	32.61	31.46	31.61	36.39	30.69	30.42	32.26	33.41	100.0		
ABW8 (NZ_QUGZ01000001.1:63017-66937) SEQ ID NO: 93	38.41	40.96	54.85	38.07	31.46	36.72	38.83	39.23	52.67	32.68	100.0	
ABW9 (NZ_CP026604.1:c5177923-5173532) SEQ ID NO: 106	33.37	39.95	36.01	33.00	36.62	39.28	44.53	40.78	34.30	29.21	34.81	100.0

[00226] In other methods, circular phylogram was prepared to assess the evolutionary relationship among the ABW1-ABW9 nucleases identified in the final round of screening. The result is illustrated in **FIG. 1**.

[00227] Following this comparison of these nucleases, the nine type V CRISPR-associated protein Cas12a (ABW) nucleases were subjected to nuclease engineering. Briefly, codon optimization was performed using the Codon Optimization Tool (IDT, https://eu.idtdna.com/CodonOpt) providing the amino acid sequence of the nuclease as an input, choosing gene as a product type, and Escherichia coli B as an organism. The IDT Codon Optimization Tool was developed to optimize a DNA or protein sequence from one organism for

expression in another by reassigning codon usage based on the frequencies of each codon's usage in the new organism. For example, valine is encoded by 4 different codons (GUG, GUU, GUC, and GUA). In human cell lines, however, the GUG codon is preferentially used (46% use vs. 18, 24, and 12%, respectively). The codon optimization tool takes this information into account and assigns valine codons with those same frequencies. In addition, the tool algorithm eliminated codons with less than 10% frequency and re-normalized the remaining frequencies to 100%. Moreover, the optimization tool reduced complexities that could interfere with manufacturing and downstream expression, such as repeats, hairpins, and extreme GC content. Exemplary engineered ABW nucleases disclosed herein are provided in **Table 4.**

Table 4: ***Sequences of exemplary engineered ABW nucleases***

	Engineered Amino Acid Sequence	**Engineered Nucleotide Sequence**
ABW1	MGHHHHHHSSGLVPRGSGTMAA FDKFIHQYQVSKTLRFALIPQG KTLENTKNNVLQEDDERQKNYE KVKPILDRIYKVFAEESLKDCS VDWNDLNACLDAYQKNPSADKR QKVKAAQDALRDEIAGYFTGKQ YANGKNKNAVKEKEQAELYKDI FSKKIFDGTVTNNKLPQVNLSA EETELLGCFDKFTTYFVGFYQN RENVFSGEDIATAIPHRIVQDN FPKFRENCRIYQDLIKNEPALK PLLQQAAAAVMAQNPKGIYQPR KSLDDIFVIPFYNHLLLQDDID YFNQILGGISGAAGQKKIQGLN ETINLFMQQHPQEADKLKKKKI RHRFIPLYKQILSDRTSFSFIP EAFSNSQEALDGIETFKKSLKK NDTFGALERLIQNLASLDLKYV YLSNKKVNEISQALYGEWHCIQ DVLKQDFSLESLIQINPQNSSN GFLATLTDEGKKRISQCRNVLG NPLPVKLADDQDKAQVKNQLDT LLAAVHYLEWFKADPDLETDPN FTVPFEKIWEELVPLLSLYSKV RNFVTKKPYSTAKFKLNFANPT LADGWDIHKESDNGALLFEKGG LYYLGIMNPKDKPNFKSYQGAE PYYQKMVYRFFPDCSKTIPKCS TQRKDVKKYFEDHPQATSYQIH DSKKEKFRQDFFEIPREIYELN NTTYGTGKSKYKKFQTQYYQKT QDKSGYQKALRKWIDFSKKFLQ TYVSTSIFDFKGLRPSKDYQDL	ATGGGCCACCATCATCATCATCATAGCAGCGGCCTGGTGCCGC GCGGCAGCGGTACCATGGCGGCGTTCGATAAGTTCATCCATCA ATATCAAGTAAGCAAAACCCTCCGTTTTGCACTTATTCCGCAG GGGAAAACCTTGGAGAATACAAAAAATAACGTACTCCAGGAAG ATGATGAGCGTCAGAAAAATTACGAAAAAGTCAAACCTATCCT TGATCGTATTTATAAGGTATTCGCTGAGGAAAGCCTGAAAGAT TGCAGCGTTGACTGGAATGACCTCAATGCATGTCTGGATGCTT ACCAAAAAAATCCTAGCGCGGATAAGCGTCAGAAGGTGAAAGC CGCGCAGGACGCGTTGCGGGACGAAATTGCCGGTTATTTTACA GGGAAACAATACGCGAACGGGAAGAACAAAAATGCCGTTAAGG AGAAAGAGCAGGCAGAATTGTATAAGGATATCTTTAGCAAAAA GATCTTTGATGGGACCGTAACGAACAACAAATTGCCACAGGTC AACCTTTCAGCCGAAGAAACAGAGTTATTAGGCTGTTTTGATA AATTCACAACATATTTCGTCGGCTTTTACCAGAACCGTGAGAA CGTATTTTCAGGGGAGGATATTGCTACAGCTATTCCGCATCGG ATCGTCCAGGATAATTTTCCTAAATTCCGGGAAAACTGTCGGA TTTATCAGGACTTAATCAAAAATGAACCTGCCCTTAAACCGCT GCTTCAGCAAGCAGCGGCCGCGGTGATGGCCCAGAATCCAAAG GGGATCTATCAACCACGTAAGAGTCTGGACGATATTTTTGTCA TTCCGTTTTATAACCATCTCCTCTTACAGGATGATATTGATTA TTTCAATCAAATCTTAGGCGGCATTTCGGGGGCAGCCGGTCAG AAAAAAATCCAGGGTTTAAATGAAACAATTAATCTGTTTATGC AACAGCACCCACAAGAAGCCGATAAGTTAAAGAAAAAAAAGAT TCGTCATCGGTTTATTCCGCTGTATAAACAAATTCTCTCTGAC CGTACGTCTTTCTCGTTCATCCCTGAAGCTTTTTCCAATTCTC AGGAAGCGTTAGACGGCATTGAGACATTCAAAAAGTCTCTTAA GAAGAATGACACATTCGGCGCGTTGGAGCGGCTGATTCAAAAT CTTGCTTCCCTGGACCTGAAATACGTGTATTTATCGAACAAGA AGGTCAATGAGATTTCGCAGGCATTATACGGCGAATGGCACTG CATCCAAGACGTCCTCAAGCAAGATTTCAGCCTTGAGAGCCTG ATCCAGATCAACCCACAAAATTCTAGCAATGGTTTCCTGGCCA CACTTACCGACGAAGGCAAGAAACGTATCTCCCAATGTCGTAA CGTACTGGGGAATCCTCTTCCAGTCAAGCTTGCGGATGATCAA

	GEFYKDVNSRCYRVTFEKIRVQ DIHEAVKNGQLYLFQLYNKDFS PKSHGLPNLHTLYWKAVFDPEN LKDPIVKLNGQAELFYRPKSNM QIIQHKTGEEIVNKKLKDGTPV PDDIYREISAYVQGKCQGNLSP EAEKWLPSVTIKKAAHDITKDR RFTEDKFFFHVPITLNYQSSGK PTAFNSQVNDFLTEHPETNIIG IDRGERNLIYAVVITPDGKILE QKSFNVIHDFDYHESLSQREKQ RVAARQAWTAIGRIKDLKEGYL SLVVHEIAQMMIKYQAVVVLEN LNTGFKRVRGGISEKAVYQQFE KMLIEKLNFLVFKDRAINQEGG VLKAYQLTDSFTSFAKLGNQSG FLFYIPSAYTSKIDPGTGFVDP FIWSHVTASEENRNEFLKGFDS LKYDAQSSAFVLHFKMKSNKQF QKNNVEGFMPEWDICFEKNEEK ISLQGSKYTAGKRIIFDSKKKQ YMECFPQNELMKALQDVGITWN TGNDIWQDVLKQASTDTGFRHR MINLIRSVLQMRSSNGATGEDY INSPVMDLDGRFFDTRAGIRDL PLDADANGAYHIALKGRMVLER IRSQKNTAIKNTDWLYAIQEER NGAPKRPAATKKAGQAKKKKAS GSGAGSPKKKRKVEDPKKKRKV **(SEQ ID NO: 3)**	GACAAAGCGCAAGTCAAAAACCAATTGGATACATTACTGGCTG CTGTACACTATCTCGAGTGGTTCAAGGCAGATCCAGACCTGGA AACAGACCCTAACTTCACTGTTCCTTTCGAAAAGATCTGGGAG GAATTGGTTCCTTTACTTTCACTGTACTCTAAAGTTCGGAATT TTGTTACAAAGAAGCCATATTCTACAGCTAAATTTAAACTGAA CTTTGCTAACCCGACATTAGCGGATGGGTGGGATATTCACAAG GAAAGTGATAACGGCGCGCTCCTGTTTGAAAAGGGTGGTTTGT ATTACTTGGGTATCATGAACCCTAAAGATAAGCCTAATTTTAA ATCCTATCAGGGTGCAGAGCCATACTATCAGAAGATGGTGTAC CGTTTTTTTCCTGACTGTTCGAAGACCATCCCAAAATGCAGCA CCCAACGTAAGGATGTAAAAAAGTACTTCGAAGACCACCCTCA AGCGACCTCATACCAGATCCACGACTCAAAGAAAGAGAAGTTT CGTCAGGATTTTTTTGAGATCCCTCGGGAGATTTACGAGCTTA ATAACACCACATACGGCACAGGTAAGTCTAAATATAAAAAATT CCAGACCCAGTATTACCAGAAGACTCAGGATAAGTCAGGCTAT CAGAAAGCACTTCGCAAATGGATTGACTTTTCCAAAAAGTTTC TTCAAACATACGTCAGTACTTCCATTTTTGATTTCAAAGGTCT CCGTCCTTCGAAGGATTATCAGGACTTAGGCGAGTTCTATAAA GACGTTAATTCGCGTTGTTACCGTGTGACGTTCGAGAAAATTC GCGTACAGGACATCCACGAAGCAGTCAAAAATGGGCAACTGTA TCTCTTCCAATTATATAATAAGGACTTCTCACCTAAAAGCCAT GGGTTGCCTAATCTTCACACTCTCTATTGGAAAGCCGTGTTCG ATCCTGAGAACTTGAAGGACCCTATCGTAAAACTTAATGGCCA AGCTGAGTTATTCTATCGGCCGAAATCCAACATGCAAATCATC CAACATAAGACCGGGGAGGAGATTGTGAACAAAAAGCTGAAGG ACGGCACCCCGGTTCCTGATGATATCTACCGCGAAATCAGTGC TTACGTCCAGGGGAAATGTCAAGGCAACTTATCCCCGGAGGCA GAGAAGTGGCTCCCAAGTGTCACAATCAAGAAAGCCGCCCATG ATATCACAAAGGATCGTCGCTTTACCGAAGATAAGTTTTTCTT TCATGTCCCTATTACACTGAACTATCAGAGTTCAGGCAAGCCG ACGGCATTCAACTCGCAAGTAAACGATTTCTTGACCGAGCACC CTGAGACAAATATCATCGGCATTGATCGGGGTGAACGTAACTT GATTTATGCCGTTGTAATCACTCCAGATGGCAAGATTCTCGAA CAGAAATCTTTTAACGTGATCCACGACTTTGATTATCATGAAT CCCTGTCCCAGCGGGAAAAACAGCGGGTAGCAGCGCGTCAGGC TTGGACAGCGATTGGTCGCATCAAGGATCTCAAGGAAGGTTAC CTGTCGCTTGTGGTGCACGAAATTGCTCAAATGATGATCAAAT ACCAAGCAGTCGTCGTATTAGAAAACCTCAACACGGGCTTTAA GCGTGTGCGCGGTGGTATCAGTGAGAAGGCCGTCTACCAACAG TTCGAAAAAATGTTGATTGAAAAATTGAACTTCCTGGTATTTA AAGATCGGGCAATCAATCAGGAAGGCGGGGTTCTCAAAGCTTA CCAGCTGACAGACTCGTTTACGTCTTTTGCAAAGTTAGGTAAC CAGTCCGGTTTCCTGTTCTACATCCCGTCCGCCTACACCAGCA AAATCGACCCTGGTACGGGCTTCGTCGATCCTTTTATCTGGTC TCACGTGACCGCTTCTGAGGAAAATCGGAATGAATTTTTAAAG GGCTTTGATAGCTTGAAATATGACGCCCAATCATCCGCCTTTG TACTGCATTTCAAGATGAAATCCAATAAGCAATTTCAGAAGAA CAATGTTGAAGGTTTCATGCCGGAATGGGATATCTGCTTCGAG AAAAACGAGGAAAAGATTTCCTTGCAGGGTAGTAAGTATACAG CCGGTAAACGCATTATTTTCGACTCCAAAAAGAAGCAATACAT GGAGTGCTTCCCGCAGAATGAGCTCATGAAAGCACTGCAGGAC GTAGGCATCACCTGGAACACGGGCAACGATATCTGGCAGGATG TCCTTAAACAAGCGAGCACAGATACAGGGTTTCGTCACCGGAT GATCAACCTGATCCGTTCAGTGCTCCAGATGCGGTCCAGTAAT GGTGCGACCGGGGAGGATTACATCAATTCACCTGTGATGGATC TGGACGGCCGTTTTTTCGACACTCGGGCGGGGATTCGTGATCT

		GCCATTGGATGCCGACGCCAACGGCGCATACCACATCGCTTTA AAAGGGCGTATGGTACTCGAACGCATTCGCTCCCAAAAGAATA CCGCGATTAAGAACACTGACTGGTTATACGCAATCCAAGAGGA ACGTAACGGCGCGCCAAAAAGGCCGGCGGCCACGAAAAAGGCC GGCCAGGCAAAAAAGAAAAAGGCTAGCGGCAGCGGCGCCGGAT CCCCAAAGAAGAAAAGGAAGGTTGAAGACCCCAAGAAAAAGAG GAAGGTGTGATAA **(SEQ ID NO: 4)**
ABW2	MGHHHHHHSSGLVPRGSGTMKE FTNQYSLTKTLRFELRPVGETA EKIEDFKSGGLKQTVEKDRERT EAYKQLKEVIDSYHRDFIEQAF ARQQTLSEEDFKQTYQLYKEAQ KEKDGETLTKQYEHLRKKIAAM FSKATKEWAVMGENNELIGKNK ESKLYQWLEKNYRAGRIEKEEF DHNAGLIEYFEKFSTYFVGFDK NRANMYSKEAKATAISFRTINE NMVKHFDNCQRLEKIKSKYPDL AEELKDFEEFFKPSYFINCMNQ SGIDYYNISAIGGKDEKDQKAN MKINLFTQKNHLKGSDKPPFFA KLYKQILSDREKSVVIDEFEKD SELTEALKNVFSKDGLINEEFF TKLKSALENFMLPEYQGQLYIR NAFLTKISANIWGSGSWGIIKD AVTQAAENNFTRKSDKEKYAKK DFYSIAELQQAIDEYIPTLENG VQNASLIEYFRKMNYKPRGSEE DAGLIEEINNNLRQAGIVLNQA ELGSGKQREENIEKIKNLLDSV LNLERFLKPLYLEKEKMRPKAA NLNKDFCESFDPLYEKLKTFFK LYNKVRNYATKKPYSKDKFKIN FDTATLLYGWSLDKETANLSVI FRKREKFYLGIINRYNSQIFNY KIAGSESEKGLERKRSLQQKVL AEEGEDYFEKMVYHLLLGASKT IPKCSTQLKEVKAHFQKSSEDY IIQSKSFAKSLTLTKEIFDLNN LRYNTETGEISSELSDTYPKKF QKGYLTQTGDVSGYKTALHKWI DFCKEFLRCYRNTEIFTFHFKD TKEYESLDEFLKEVDSSGYEIS FDKIKASYINEKVNAGELYLFE IYNKDFSEYSKGKPNLHTIYWK SLFETQNLLDKTAKLNGKAEIF FRPRSIKHNDKIIHRAGETLKN KNPLNEKPSSRFDYDITKDRRF TKDKFFLHCPITLNFKQDKPVR FNEQVNLYLKDNPDVNIIGIDR GERHLLYYTLINQNGEILQQGS LNRIGEEESRPTDYHRLLDERE KQRQQARETWKAVEGIKDLKAG YLSRVVHKLAGLMVQNNAIVVL EDLNKGFKRGRFAVEKQVYQNF EKALIQKLNYLVFKEVNSKDAP	ATGGGCCACCATCATCATCATCATAGCAGCGGCCTGGTGCCGC GCGGCAGCGGTACCATGAAGGAGTTTACCAACCAATATTCCTT AACCAAGACCCTGCGGTTCGAGTTGCGGCCAGTCGGCGAAACA GCAGAAAAGATCGAAGATTTTAAATCGGGCGGGCTCAAGCAAA CAGTGGAAAAGGATCGTGAGCGTACAGAAGCGTATAAGCAGTT GAAAGAGGTTATTGACTCCTATCATCGTGACTTCATTGAGCAA GCTTTTGCGCGCCAGCAGACGCTGTCCGAGGAGGATTTTAAAC AAACATATCAACTGTACAAAGAGGCCCAGAAAGAGAAGGATGG GGAAACATTAACAAAGCAGTACGAGCATTTACGGAAGAAAATC GCAGCTATGTTCAGCAAGGCTACGAAGGAATGGGCCGTTATGG GGGAGAATAACGAATTGATCGGGAAAAACAAAGAGTCAAAGTT GTATCAGTGGCTGGAGAAGAACTACCGCGCAGGTCGCATCGAA AAAGAGGAATTCGACCATAATGCGGGCTTAATCGAATACTTCG AGAAATTTTCCACATATTTCGTAGGTTTTGACAAAAATCGTGC GAATATGTATTCAAAGGAGGCAAAGGCGACCGCAATTTCCTTC CGGACGATTAATGAGAACATGGTCAAGCATTTCGATAATTGCC AGCGGCTCGAGAAGATTAAATCTAAATATCCTGATTTGGCCGA GGAGCTGAAGGATTTTGAGGAGTTTTTTAAACCTAGCTATTTC ATTAATTGTATGAATCAATCGGGTATCGACTACTACAATATCA GCGCGATCGGCGGTAAGGATGAAAAGGATCAGAAAGCGAATAT GAAGATCAACCTTTTCACGCAAAAAAATCATTTAAAGGGCAGT GATAAACCACCATTTTTTGCTAAGCTCTACAAGCAAATTTTGA GTGACCGGGAGAAGTCCGTGGTAATCGACGAGTTCGAAAAGGA CAGCGAATTGACAGAGGCACTCAAAAACGTGTTTTCCAAGGAC GGTTTGATCAATGAGGAGTTTTTTACAAAGTTAAAAAGTGCAT TAGAAAATTTTATGTTGCCTGAATATCAAGGTCAACTCTACAT CCGTAACGCTTTCCTTACGAAGATCAGCGCAAACATTTGGGGC TCTGGTTCTTGGGGCATCATCAAGGACGCAGTTACCCAGGCTG CGGAAAACAATTTCACGCGTAAGTCTGACAAGGAAAAGTATGC CAAGAAAGACTTCTATTCCATTGCTGAACTCCAGCAGGCTATT GATGAATACATTCCTACTCTGGAGAACGGGGTTCAAAACGCAT CACTCATCGAGTACTTTCGCAAAATGAATTACAAACCACGCGG TTCTGAAGAAGACGCAGGCTTGATCGAAGAAATTAATAACAAC CTGCGTCAGGCTGGGATCGTCCTGAATCAAGCCGAGCTGGGGT CTGGTAAGCAGCGGGAAGAGAATATTGAAAAAATTAAGAACTT ATTAGATTCGGTTTTGAATCTCGAACGTTTCTTAAAGCCACTT TACTTGGAGAAAGAGAAAATGCGTCCAAAAGCTGCTAACCTGA ATAAGGATTTTTGTGAGTCATTTGATCCACTTTACGAGAAACT GAAAACGTTTTTCAAGCTCTACAATAAAGTACGTAACTACGCA ACAAAGAAACCATACTCAAAGGACAAATTTAAGATCAATTTTG ATACCGCTACGTTATTATATGGGTGGAGTTTGGATAAGGAAAC CGCGAATCTCAGCGTCATTTTCCGTAAACGCGAAAAATTCTAT TTGGGTATCATCAACCGGTACAATAGCCAGATTTTCAATTATA AGATTGCGGGCAGTGAGAGCGAGAAAGGGTTAGAGCGTAAGCG GTCGCTGCAGCAAAAGGTGCTTGCAGAGGAGGGTGAAGATTAT TTTGAGAAAATGGTATACCACCTGCTGCTTGGCGCGTCGAAAA CTATTCCGAAATGCTCGACACAGTTGAAAGAAGTAAAAGCACA CTTTCAAAAGTCATCAGAAGATTATATTATCCAATCCAAATCA TTTGCAAAGTCATTAACATTAACAAAAGAGATCTTTGACTTAA

	GHYLKAYQLTAPFISFEKLGTQ SGFLFYVRAWNTSKIDPATGFT DQIKPKYKNQKQAKDFMSSFDS VRYNRKENYFEFEADFEKLAQK PKGRTRWTICSYGQERYSYSPK ERKFVKHNVTQNLAELFNSEGI SFDSGQCFKDEILKVEDASFFK SIIFNLRLLLKLRHTCKNAEIE RDFIISPVKGNNSSFFDSRIAE QENITSIPQNADANGAYNIALK GLMNLHNISKDGKAKLIKDEDW IEFVQKRKFAAAKRPAATKKAG QAKKKKASGSGAGSPKKKRKVE DPKKKRKV **(SEQ ID NO: 16)**	ATAATCTGCGGTATAACACAGAAACGGGCGAAATTAGTTCCGA GCTTTCTGATACATATCCGAAGAAGTTCCAGAAGGGGTATCTC ACACAAACAGGCGACGTTTCGGGTTACAAAACTGCTCTGCATA AGTGGATTGATTTCTGCAAAGAGTTCTTGCGTTGCTATCGTAA TACGGAGATCTTCACGTTCCATTTCAAGGACACGAAGGAGTAC GAGTCGTTAGATGAGTTCTTGAAAGAAGTGGATAGTTCAGGTT ATGAGATTTCATTCGATAAGATCAAAGCCTCTTATATCAACGA GAAGGTTAATGCAGGCGAGCTGTACTTGTTCGAGATCTATAAT AAAGATTTCTCCGAGTATTCCAAAGGTAAGCCAAATCTGCATA CCATTTATTGGAAAAGTCTCTTCGAGACTCAAAACTTGCTGGA TAAAACAGCGAAACTCAACGGCAAGGCAGAGATCTTCTTCCGG CCACGTTCGATCAAACACAACGACAAAATCATCCACCGTGCGG GCGAAACACTTAAGAATAAAAACCCGCTCAATGAAAAGCCTAG TTCGCGTTTCGATTACGATATTACGAAAGATCGTCGTTTTACG AAAGACAAATTTTTTTTACACTGCCCTATTACGTTAAACTTTA AGCAGGACAAGCCTGTTCGCTTTAATGAACAAGTCAACTTATA CTTAAAAGACAATCCAGACGTGAATATTATCGGTATCGATCGT GGTGAGCGTCACTTGCTTTATTACACTTTGATCAATCAGAATG GTGAGATCTTACAGCAGGGTTCACTTAATCGCATTGGTGAGGA AGAATCTCGGCCTACGGACTACCATCGGTTACTCGATGAGCGT GAAAAGCAGCGTCAACAAGCACGGGAGACGTGGAAAGCAGTAG AAGGGATTAAGGACTTAAAAGCTGGGTATCTTTCACGGGTTGT ACATAAACTTGCAGGTTTAATGGTACAAAACAACGCAATTGTC GTTCTGGAAGATCTTAACAAGGGTTTTAAGCGCGGTCGTTTCG CTGTTGAGAAACAGGTGTACCAGAACTTCGAAAAAGCACTTAT TCAAAAGCTTAACTATTTAGTGTTCAAGGAGGTCAACTCTAAA GACGCCCCTGGCCACTATTTGAAGGCATATCAGCTTACGGCCC CTTTCATCTCGTTCGAAAAATTGGGTACTCAGAGCGGTTTCCT TTTTTATGTGCGCGCATGGAATACCTCGAAGATCGACCCGGCG ACGGGTTTTACCGACCAAATCAAACCAAAGTATAAAAACCAAA AACAAGCTAAAGACTTCATGTCAAGCTTCGACTCTGTCCGGTA CAACCGCAAGGAAAATTATTTTGAATTCGAGGCGGACTTTGAA AAACTGGCACAGAAACCTAAGGGGCGCACCCGCTGGACGATTT GTTCCTATGGCCAGGAACGGTACTCTTACTCCCCAAAAGAACG GAAGTTTGTAAAGCACAACGTTACACAAAATCTTGCTGAGCTT TTTAATTCAGAGGGTATCTCGTTCGACTCCGGGCAGTGTTTCA AGGATGAGATCCTGAAGGTCGAGGATGCCAGTTTCTTTAAGTC TATTATTTTCAATCTTCGCCTCCTTCTCAAGCTTCGTCACACT TGCAAGAACGCCGAGATCGAACGTGATTTCATCATTTCTCCTG TCAAGGGGAACAATTCGTCCTTTTTTGACTCCCGTATTGCCGA ACAAGAAAATATCACCAGCATTCCACAGAATGCTGATGCAAAC GGTGCATACAACATCGCGCTGAAGGGCCTGATGAACCTCCATA ATATCTCTAAGGACGGCAAGGCAAAATTAATTAAGGATGAAGA TTGGATCGAATTTGTCCAAAAACGCAAGTTCGCGGCCGCAAAA AGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAA AGGCTAGCGGCAGCGGCGCCGGATCCCCAAAGAAGAAAAGGAA GGTTGAAGACCCCAAGAAAAAGAGGAAGGTGTGATAA **(SEQ ID NO: 17)**
ABW3	MGHHHHHHSSGLVPRGSLQMKT LSDFTNLFPLSKTLRFKLIPIG NTLKNIEASGILDEDRHRAESY VKVKAIIDEYHKAFIDRVLSDT CLQTESIGKHNSLEEFFFYYQI GAKSEQQKKTFKKIQDALRKQI ADSLTKDKHFSRIDKKELIQED LIQFVRDGEDAAEKTSLISEFQ	ATGGGCCACCATCATCATCATCATAGCAGCGGCCTGGTGCCGC GCGGCAGCCTGCAGATGAAGACCTTGTCTGATTTTACCAATCT GTTCCCTTTATCTAAGACTCTCCGTTTCAAGCTGATTCCAATC GGCAACACGCTCAAGAACATTGAAGCTAGTGGCATCCTTGACG AGGATCGCCACCGCGCGGAGTCCTATGTCAAGGTCAAGGCCAT CATCGACGAATATCATAAAGCTTTCATCGATCGGGTCCTGTCG GATACTTGCCTCCAGACGGAATCTATCGGCAAACACAACAGTC TCGAGGAATTCTTTTTCTACTACCAAATTGGTGCAAAAAGTGA

	NFTVYFTGFHENRQNMYSPDEK	ACAGCAGAAAAAGACGTTTAAAAAGATTCAAGACGCCTTGCGC
	STAIAYRLINENLPKFVDNMKV	AAACAAATCGCAGATAGCCTCACCAAGGACAAACATTTTTCAC
	FDRIAASELASCFDELYHNFEE	GGATTGATAAAAAAGAATTGATCCAAGAGGATTTGATCCAGTT
	YLQVERLHDIFSLDYFNLLLTQ	TGTGCGCGATGGGGAGGATGCCGCTGAAAAGACGTCTCTGATT
	KHIDVYNALIGGKATETGEKIK	TCCGAATTTCAAAATTTCACAGTTTATTTTACCGGGTTTCATG
	GLNEYINLYNQRHKQEKLPKFK	AGAATCGCCAGAACATGTACAGTCCGGACGAGAAGTCCACGGC
	MLFKQILTDREAISWLPRQFDD	CATCGCATATCGCTTAATTAACGAGAATCTCCCAAAATTCGTA
	NSQLLSAIEQCYNHLSTYTLKD	GACAACATGAAAGTTTTTGACCGTATCGCGGCGTCCGAATTGG
	GSLKYLLENLHTYDTEKIFIRN	CATCGTGTTTCGACGAATTATACCACAACTTCGAGGAATACCT
	DSLLTEISQRHYGSWSILPEAI	CCAAGTGGAGCGGTTACATGATATCTTTAGTTTGGACTATTTC
	KRHLERANPQKRRETYEAYQSR	AATCTGCTTCTCACGCAGAAACATATCGACGTCTATAATGCTC
	IEKAFKAYPGFSIAFLNGCLTE	TGATCGGTGGGAAGGCAACCGAAACCGGGGAAAAGATCAAGGG
	TGKESPSIESYFESLGAVETET	CTTAAATGAATACATCAATCTCTACAATCAACGTCACAAGCAG
	SQQENWFARIANAYTDFREMQN	GAAAAACTGCCAAAATTCAAGATGTTATTCAAGCAAATTCTTA
	RLHATDVPLAQDAEAVARIKKL	CCGACCGTGAGGCAATCAGCTGGTTGCCACGCCAATTTGACGA
	LDALKGLQLFIKPLLDTGEEAE	TAATAGTCAGTTACTCTCAGCCATTGAACAGTGTTATAACCAC
	KDERFYGDFTEFWNELDTITPL	CTTTCGACCTACACACTCAAGGATGGGTCACTCAAATACCTGT
	YNMVRNYLTRKPYSEEKIKLNF	TAGAAAACCTGCATACATACGATACTGAAAAGATCTTCATCCG
	QNPTLLNGWDLNKEVDNTSVIL	CAATGACAGTTTACTTACGGAAATCTCCCAACGGCATTACGGT
	RRNGRYYLAIMHRNHRRVFSQY	TCGTGGTCGATTTTACCAGAAGCTATCAAACGTCATCTCGAGC
	PGTERGDCYEKMEYKLLPGANK	GCGCGAACCCGCAAAAACGGCGCGAAACATACGAGGCCTATCA
	MLPKVFFSKSRIDEFNPSEELL	ATCTCGCATTGAGAAGGCCTTTAAGGCATATCCGGGGTTTTCA
	ARYQQGTHKKGENFNLHDCHAL	ATTGCTTTCCTCAATGGGTGTTTAACAGAGACAGGTAAGGAGT
	IDFFKDSIEKHEEWRNFHFKFS	CGCCATCCATCGAAAGCTATTTTGAAAGTCTGGGTGCTGTCGA
	DTSSYTDMSGFYREIETQGYKL	AACAGAGACCTCTCAGCAGGAAAACTGGTTTGCCCGCATCGCA
	SFVPVACEYIDELVRDGKIFLF	AACGCTTATACGGACTTTCGTGAAATGCAAAATCGGCTGCACG
	QIYNKDFSTYSKGKPNMHTLYW	CCACTGACGTGCCGTTGGCTCAAGACGCTGAGGCAGTGGCCCG
	EMLFDERNLMNVVYKLNGQAEI	GATCAAGAAGCTGTTAGATGCACTGAAAGGCCTGCAATTATTC
	FFRKASLSARHPEHPAGLPIKK	ATTAAGCCTCTTTTGGATACTGGCGAAGAAGCAGAGAAAGATG
	KQAPTEESCFPYDLIKNKRYTV	AACGGTTCTATGGGGACTTTACCGAATTCTGGAACGAGTTAGA
	DQFQFHVPITINFKATGTSNIN	CACTATCACGCCATTGTACAATATGGTACGGAACTATCTCACG
	PSVTDYIRTADDLHIIGIDRGE	CGTAAGCCTTATAGTGAAGAAAAAATCAAGCTCAATTTCCAGA
	RHLLYLVVIDSQGRICEQFSLN	ATCCGACATTACTGAACGGTTGGGATTTGAACAAAGAGGTAGA
	EIVTQYQGHQYRTDYHALLQKK	TAATACATCTGTCATCCTCCGCCGGAATGGTCGTTATTATCTT
	EDERQKARQSWQSIENIKELKE	GCCATCATGCACCGCAACCACCGGCGTGTATTTTCACAGTATC
	GYLSQVVHKVSELMIKYKAIVV	CAGGCACAGAACGTGGCGATTGTTATGAGAAAATGGAATATAA
	LEDLNAGFKRSRQKVEKQVYQK	ACTGCTTCCGGGCGCCAACAAGATGCTCCCAAAAGTCTTCTTC
	FEKMLIDKLNYLVFKTAEADQP	TCTAAATCACGCATCGATGAATTCAACCCTAGCGAAGAATTAT
	GGLLHAYQLTNKFESFKKMGKQ	TAGCACGTTACCAGCAAGGTACCCACAAGAAGGGTGAGAATTT
	SGFLFYIPAWNTSKIDPTTGFV	TAATTTACACGACTGCCATGCCTTGATTGATTTTTTTAAAGAC
	NLFDTRYENVDKSRAFFGKFDS	TCTATTGAGAAACATGAAGAATGGCGTAACTTTCATTTTAAAT
	IRYRADKGTFEWTFDYNNFHKK	TTAGTGATACGTCCAGTTACACCGACATGAGCGGCTTTTATCG
	AEGTRSSWCLSSHGNRVRTFRN	TGAAATCGAAACACAGGGTTACAAGTTGTCATTTGTGCCAGTG
	PAKNNQWDNEEIDLTQAFRDLF	GCGTGTGAATACATCGATGAGTTGGTACGTGATGGCAAAATCT
	EAWGIEITSNLKEAICNQSEKK	TTTTGTTCCAGATCTATAATAAGGACTTTTCGACCTACTCTAA
	FFSELFELFKLMIQLRNSVTGT	GGGCAAGCCAAATATGCACACTCTTTATTGGGAAATGCTTTTC
	NIDYMVSPVENHYGTFFDSRTC	GACGAGCGGAACCTGATGAACGTGGTGTATAAACTCAATGGCC
	DSSLPANADANGAYNIARKGLM	AAGCAGAGATCTTTTTTCGTAAAGCATCACTGAGCGCACGTCA
	LARRIQATPENDPISLTLSNKE	CCCTGAGCACCCGGCAGGGTTGCCAATTAAAAAAAAACAGGCC
	WLRFAQGLDETTTYEAAAKRPA	CCGACGGAAGAATCTTGTTTCCCATATGATCTCATTAAGAATA
	ATKKAGQAKKKKASGSGAGSPK	AGCGGTATACAGTTGACCAGTTTCAGTTTCACGTGCCAATTAC
	KKRKVEDPKKKRKV **(SEQ ID NO: 29)**	TATTAATTTTAAAGCAACTGGGACTTCAAATATCAACCCGTCG
		GTCACTGATTATATTCGTACGGCCGATGACCTCCATATCATTG
		GCATTGATCGCGGTGAGCGCCATTTACTTTATTTAGTGGTGAT
		TGACTCACAAGGGCGCATCTGTGAACAGTTTTCCTTAAACGAG
		ATCGTAACGCAATACCAAGGTCACCAGTACCGTACAGATTATC

		ATGCTCTCTTGCAGAAAAAAGAGGATGAACGGCAAAAAGCTCG CCAGTCTTGGCAATCGATCGAAAACATCAAGGAATTAAAAGAG GGGTATCTGAGCCAAGTAGTGCACAAGGTTTCTGAACTGATGA TCAAATATAAAGCAATTGTGGTGTTGGAAGATTTAAATGCTGG GTTCAAGCGGAGTCGGCAGAAGGTTGAAAAGCAAGTGTATCAA AAATTTGAGAAGATGCTGATCGACAAACTTAACTATCTTGTGT TCAAGACCGCAGAAGCTGACCAACCTGGCGGCCTCCTGCACGC ATACCAATTAACAAATAAATTTGAGTCATTCAAGAAAATGGGG AAGCAAAGTGGCTTCCTCTTCTACATTCCTGCATGGAACACGT CTAAAATCGACCCGACCACGGGCTTTGTCAACCTTTTTGATAC CCGGTATGAGAACGTAGACAAATCCCGTGCCTTCTTCGGCAAA TTCGATAGCATCCGCTACCGTGCGGACAAGGGCACGTTCGAGT GGACGTTCGATTATAATAACTTTCACAAAAAGGCCGAAGGTAC GCGGTCGAGCTGGTGTTTGTCTTCTCATGGTAACCGGGTCCGT ACTTTCCGCAATCCTGCGAAAAACAACCAATGGGACAACGAAG AGATCGACTTAACACAAGCGTTCCGCGATCTGTTTGAAGCTTG GGGGATCGAGATCACTTCGAACTTAAAAGAGGCCATTTGCAAC CAGTCTGAGAAGAAATTCTTTTCTGAGCTTTTCGAACTGTTCA AACTTATGATCCAGCTGCGGAACTCAGTGACAGGCACGAATAT CGACTATATGGTGAGCCCAGTCGAGAATCACTACGGCACGTTC TTCGATTCGCGCACATGCGATTCGTCTCTGCCGGCTAACGCTG ACGCTAATGGTGCTTATAATATTGCCCGTAAGGGGTTAATGCT GGCTCGCCGCATTCAGGCTACCCCTGAGAATGATCCGATCTCC TTAACATTGAGCAACAAAGAGTGGTTACGCTTTGCACAGGGGC TCGATGAGACAACAACCTACGAGGCGGCCGCAAAAAGGCCGGC GGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGCTAGC GGCAGCGGCGCCGGATCCCCAAAGAAGAAAAGGAAGGTTGAAG ACCCCAAGAAAAAGAGGAAGGTGTGATAA **(SEQ ID NO: 30)**
ABW4	MGHHHHHHSSGLVPRGSGTMKN MESFINLYPVSKTLRFELKPIG KTLETFSRWIEELKEKEAIELK ETGNLLAQDEHRAESYKKVKKI LDEYHKWFITESLQNTKLNGLD VFYHNYMLPKKEDHEKKAFASC QDNLRKQIVNAFRQETGLFNKL SGKELFKDSKEEVALLKAIVPY FDNKTLENIGVKSNEGALLLIE EFKDFTTYFGGFHENRKNMYSD EAKSTAVAFRLIHENLPRFIDN KKVFEEKIMNSELKDKFPEILK ELEQILQVNEIEEMFQLDYFND TLIQNGIDVYNHLIGGYAEEGK KKIQGLNEHINLYNQIQKEKNK RIPRLKPLYKQILSDRETASFV IEAFENDGELLESLEKSYRLLQ QEVFTPEGKEGLANLLAAIAES ETHKIFLKNDLGLTEISQQIYE SWSLIEEAWNKQYDNKQKKVTE TETYVDNRKKAFKSIKSFSIAE VEEWVKALGNEKHKGKSVATYF KSLGKTDEKVSLIEQVENNYNI IKDLLNTPYPPSKDLAQQKDDV EKIKNYLDSLKALQRFIKPLLG SGEESDKDAHFYGEFTAFWDVL DKVTPLYNKVRNYMTKKPYSTE	ATGGGCCACCATCATCATCATCATAGCAGCGGCCTGGTGCCGC GCGGCAGCGGTACCATGAAGAACATGGAGTCTTTTATTAATTT ATATCCGGTTTCGAAAACTTTACGTTTTGAGTTAAAGCCTATT GGCAAAACACTCGAAACTTTCTCCCGCTGGATCGAAGAGTTGA AAGAGAAAGAGGCTATTGAGCTGAAAGAAACTGGCAACCTGTT GGCGCAGGATGAGCATCGGGCCGAGTCTTATAAGAAGGTCAAA AAAATTCTTGACGAATATCATAAATGGTTCATCACTGAAAGCC TCCAGAACACAAAGTTAAATGGGTTGGACGTTTTTTATCATAA CTATATGCTCCCGAAGAAAGAGGACCATGAGAAGAAAGCTTTT GCTTCGTGTCAAGATAATCTCCGTAAGCAAATTGTAAACGCGT TTCGTCAAGAAACCGGTTTATTTAACAAACTGTCAGGCAAAGA ACTGTTTAAAGATTCGAAGGAAGAGGTTGCACTGTTGAAAGCC ATTGTACCGTATTTCGATAACAAGACTCTGGAAAACATTGGTG TTAAGAGTAATGAAGGGGCTCTCCTTTTAATTGAAGAGTTCAA GGATTTTACCACGTATTTCGGTGGCTTCCATGAGAATCGCAAA AATATGTATAGCGACGAAGCAAAATCAACAGCGGTTGCCTTTC GTCTTATTCACGAAAATTTGCCGCGCTTCATTGACAATAAGAA GGTCTTCGAAGAGAAAATCATGAATAGTGAATTAAAGGATAAA TTTCCAGAGATTTTGAAGGAGCTGGAACAGATTCTGCAAGTCA ACGAGATTGAAGAGATGTTTCAGCTCGACTATTTTAACGACAC ATTGATCCAGAATGGCATCGATGTCTATAACCATTTGATCGGC GGCTACGCCGAGGAAGGCAAGAAAAAAATTCAAGGGCTTAACG AGCATATTAACCTCTATAACCAGATCCAGAAGGAGAAGAATAA GCGTATCCCGCGGCTGAAACCACTCTATAAGCAAATTTTGAGT GATCGCGAAACCGCCTCATTTGTTATCGAGGCGTTTGAGAACG ATGGCGAGTTATTAGAATCATTGGAGAAGTCATATCGCTTACT GCAGCAGGAGGTCTTTACGCCTGAAGGTAAAGAAGGTCTGGCG

	KFKLNFENSYFLNGWAQDYETK AGLIFLKDGNYFLAINNKKLDE KEKKQLKTNYEKNPAKRIILDF QKPDNKNIPRLFIRSKGDNFAP AVEKYNLPISDVIDIYDEGKFK TEYRKINEPEYLKSLHKLIDYF KLGFSKHESYKHYSFSWKKTHE YENIAQFYHDVEVSCYQVLDEN INWDSLMEYVEQNKLYLFQIYN KDFSPNSKGTPNMHTLYWKMLF NPDNLKDVVYKLNGQAEVFYRK ASIKKENKIVHKANDPIDNKNE LNKKKQNTFEYDIVKDKRYTVD KFQFHVPITLNFKAEGLNNLNS KVNEYIKECDDLHIIGIDRGER HLLYLSLIDMKGNIVKQFSLNE IVNEHKGNTYRTNYHNLLDKRE KEREKERESWKTIETIKELKEG YISQVVHKITQLMIEYNAIVVL EDLNFGFKRGRFKVEKQVYQKF EKMLIDKLNYLVDKKKEANESG GTLKAYQLTDSYADFMKYKKKQ CGFLFYVPAWNTSKIDPTTGFV NLFDTHYVNVSKAQEFFSKFKS IRYNAANNYFEFEVTDYFSFSG KAEGTKQNWIICTHGTRIINFR NPEKNSQWDNKEVVITDEFKKL FEKHGIDYKNSSDLKGQIASQS EKAFFHNEKKDTKDPDGLLQLF KLALQMRNSFIKSEEDYLVSPV MNDEGEFFDSRKAQPNQPENAD ANGAYNIAMKGKWVVKQIRESE DLDKLKLAISNKEWLNFAQRSA AAKRPAATKKAGQAKKKKASGS GAGSPKKKRKVEDPKKKRKV (**SEQ ID NO: 42**)	AATTTACTCGCAGCAATCGCTGAAAGCGAGACACACAAGATCT TTCTGAAGAACGACTTGGGTCTCACCGAGATCTCTCAACAAAT TTATGAATCATGGTCGCTGATTGAAGAGGCATGGAATAAACAA TATGACAACAAACAGAAGAAAGTTACGGAGACAGAGACATATG TGGACAATCGGAAAAAGGCTTTCAAGTCCATCAAGAGCTTTAG CATCGCAGAGGTTGAGGAATGGGTGAAAGCACTTGGGAATGAG AAACACAAGGGCAAAAGCGTGGCAACCTATTTTAAAAGTCTCG GGAAGACTGACGAAAAAGTTAGCCTTATTGAACAGGTAGAGAA CAATTATAATATCATCAAGGACCTTTTGAACACACCGTATCCT CCTTCGAAGGACTTGGCCCAGCAAAAAGATGACGTTGAAAAAA TCAAAAATTATTTGGACTCTCTGAAGGCCCTCCAGCGGTTCAT TAAGCCATTGTTGGGTAGCGGGGAGGAATCCGATAAAGATGCG CACTTTTATGGTGAGTTTACCGCTTTCTGGGATGTGCTCGACA AAGTAACCCCACTCTACAATAAAGTCCGCAACTATATGACTAA GAAACCTTATAGCACAGAGAAATTTAAGCTGAATTTTGAAAAT AGTTACTTTTTGAATGGTTGGGCACAGGACTACGAGACAAAAG CGGGGCTTATCTTCTTGAAGGACGGCAATTACTTCCTTGCCAT CAATAATAAGAAATTAGATGAAAAGGAGAAAAAACAGCTCAAG ACTAATTATGAGAAGAATCCTGCGAAGCGTATCATCTTAGACT TTCAGAAGCCAGACAATAAGAACATTCCTCGCTTGTTCATTCG CAGTAAAGGCGACAATTTCGCTCCTGCAGTAGAAAAGTATAAT CTTCCGATCTCTGACGTTATTGACATCTATGACGAGGGGAAGT TTAAGACTGAGTATCGCAAAATTAACGAGCCGGAATATCTCAA ATCTCTCCATAAGCTGATTGACTACTTCAAACTTGGGTTCTCC AAGCATGAATCCTACAAGCATTATTCTTTTTCATGGAAGAAAA CACATGAGTATGAGAACATCGCCCAGTTTTACCACGACGTGGA GGTCTCTTGCTATCAGGTGCTCGACGAAAATATTAACTGGGAT TCCCTCATGGAGTATGTAGAACAGAACAAATTGTACTTGTTCC AGATTTATAACAAAGACTTCTCCCCAAACTCGAAAGGCACTCC GAATATGCACACTTTGTACTGGAAGATGTTGTTTAATCCGGAT AATCTTAAGGACGTGGTCTATAAGCTGAACGGTCAGGCTGAAG TATTCTACCGGAAGGCGAGTATTAAGAAAGAAAACAAGATTGT CCACAAGGCGAACGACCCTATTGACAATAAAAACGAGTTGAAT AAGAAAAAGCAAAATACATTTGAATACGACATCGTCAAAGATA AACGGTATACAGTGGATAAGTTTCAATTCCATGTTCCTATCAC GCTCAACTTTAAAGCTGAAGGCCTGAATAACTTGAATAGCAAA GTTAACGAATACATCAAAGAGTGTGACGACCTTCACATTATTG GCATCGACCGGGGTGAACGGCACCTCTTGTATCTGAGCCTCAT CGATATGAAAGGTAACATTGTAAAGCAATTTAGTCTTAACGAG ATCGTTAATGAGCACAAGGGGAACACGTACCGCACGAACTATC ATAACCTCTTGGACAAACGTGAAAAGGAACGTGAAAAAGAGCG CGAGTCATGGAAAACCATTGAGACCATCAAAGAGCTGAAAGAA GGCTATATTAGTCAAGTAGTACATAAAATCACTCAGTTAATGA TCGAATATAATGCGATCGTTGTACTCGAAGACCTGAATTTCGG CTTCAAACGCGGCCGGTTCAAGGTGGAGAAGCAAGTGTATCAA AAATTTGAGAAGATGTTAATTGATAAACTGAACTACTTGGTCG ATAAGAAGAAGGAAGCCAATGAGAGTGGCGGGACACTCAAAGC CTACCAGCTTACCGATAGTTACGCTGACTTCATGAAGTACAAG AAAAAGCAATGCGGCTTCCTGTTTTATGTCCCGGCCTGGAACA CTTCCAAAATCGATCCTACTACTGGGTTCGTGAATCTGTTTGA CACACATTATGTCAATGTTAGTAAGGCCCAGGAATTTTTCTCG AAATTCAAGTCAATTCGCTACAACGCGGCCAACAACTATTTCG AGTTTGAAGTAACAGATTATTTTTCCTTCAGTGGTAAAGCTGA GGGCACCAAGCAGAATTGGATCATTTGCACCCATGGCACCCGC ATTATCAATTTTCGTAACCCGGAAAAAAATTCGCAGTGGGATA ATAAGGAAGTAGTGATCACAGATGAATTCAAGAAACTGTTTGA

		GAAGCACGGCATTGACTACAAAAATAGTTCCGACCTCAAGGGG CAGATCGCCTCTCAATCGGAGAAGGCGTTTTTTCATAACGAAA AAAAAGATACAAAGGACCCAGATGGCCTTCTGCAGCTTTTTAA ACTGGCGCTGCAGATGCGGAACTCTTTCATTAAGAGCGAAGAG GACTACTTAGTATCTCCTGTGATGAACGACGAAGGTGAATTCT TTGACTCGCGCAAAGCCCAGCCTAATCAGCCAGAGAACGCTGA TGCTAATGGGGCGTACAATATTGCAATGAAAGGGAAATGGGTT GTTAAGCAAATCCGCGAATCGGAGGACCTCGACAAGCTGAAAC TGGCAATCTCAAATAAAGAATGGTTGAACTTCGCCCAGCGCTC CGCGGCCGCAAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAG GCAAAAAAGAAAAAGGCTAGCGGCAGCGGCGCCGGATCCCCAA AGAAGAAAAGGAAGGTTGAAGACCCCAAGAAAAAGAGGAAGGT GTGATAA **(SEQ ID NO: 43)**
ABW5	MGHHHHHHSSGLVPRGSGTMKN ILEQFVGLYPLSKTLRFELKPL GKTLEHIEKKGLIAQDEQRAEE YKLVKDIIDRYHKAFIHMCLKH FKLKMYSEQGYDSLEEYRKLAS ISKRNEKEEQQFDKVKENLRKQ IVDAFKNGGSYDDLFKKELIQK HLPRFIEGEEEKRIVDNFNKFT TYFTGFHENRKNMYSDEKESTA IAYRLIHENLPLFLDNMKSFAK IAESEVAARFTEIETAYRTYLN VEHISELFTLDYFSTVLTQEQI EVYNNIIGGRVDDDNVKIQGLN EYVNLYNQQQKDRSKRLPLLKS LYKMILSDRIAISWLPEEFKSD KEMIEAINNMHDDLKDILAGDN EDSLKSLLQHIGQYDLSKIYIA NNPGLTDISQQMFGCYDVFTNG IKQELRNSITPSKKEKADNEIY EERINKMFKSEKSFSIAYLNSL PHPKTDAPQKNVEDYFALLGTC NQNDEQPINLFAQIEMARLVAS DILAGRHVNLNQSENDIKLIKD LLDAYKALQHFVKPLLGSGDEA EKDNEFDARLRAAWNALDIVTP LYNKVRNWLTRKPYSTEKIKLN FENAQLLGGWDQNKEPDCTSVL LRKDGMYYLAIMDKKANHAFDC DCLPSDGACFEKIDYKLLPGAN KMLPKVFFSKSRIKEFSPSESI IAAYKKGTHKKGPNFSLSDCHR LIDFFKASIDKHEDWSKFRFRF SDTKTYEDISGFYREVEQQGYM LGFRKVSEAFVNKLVDEGKLYL FHIWNKDFSKHSKGTPNLHTIY WKMLFDEKNLTDVIYKLNGQAE VFYRKKSLDLNKTTTHKAHAPI TNKNTQNAKKGSVFDYDIIKNR RYTVDKFQFHVPITLNFKATGR NYINEHTQEAIRNNGIEHIIGI DRGERHLLYLSLIDLKGNIVKQ MTLNDIVNEYNGRTYATNYKDL LATREGERTDARRNWQKIENIK	ATGGGCCACCATCATCATCATCATAGCAGCGGCCTGGTGCCGC GCGGCAGCGGTACCATGAAGAACATCTTAGAGCAGTTTGTCGG CTTATATCCGTTGTCTAAAACACTTCGGTTTGAGCTTAAACCT TTGGGTAAGACGTTGGAACATATTGAGAAAAAAGGCTTGATTG CCCAAGACGAACAGCGGGCGGAGGAGTACAAATTGGTTAAAGA TATTATTGATCGCTACCACAAGGCTTTTATTCATATGTGCTTA AAACATTTTAAGCTCAAGATGTACAGTGAACAAGGGTATGATA GCTTGGAGGAGTACCGCAAGCTTGCGTCAATTTCCAAACGCAA CGAGAAAGAGGAGCAGCAATTTGACAAAGTCAAGGAAAATCTT CGTAAGCAAATTGTCGACGCGTTTAAAAATGGCGGGAGTTATG ATGATCTGTTTAAGAAAGAATTGATCCAGAAACACCTCCCACG TTTTATTGAGGGTGAAGAAGAAAAACGTATCGTTGACAACTTC AACAAGTTCACGACCTATTTTACTGGTTTTCATGAAAATCGCA AGAATATGTATAGTGACGAAAAGGAATCGACGGCTATTGCTTA TCGTCTCATTCACGAAAACTTGCCATTGTTTTTGGATAACATG AAGAGCTTCGCTAAGATCGCCGAATCGGAAGTGGCTGCTCGTT TTACCGAAATCGAAACCGCTTACCGGACATACTTGAACGTAGA ACACATTAGTGAACTGTTCACCCTCGACTATTTTAGCACGGTT TTGACGCAAGAACAAATCGAAGTATATAATAACATTATCGGCG GGCGCGTCGACGACGACAACGTAAAGATCCAAGGGTTGAATGA GTACGTAAATTTATATAATCAGCAGCAGAAGGACCGGTCTAAG CGCTTACCGCTTCTTAAGTCCCTCTACAAAATGATCTTATCCG ATCGTATTGCAATTTCGTGGTTACCTGAGGAGTTCAAATCCGA TAAGGAGATGATTGAAGCAATTAACAACATGCATGACGACCTG AAGGACATTCTGGCAGGCGACAACGAAGACTCGCTTAAGTCCT TACTGCAGCATATTGGCCAATACGATCTCTCGAAAATCTACAT TGCGAACAATCCGGGCCTGACAGATATCTCACAACAAATGTTC GGGTGTTATGACGTCTTTACTAATGGGATCAAGCAGGAGCTCC GGAACAGTATTACCCCTTCAAAAAAGGAGAAAGCCGATAACGA AATCTACGAGGAGCGGATTAACAAAATGTTTAAAAGTGAGAAG AGTTTCTCAATTGCCTACCTGAATTCGTTGCCGCACCCAAAGA CGGATGCGCCTCAAAAAAATGTTGAGGATTATTTTGCTCTCCT GGGGACTTGCAATCAAAACGATGAACAGCCGATTAATTTGTTT GCCCAAATTGAGATGGCACGCTTAGTCGCCTCTGATATTCTCG CAGGCCGGCACGTTAATTTGAACCAATCTGAGAATGATATCAA GTTAATCAAGGATCTGTTAGATGCTTACAAGGCTCTGCAGCAT TTCGTCAAACCACTCCTTGGCTCGGGTGACGAGGCTGAGAAAG ATAACGAGTTCGATGCACGCCTCCGTGCGGCTTGGAATGCGTT GGACATTGTTACACCACTCTATAACAAGGTTCGGAACTGGCTG ACCCGCAAACCATATTCTACAGAAAAAATCAAGCTTAATTTCG AAAACGCCCAACTTCTGGGGGGTTGGGATCAGAACAAAGAACC GGATTGCACATCAGTCCTCCTTCGGAAGGATGGGATGTACTAT TTAGCGATCATGGATAAAAAGGCGAATCACGCCTTTGACTGTG

	EIKEGYLSQVVHILSKMMVDYK AIVVLEDLNTGFMRNRQKIERQ VYEKFEKMLIDKLNCYVDKQKD ADETGGALHPLQLTNKFESFRK LGKQSGWLFYIPAWNTSKIDPV TGFVNMLDTRYENADKARCFFS KFDSIRYNADKDWFEFAMDYSK FTDKAKDTYTWWTLCSYGTRIK TFRNPAKNNLWDNEEVVLTDEF KKVFAAAGIDVHENLKEAICAL TDKKYLEPLMRLMTLLVQMRNS ATNSETDYLLSPVADESGMFYD SREGKETLPKDADANGAYNIAR KGLWTIRRIQATNCEEKVNLVL SNREWLQFAQQKPYLNDAAAKR PAATKKAGQAKKKKASGSGAGS PKKKRKVEDPKKKRKV **(SEQ ID NO: 55)**	ACTGCTTACCGTCTGACGGGGCCTGTTTCGAGAAAATTGACTA CAAGCTGCTCCCGGGCGCGAATAAAATGTTGCCGAAAGTTTTT TTTTCTAAAAGCCGCATCAAAGAATTTTCCCCTTCGGAATCGA TCATCGCTGCTTATAAAAAGGGGACTCATAAAAAAGGGCCGAA TTTCAGTCTCTCTGATTGTCATCGCTTGATTGACTTTTTTAAG GCTAGCATTGATAAGCACGAAGATTGGTCAAAATTTCGTTTTC GCTTCTCAGATACCAAAACGTATGAAGACATCAGTGGTTTCTA CCGTGAAGTAGAACAGCAAGGCTATATGCTGGGTTTTCGTAAA GTCTCTGAGGCCTTTGTGAATAAACTCGTTGATGAAGGTAAGT TATACTTATTCCATATCTGGAACAAAGACTTTAGTAAGCACTC CAAAGGTACACCTAATCTCCACACTATTTATTGGAAAATGCTC TTCGATGAGAAAAATCTCACTGACGTCATCTACAAACTGAATG GGCAGGCTGAAGTATTCTACCGTAAAAAAAGTCTGGATCTTAA TAAGACAACTACTCACAAGGCACATGCCCCAATCACCAATAAA AATACCCAAAACGCAAAGAAGGGTAGTGTTTTCGATTACGATA TCATCAAAAATCGTCGCTACACAGTGGACAAATTCCAGTTCCA CGTCCCTATCACCTTAAATTTTAAGGCAACAGGTCGTAATTAC ATTAATGAGCACACTCAAGAGGCAATCCGTAATAATGGCATCG AACATATCATTGGCATCGACCGTGGGGAGCGTCACTTGCTTTA CTTGTCGCTCATTGATCTGAAGGGTAATATCGTCAAGCAGATG ACCCTTAATGATATTGTCAATGAATATAATGGTCGGACTTATG CGACGAACTACAAGGACTTGCTGGCAACACGGGAGGGTGAGCG TACGGACGCTCGGCGCAACTGGCAGAAGATTGAAAATATTAAA GAAATCAAGGAAGGTTACCTTAGCCAGGTGGTGCACATCTTGA GTAAAATGATGGTCGACTACAAGGCTATCGTTGTTCTGGAAGA CTTGAATACAGGCTTCATGCGGAATCGTCAAAAAATCGAACGT CAAGTATATGAGAAGTTCGAAAAAATGTTAATTGACAAGCTGA ACTGCTATGTTGACAAACAAAAGGATGCTGACGAGACGGGCGG TGCCCTCCACCCGCTGCAGCTGACAAACAAATTTGAGTCGTTT CGTAAGTTAGGTAAGCAGAGTGGTTGGCTTTTTTACATCCCAG CATGGAACACTTCGAAAATCGACCCAGTTACTGGGTTCGTGAA CATGTTAGACACGCGCTACGAGAACGCCGATAAGGCGCGGTGT TTCTTCTCGAAATTCGATTCCATCCGGTATAACGCTGACAAAG ATTGGTTTGAGTTTGCTATGGATTACAGTAAGTTCACTGATAA AGCGAAAGATACTTACACGTGGTGGACTCTGTGTTCCTATGGG ACGCGTATTAAAACTTTTCGTAATCCGGCTAAGAATAATTTGT GGGATAATGAGGAGGTTGTCCTTACTGATGAGTTCAAGAAAGT TTTCGCAGCGGCAGGTATTGATGTCCATGAGAACCTTAAGGAA GCGATCTGTGCTCTGACAGATAAAAAGTATCTTGAACCACTCA TGCGTCTCATGACCCTGCTCGTTCAAATGCGGAACTCTGCTAC TAACTCCGAAACAGACTATTTACTTTCACCAGTTGCTGACGAG TCAGGGATGTTCTATGACTCCCGCGAAGGGAAGGAAACACTGC CAAAAGATGCGGACGCCAACGGTGCATATAACATTGCCCGTAA GGGCCTCTGGACCATCCGGCGGATTCAAGCCACCAACTGTGAG GAGAAAGTTAACTTAGTCCTCAGTAATCGTGAATGGTTGCAGT TTGCCCAGCAGAAACCATATCTGAATGATGCGGCCGCAAAAAG GCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG GCTAGCGGCAGCGGCGCCGGATCCCCAAAGAAGAAAAGGAAGG TTGAAGACCCCAAGAAAAAGAGGAAGGTGTGATAA **(SEQ ID NO: 56)**
ABW 6	MGHHHHHHSSGLVPRGSGTMIY RENFKRKKEKIEMNTGFNDFTN LSSVTKTLCNRLIPTEITAKYI KEHGVIEADQERNMMSQELKNI LNDFYRSFLNENLVKVHELDFK PLFTEMKKYLETKDNKEALEKA	ATGGGCCACCATCATCATCATCATAGCAGCGGCCTGGTGCCGC GCGGCAGCGGTACCATGATCTACCGTGAGAATTTTAAGCGGAA AAAGGAGAAGATTGAAATGAACACTGGGTTTAATGACTTCACT AATTTGAGTTCCGTGACCAAGACGTTATGCAACCGGTTGATCC CAACAGAAATTACCGCAAAGTACATTAAGGAGCATGGGGTAAT TGAGGCGGACCAAGAACGGAACATGATGAGTCAAGAGCTGAAA

	QDDMRKAIHDIFESDDRYKKMF	AATATCTTGAATGACTTTTACCGGAGTTTCCTGAACGAGAACC
	KAEITASILPEFILHNGAYSAE	TTGTGAAGGTGCACGAACTTGATTTCAAGCCGTTATTCACCGA
	EKEEKMQVVKMFNGFMTSFSAF	GATGAAAAAGTACCTCGAAACAAAAGATAACAAGGAAGCACTC
	FTNRENCFSKEKISSSACYRIV	GAAAAGGCCCAGGACGACATGCGGAAGGCAATCCATGATATCT
	DDNAKIHFDNIRIYKNIANKFD	TTGAAAGTGATGACCGCTACAAAAAAATGTTCAAGGCTGAGAT
	YEIEMIEKIEEAAGGADIRNIF	CACGGCGTCGATTTTGCCTGAATTCATTCTTCATAACGGGGCA
	SYNFDHFAFNHFVSQDDISFYN	TATTCAGCCGAAGAAAAGGAGGAGAAAATGCAAGTAGTCAAGA
	YVVGGINKFMNLYCQATKEKLS	TGTTCAATGGCTTTATGACGTCTTTCTCAGCATTCTTTACGAA
	PYKLRHLHKQILCIEESLYDVP	TCGTGAGAATTGTTTCTCCAAAGAAAAGATCAGCTCCTCCGCA
	AKFNCDEDVYAAVNDFLNNVRT	TGTTACCGTATTGTTGATGACAACGCGAAAATCCATTTCGATA
	KSVIERLQMLGKNADSYDLDKI	ACATTCGTATTTATAAAAATATCGCCAACAAGTTCGATTATGA
	YISKKHFTNISQTLYRDFSVIN	AATTGAAATGATCGAGAAGATCGAAGAGGCGGCGGGGGGTGCC
	TALTMSYIDTLPGKGKTKEKKA	GACATTCGTAATATCTTCTCGTACAACTTTGACCACTTTGCAT
	ASMAKNTELISLGEIDKLVDKY	TCAATCATTTCGTTAGTCAAGATGATATCTCATTCTACAATTA
	NLCPDKAASTRSLIRSISDIVA	TGTTGTTGGTGGTATTAACAAGTTTATGAACTTGTATTGTCAA
	DYKANPLTMNSGIPLAENETEI	GCCACCAAAGAGAAATTATCGCCTTATAAACTGCGTCACCTTC
	AVLKEAIEPFMDIFRWCAKFKT	ACAAACAGATTCTGTGTATTGAGGAAAGCCTCTATGACGTGCC
	DEPVDKDTDFYTELEDINDEIH	AGCGAAGTTTAATTGTGATGAGGACGTATATGCAGCTGTCAAC
	SIVSLYNRTRNYVTKKPYNTDK	GATTTTCTTAATAACGTTCGGACGAAATCAGTAATTGAACGCT
	FGLYFGTSSFASGWSESKEFTN	TGCAAATGCTCGGCAAAAATGCAGACAGTTACGACCTGGATAA
	NAILLAKDDKFYLGVFNAKNKP	AATTTATATCTCTAAAAAGCACTTCACCAATATCTCTCAAACT
	AKSIIKGHDTIQDGDYKKMVYS	TTATATCGCGACTTCTCTGTGATCAACACTGCCCTCACTATGT
	LLTGPNKMLPHMFISSSKAVPV	CTTATATCGATACTCTTCCGGGTAAGGGGAAAACCAAGGAAAA
	YGLTDELLSDYKKGRHLKTSKN	AAAGGCAGCATCGATGGCCAAAAACACCGAACTTATTTCGTTA
	FDIDYCHKLIDYFKHCLALYTD	GGCGAAATTGATAAGTTGGTGGATAAATATAACCTCTGTCCAG
	WDCFNFKFSDTESYNDIGEFYK	ATAAGGCAGCTAGCACTCGTAGCCTCATTCGGTCTATTAGCGA
	EVAEQGYYMNWTYIGSDDIDSL	CATCGTCGCTGACTACAAGGCAAACCCTCTTACAATGAATAGT
	QENGQLYLFQIYNKDFSEKSFG	GGGATTCCGTTGGCAGAGAACGAGACAGAAATCGCGGTGTTAA
	KPSKHTAILRSLFSDENVADPV	AAGAGGCGATCGAGCCTTTTATGGATATCTTCCGGTGGTGTGC
	IKLCGGTEVFFRPKSIKTPVVH	TAAGTTTAAAACCGACGAGCCTGTCGATAAGGATACAGATTTC
	KKGSILVSKTYNAQEMDENGNI	TACACGGAGTTAGAAGACATTAACGATGAAATCCATAGTATTG
	ITVRKCVPDDVYMELYGYYNNS	TCAGTCTTTATAACCGGACCCGGAATTATGTCACTAAAAAGCC
	GTPLSAEALKYKDIVDHRTAPY	GTACAACACAGATAAGTTCGGTCTGTATTTTGGCACTTCGTCG
	DIIKDRRYTEDEFFINMPVSLN	TTCGCATCGGGTTGGAGCGAGAGCAAAGAGTTTACTAACAACG
	YKAENRRVNVNEMALKYIAQTK	CAATTTTGTTAGCCAAGGATGACAAGTTTTACCTCGGCGTGTT
	DTYIIGIDRGERNLLYVSVIDT	CAACGCAAAAAACAAGCCAGCAAAATCGATTATCAAAGGGCAT
	DGNIVEQKSLNIINNVDYQAKL	GACACAATCCAAGATGGTGATTATAAGAAAATGGTGTATTCAC
	KQVEIMRKLARQNWKQGVKIAD	TGCTCACCGGGCCAAATAAGATGCTTCCTCACATGTTTATCTC
	LKKGYLSQAVHEVAELVIKYNG	GAGCAGTAAAGCGGTTCCTGTTTACGGGCTCACTGACGAGCTT
	IVVMEDLNSRFKEKRSKIERGV	CTCAGCGACTATAAGAAAGGTCGCCACCTTAAGACATCCAAGA
	YQQFETSLIKTLNYLTFKDRKP	ATTTCGACATTGATTACTGTCACAAACTTATCGATTACTTCAA
	LEAGGIANGYQLTYIPESLKNV	ACATTGTCTCGCTTTGTATACTGATTGGGATTGCTTCAACTTC
	GSQCGCILYVPAAYTSKIDPTT	AAATTCTCTGATACGGAGTCCTACAATGATATCGGCGAGTTCT
	GFVTLFKFKDISSEKAKTDFIG	ACAAAGAGGTTGCCGAGCAAGGCTACTACATGAACTGGACATA
	RFDCIRYDAEKDLFAFEFDYDN	TATCGGGTCGGACGATATCGATTCGCTGCAGGAAAACGGCCAG
	FETYETCARTKWCAYTYGTRVK	CTCTATCTTTTTCAAATTTATAACAAAGATTTCAGCGAAAAGT
	KTFRNRKFVSEVIIDITEEIKK	CATTCGGTAAACCGTCTAAACATACGGCCATCCTGCGTAGCTT
	TLAATDINWIDSHDIKQEIIDY	ATTCAGCGATGAAAACGTGGCCGACCCAGTCATTAAACTGTGT
	ALSSHIFEMFKLTVQMRNSLCE	GGGGGGACCGAAGTTTTTTTCCGGCCGAAGTCTATTAAGACAC
	SKDREYDKFVSPILNASGKFFD	CAGTAGTACATAAAAAAGGCAGCATCCTCGTATCCAAAACCTA
	TDAADKSLPIEADANDAYGIAM	TAACGCACAAGAAATGGACGAGAATGGTAATATCATCACCGTG
	KGLYNVLQVKNNWAEGEKFKFS	CGGAAGTGTGTTCCAGACGACGTCTATATGGAGCTCTACGGCT
	RLSNEDWFNFMQKRAAAKRPAA	ATTACAACAACTCTGGGACGCCTCTGTCCGCCGAAGCTTTGAA
	TKKAGQAKKKKASGSGAGSPKK	ATACAAGGATATTGTGGACCACCGCACGGCTCCGTACGACATT
	KRKVEDPKKKRKV **(SEQ ID NO: 68)**	ATCAAGGACCGGCGTTACACCGAAGACGAATTTTTCATCAACA TGCCGGTGTCATTGAATTATAAAGCGGAAAACCGCCGTGTTAA

		TGTGAACGAAATGGCCTTAAAATACATCGCACAGACCAAGGAC ACCTACATCATTGGCATCGATCGGGGCGAACGTAATCTGTTGT ATGTGAGCGTTATCGATACTGACGGCAATATCGTTGAGCAAAA GAGTCTCAATATCATCAATAACGTGGATTATCAAGCCAAATTA AAGCAAGTGGAAATCATGCGTAAACTGGCCCGTCAGAATTGGA AGCAGGGGGTAAAGATTGCAGACCTGAAAAAGGGCTACCTGTC ACAAGCGGTACATGAAGTCGCGGAACTTGTAATTAAATACAAC GGGATTGTTGTAATGGAGGACTTAAACTCCCGCTTCAAAGAGA AGCGTTCTAAAATTGAACGCGGCGTCTACCAACAGTTTGAGAC ATCATTAATCAAGACATTGAATTATTTGACGTTCAAAGATCGC AAACCGTTAGAAGCCGGGGGCATTGCGAATGGTTATCAATTAA CTTATATTCCGGAGTCTCTTAAAAATGTGGGCTCTCAGTGCGG CTGTATCTTGTATGTGCCAGCAGCCTACACCTCGAAGATCGAC CCTACCACTGGTTTCGTCACCTTGTTCAAATTCAAAGACATTT CGAGCGAGAAAGCTAAAACGGATTTTATTGGTCGGTTCGACTG CATCCGTTATGATGCAGAAAAGGACCTTTTCGCATTTGAATTC GATTATGACAACTTTGAGACTTATGAGACTTGTGCGCGTACCA AATGGTGTGCATATACATACGGGACTCGGGTGAAGAAAACTTT CCGGAATCGGAAATTCGTGTCAGAGGTGATCATCGACATCACT GAAGAGATCAAGAAGACCCTTGCAGCGACCGATATTAATTGGA TTGACAGTCACGACATCAAACAAGAGATCATCGACTATGCCCT TAGCAGCCATATTTTTGAAATGTTCAAATTAACGGTACAGATG CGTAACAGCCTTTGCGAGAGTAAAGATCGCGAGTACGACAAGT TCGTCTCACCTATTCTCAACGCGTCGGGCAAATTTTTCGACAC CGATGCCGCTGATAAAAGTCTGCCTATTGAAGCTGATGCGAAC GATGCGTATGGTATTGCTATGAAAGGGTTGTATAATGTTTTAC AAGTCAAAAACAACTGGGCGGAGGGCGAGAAATTTAAGTTCTC CCGTTTAAGCAACGAAGATTGGTTCAACTTCATGCAAAAGCGG GCGGCCGCAAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGG CAAAAAAGAAAAAGGCTAGCGGCAGCGGCGCCGGATCCCCAAA GAAGAAAAGGAAGGTTGAAGACCCCAAGAAAAAGAGGAAGGTG TGATAA **(SEQ ID NO: 69)**
ABW 7	MGHHHHHHSSGLVPRGSLQMTM DYGNGQFERRAPLTKTITLRLK PIGETRETIREQKLLEQDAAFR KLVETVTPIVDDCIRKIADNAL CHFGTEYDFSCLGNAISKNDSK AIKKETEKVEKLLAKVLTENLP DGLRKVNDINSAAFIQDTLTSF VQDDADKRVLIQELKGKTVLMQ RFLTTRITALTVWLPDRVFENF NIFIENAEKMRILLDSPLNEKI MKFDPDAEQYASLEFYGQCLSQ KDIDSYNLIISGIYADDEVKNP GINEIVKEYNQQIRGDKDESPL PKLKKLHKQILMPVEKAFFVRV LSNDSDARSILEKILKDTEMLP SKIIEAMKEADAGDIAVYGSRL HELSHVIYGDHGKLSQIIYDKE SKRISELMETLSPKERKESKKR LEGLEEHIRKSTYTFDELNRYA EKNVMAAYIAAVEESCAEIMRK EKDLRTLLSKEDVKIRGNRHNT LIVKNYFNAWTVFRNLIRILRR KSEAEIDSDFYDVLDDSVEVLS LTYKGENLCRSYITKKIGSDLK	ATGGGCCACCATCATCATCATCATAGCAGCGGCCTGGTGCCGCG CGGCAGCCTGCAGATGACAATGGATTACGGTAACGGTCAATTTG AGCGGCGCGCCCCGCTCACCAAGACAATCACTCTCCGGTTGAAA CCGATCGGGGAGACCCGTGAGACGATTCGCGAGCAAAAGCTCCT CGAACAAGATGCTGCATTCCGTAAACTTGTTGAAACTGTCACCC CTATCGTGGATGATTGTATCCGGAAAATTGCTGACAACGCTTTG TGTCATTTTGGCACGGAATATGATTTCTCCTGTTTAGGTAATGC CATCTCAAAAAATGACAGCAAAGCGATTAAGAAAGAGACCGAAA AAGTAGAGAAGCTGTTGGCCAAGGTTCTGACAGAGAACTTGCCA GACGGTCTGCGTAAAGTCAACGATATTAACAGCGCGGCTTTTAT TCAGGACACACTGACATCATTCGTCCAGGACGATGCTGACAAAC GTGTGTTAATTCAAGAGTTAAAGGGCAAAACTGTGTTAATGCAA CGCTTTTTAACAACCCGGATTACTGCATTGACTGTATGGCTCCC TGACCGGGTGTTTGAGAACTTCAACATTTTTATCGAAAATGCTG AAAAGATGCGCATCTTGCTCGACTCACCATTGAATGAAAAGATC ATGAAGTTCGATCCGGATGCTGAACAATACGCGAGTTTGGAATT CTATGGTCAATGTCTGTCCCAGAAGGATATTGATTCGTACAACC TCATCATTTCCGGGATTTATGCCGATGATGAGGTCAAGAACCCA GGTATCAATGAAATTGTTAAGGAATACAACCAGCAAATTCGCGG GGATAAGGATGAGTCACCTTTACCTAAACTGAAAAAGTTGCATA AACAAATTTTGATGCCTGTCGAGAAGGCATTTTTCGTTCGGGTA CTCAGTAATGATTCTGATGCTCGTTCAATTTTAGAAAAAATCTT GAAGGATACTGAGATGTTGCCTTCTAAGATCATTGAAGCGATGA AAGAAGCAGACGCTGGGGACATCGCTGTATATGGTTCACGTTTG

	PEIATYGSALRPNSRWWSPGEK	CACGAGTTAAGCCACGTAATCTATGGCGATCACGGGAAGCTCTC
	FNVKFHTIVRRDGRLYYFILPK	TCAGATTATCTATGATAAGGAGTCGAAACGCATCAGCGAGCTCA
	GAKPVELEDMDGDIECLQMRKI	TGGAAACGTTATCGCCTAAGGAGCGCAAAGAGTCAAAGAAACGC
	PNPTIFLPKLVFKDPEAFFRDN	TTGGAGGGTCTGGAAGAACATATCCGGAAGTCGACATATACCTT
	PEADEFVFLSGMKAPVTITRET	CGACGAGCTTAATCGTTATGCGGAAAAGAACGTCATGGCTGCCT
	YEAYRYKLYTVGKLRDGEVSEE	ACATCGCGGCCGTGGAGGAAAGCTGCGCCGAAATTATGCGTAAG
	EYKRALLQVLTAYKEFLENRMI	GAGAAGGACTTACGCACGCTTCTTAGTAAGGAGGATGTCAAGAT
	YADLNFGFKDLEEYKDSSEFIK	TCGTGGTAATCGCCACAATACGTTAATTGTTAAGAACTACTTCA
	QVETHNTFMCWAKVSSSQLDDL	ATGCCTGGACTGTCTTCCGGAATTTGATCCGCATCCTCCGGCGG
	VKSGNGLLFEIWSERLESYYKY	AAATCCGAGGCGGAGATCGACTCAGATTTCTATGACGTCTTGGA
	GNEKVLRGYEGVLLSILKDENL	TGACTCTGTGGAAGTTTTATCGCTCACATATAAAGGTGAAAACT
	VSMRTLLNSRPMLVYRPKESSK	TGTGCCGGTCTTACATTACGAAGAAGATCGGGAGCGATTTAAAG
	PMVVHRDGSRVVDRFDKDGKYI	CCAGAGATTGCTACCTATGGTTCCGCCTTGCGCCCTAATTCACG
	PPEVHDELYRFFNNLLIKEKLG	GTGGTGGTCACCGGGCGAGAAGTTTAACGTAAAGTTCCACACCA
	EKARKILDNKKVKVKVLESERV	TTGTTCGCCGGGACGGTCGCCTTTATTATTTCATCTTGCCGAAA
	KWSKFYDEQFAVTFSVKKNADC	GGTGCCAAACCTGTCGAGCTCGAAGATATGGATGGGGACATCGA
	LDTTKDLNAEVMEQYSESNRLI	ATGCTTGCAAATGCGCAAGATTCCGAATCCGACTATTTTCCTTC
	LIRNTTDILYYLVLDKNGKVLK	CAAAATTGGTTTTCAAGGACCCAGAGGCCTTCTTCCGCGACAAT
	QRSLNIINDGARDVDWKERFRQ	CCAGAGGCAGATGAATTCGTTTTTCTTTCGGGTATGAAAGCTCC
	VTKDRNEGYNEWDYSRTSNDLK	AGTGACCATCACGCGTGAAACCTATGAGGCGTATCGCTACAAAC
	EVYLNYALKEIAEAVIEYNAIL	TTTATACAGTTGGGAAGTTACGCGACGGTGAAGTGAGCGAAGAA
	IIEKMSNAFKDKYSFLDDVTFK	GAGTATAAACGTGCGTTGTTACAAGTATTGACCGCCTATAAGGA
	GFETKLLAKLSDLHFRGIKDGE	ATTCTTAGAGAATCGGATGATCTACGCAGATCTGAACTTTGGCT
	PCSFTNPLQLCQNDSNKILQDG	TTAAAGATCTCGAAGAATACAAAGACTCGTCAGAATTTATCAAA
	VIFMVPNSMTRSLDPDTGFIFA	CAAGTCGAAACTCACAACACTTTTATGTGCTGGGCTAAGGTCAG
	INDHNIRTKKAKLNFLSKFDQL	TAGCAGTCAGCTCGACGACCTGGTCAAGAGCGGGAACGGGTTAC
	KVSSEGCLIMKYSGDSLPTHNT	TGTTCGAAATCTGGTCAGAACGGTTGGAGTCCTATTACAAATAT
	DNRVWNCCCNHPITNYDRETKK	GGCAACGAGAAGGTGCTGCGTGGGTACGAGGGCGTTCTTTTGAG
	VEFIEEPVEELSRVLEENGIET	TATCCTTAAGGACGAGAACCTCGTGAGCATGCGGACGCTGCTTA
	DTELNKLNERENVPGKVVDAIY	ATTCTCGGCCGATGCTCGTCTACCGCCCTAAAGAATCATCCAAG
	SLVLNYLRGTVSGVAGQRAVYY	CCGATGGTCGTTCACCGGGACGGTAGCCGCGTCGTTGATCGGTT
	SPVTGKKYDISFIQAMNLNRKC	CGATAAGGATGGGAAGTATATTCCACCAGAGGTACACGACGAAT
	DYYRIGSKERGEWTDFVAQLIN	TATACCGGTTCTTTAACAATTTGCTTATTAAGGAAAAGCTCGGC
	AAAKRPAATKKAGQAKKKKASG	GAGAAAGCGCGCAAAATTTTAGACAACAAAAAAGTAAAAGTAAA
	SGAGSPKKKRKVEDPKKKRKV	GGTATTGGAATCTGAACGTGTAAAGTGGTCAAAGTTTTATGATG
	(SEQ ID NO: 81)	AACAGTTTGCAGTTACATTCTCTGTTAAAAAGAATGCAGACTGT
		CTGGATACCACGAAAGATCTCAATGCCGAAGTTATGGAGCAGTA
		TTCCGAATCGAACCGGCTTATCCTGATCCGCAATACCACTGACA
		TCTTGTATTATCTTGTACTTGATAAGAATGGGAAAGTGCTGAAA
		CAACGCTCATTGAATATCATTAACGACGGGGCTCGCGACGTTGA
		TTGGAAAGAGCGTTTTCGGCAGGTAACAAAAGATCGTAACGAAG
		GCTATAACGAGTGGGACTACTCGCGGACTAGCAACGATTTGAAA
		GAGGTCTATCTGAATTATGCATTGAAGGAGATTGCCGAAGCGGT
		AATCGAATACAACGCAATTTTGATTATTGAAAAAATGTCGAATG
		CCTTCAAGGATAAGTACTCCTTTTTGGATGATGTTACCTTCAAA
		GGTTTTGAGACCAAACTTCTTGCGAAGCTCTCTGACTTGCATTT
		CCGGGGTATTAAAGATGGGGAGCCATGTTCGTTTACGAACCCGT
		TACAGTTATGTCAGAACGACTCAAACAAAATTTTACAAGACGGT
		GTGATTTTCATGGTCCCTAACAGCATGACGCGCAGTCTGGACCC
		TGACACTGGGTTCATTTTTGCGATTAACGATCACAACATCCGCA
		CTAAGAAAGCGAAGTTAAACTTCCTTAGTAAATTCGATCAGCTG
		AAAGTGTCATCAGAGGGCTGTTTAATCATGAAATATTCGGGGGA
		CTCCCTTCCTACACACAACACAGATAATCGTGTATGGAACTGTT
		GTTGCAATCACCCGATCACCAACTACGACCGCGAGACGAAAAAG
		GTCGAATTCATCGAGGAGCCAGTGGAAGAGTTGAGTCGCGTCTT
		AGAAGAGAATGGGATTGAGACAGATACGGAACTTAACAAGCTTA

		ACGAGCGCGAGAATGTTCCGGGCAAGGTAGTAGATGCCATCTAT TCTCTGGTGTTGAATTACTTGCGTGGTACCGTGTCCGGCGTTGC AGGCCAACGGGCGGTCTACTATTCCCCTGTGACGGGGAAAAAAT ATGATATTTCGTTTATCCAAGCAATGAATCTGAATCGTAAGTGC GATTACTACCGGATCGGGAGCAAAGAACGCGGCGAATGGACGGA TTTTGTAGCGCAGTTAATTAACGCGGCCGCAAAAAGGCCGGCGG CCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGCTAGCGGC AGCGGCGCCGGATCCCCAAAGAAGAAAAGGAAGGTTGAAGACCC CAAGAAAAAGAGGAAGGTGTGATAA **(SEQ ID NO: 82)**
ABW8	MGHHHHHHSSGLVPRGSGTMCY DLNNIKTKLREREVETMGNNMD NSFEPFIGGNSVSKTLRNELRV GSEYTGKHIKECAIIAEDAVKA ENQYIVKEMMDDFYRDFINRKL DALQGINWEQLFDIMKKAKLDK SNKVSKELDKIQESTRKEIGKI FSSDPIYKDMLKADMISKILPE YIVDKYGDAASRIEAVKVFYGF SGYFIDFWASRKNVFSDKNIAS AIPHRIVNVNARIHLDNITAFN RIAEIAGDEVAGIAEDACAYLQ NMSLEDVFTGACYGEFICQKDI DRYNNICGVINQHMNQYCQNKK ISRSKFKMERLHKQILCRSESG FEIPIGFQTDGEVIDAINSFST ILEEKDILDRLRTLSQEVTGYD MERIYVSSKAFESVSKYIDHKW DVIASSMYNYFSGAVRGKDDKK DVKIQTEIKKIKSCSLLDLKKL VDMYYKMDGMCLEHEATEYVAG ITEILVDFNYKTFDMDDSVKMI QNEHMINEIKEYLDTYMSIYHW AKDFMIDELVDRDMEFYSELDE IYYDLSDIVPLYNKVRNYVTQK PYSQDKIKLNFGSPTLANGWSK SKEFDNNVVVLLRDEKIYLAIL NVGNKPSKDIMAGEDRRRSDTD YKKMNYYLLPGASKTLPHVFIS SNAWKKSHGIPDEIMYGYNQNK HLKSSPNFDLEFCRKLIDYYKE CIDSYPNYQIFNFKFAATETYN DISEFYKDVERQGYKIEWSYIS EDDINQMDRDGQIYLFQIYNKD FAPNSKGMQNLHTLYLKNIFSE ENLSDVVIKLNGEAELFFRKSS IQHKRGHKKGSVLVNKTYKTTE KTENGQGEIEVIESVPDQCYLE LVKYWSEGGVGQLSEEASKYKD KVSHYAATMDIVKDRRYTEDKF FIHMPITINFKADNRNNVNEKV LKFIAENDDLHVIGIDRGERNL LYVSVIDSRGRIVEQKSFNIVE NYESSKNVIRRHDYRGKLVNKE HYRNEARKSWKEIGKIKEIKEG YLSQVIHEISKLVLKYNAIIVM EDLNYGFKRGRFKVERQVYQKF	ATGGGCCACCATCATCATCATCATAGCAGCGGCCTGGTGCCGC GCGGCAGCGGTACCATGTGCTACGACTTAAACAACATCAAGAC AAAGTTACGTGAACGCGAAGTCGAAACTATGGGCAATAACATG GATAATAGCTTCGAGCCTTTTATTGGCGGTAATAGTGTCTCTA AAACACTTCGGAATGAGCTGCGTGTAGGTTCCGAATATACTGG TAAACACATTAAAGAGTGCGCGATCATTGCAGAGGACGCCGTG AAGGCGGAGAACCAGTACATCGTAAAAGAGATGATGGACGACT TTTACCGTGACTTCATTAATCGCAAACTTGACGCCTTGCAGGG TATTAATTGGGAGCAGCTTTTTGACATTATGAAGAAGGCGAAA TTGGATAAGTCGAATAAAGTCAGCAAAGAGTTAGACAAGATTC AAGAGTCTACGCGGAAAGAAATCGGGAAAATCTTCTCATCCGA TCCAATCTATAAAGACATGCTCAAAGCGGACATGATCAGCAAA ATTCTGCCAGAGTATATTGTCGACAAATACGGTGATGCAGCCT CGCGGATCGAAGCTGTAAAGGTGTTTTACGGCTTTTCGGGTTA TTTTATCGACTTCTGGGCATCGCGCAAGAACGTCTTCTCAGAT AAGAACATCGCGTCGGCCATTCCGCACCGGATTGTCAATGTGA ACGCTCGGATCCATCTGGACAACATCACGGCCTTCAACCGTAT CGCAGAAATTGCAGGGGATGAAGTCGCCGGCATTGCTGAAGAT GCTTGTGCTTACCTGCAGAATATGAGCTTAGAGGATGTATTCA CGGGGGCCTGCTACGGTGAGTTCATCTGTCAGAAGGATATTGA TCGTTACAATAACATTTGCGGTGTTATCAACCAGCACATGAAT CAATACTGCCAAAACAAAAAGATCTCACGCTCAAAATTTAAGA TGGAACGTCTGCACAAACAGATCTTATGTCGCTCTGAGAGTGG TTTTGAGATCCCGATTGGGTTTCAAACCGACGGGGAGGTAATC GATGCTATCAACTCCTTTTCTACGATTCTTGAAGAGAAAGATA TCTTGGATCGTCTGCGCACTTTGTCGCAGGAGGTAACAGGTTA TGACATGGAGCGTATCTATGTAAGTTCCAAGGCGTTTGAGTCT GTATCAAAGTACATCGATCACAAATGGGACGTAATTGCTTCTT CCATGTACAATTACTTTTCTGGGGCTGTTCGTGGGAAGGACGA CAAGAAAGATGTCAAGATTCAGACGGAAATTAAAAAGATTAAG TCATGTTCGTTATTGGACCTCAAAAAGCTGGTAGATATGTATT ATAAAATGGATGGGATGTGTTTAGAGCACGAAGCGACGGAGTA CGTGGCAGGTATTACGGAGATCCTGGTTGACTTTAACTATAAG ACCTTCGACATGGATGATTCCGTTAAGATGATTCAAAATGAGC ACATGATTAATGAAATTAAAGAATATTTAGATACCTATATGTC TATCTATCATTGGGCGAAGGACTTTATGATCGATGAGCTCGTA GATCGCGACATGGAATTCTACAGTGAGCTCGATGAAATCTATT ATGATTTGTCCGACATCGTACCACTGTATAATAAAGTCCGCAA CTACGTCACGCAAAAACCGTATTCCCAGGATAAAATCAAGTTA AACTTTGGCAGCCCAACCTTAGCAAACGGTTGGAGCAAGTCGA AAGAATTTGATAACAACGTTGTAGTATTGTTGCGTGACGAAAA GATTTATCTGGCCATCTTAAATGTGGGGAATAAACCGTCAAAG GATATCATGGCGGGCGAAGACCGTCGTCGCTCCGATACTGATT ACAAGAAAATGAATTACTATCTGCTCCCTGGGGCAAGCAAAAC CCTGCCACACGTTTTTATCTCTTCAAATGCATGGAAGAAATCC CACGGTATCCCTGACGAGATTATGTACGGCTATAACCAAAATA AGCATTTAAAATCTTCGCCAAACTTCGACTTAGAGTTTTGTCG

	ETMLINKLAYLVDKSRAVDEPG GLLKGYQLTYVPDNLGELGSQC GIIFYVPAAYTSKIDPVTGFVD VFDFKAYSNAEARLDFINKLDC IRYDAPRNKFEIAFDYGNFRTH HTTLAKTSWTIFIHGDRIKKER GSYGWKDEIIDIEARIRKLFED TDIEYADGHNLIGDINELESPI QKKFVGELFDIIRFTVQLRNSK SEKYDGTEKEYDKIISPVMDEE GVFFTTDSYIRADGTELPKDAD ANGAYCIALKGLYDVLAVKKYW KEGEKFDRKLLAITNYNWFDFI QNRRFAAAKRPAATKKAGQAKK KKASGSGAGSPKKKRKVEDPKK KRKV **(SEQ ID NO: 94)**	CAAGCTGATCGATTATTACAAAGAATGTATTGACAGCTATCCT AACTATCAGATCTTCAATTTCAAATTCGCCGCTACGGAAACTT ACAACGATATTTCGGAGTTCTACAAAGATGTTGAACGTCAGGG GTACAAGATTGAATGGTCGTACATTTCCGAGGACGATATTAAT CAGATGGATCGTGACGGCCAGATTTATCTTTTTCAAATCTACA ACAAGGATTTTGCCCCAAACTCTAAGGGCATGCAGAATTTACA TACACTCTATTTAAAAAATATTTTTTCAGAGGAAAACCTCTCT GATGTCGTCATTAAACTGAATGGCGAGGCTGAGCTCTTCTTCC GCAAGAGCTCGATCCAACATAAACGCGGTCATAAGAAGGGTAG TGTGTTGGTAAATAAGACCTATAAAACCACAGAAAAAACTGAA AATGGTCAAGGCGAAATTGAAGTAATCGAGAGCGTGCCGGACC AGTGTTACCTGGAGCTTGTTAAGTACTGGTCAGAGGGTGGTGT AGGTCAGTTGTCAGAAGAGGCTTCCAAATACAAAGATAAAGTC AGCCACTACGCTGCAACAATGGATATTGTCAAGGACCGGCGGT ACACGGAGGATAAGTTCTTTATTCACATGCCGATTACGATTAA TTTTAAAGCTGATAACCGGAACAATGTCAACGAGAAAGTGCTG AAGTTTATTGCAGAAAACGATGATCTCCACGTTATTGGTATTG ACCGTGGGGAACGTAATCTCCTGTACGTCTCAGTAATTGATTC ACGTGGGCGTATTGTTGAGCAGAAGTCGTTTAATATTGTTGAG AATTACGAGAGCAGTAAAAATGTGATCCGCCGCCATGATTATC GTGGGAAATTAGTAAATAAAGAGCACTATCGTAATGAGGCACG TAAGAGCTGGAAAGAAATCGGCAAAATCAAGGAGATCAAAGAA GGTTATCTCAGTCAAGTTATCCATGAGATTAGTAAGTTGGTAT TAAAGTATAACGCCATCATCGTGATGGAAGATCTTAATTATGG CTTCAAACGCGGGCGGTTTAAAGTCGAGCGGCAGGTATACCAG AAGTTCGAGACCATGCTTATTAACAAATTAGCCTACTTAGTGG ACAAATCACGCGCGGTAGACGAACCGGGTGGGTTATTAAAAGG CTACCAGCTGACATACGTGCCAGATAACTTGGGTGAACTGGGG TCCCAGTGCGGGATCATTTTTTATGTGCCAGCAGCATACACTT CGAAAATCGATCCTGTTACGGGCTTTGTAGACGTGTTTGATTT TAAGGCATACTCCAATGCCGAAGCACGTTTAGATTTCATCAAT AAACTGGACTGCATCCGGTATGACGCGCCGCGTAACAAGTTTG AAATTGCTTTCGACTACGGTAACTTCCGGACTCATCATACAAC CCTTGCAAAGACTAGCTGGACTATTTTTATTCACGGCGACCGT ATTAAAAAGGAGCGCGGTTCTTACGGCTGGAAGGACGAAATTA TCGATATCGAGGCCCGTATTCGTAAGCTGTTTGAAGACACAGA CATCGAATACGCCGATGGTCACAATTTGATCGGTGACATTAAC GAGCTCGAGAGTCCAATTCAAAAGAAATTCGTTGGTGAGCTGT TCGACATTATCCGTTTCACTGTCCAACTGCGCAACAGCAAAAG TGAGAAATATGACGGCACCGAAAAGGAGTATGACAAAATTATT TCGCCGGTAATGGACGAGGAGGGGGTTTTCTTTACAACCGACA GTTATATCCGCGCAGATGGTACTGAATTACCTAAAGATGCTGA TGCTAACGGGGCCTATTGTATCGCGCTGAAGGGTCTTTACGAC GTGCTCGCGGTAAAGAAATATTGGAAGGAGGGGGAGAAGTTCG ATCGGAAGTTACTTGCCATCACCAATTACAACTGGTTTGATTT CATTCAGAATCGTCGCTTCGCGGCCGCAAAAAGGCCGGCGGCC ACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGCTAGCGGCA GCGGCGCCGGATCCCCAAAGAAGAAAAGGAAGGTTGAAGACCC CAAGAAAAAGAGGAAGGTGTGATAA **(SEQ ID NO: 95)**
ABW9	MGHHHHHHSSGLVPRGSGTMSD RLDVLTNQYPLSKTLRFELKPV GATADWIRKHNVIRYHNGKLVG KDAIRFQNYKYLKKMLDEMHRL FLQQALVLEPNSNQAQELTALL RAIENNYCNNNDLLAGDYPSLS TDKTIKISNGLSKLTTDLFDKK	ATGGGGCATCACCACCACCACCACTCGTCGGGTCTTGTTCCAC GTGGTTCTGGTACCATGTCTGATCGCCTGGACGTGCTTACTAA CCAATACCCATTATCGAAAACTTTGCGCTTCGAATTGAAGCCG GTTGGAGCCACAGCTGACTGGATTCGCAAACACAACGTTATCC GCTATCATAATGGTAAACTGGTTGGAAAGGATGCGATCCGTTT TCAAAATTATAAGTATCTGAAGAAAATGCTTGATGAGATGCAT CGCTTATTTCTTCAGCAAGCACTGGTGTTGGAGCCAAATAGCA

	FEDWAYQYKEDMPNFWRQDIAE	ACCAGGCGCAGGAGTTGACCGCACTGCTGCGTGCTATTGAGAA
	LEQKLQVSANAKDQKFYKGIIK	TAATTATTGCAACAACAACGACCTGCTGGCGGGCGATTATCCC
	KLKNKIQKSELKAETHKGLYSP	AGCCTCTCTACCGATAAGACCATTAAAATCAGCAACGGCCTTA
	TESLQLLEWLVRRGDIKLTYLE	GCAAGCTGACCACGGATCTGTTCGATAAGAAGTTCGAAGACTG
	IGKENEKLNELVPLVELKDIHR	GGCATACCAATACAAAGAAGATATGCCCAATTTCTGGCGTCAA
	NFNNFATYLSGFSKNRENVYST	GATATTGCGGAATTAGAGCAAAAGCTTCAGGTGAGTGCGAACG
	KFDRRSGYKATSVIARTFEQNL	CAAAAGATCAAAAGTTCTACAAAGGGATCATCAAGAAGCTGAA
	MFCLGNIAKWHKVTEFINQANN	GAATAAGATCCAGAAGTCTGAACTGAAAGCGGAAACGCACAAG
	YELLQEHGIDWNKQIAALEHKL	GGCTTATACTCACCTACGGAGTCACTGCAACTGCTGGAGTGGC
	DVCLAEFFALNNFSQTLAQQGI	TGGTACGTCGTGGCGATATTAAACTGACTTACTTAGAGATTGG
	EKYNQVLAGIAEIAGQPKTQGL	TAAAGAGAACGAGAAACTTAATGAACTGGTCCCGCTGGTCGAA
	NELINLARQKLSAKRSQLPTLQ	CTTAAGGACATTCATCGCAATTTCAATAATTTCGCCACATATC
	LLYKQILSKGDKPFIDDFKSDQ	TTTCTGGCTTCAGCAAGAATCGTGAGAATGTGTACTCAACCAA
	ELIAELNEFVSSQIHGEHGAIK	ATTTGATCGTCGTTCGGGTTATAAAGCCACCAGTGTAATCGCA
	LINHELESFINEARAAQQQIYV	CGCACGTTCGAACAGAATTTAATGTTCTGTCTTGGTAACATTG
	PKDKLTELSLLLTGSWQAINQW	CCAAGTGGCACAAGGTGACAGAATTCATCAACCAGGCGAACAA
	RYKLFDQKQLDKQQKQYSFSLA	TTACGAGCTCCTGCAGGAGCACGGCATCGATTGGAATAAGCAA
	QVERWLATEVEQQNFYQTEKER	ATTGCCGCGCTGGAACACAAACTGGACGTGTGTCTCGCAGAGT
	QQHKDTQPANVTTSSDGHSILT	TCTTCGCGCTTAATAACTTCTCACAAACCCTTGCACAACAGGG
	AFEQQVQTLLTNICVAAEKYRQ	TATCGAAAAGTATAACCAGGTCTTGGCCGGCATCGCCGAGATT
	LSDNLTAIDKQRESESSKGFEQ	GCAGGCCAACCCAAGACCCAGGGCCTGAACGAACTCATTAACC
	IAVIKTLLDACNELNHFLARFT	TGGCCCGTCAGAAATTGTCTGCCAAACGCTCACAACTGCCTAC
	VNKKDKLPEDRAEFWYEKLQAY	GTTGCAACTCCTTTACAAACAAATCTTAAGCAAGGGTGATAAG
	IDAFPIYELYNKVRNYLSKKPF	CCATTCATCGACGATTTTAAAAGCGACCAAGAGTTGATCGCCG
	STEKVKINFDNSHFLSGWTADY	AATTAAATGAGTTTGTAAGCAGCCAGATTCACGGAGAGCATGG
	ERHSALLFKFNENYLLGVVNEN	TGCAATCAAATTAATTAATCACGAACTTGAAAGCTTTATCAAT
	LSSEEEEKLKLVGGEEHAKRFI	GAAGCCCGTGCAGCGCAGCAACAGATTTATGTGCCCAAGGACA
	YDFQKIDNSNPPRVFIRSKGSS	AGCTTACCGAATTAAGTCTTCTCTTAACGGGCAGTTGGCAAGC
	FAPAVEKYQLPIGDIIDIYDQG	TATTAATCAATGGCGTTACAAACTGTTCGACCAGAAACAGCTG
	KFKTEHKKKNEAEFKDSLVRLI	GATAAACAACAGAAACAATATTCATTTAGCCTGGCCCAGGTTG
	DYFKLGFSRHDSYKHYPFKWKA	AACGCTGGCTGGCAACTGAGGTTGAGCAACAAAACTTCTACCA
	SHQYSDIAEFYAHTASFCYTLK	AACCGAAAAGGAGCGCCAGCAGCATAAAGATACGCAGCCGGCG
	EENINFNVLRELSSAGKVYLFE	AACGTCACCACCAGCAGCGATGGACACAGCATTTTAACAGCAT
	IYNKDFSKNKRGQGRDNLHTSY	TTGAGCAACAGGTGCAGACCTTATTAACCAACATCTGTGTTGC
	WKLLFSAENLKDVVLKLNGQAE	TGCCGAGAAATATCGCCAATTAAGTGATAATCTCACAGCCATC
	IFYRPASLAETKAYTHKKGEVL	GATAAACAACGCGAGAGCGAATCAAGTAAGGGATTCGAGCAAA
	KHKAYSKVWEALDSPIGTRLSW	TCGCGGTGATTAAAACCTTGCTGGACGCGTGTAACGAGCTGAA
	DDALKIPSITEKTNHNNQRVVQ	TCACTTTCTGGCACGCTTCACGGTCAACAAGAAGGACAAACTC
	YNGQEIGRKAEFAIIKNRRYSV	CCCGAAGATCGCGCAGAATTTTGGTATGAAAAGTTACAAGCGT
	DKFLFHCPITLNFKANGQDNIN	ACATTGACGCGTTTCCGATCTACGAGCTGTATAATAAAGTGCG
	ARVNQFLANNKKINIIGIDRGE	TAATTACTTAAGCAAGAAGCCGTTTAGCACTGAGAAAGTCAAA
	KHLLYISVINQQGEVLHQESFN	ATTAATTTTGACAATTCCCATTTCCTGTCGGGTTGGACGGCGG
	TITNSYQTANGEKRQVVTDYHQ	ACTATGAGCGTCACAGCGCCTTATTATTCAAATTTAATGAAAA
	KLDMSEDKRDKARKSWSTIENI	TTACCTGCTGGGTGTAGTGAATGAGAACTTAAGCAGCGAGGAA
	KELKAGYLSHVVHRLAQLIIEF	GAAGAAAAGCTGAAGCTCGTGGGCGGCGAAGAACATGCCAAGC
	NAIVALEDLNHGFKRGRFKIEK	GCTTCATTTATGATTTTCAGAAAATCGACAACTCAAACCCACC
	QVYQKFEKALIDKLSYLAFKDR	GCGCGTTTTCATTCGTAGCAAGGGGTCATCGTTCGCACCTGCG
	TSCLETGHYLNAFQLTSKFKGF	GTCGAAAAGTATCAGTTACCGATTGGCGATATCATTGACATTT
	NNLGKQSGILFYVNADYTSTTD	ACGATCAGGGTAAATTTAAGACAGAACACAAGAAGAAGAATGA
	PLTGYIKNVYKTYSSVKDSTEF	GGCCGAGTTTAAAGACAGTCTGGTACGTTTGATCGATTATTTT
	WQRFNSIRYIASENRFEFSYDL	AAGCTGGGCTTCTCTCGCCATGACAGCTATAAGCACTACCCAT
	ADLKQKSLESKTKQTPLAKTQW	TCAAGTGGAAAGCCAGTCATCAATATAGCGACATTGCGGAATT
	TVSSHVTRSYYNQQTKQHELFE	TTACGCTCATACCGCCTCATTTTGTTACACGCTTAAGGAAGAA
	VTARIQQLLSKAEISYQHQNDL	AACATCAATTTTAACGTTCTGCGTGAGTTGTCGTCGGCGGGCA
	IPALASCQSKALHKELIWLFNS	AAGTATATCTCTTCGAAATTTACAATAAGGATTTCTCAAAGAA
	ILTMRVTDSSKPSATSENDFIL	CAAGCGCGGCCAAGGACGCGACAACTTGCATACCAGTTATTGG

	SPVAPYFDSRNLNKQLPENGDA NGAYNIARKGIMLLERIGDFVP EGNKKYPDLLIRNNDWQNFVQR PEMVNKQKKKLVKLKTEYSNGS LFNDLAFKAAAKRPAATKKAGQ AKKKKASGSGAGSPKKKRKVED PKKKRKV **(SEQ ID NO: 107)**	AAGTTGCTGTTCTCGGCTGAGAACCTGAAGGATGTTGTGCTGA AATTAAACGGCCAAGCGGAGATCTTTTACCGCCCAGCGTCTTT GGCCGAAACCAAGGCCTACACCCATAAGAAAGGGGAAGTACTG AAACATAAGGCTTATAGCAAAGTGTGGGAAGCCCTGGATTCTC CCATTGGCACCCGCCTGAGCTGGGACGATGCTTTAAAGATCCC GTCTATTACCGAGAAGACCAATCACAATAATCAGCGTGTTGTC CAGTACAACGGCCAAGAAATTGGCCGCAAAGCGGAGTTCGCTA TTATCAAGAACCGCCGTTATTCCGTCGATAAATTCCTCTTTCA CTGCCCGATTACACTCAACTTCAAGGCGAACGGCCAGGACAAC ATTAACGCACGCGTTAATCAATTCCTGGCAAATAACAAGAAGA TCAACATTATTGGAATTGACCGTGGTGAAAAGCATTTACTGTA TATCAGCGTGATTAATCAACAAGGCGAAGTCCTGCATCAGGAA AGCTTCAATACAATCACGAATTCATATCAGACCGCCAATGGCG AGAAACGCCAAGTAGTCACTGACTATCACCAGAAGTTGGACAT GAGCGAGGACAAACGCGATAAAGCACGTAAGAGCTGGAGTACA ATCGAAAATATCAAAGAGCTGAAGGCGGGGTATCTGAGCCACG TTGTACATCGCCTCGCGCAACTGATTATCGAATTTAATGCCAT TGTTGCGTTGGAAGATCTTAACCACGGGTTCAAACGCGGACGT TTTAAAATCGAAAAGCAAGTGTATCAGAAGTTCGAAAAGGCGC TGATCGACAAATTGAGCTACTTAGCGTTTAAGGATCGCACGTC GTGTCTGGAAACTGGACATTACTTGAATGCCTTTCAATTAACC TCAAAGTTCAAAGGCTTTAACAACCTTGGCAAGCAATCCGGGA TTTTGTTCTACGTTAACGCCGATTACACGAGCACCACGGATCC CTTAACAGGCTATATTAAGAACGTATACAAAACCTACTCCTCG GTGAAGGATTCGACCGAATTTTGGCAGCGCTTTAACTCTATCC GCTATATTGCGAGCGAGAACCGTTTTGAATTTAGCTACGACTT AGCGGACCTGAAACAGAAGTCGCTCGAGAGTAAAACCAAACAG ACCCCTCTCGCCAAGACCCAATGGACGGTCTCTAGCCACGTTA CCCGTTCCTATTACAACCAGCAGACGAAGCAACATGAGTTATT CGAAGTGACAGCGCGCATTCAGCAATTGCTTAGCAAAGCAGAA ATCAGCTATCAACATCAAAACGACTTGATCCCTGCGTTAGCAT CATGTCAAAGTAAGGCGTTACACAAGGAGTTGATTTGGCTGTT CAACAGCATCCTGACTATGCGCGTCACGGACTCAAGCAAACCG TCCGCGACCTCGGAGAATGATTTTATCCTGAGCCCGGTAGCGC CGTACTTCGACTCCCGCAATCTGAATAAGCAGCTGCCGGAAAA CGGCGACGCGAACGGCGCATACAATATCGCTCGTAAAGGTATC ATGCTTCTGGAACGTATCGGGGACTTCGTCCCGGAAGGTAACA AGAAGTACCCCGATTTACTGATCCGCAATAATGACTGGCAGAA TTTTGTACAACGCCCGGAGATGGTGAACAAGCAGAAGAAGAAA CTCGTGAAGTTGAAAACGGAATACTCTAATGGCAGCCTCTTCA ATGATTTGGCGTTTAAGGCCGCAGCTAAGCGCCCCGCCGCGAC TAAGAAAGCGGGTCAAGCGAAGAAGAAGAAAGCGTCGGGGTCG GGAGCGGGCAGTCCGAAGAAGAAGCGTAAAGTAGAGGATCCGA AGAAGAAACGCAAAGTATAATAA**(SEQ ID NO: 108)**

[00228] In another exemplary method, the nine targeted type V CRISPR-associated protein Cas12a (referred to as ABW) nucleases were further engineered to create novel variants ABW1-ABW9 for each of the targeted nucleases and then compared to native amino acid sequences of three Cas12a (Cpf1) nucleases from different organisms. Exemplary results are provided in **Tables 5-6** below:

[00229] **Table 5** represents the percent identity between amino acid sequences of engineered ABW nucleases and native Cas12a nucleases. Percent identity between sequences was assessed by alignment and comparison in a BLAST, using a blastp algorithm. NCBI references are provided for each Cas12a sequence. As demonstrated below, arrows indicate decrease (↓) and increase (↑) in sequence similarity after this round of engineering.

	Percent Identity Amino Acid Between Sequences		
	AsCpf1 (WP_021736722.1)	**FnCpf1** (WP_003040289.1)	**EeCpf1** (WP_055225123.1)
ABW1	48.81	34.75	↑ 32.29
ABW2	34.14	37.25	↓ 30.15
ABW3	↓ 33.60	↓ 42.72	35.66
ABW4	↓ 34.63	41.17	33.65
ABW5	↓ 32.95	↓ 42.73	↓ 35.00
ABW6	32.64	33.28	52.45
ABW7	↑ 23.36	22.80	↑ 26.54
ABW8	31.65	35.39	48.69
ABW9	30.67	32.36	↑ 34.17

[00230] **Table 6:** ***Percent identity between amino acid sequences of engineered novel ABW nucleases and native Cas12a nucleases.*** Percent identity between sequences was assessed using alignment and pairwise comparison in CLC Main Workbench 7.9.1. NCBI references are provided for each Cas12a sequence.

	Percent Identity Amino Acid Between Sequences											
	AsCpf1	**FnCpf1**	**EeCpf1**	**ABW1**	**ABW2**	**ABW3**	**ABW4**	**ABW5**	**ABW6**	**ABW7**	**ABW8**	**ABW9**
AsCpf1 (WP_021736722.1)	100.00											
FnCpf1 (WP_003040289.1)	29.15	100.00										
EeCpf1 (WP_055225123.1)	29.04	31.22	100.00									
ABW1	46.65	29.27	27.38	100.0								
ABW2	28.54	32.33	25.07	32.98	100.0							
ABW3	27.72	39.56	29.02	31.86	36.65	100.0						

ABW4	27.55	37.89	27.10	33.42	37.38	48.89	100.0					
ABW5	26.82	40.30	28.37	30.59	35.79	55.64	48.25	100.00				
ABW6	27.30	26.82	48.89	30.87	28.07	31.10	29.28	29.74	100.0			
ABW7	16.12	16.04	16.53	20.68	20.36	20.11	19.91	19.22	20.23	100.0		
ABW8	26.71	28.37	45.97	31.51	29.06	32.37	30.90	32.57	46.02	20.03	100.0	
ABW9	23.50	27.02	22.23	27.39	30.67	29.61	34.31	30.84	25.47	16.48	26.07	100.00

[00231] The nucleotide sequences of the engineered ABW1-ABW9 nucleases were also compared to nucleotide sequences of two engineered control nucleases: Cas12a (Cpf1) and MAD7 (positive control). Sequences of engineered AsCpf1 and FnCpf1 was obtained from Zetsche *et al.* (2015) Cell.;163(3):759-71, the disclosure of which is incorporated herein. The results are provided in **Table 7** below:

[00232] **Table 7** represents the percent identity between nucleotide sequences of engineered ABW nucleases and engineered Cas12a nucleases. Percent identity was assessed by alignment and pairwise comparison in CLC Main Workbench 7.9.1.

	Percent Identity Between Nucleotide Sequences											
	AsCpf1	**FnCpf1**	**MAD7**	**ABW1**	**ABW2**	**ABW3**	**ABW4**	**ABW5**	**ABW6**	**ABW7**	**ABW8**	**ABW9**
AsCpf1	100.00											
FnCpf1	51.08	100.00										
Positive Control	39.44	37.68	100.00									
ABW1	43.19	51.02	36.66	100.00								
ABW2	40.44	37.72	34.55	40.49	100.00							
ABW3	41.34	37.38	36.68	39.59	45.57	100.00						
ABW4	42.05	38.11	36.79	40.95	47.66	53.49	100.00					
ABW5	41.37	36.96	36.69	39.12	45.57	57.06	52.96	100.00				
ABW6	41.39	39.04	47.21	40.64	38.35	37.95	39.45	38.27	100.00			
ABW7	33.27	31.99	30.78	34.30	33.80	34.90	33.65	34.21	35.00	100.00		
ABW8	41.05	38.80	46.36	40.82	39.60	39.96	39.76	40.71	54.90	35.02	100.00	
ABW9	35.17	32.86	32.64	34.65	34.58	36.13	37.96	35.55	35.22	28.41	36.77	100.00

Example 2

[00233] In some methods, codon optimization, as described in Example 1, can lower nucleotide sequence similarity in most cases; however, it does not change the amino acid sequence of the protein. Further engineering was applied to sequences to improve the activity of the nucleases outside their native context. The native sequences of nine type V CRISPR-associated protein Cas12a/Cpf1 (ABW) nucleases were engineered to include glycine, 6x Histidine, and 3x nuclear localization signal tags.

[00234] These Gly-6xHis tag were applied for several reasons including: 1) a 6xHis tag can be used in protein purification to allow binding to the chromatographic columns for purification, and 2) the N-terminal glycine allows further, site-specific, chemical modifications that permit advanced protein engineering. Further, the Gly-6xHis was designed for easy removal, if desired, by digestion with Tobacco Etch Virus (TEV) protease. For these constructs, the Gly-6xHis tag was positioned on the N-terminus. Gly-6xHis tags are further described in Martos-Maldonado et al., *Nat Commun.* (2018) 17;9(1):3307, the disclosure of which is incorporated herein by reference.

[00235] The NLS (Nuclear Localization Signal) fragments were added to improve transport to the nucleus. NLS fragments used in these examples were successfully added to Cas9 constructs, as previously described.

[00236] Using the engineered ABW nuclease sequence, at least 10 variants were developed for each of the nine engineered ABW nucleases. The nucleotide sequence of each ABW engineered novel variant was compared to the corresponding ABW engineered nucleotide sequence. Exemplary sequence comparisons are provided in **Tables 8-16** below. Note that that the sequences provided in Tables 7-15 do not exhaust all possible sequences as only 10 variants were selected for each ABW nuclease.

[00237] **Table 8** represents the percent identity between nucleotide sequences of engineered ABW1 nuclease and engineered ABW1 nuclease variants 2-10. The percent identity between sequences illustrates resulted from alignment and pairwise comparison in CLC Main Workbench 7.9.1.

	Percent Identity Between Nucleotide Sequences of ABW1 Engineered Variants									
	ABW1 (SEQ ID NO: 4)	**Variant 2** (SEQ ID NO: 5)	**Variant 3** (SEQ ID NO: 6)	**Variant 4** (SEQ ID NO: 7)	**Variant 5** (SEQ ID NO: 8)	**Variant 6** (SEQ ID NO: 9)	**Variant 7** (SEQ ID NO: 10)	**Variant 8** (SEQ ID NO: 11)	**Variant 9** (SEQ ID NO: 12)	**Variant 10** (SEQ ID NO: 13)
variant #6	100.00									

variant #10	78.94	100.00								
variant #3	78.99	78.84	100.00							
ABW1	78.53	78.62	78.53	100.00						
variant #8	78.84	77.57	79.01	78.84	100.00					
variant #9	79.23	78.50	78.77	78.94	78.65	100.00				
variant #5	78.57	78.11	77.72	78.28	78.57	78.84	100.00			
variant #2	78.28	78.21	78.97	78.79	78.53	78.75	78.87	100.00		
variant #4	78.84	78.36	77.84	79.14	78.67	78.38	78.31	79.40	100.00	
variant #7	78.28	78.36	79.09	79.11	78.06	79.04	78.48	78.31	78.16	100.00

[00238] **Table 9** represents the percent identity between nucleotide sequences of engineered ABW2 nuclease and engineered ABW2 nuclease variants 2-10. Percent identity between sequences is illustrated from alignment and pairwise comparison in CLC Main Workbench 7.9.1.

	Percent Identity Between Nucleotide Sequences of ABW2 Engineered Variants									
	ABW2 (SEQ ID NO: 17)	**Variant 2** (SEQ ID NO: 18)	**Variant 3** (SEQ ID NO: 19)	**Variant 4** (SEQ ID NO: 20)	**Variant 5** (SEQ ID NO: 21)	**Variant 6** (SEQ ID NO: 22)	**Variant 7** (SEQ ID NO: 23)	**Variant 8** (SEQ ID NO: 24)	**Variant 9** (SEQ ID NO: 25)	**Variant 10** (SEQ ID NO: 26)
variant #8	100.00									
variant #9	79.84	100.00								
ABW2	79.23	77.97	100.00							
variant #10	79.60	78.41	78.51	100.00						
variant #5	78.63	78.19	78.51	78.89	100.00					
variant #3	78.94	78.24	78.04	78.26	79.04	100.00				
variant #4	78.92	78.17	79.14	78.85	78.92	78.99	100.00			
variant #6	78.68	78.43	78.55	78.48	78.14	79.31	78.48	100.00		
variant #2	78.53	77.87	78.02	78.75	78.26	78.68	78.41	79.26	100.00	
variant #7	78.34	78.07	78.85	78.31	78.85	78.63	78.85	79.18	77.90	100.00

[00239] **Table 10** represents the percent identity between nucleotide sequences of engineered ABW3 nuclease and engineered ABW3 nuclease variants 2-10. Percent identity between sequences is illustrated from alignment and pairwise comparison in CLC Main Workbench 7.9.1.

	Percent Identity Between Nucleotide Sequences of ABW3 Engineered Variants									
	ABW3 (SEQ ID NO: 30)	**Variant 2** (SEQ ID NO: 31)	**Variant 3** (SEQ ID NO: 32)	**Variant 4** (SEQ ID NO: 33)	**Variant 5** (SEQ ID NO: 34)	**Variant 6** (SEQ ID NO: 35)	**Variant 7** (SEQ ID NO: 36)	**Variant 8** (SEQ ID NO: 37)	**Variant 9** (SEQ ID NO: 38)	**Variant 10** (SEQ ID NO: 39)
variant #8	100.00									
variant #10	79.00	100.00								
variant #6	77.73	79.20	100.00							
variant #4	78.31	78.06	78.54	100.00						
variant #3	78.13	77.93	79.60	78.82	100.00					
ABW3	78.49	77.14	77.70	78.13	78.13	100.00				
variant #7	79.48	78.61	78.59	78.39	78.29	79.05	100.00			
variant #2	78.61	78.44	78.31	79.25	78.08	78.44	78.06	100.00		
variant #5	78.59	77.90	77.78	77.32	78.23	78.36	78.46	77.75	100.00	
variant #9	78.69	78.56	78.34	77.45	78.41	78.29	78.64	78.46	79.38	100.00

[00240] **Table 11** represents the percent identity between nucleotide sequences of engineered ABW4 nuclease and engineered ABW4 nuclease variants 2-10. Percent identity between sequences is illustrated from alignment and pairwise comparison in CLC Main Workbench 7.9.1.

	Percent Identity Between Nucleotide Sequences of ABW4 Engineered Variants									
	ABW4 (SEQ ID NO: 43)	**Variant 2** (SEQ ID NO: 44)	**Variant 3** (SEQ ID NO: 45)	**Variant 4** (SEQ ID NO: 46)	**Variant 5** (SEQ ID NO: 47)	**Variant 6** (SEQ ID NO: 48)	**Variant 7** (SEQ ID NO: 49)	**Variant 8** (SEQ ID NO: 50)	**Variant 9** (SEQ ID NO: 51)	**Variant 10** (SEQ ID NO: 52)
variant #2	100.00									
variant #6	79.57	100.00								
variant #5	79.35	80.08	100.00							
variant #4	80.01	79.59	79.01	100.00						
variant #9	79.74	79.08	79.59	78.49	100.00					
ABW4	79.03	78.74	78.86	78.93	78.91	100.00				
variant #7	79.23	78.54	79.20	79.11	79.67	79.23	100.00			
variant #3	79.20	79.35	79.08	78.93	78.64	79.35	78.74	100.00		
variant #10	78.98	79.18	79.55	79.57	79.40	78.98	78.59	78.91	100.00	

variant #8	79.37	78.79	79.33	78.89	78.89	79.25	78.57	78.62	78.98	100.00

[00241] **Table 12** represents the percent identity between nucleotide sequences of engineered ABW5 nuclease and engineered ABW5 nuclease variants 2-10. Percent identity between sequences is illustrated from alignment and pairwise comparison in CLC Main Workbench 7.9.1.

	Percent Identity Between Nucleotide Sequences of ABW5 Engineered Variants									
	ABW5 (SEQ ID NO: 56)	**Variant 2** (SEQ ID NO: 57)	**Variant 3** (SEQ ID NO: 58)	**Variant 4** (SEQ ID NO: 59)	**Variant 5** (SEQ ID NO: 60)	**Variant 6** (SEQ ID NO: 61)	**Variant 7** (SEQ ID NO: 62)	**Variant 8** (SEQ ID NO: 63)	**Variant 9** (SEQ ID NO: 64)	**Variant 10** (SEQ ID NO: 65)
variant #3	100.00									
variant #5	79.43	100.00								
variant #8	79.15	79.58	100.00							
variant #4	78.75	78.85	79.33	100.00						
variant #6	78.72	79.36	79.26	79.03	100.00					
variant #10	79.23	78.85	79.48	79.41	79.79	100.00				
variant #2	78.77	78.29	78.19	78.98	79.56	78.57	100.00			
ABW5	77.89	77.58	78.95	78.34	78.65	77.36	79.1	100.00		
variant #7	79.18	77.71	78.88	78.9	78.55	78.29	79.13	78.95	100.00	
variant #9	78.93	78.44	78.42	78.34	79.41	79.13	78.57	78.88	79.38	100.00

[00242] **Table 13** represents the percent identity between nucleotide sequences of engineered ABW6 nuclease and engineered ABW6 nuclease variants 2-10. Percent identity between sequences is illustrated from alignment and pairwise comparison in CLC Main Workbench 7.9.1.

	Percent Identity Between Nucleotide Sequences of ABW6 Engineered Variants									
	ABW6 (SEQ ID NO: 69)	**Variant 2** (SEQ ID NO: 70)	**Variant 3** (SEQ ID NO: 71)	**Variant 4** (SEQ ID NO: 72)	**Variant 5** (SEQ ID NO: 73)	**Variant 6** (SEQ ID NO: 74)	**Variant 7** (SEQ ID NO: 75)	**Variant 8** (SEQ ID NO: 76)	**Variant 9** (SEQ ID NO: 77)	**Variant 10** (SEQ ID NO: 78)
variant #5	100.00									
variant #2	79.88	100.00								
variant #4	79.60	79.88	100.00							
ABW6	79.88	79.03	78.98	100.00						

variant #6	79.35	79.13	79.38	78.50	100.00					
variant #10	79.28	79.00	78.50	79.03	79.85	100.00				
variant #7	79.25	78.68	78.80	79.18	79.13	79.58	100.00			
variant #3	77.55	79.38	79.73	79.20	79.08	78.13	78.35	100.00		
variant #8	78.65	78.53	79.20	77.95	77.88	78.10	78.43	78.78	100.00	
variant #9	78.88	79.28	79.50	79.00	79.83	78.70	78.48	78.78	79.20	100.00

[00243] **Table 14** represents the percent identity between nucleotide sequences of engineered ABW7 nuclease and engineered ABW7 nuclease variants 2-10. Percent identity between sequences is illustrated from alignment and pairwise comparison in CLC Main Workbench 7.9.1.

	Percent Identity Between Nucleotide Sequences of ABW7 Engineered Variants									
	ABW7 (SEQ ID NO: 82)	**Variant 2** (SEQ ID NO: 83)	**Variant 3** (SEQ ID NO: 84)	**Variant 4** (SEQ ID NO: 85)	**Variant 5** (SEQ ID NO: 86)	**Variant 6** (SEQ ID NO: 87)	**Variant 7** (SEQ ID NO: 88)	**Variant 8** (SEQ ID NO: 89)	**Variant 9** (SEQ ID NO: 90)	**Variant 10** (SEQ ID NO: 91)
variant #2	100.00									
variant #8	79.80	100.00								
variant #7	78.01	78.70	100.00							
variant #4	78.34	77.68	78.6	100.00						
variant #9	78.47	78.93	78.34	78.24	100.00					
variant #6	77.85	78.32	77.85	78.39	78.47	100.00				
variant #5	78.91	78.11	78.16	79.34	78.65	78.52	100.00			
variant #3	78.32	77.80	78.03	78.75	77.75	78.14	78.68	100.00		
variant #10	77.24	78.27	77.93	77.78	78.09	78.24	78.47	78.27	100.00	
ABW7	78.27	76.98	77.88	77.83	77.44	78.09	77.85	78.65	76.98	100.00

[00244] **Table 15** represents the percent identity between nucleotide sequences of engineered ABW8 nuclease and engineered ABW8 nuclease variants 2-10. Percent identity between sequences is illustrated from alignment and pairwise comparison in CLC Main Workbench 7.9.1.

	Percent Identity Between Nucleotide Sequences of ABW8 Engineered Variants									
	ABW8 (SEQ ID NO: 95)	**Variant 2**	**Variant 3**	**Variant 4**	**Variant 5**	**Variant 6**	**Variant 7**	**Variant 8**	**Variant 9**	**Variant 10**

		(SEQ ID NO: 96)	(SEQ ID NO: 97)	(SEQ ID NO: 98)	(SEQ ID NO: 99)	(SEQ ID NO: 100)	(SEQ ID NO: 101)	(SEQ ID NO: 102)	(SEQ ID NO: 103)	(SEQ ID NO: 104)
ABW8	100.00									
variant #6	79.64	100.00								
variant #3	79.27	79.32	100.00							
variant #10	78.39	78.54	79.15	100.00						
variant #8	78.52	79.66	79.73	78.91	100.00					
variant #9	78.83	79.64	79.37	79.46	79.59	100.00				
variant #7	79.32	78.95	79.46	77.98	79.93	78.47	100.00			
variant #2	78.81	79.32	79.17	78.32	79.08	79.68	79.32	100.00		
variant #4	79.03	79.56	78.91	78.20	79.15	79.34	79.64	80.02	100.00	
variant #5	78.73	79.42	78.59	78.86	79.76	79.98	79.15	78.81	78.35	100.00

[00245] **Table 16** represents the percent identity between nucleotide sequences of engineered ABW9 nuclease and engineered ABW9 nuclease variants 2-10. Percent identity between sequences is illustrated from alignment and pairwise comparison in CLC Main Workbench 7.9.1.

	Percent Identity Between Nucleotide Sequences of ABW9 Engineered Variants									
	ABW9 (SEQ ID NO: 108)	**Variant 2** (SEQ ID NO: 109)	**Variant 3** (SEQ ID NO: 110)	**Variant 4** (SEQ ID NO: 111)	**Variant 5** (SEQ ID NO: 112)	**Variant 6** (SEQ ID NO: 113)	**Variant 7** (SEQ ID NO: 114)	**Variant 8** (SEQ ID NO: 115)	**Variant 9** (SEQ ID NO: 116)	**Variant 10** (SEQ ID NO: 117)
variant #3	100.00									
variant #4	78.96	100.00								
variant #7	78.59	78.02	100.00							
variant #9	78.50	78.02	78.56	100.00						
variant #5	77.21	78.32	77.14	77.67	100.00					
variant #10	77.71	77.65	77.54	78.13	78.61	100.00				
variant #2	77.80	76.58	78.32	77.17	77.69	77.54	100.00			
variant #8	77.10	78.37	77.28	78.13	78.26	77.69	77.91	100.00		
variant #6	77.28	77.78	77.14	77.04	77.62	77.69	77.76	77.08	100.00	
ABW9	75.94	76.27	75.90	75.51	75.55	75.07	76.58	76.34	75.88	100.00

Example 3

[00246] In another exemplary method, it is understood that a CRISPR-Cas genome editing system requires at least 2 components: a guide RNA (gRNA) and CRISPR-associated (Cas) nuclease. Guide RNA is a specific RNA sequence that recognizes the targeted DNA region of interest and directs the Cas nuclease to this region for editing. gRNA is made up of two parts: crispr RNA (crRNA), a 17-20 nucleotide sequence complementary to the target DNA, and a tracr RNA, which serves as a binding scaffold for the Cas nuclease in order to facilitate editing. The crRNA part of the gRNA is customizable and this feature enables specificity in every CRISPR experiment. In one method, predicted crRNA sequence of the gRNA for nucleases ABW1-ABW9, MAD7 (positive control), and AsCas12a are provided in **Table 17** below:

Table 17: ***Predicted crRNA Sequences***

	Organism of origin	**predicted crRNA_sequence**	**Spacer length**	**cRNA length**	**SEQ ID NO:**
ABW1	Acidaminococcus massiliensis Marseille-P2828	GUCUAAAAGACCAUAUGAAUUUCUACUUUCGUAGAUN NNNNNNNNNNNNNNNNNNNNNNNNNN	28	36	129
ABW2	Sedimentisphaera cyanobacteriorum strain L21-RPul-D3	GUCUAAAGGCCUUAUAAAAUUUCUACUGUCGUAGAUN NNNNNNNNNNNNNNNNNNNNNNNNNN	27	36	130
ABW3	Barnesiella sp. An22	GUCUAUACAGACACUUUAAUUUCUACUAUUGUAGAUN NNNNNNNNNNNNNNNNNNNNNNNNNN	28	36	131
ABW4	Bacteroidetes bacterium HGW-Bacteroidetes-6	GUCUGAAAGACAAGUAUAAUUUCUACUAUUGUAGAUN NNNNNNNNNNNNNNNNNNNNNNNNNN	27	36	132
ABW5	Parabacteroides distasonis strain 8-P5	GGCUAUAAGCCUUGUAUAAUUUCUACUAUUGUAGAUN NNNNNNNNNNNNNNNNNNNNNNNNNN	27	36	133
ABW6	Collinsella tanakaei	GUUGAAACUGUAAGCGGAAUGUCUACUUGGGUAGAUN NNNNNNNNNNNNNNNNNNNNNNNNNN	27	36	134
ABW7	Lachnospiraceae bacterium MC2017	GCAUGAGAACCAUGCAUUUCUAAGGUACUCCAAAACN NNNNNNNNNNNNNNNNNNNNNNNNNN	29	36	135
ABW8	Coprococcu ssp. AF16-5	GUUGAGUAACCUUAAAUAAUUUCUACUGUUGUAGAUN NNNNNNNNNNNNNNNNNNNNNNNNNN	26	36	136
ABW9	Catenovulum sp. CCB-QB4	AUCUACAACAGUAGAAAUUUAAGCUAAGGCUUAGACN NNNNNNNNNNNNNNNNNNNNNNNNNN	27	36	137
MAD7	Eubacterium rectale	GUCAAAAGACCUUUUUAAUUUCUACUCUUGUAGAUNN NNNNNNNNNNNNNNNNNNNN	21	35	138
AsCpf1	Acidaminococcus sp. BV3L6	UAAUUUCUACUCUUGUAGAUNNNNNNNNNNNNNNNNN NNNNNNN	24	20	139

Example 4

[00247] In another exemplary method, cleavage efficiency of ABW nucleases was tested *in vitro*. As efficacy of *in vitro* cleavage efficiency is a predictor of *in vivo* cleavage, it is important prior to testing the ABW nucleases to determine which nucleases would be predicted to be the more effective prior to delivering to the nucleases to test in cells.

[00248] In one exemplary method, to prepare partially cognate DNA substrates for the *in vitro* cleavage assay, DNMT1 target sequences and partially cognate target sequences were cloned to a plasmid, pRG2 plasmid. Before testing for *in vitro* DNA cleavage, the target plasmids were linearized and purified. An aliquot of linearized products was incubated with purified ABW1, ABW2, ABW3, ABW4, ABW5, ABW6, ABW7, ABW8, or ABW9 and 35 ng (165.3 nM) DNMT1 in combination with predicted-cognate gRNAs, Cas12a gRNA, or as referenced herein, "split" gRNA prepared using STAR. After incubation, products were loaded in 1.5% agarose gel for analysis. The sequences of gRNAs used the DNMT1 *in vitro* cleavage assays are provided in **FIG. 2**. Images of the 1.5% agarose gels illustrating DNMT1 cleavage are provided in **FIGS. 3-5**. Data illustrated in these figures is provided as an overview in **Table 18** below.

[00249] These experiments indicate that ABW nucleases: ABW1, ABW2, ABW3, ABW4, ABW5, and ABW8 effectively cleaved the gRNAs tested. ABW9 cleaved only Cas12a gRNA whereas ABW6 and ABW7 failed to cleave any of the gRNAs tested.

[00250] CRISPR nucleases, in general, can differ in properties such as activity and specificity. Although ABW6 and ABW7 failed to demonstrate activity in the *in vitro* experiment, the nucleases behave differently under different conditions but retain genome editing properties in other settings.

[00251] Table 18: ***DNMT1 Amplicon in vitro Cleavage Assay Overview***

Nuclease (engineered and synthetized variant)	**gRNAs**		
	predicted – cognate gRNA	**Cas12a gRNA** (SEQ ID NO: 127)	**STAR gRNA** (SEQ ID NO: 128)
ABW1	✔	✔	✔
ABW2	✔	✔	✔
ABW3	✔	✔	✔
ABW4	✔	✔	✔
ABW5	✔	✔	✔
ABW6			

ABW7			
ABW8	✔	✔	✔
ABW9		✔	

[00252] In another exemplary method, the *in vitro* DNA cleavage assay was repeated using a time-course assay using a known nuclease as the active nuclease reference. The pRG2 plasmid having the DNMT1 target sequence was linearized and purified before testing. An aliquot of linearized products was incubated with a control gRNA (UAAUUUCUACUCUUGUAGAUCUGAUGGUCCAUGUCUGUUA; SEQ ID NO: 149) and one of the following purified nucleases: Cas12a Ultra, LbaCas12a, control nuclease (MAD7), ABW1, ABW5, ABW8, M21, M44. The ratio of nuclease:gRNA:target DNA per incubation is provided in **Table 19**.

[00253] **Table 19** represents the ratio of nuclease: gRNA: target DNA per assay incubation.

Nuclease (Cas12a Ultra, LbaCas12a, MAD7, ABW1, ABW5, ABW8, M21, or M44)	**gRNA** (SEQ ID NO: 149)	**target DNA** (DNMT1)
20	60	1
10	30	1
5	15	1
2.5	7.5	1
1.25	3.75	1
0.625	1.875	1

[00254] After incubation, products were loaded in 1.5% agarose gel for analysis and separation. Exemplary results are illustrated in **FIGS. 6A and 6B**.

Example 5

[00255] In another exemplary method, cleavage efficiency of ABW nucleases was tested *in vivo* in an exemplary eukaryotic cell population. Using these methods, the assay is based on the *in vitro* DNA cleavage assay. Jurkat cells, an acceptable immortalized line of human T lymphocyte cells, were cultivated under exemplary conditions in RPMI 1640 media with 10% Fetal Bovine Serum (FBS) and split regularly before being harvested for the transfection. Two target loci, DNMT1 and TRAC43, were chosen in genomic Jurkat's DNA as targets. Nucleases ABW1, 2, 3, 4, 5, 8, and control nuclease were diluted in the storage buffer (*e.g.* NaCl 300 mM, Na-phosphate 50 mM, EDTA 0.1 mM, DTT 1 mM, and glycerol 10%) to 20 mg/mL.

Analogically, the gRNAs were diluted in the nuclease-free water to 100 μM. The RNA-protein complexes (RNPs) were prepared by mixing 1 μL nuclease solution and 1.5 μL gRNA solution. Complexes were formed in 96-well V-bottom plate during 10 minute incubations at room temperature.

[00256] Cells were counted and their viability was estimated in the NucleoCounter NC-200. Harvested cells were resuspended in the transfection buffer (SF from SF Cell Line 96-well Nucleofector Kit, Lonza) at 100 x 10^5 cells/mL concentration. 20 μL of that solution was added to the well with formed RNPs, mixed by pipetting, and transferred to 96-well Nucleocuvette plate (Lonza). Cells were electroporated. 80 μL of fresh RPMI 1640 media with 10% FBS were added to the Nucleocuvette plate immediately after the electroporation. The solution was mixed and 50 μL was transferred to the 96-well flat-bottom cultivation plate with 150 μL of fresh media. Cells were cultivated for 72 hours before being harvested for DNA extraction.

[00257] Cells were harvested by centrifugation 1000 x g for 10 minutes and washed with buffer (PBS). The supernatant was carefully removed, and the cell pellet was treated with 20 μL preheated QuickExtract DNA Extraction Solution (Lucigen). The plate was placed in the thermocycler (Biorad) and the temperature treatment (*e.g.* 15 minutes at 65°C, 15 minutes at 68°C, 10 minutes at 95°C, cold down to 4°C) was applied. Cell debris was harvested by centrifugation, and the supernatant containing genomic DNA was collected. DNA fragments containing target sites were amplified in the PCR reaction and DNA was prepared for sequencing. The Next Generation Sequencing (NGS) data are presented in **FIGS. 7A-7C**.

Example 6

[00258] In another exemplary method, a T7 assay was performed on the genomic DNA from Jurkat cells. The T7 endonuclease catalyzes cleavage DNA mismatches and non-β DNA structures like junctions and cruciform. When DNA cleavage occurs, random nucleotides are inserted or deleted, and a double-strand break is repaired. Thermal denaturation causes separation of strands and cooling down allows a DNA duplex to reassemble. If the edited strand reassembles with an unedited strand, a mismatch(s) appears. T7 endonuclease cleaves the mismatch, and DNA fragments can be visualized on the agarose gel in order to verify the process. In these examples, the targeted DNA was amplified in a PCR reaction. PCR products were purified and temperature treated (*e.g.* 10 min at 95°C, gradual cooling to 85°C - 5 cycles, -2°C per cycle, gradual cooling to 25°C – 200 cycles, -0.3°C per cycle) to create the

heteroduplexes with mismatches. DNA was divided into two tubes, and one aliquot was treated with T7 endonuclease. Both DNA samples were analyzed on an agarose gel.

[00259] It was observed that using these exemplary conditions ABW1, ABW2, ABW3, ABW4, ABW5, and ABW8 demonstrated editing of the DNMT1 gene in Jurkat cells (**FIG. 8**), and ABW1, ABW2, ABW3, ABW4 and ABW8 demonstrated TRAC gene cleavage (**FIG. 9**).

Example 7

[00260] In another exemplary method, cleavage efficiency of ABW nucleases was tested *in vitro*. As efficacy of *in vitro* cleavage efficiency is a predictor of *in vivo* cleavage, it is important prior to testing the ART nucleases to determine which nucleases would be predicted to be the more effective prior to delivering to the nucleases to test in cells.

[00261] In some exemplary methods, cleavage efficiency of ABW nucleases was tested *in vivo* in *Escherichia coli* (*E. coli*). In these methods, the assay was based on *in vivo* depletion assay in *E. coli*. First, a glycerol stock of *E. coli* MG1655 harboring a plasmid that expresses the ART nuclease was removed from -80°C freezer and take 20 μL cells into 2 of 4 mL LB (lysogeny broth) medium with 34 μg/mL chloramphenicol in 15 mL tubes. The cells were cultured at 30°C and 200 rpm for overnight. Then, 4 mL overnight culture was put into 200 mL LB medium with 34 μg/mL chloramphenicol into 2 of 1 L flasks The cells were cultured at 30°C and 200 rpm until OD_{600} reached 0.5-0.6. The flasks were put into a shaking water bath incubator at 42°C and 200 rpm for 15 minutes. Then, the flasks were put in the ice with manually slow shaking and were kept in the ice for 15 minutes. After that, the cells were transferred from flasks to 50 mL tubes (4 tubes for 200 mL cells) and centrifuged at 8000 rpm and 4°C for 5 minutes to remove supernatant. Then, 50 mL ice-cold 10% glycerol were added for 200 mL culture and the cells were resuspended. The resuspended cells were centrifuged at 8000 rpm and 4°C for 5 minutes to remove supernatant and 2 mL ice-cold 10% glycerol was added. Cells were resuspended with pipette gently and divided into 50 μL of the competent cells. The mixtures was then aliquoted into 72 chilled 0.1 cm electroporation cuvettes (Bio-rad).

[00262] The 24 gRNAs and one non-targeting control gRNA were diluted in the nuclease-free water to 25 ng/ul. gRNA_EC1 to gRNA_EC23 were targeted 18 target loci which are *galK*, *lpd*, *accA*, *cynT*, *cynS*, *adhE*, *oppA*, *fabI*, *ldhA*, *pntA*, *pta*, *accD*, *pheA*, *accB*, *accC*, *aroE*, *aroB*, and *aroK* genes. 2 μL (50 ng) chilled plasmids were put into the electroporation cuvettes and the electroporation were done at 1800 V. Then, 950 μL LB medium were added into the cuvette and

mixed, then the cells were taken out into a 96-deep well plate (Light labs). The 96-deep well plate with cells were put at 30°C and 200 rpm for 2 hours.

[00263] Dilutions were made at 10^0, 10^1, and 10^2 for the recovered cells after 2 hours of culture. Then, 10 µL of cells were put into 90 µL ddH_2O and mixed with pipette. After dilution, 8 µL of cells were taken from each dilution and placed by pipette onto a LB agar plate 34 µg/mL chloramphenicol and 100 µg/mL carbenicillin and allowed to dry without covers for several minutes. Then the covers were put back onto the plates and the plates were returned to culture at 30°C for overnight. The next day, results were checked by counting the number of colonies.

[00264] Exemplary depletion assay outcomes using constructs disclosed herein, ABW1, ABW2, ABW3, ABW4, ABW5, ABW6, ABW7, ABW8, and ABW9 are provided in **FIGS. 10-18** where the data depict percent cutting efficiency = 1- (# of colonies plate with on-target gRNA/ # of colonies on plate with non-target gRNA)*100% . In this example, ABW8 had the highest microbial activity of the nucleases tested followed by ABW3, then ABW7, then ABW9 compared to the remaining nucleases which showed some activity (*i.e.*, ABW8>ABW3>ABW7>ABW9> ABW1, ABW2, ABW4, ABW5, and ABW6).

Example 8

[00265] In another exemplary method, ribonucleoproteins (RNPs) are produced by complexing of a single gRNA or STAR gRNA with nucleases disclosed herein (*e.g.*, ABW nucleases but can be other nucleases). Single or STAR gRNAs are synthesized as described herein. Recombinant ABW are produced by expression of a E. coli codon optimized and 6XHis-tagged ABW nuclease in E. coli and purified by standard methods. Recombinant ABW nuclease is stored in a 25 mM Tris-HCl pH 7.4, 300 mM NaCl, 0.1 mM EDTA, 1 mM DTT, and 50% (v/v) glycerol buffer at -80°C prior to use. Single gRNAs or STAR gRNAs are resuspended in IDTE (10 mM Tris, 0.1 mM EDTA) pH 7.5 buffer to produce a 100 µM stock and stored at -80°C prior to use. Just before nucleoporation, recombinant ABW is diluted in a working buffer consisting of 20 mM HEPES and 150 mM KCl pH 7.5 and gRNAs are diluted to a final working concentration with IDTE pH 7.5 buffer (annealed first if STAR). Following dilutions of the ABW nuclease and gRNA, both are then mixed 1:1 by volume (2:1 gRNA to nuclease ratio) at 37°C for 10 minutes to form RNPs. Following complexing, RNPs are resuspended in the appropriate nucleoporation buffer and delivered via an optimized nucleoporator program and assessed.

****

The foregoing discussion of the disclosure has been presented for purposes of illustration and description. The foregoing is not intended to limit the disclosure to the form or forms disclosed herein. Although the description of the disclosure has included description of one or more embodiments and certain variations and modifications, other variations and modifications are within the scope of the disclosure, *e.g.*, as can be within the skill and knowledge of those in the art, after understanding the present disclosure. It is intended to obtain rights which include alternative embodiments to the extent permitted, including alternate, interchangeable and/or equivalent structures, functions, ranges or steps to those claimed, whether or not such alternate, interchangeable and/or equivalent structures, functions, ranges or steps are disclosed herein, and without intending to publicly dedicate any patentable subject matter.

WHAT IS CLAIMED IS:

1. A nucleic acid-guided nuclease system comprising:

(a) an engineered nucleic acid-guided nuclease, wherein the engineered nucleic acid-guided nuclease comprises a polypeptide sequence comprising at least 85% homology to the polypeptide represented by SEQ ID NO: 94 (ABW8), 29 (ABW3), 81 (ABW7), 107 (ABW9), 3 (ABW1), 16 (ABW2), 42 (ABW4), 55 (ABW5), and 68 (ABW6); or at least 85% homology to the polynucleotide encoding a polypeptide comprising a polynucleotide represented by SEQ ID NO: 95-104 (ABW8 variants 1-10), 30-39 (ABW3 variants 1-10), 82-91 (ABW7 variants 1-10), 108-117 (ABW9 variants 1-10), 4-13 (ABW1 variants 1-10), 17-26 (ABW2 variants 1-10), 43-52 (ABW4 variants 1-10), 56-65 (ABW5 variants 1-10), and 69-78 (ABW6 variants 1-10); and

(b) a guide polynucleotide for complexing with the nucleic acid-guided nuclease.

2. The system according to claim 1, wherein the target region is within a eukaryotic cell.

3. The system according to claim 1, wherein the target region is within a bacterial cell.

4. The system according to claim 1, wherein the target region is within a plant cell.

5. The system according to claim 1, wherein the target region is within a mammalian cell.

6. The system according to any one of claims 1-5, wherein the gRNA comprises STAR gRNA.

7. The system according to any one of claims 1-5, wherein the gRNA comprises split gRNA.

8. A nucleic acid-guided nuclease system comprising:

(a) a nucleic acid-guided nuclease; and

(b) an engineered guide polynucleotide (gRNA) for complexing with the nucleic acid-guided nuclease, wherein the engineered guide polynucleotide comprises a polynucleotide represented by SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127, or 128.

9. The system according to claim 8, wherein the target region is within a eukaryotic cell.

10. The system according to claim 8, wherein the target region is within a bacterial cell.

11. The system according to claim 8, wherein the target region is within a plant cell.

12. The system according to claim 8, wherein the target region is within a mammalian cell.

13. The system according to claim 8, wherein the target region is within a human cell.

14. The system according to any one of claims 8-13, wherein the gRNA comprises STAR gRNA.

15. The system according to any one of claims 8-13, wherein the gRNA comprises split gRNA.

16. A nucleic acid-guided nuclease system comprising:

(a) an engineered nucleic acid-guided nuclease, wherein the engineered nucleic acid-guided nuclease comprises a polypeptide sequence comprising at least 85% homology to the polypeptide represented by SEQ ID NO: 94 (ABW8), 29 (ABW3), 81 (ABW7), 107 (ABW9), 3 (ABW1), 16 (ABW2), 42 (ABW4), 55 (ABW5), and 68 (ABW6), and

(b) an engineered guide polynucleotide (gRNA) for complexing with the nucleic acid-guided nuclease, wherein the engineered guide polynucleotide comprises a polynucleotide represented by SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 or 128.

17. The system according to claim 16, wherein the target region is within a eukaryotic cell.

18. The system according to claim 16, wherein the target region is within a bacterial cell.

19. The system according to claim 16, wherein the target region is within a plant cell.

20. The system according to claim 16, wherein the target region is within a mammalian cell.

21. A method of modifying a target region, the method comprising:

(a) contacting a sample having a targeted genomic region with:

(i) an engineered nucleic acid-guided nuclease according to claim 1; and

(ii) a guide nucleic acid according to any one of claims 1 or 8;

and

(b) allowing the nuclease and the guide nucleic acid to modify the target region.

22. The method according to claim 21, wherein the sample is further contacted with (iii) an editing sequence having a change in sequence relative to the sequence of the targeted genomic region.

23. The method according to claim 21, wherein modifying the targeted genomic region comprises editing the targeted genomic region.

24. The method according to claim 21, wherein the guide nucleic acid (ii) and the editing sequence (iii) are provided as a single nucleic acid.

25. The method according to claim 21, wherein the editing sequence further comprises a mutation in a protospacer adjacent motif (PAM) site.

26. The method according to claim 21, wherein the target region is within a eukaryotic cell.

27. The method according to claim 21, wherein the target region is within a bacterial cell.

28. The method according to claim 21, wherein the target region is within a plant cell.

29. The method according to claim 21, wherein the target region is within a mammalian cell.

30. A kit comprising:

(a) an engineered nucleic acid-guided nuclease, wherein the engineered nucleic acid-guided nuclease comprises a polypeptide sequence comprising at least 85% homology to the polypeptide represented by SEQ ID NO: 95-104 (ABW8 variants 1-10), 30-39 (ABW3 variants 1-10), 82-91 (ABW7 variants 1-10), 108-117 (ABW9 variants 1-10), 4-13 (ABW1 variants 1-10), 17-26 (ABW2 variants 1-10), 43-52 (ABW4 variants 1-10), 56-65 (ABW5 variants 1-10), and 69-78 (ABW6 variants 1-10) and, optionally,

(b) an engineered guide polynucleotide for complexing with the nucleic acid-guided nuclease; optionally, wherein the engineered guide polynucleotide comprises a polynucleotide represented by SEQ ID NO: 125, 120, 124, 126, 118, 119, 121, 122, 123, 127 or 128; and

(c) a container.

31. The kit according to claim 30, wherein the kit is of use to edit a prokaryote genome, a plant genome, a eukaryotic genome or yeast genome.

32. The kit according to claim 30, wherein the kit is of use to edit a eukaryotic genome of a human or other mammal.

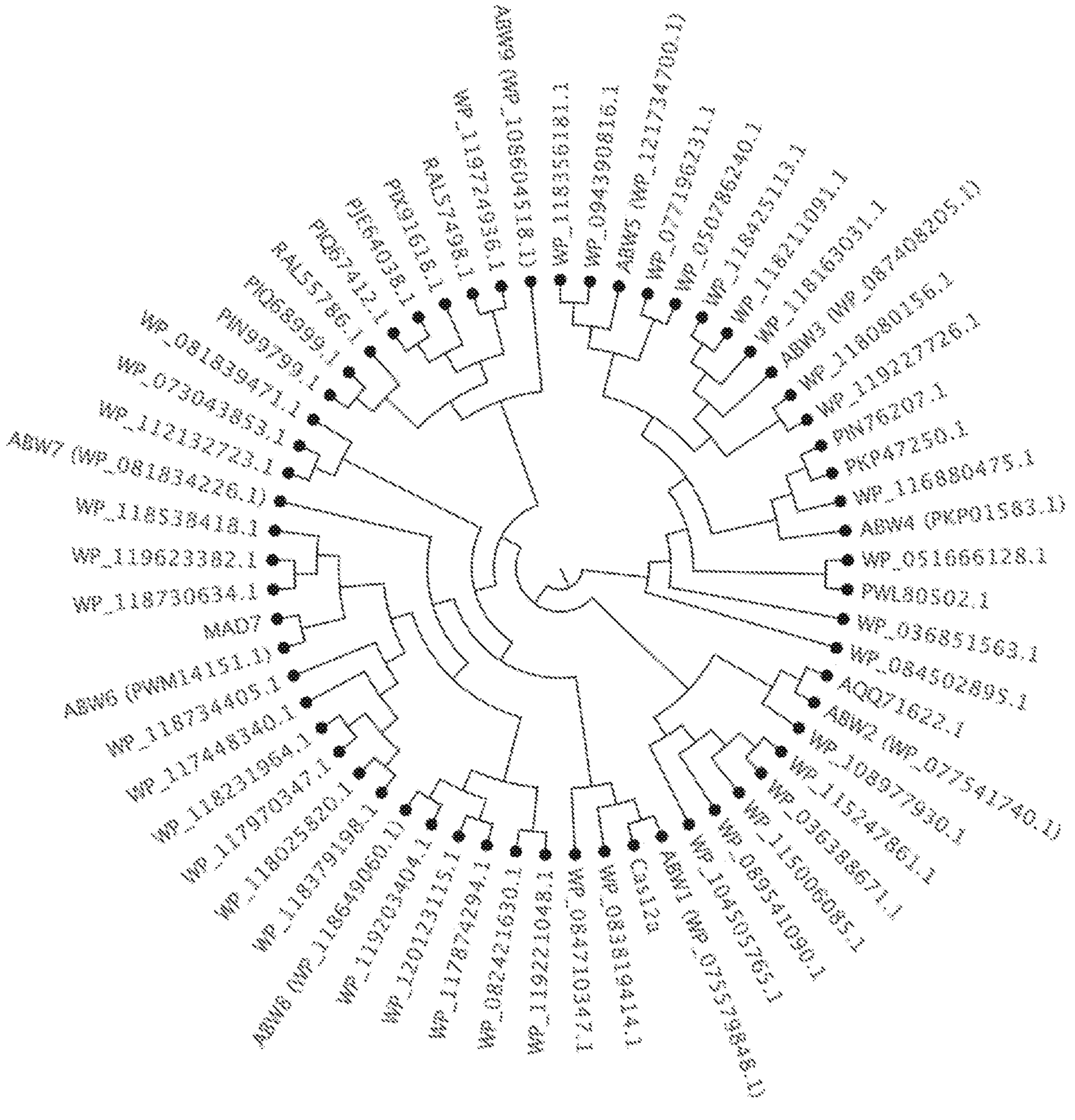

WP_118336181.1
WP_094390816.1
ABW5 (WP_121734702.1)
WP_077196231.1
WP_050786240.1
WP_118423113.1
WP_118211091.1
WP_118163031.1
ABW3 (WP_087408205.1)
WP_118080156.1
WP_119227726.1
PWT6207.1
PKP47250.1
WP_116880475.1
ABW4 (PKP01583.1)
WP_051666128.1
PWL80502.1
WP_036851563.1
WP_084502895.1
AQQ71622.1
ABW2 (WP_017541740.1)
WP_108977930.1
WP_115247691.1
WP_036388671.1
WP_115006085.1
WP_089541090.1
WP_104505765.1
ABW1 (WP_075579846.1)
OSS12
WP_083810414.1
WP_044710347.1
WP_119221048.1
WP_082421630.1
WP_117874294.1
WP_120123115.1
WP_119203404.1
ABW8 (WP_118649060.1)
WP_118379198.1
WP_118025820.1
WP_117970347.1
WP_118231964.1
WP_117448340.1
WP_118734405.1
ABW6 (PWM14151.1)
MAD7
WP_118730634.1
WP_119623382.1
WP_118538418.1
ABW7 (WP_081834228.1)
WP_112132723.1
WP_073043853.1
WP_081839471.1
PWN99799.1
PKD68999.1
RAL57786.1
PKD74212.1
PKB64038.1
PKB71618.1
RAL57498.1
WP_119724930.1
WP_119504501.1
(WP_105045018.1)

FIG. 1

sequences of gRNAs used DNMT1 *in vitro* cleavage assay

predicted – cognate gRNAs:

ABW1: GUCUAAAAGACCAUAUGAAUUUCUACUUUCGUAGAUCUGAUGGUCCAUGUCUGUUA **SEQ ID NO: 140**

ABW2: GUCUAAAGGCCUUAUAAAAUUUCUACUGUCGUAGAUCUGAUGGUCCAUGUCUGUUA **SEQ ID NO: 141**

ABW3: GUCUAUACAGACACUUUAAUUUCUACUAUUGUAGAUCUGAUGGUCCAUGUCUGUUA **SEQ ID NO: 142**

ABW4: GUCUGAAAGACAAGUAUAAUUUCUACUAUUGUAGAUCUGAUGGUCCAUGUCUGUUA **SEQ ID NO: 143**

ABW5: GGCUAUAAGCCUUGUAUAAUUUCUACUAUUGUAGAUCUGAUGGUCCAUGUCUGUUA **SEQ ID NO: 144**

ABW6: GUUGAAACUGUAAGCGGAAUGUCUACUUGGGUAGAUCUGAUGGUCCAUGUCUGUUA **SEQ ID NO: 145**

ABW7: GCAUGAGAACCAUGCAUUUCUAAGGUACUCCAAAACCUGAUGGUCCAUGUCUGUUA **SEQ ID NO: 146**

ABW8: GUUGAGUAACCUUAAAUAAUUUCUACUGUUGUAGAUCUGAUGGUCCAUGUCUGUUA **SEQ ID NO: 147**

ABW9: AUCUACAACAGUAGAAAUUUAAGCUAAGGCUUAGACCUGAUGGUCCAUGUCUGUUA **SEQ ID NO: 148**

Control gRNA: UAAUUUCUACUCUUGUAGAUCUGAUGGUCCAUGUCUGUUA **SEQ ID NO: 149**

STAR gRNA: UAAUUUCUACUC – UUGUAGAUCUGAUGGUCCAUGUCUGUUA **SEQ ID NO: 150**

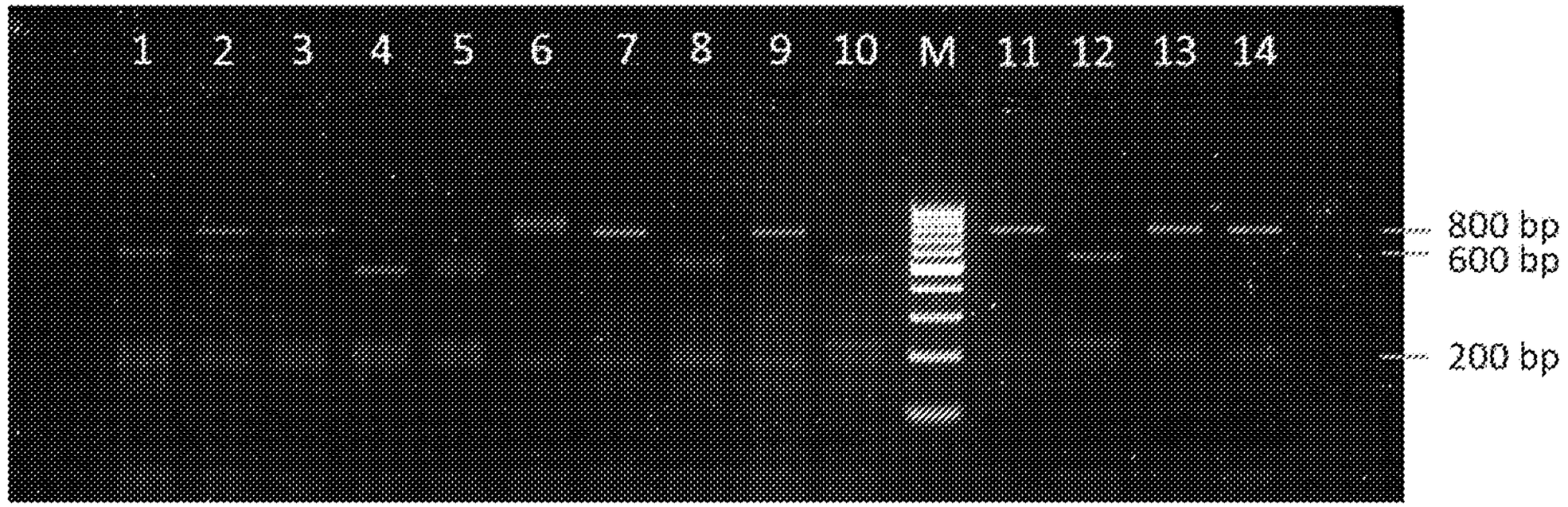
dnmt1_s1_1 dnmt1_s1_2

* * *DNMT1 amplicon targeting fragments are underlined

** Positive Control

FIG. 2

DNMT1 amplicon *in vitro* cleavage assay
with predicted - cognate gRNAs



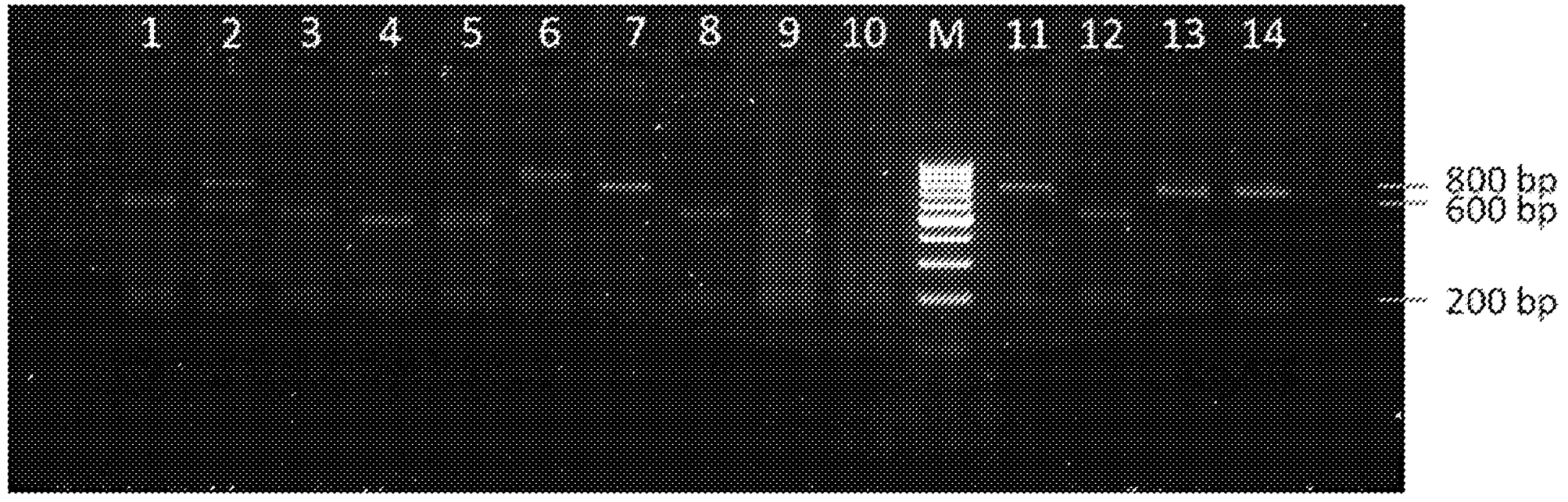
legend:

1. ABW 1
2. ABW2
3. ABW3
4. ABW4
5. ABW5
6. ABW6
7. ABW7
8. ABW8
9. ABW9
10. nuclease positive control
M. size marker (100bp)
11. uncut DNA
12. cutting positive control
13. cutting negative control (nuclease only)
14. cutting negative control (gRNA only)

*active nucleases are underlined

FIG. 3

DNMT1 amplicon *in vitro* cleavage assay with Control gRNA



legend:

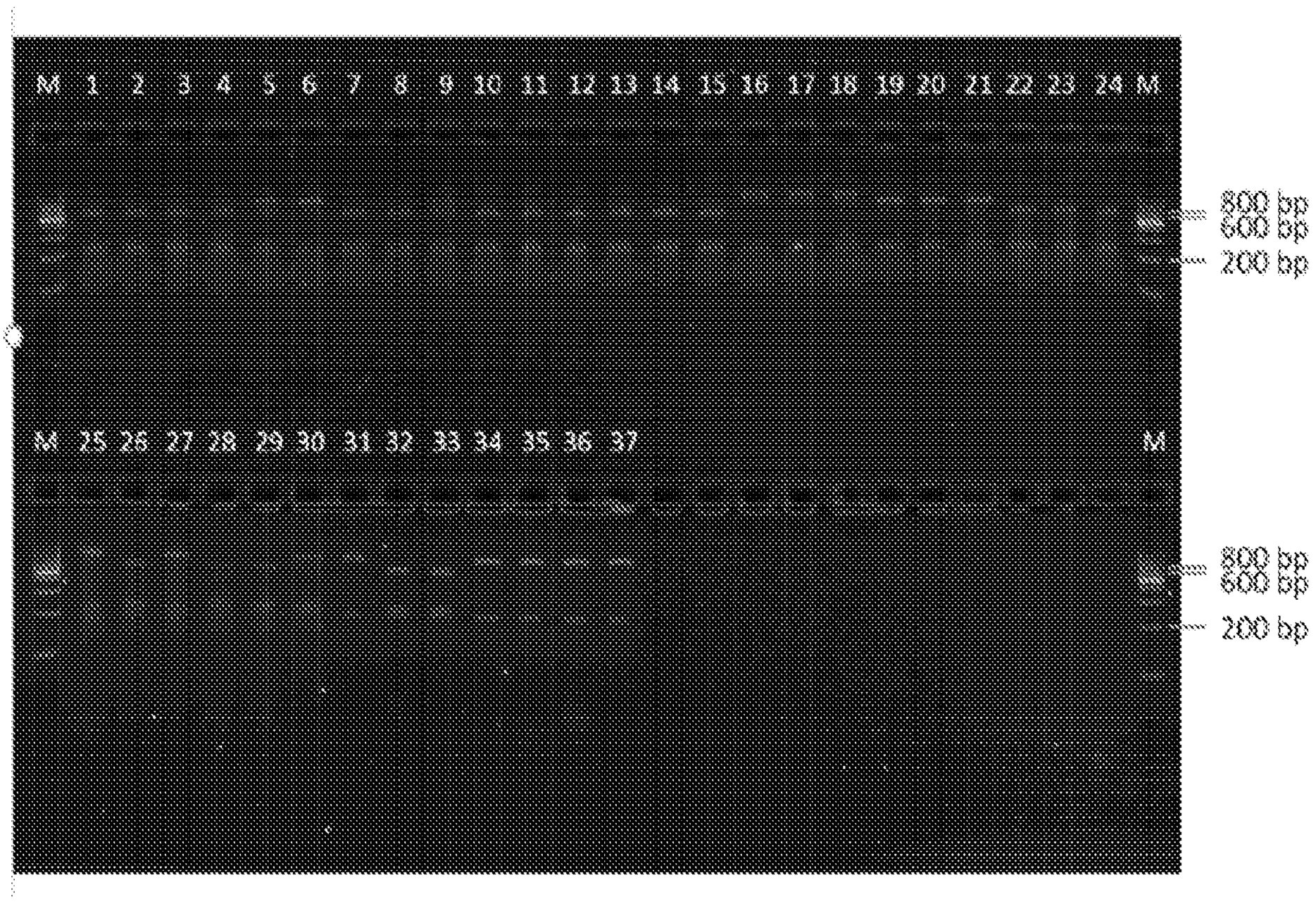
1. ABW 1
2. ABW2
3. ABW3
4. ABW4
5. ABW5
6. ABW6
7. ABW7
8. ABW8
9. ABW9
10. nuclease positive control

M. size marker (100bp)

11. uncut DNA
12. cutting positive control
13. cutting negative control (nuclease only)
14. cutting negative control (gRNA only)

*active nucleases are underlined

FIG. 4



legend:

M. size marker (100bp)

1. ABW1 + cognate gRNA
2. ABW1 + control gRNA
3. ABW1 + STAR gRNA (s1_1_2)
4. ABW2 + cognate gRNA
5. ABW2 + control gRNA
6. ABW2 + STAR gRNA (s1_1_2)
7. ABW3 + control gRNA
8. ABW3 + control gRNA
9. ABW3 + STAR gRNA (s1_1_2)
10. ABW4 + cognate gRNA
11. ABW4 + control gRNA
12. ABW4 + STAR gRNA (s1_1_2)
13. ABW5 + cognate gRNA
14. ABW5 + control gRNA
15. ABW5 + STAR gRNA (s1_1_2)
16. ABW6 + cognate gRNA
17. ABW6 + control gRNA
18. ABW6 + STAR gRNA (s1_1_2)
19. ABW7 + cognate gRNA
20. ABW7 + control gRNA
21. ABW7 + STAR gRNA (s1_1_2)
22. ABW8 + cognate gRNA
23. ABW8 + control gRNA
24. ABW8 + STAR gRNA (s1_1_2)
25. ABW9 + cognate gRNA
26. ABW9 + control gRNA
27. ABW9 + STAR gRNA (s1_1_2)
28. nuclease positive control + cognate gRNA
29. nuclease positive control + control gRNA
30. nuclease positive control + STAR gRNA (s1_1_2)
31. uncut DNA
32. cutting positive control + control gRNA
33. cutting positive control + STAR gRNA (s1_1_2)
34. cutting negative control (STAR s1_1)
35. cutting negative control (STAR s1_2)
36. cutting negative control (nuclease only)
37. cutting negative control (gRNA only)

*active nucleases are underlined

FIG. 5



molar ratio
nuclease : gRNA : target DNA
20 : 60 : 1
10 : 30 : 1
5 :10 : 1
2.5 : 7.5 : 1
1.25 : 3.75 : 1
0.625 : 1.875 : 1

M = GeneRuler Low Range DNA Ladder

FIG. 6A

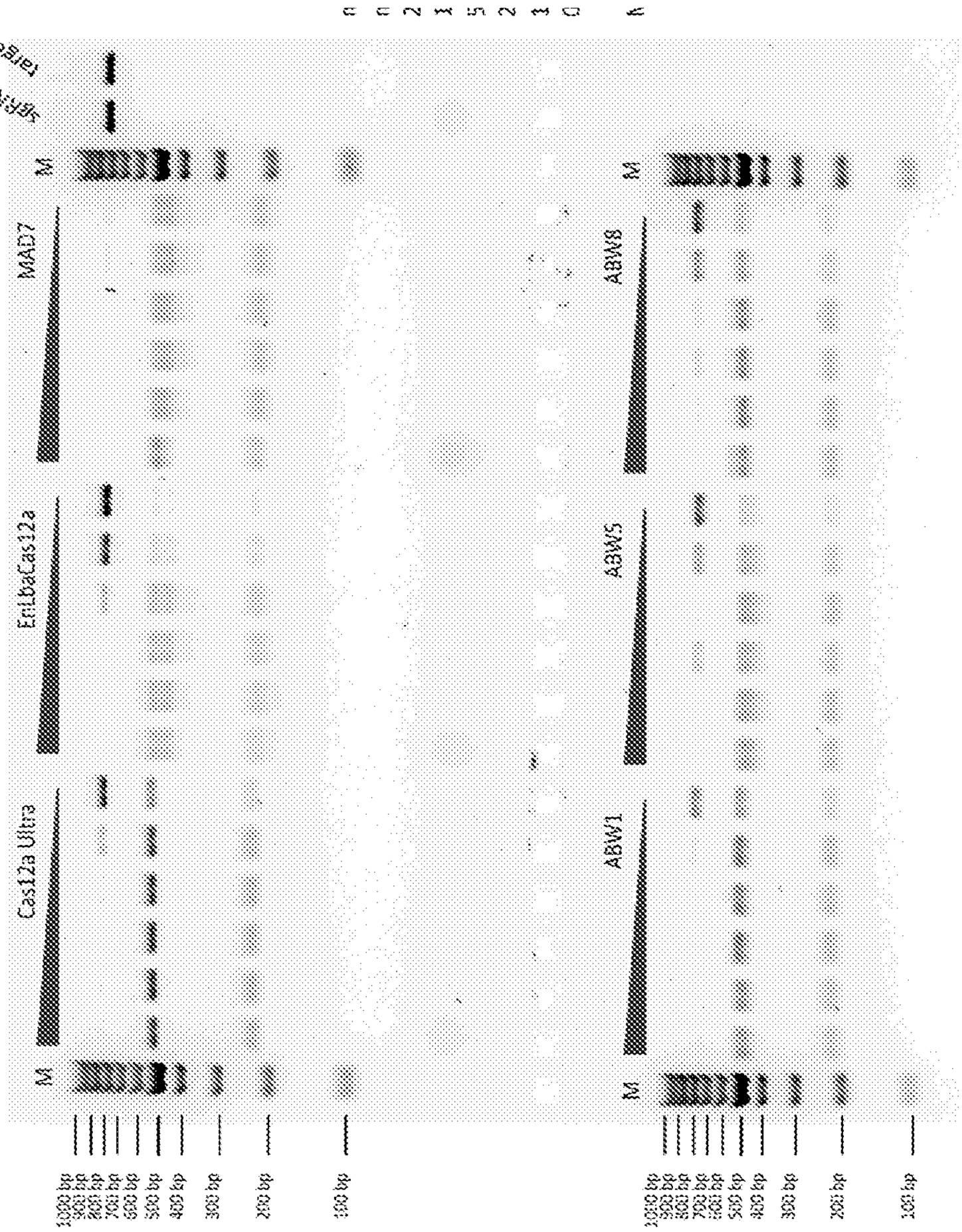
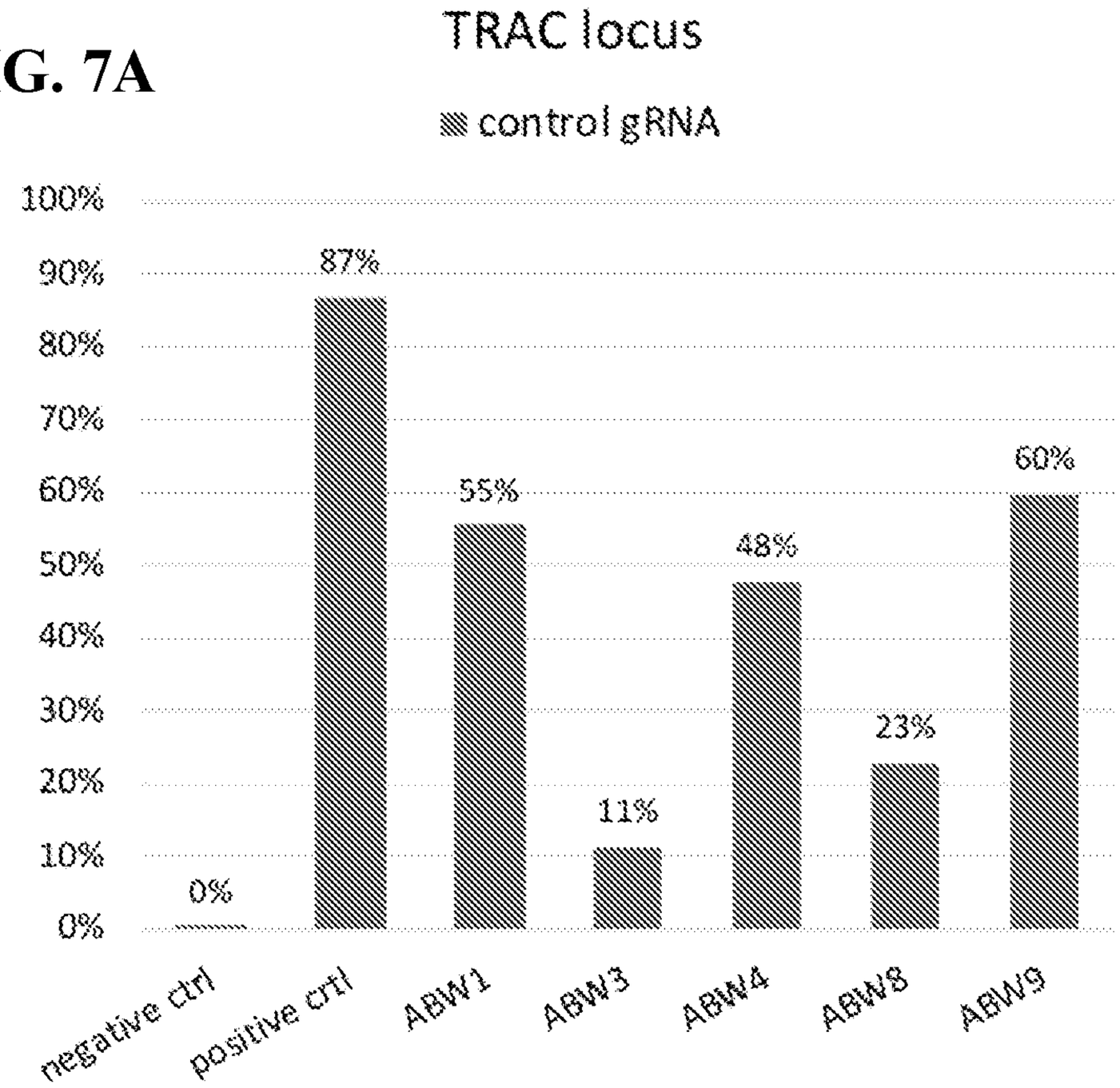


FIG. 6B

molar ratio
nuclease : gRNA : target DNA
20 : 60 : 1
10 : 30 : 1
5 :10 : 1
2.5 : 7.5 : 1
1.25 : 3.75 : 1
0.625 : 1.875 : 1

M = GeneRuler Low Range DNA Ladder

FIG. 7A



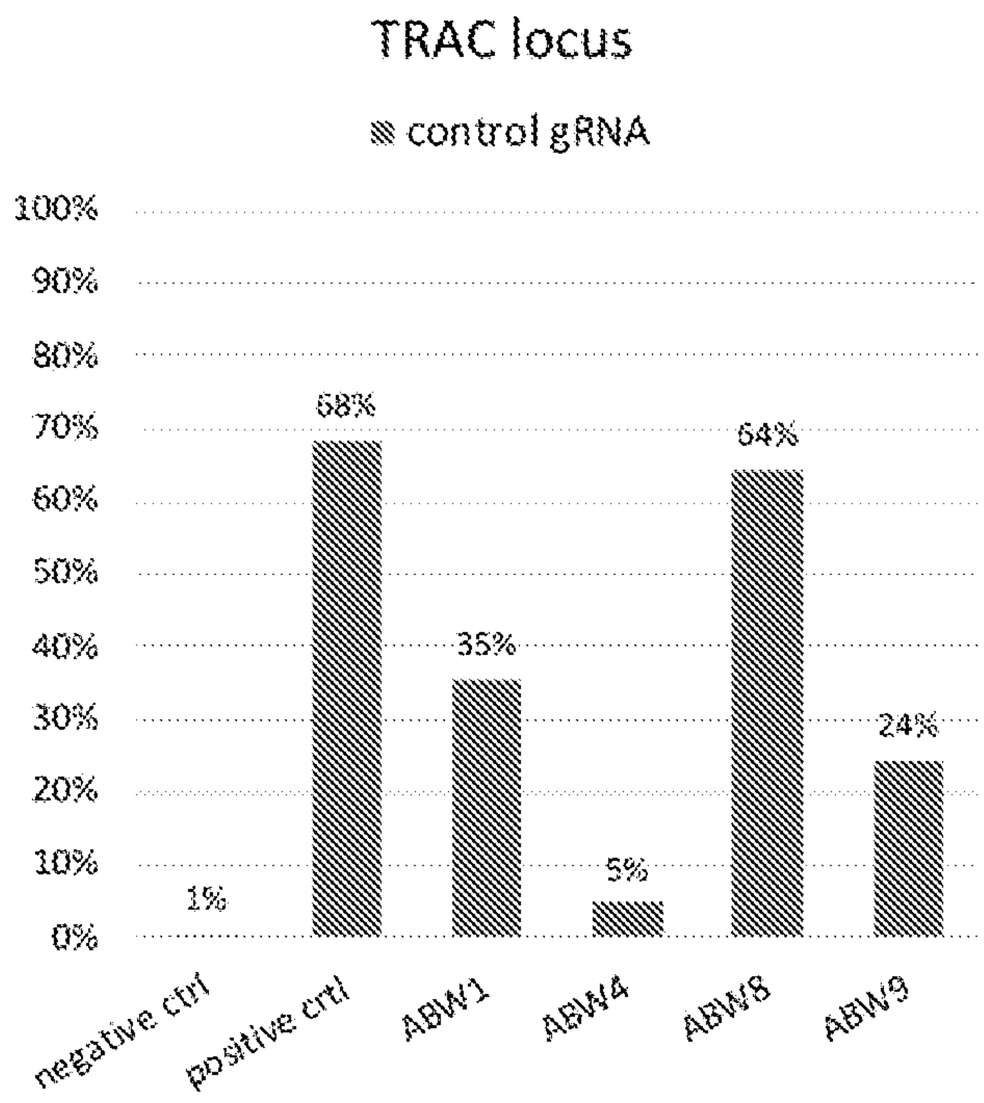


FIG. 7B

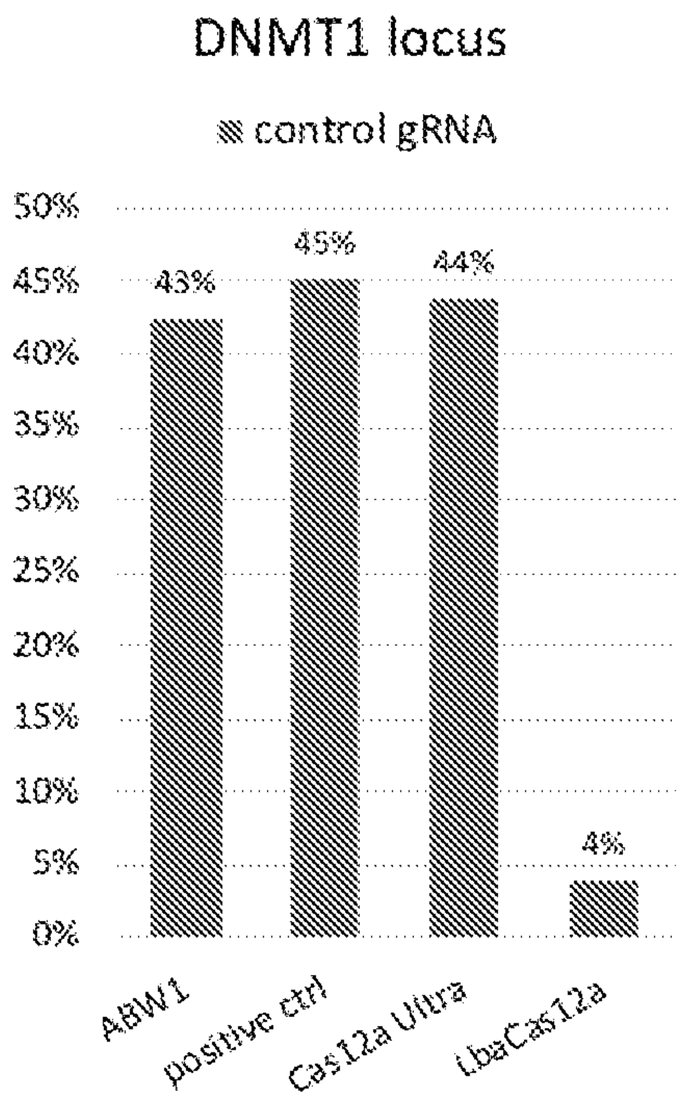
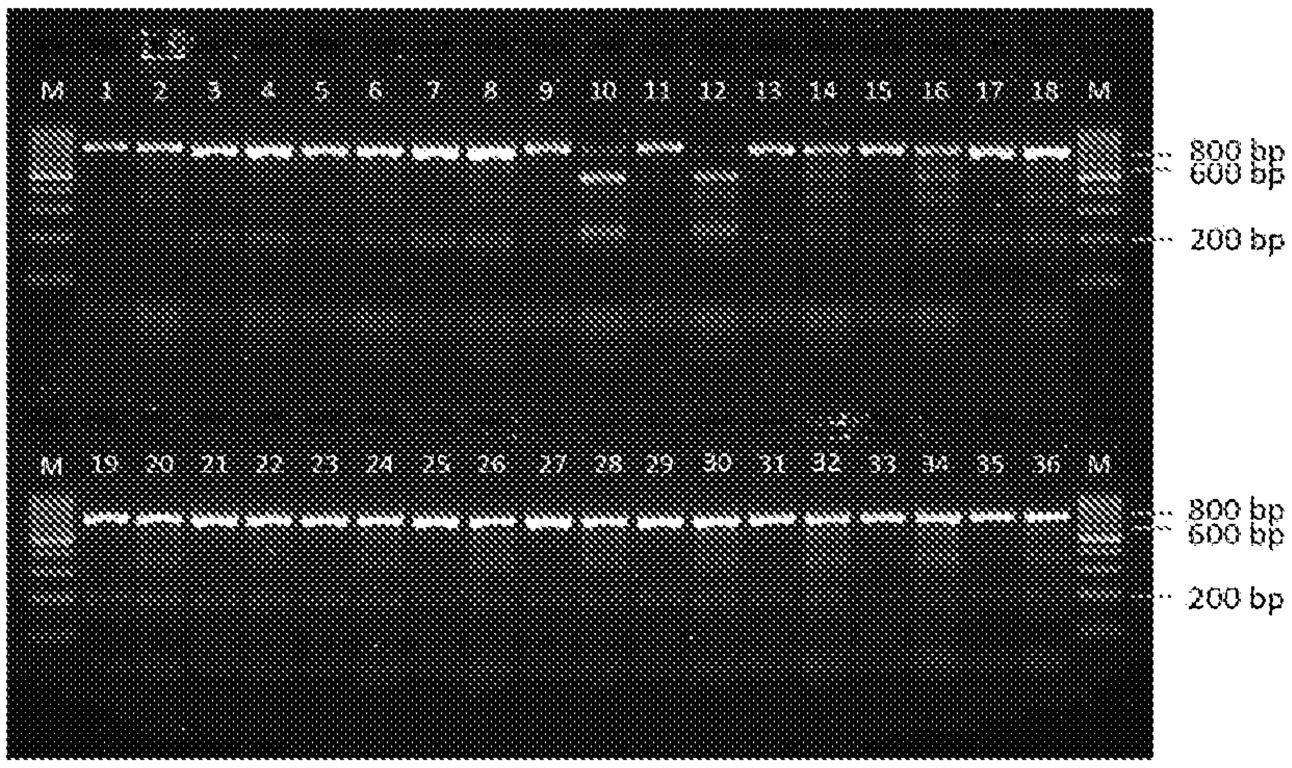


FIG. 7C

FIG. 8

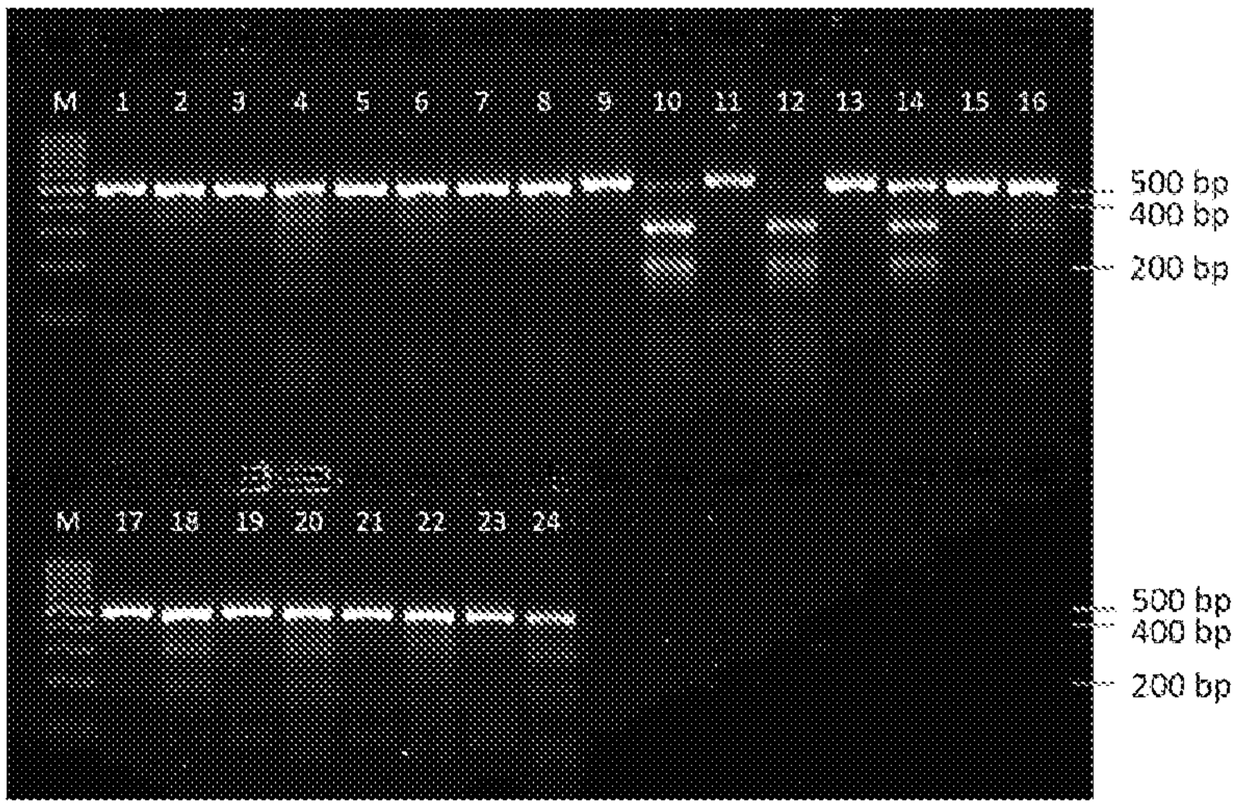


*active nucleases are underlined

M size marker (100bp)

1. editing negative control #1
2. T7 treated editing negative control #1
3. editing negative control #2
4. T7 treated editing negative control #2
5. editing negative control #3
6. T7 treated editing negative control #3
7. editing negative control #4
8. T7 treated editing negative control #4
9. editing positive control #1
10. T7 treated editing positive control #1
11. editing positive control #2
12. T7 treated editing positive control #2
13. ABW1 + cognate gRNA
14. T7 treated ABW1 + cognate gRNA
15. ABW1 + control gRNA
16. T7 treated ABW1 + control gRNA
17. ABW2 + cognate gRNA
18. T7 treated ABW2 + cognate gRNA
19. ABW2 + control gRNA
20. T7 treated ABW2 + control gRNA
21. ABW3 + cognate gRNA
22. T7 treated ABW3 + cognate gRNA
23. ABW3 + control gRNA
24. T7 treated ABW3 + control gRNA
25. ABW4 + cognate gRNA
26. T7 treated ABW4 + cognate gRNA
27. ABW4 + control gRNA
28. T7 treated ABW4 + control gRNA
29. ABW5 + cognate gRNA
30. T7 treated ABW5 + cognate gRNA
31. ABW5 + control gRNA
32. T7 treated ABW5 + control gRNA
33. ABW8 + cognate gRNA
34. T7 treated ABW8 + cognate gRNA
35. ABW8 + control gRNA
36. T7 treated ABW8 + control gRNA

FIG. 9



*active nucleases are underlined

M size 100bp)

1. editing negative control #1
2. T7 treated editing negative control #1
3. editing negative control #2
4. T7 treated editing negative control #2
5. editing negative control #3
6. T7 treated editing negative control #3
7. editing negative control #4
8. T7 treated editing negative control #4
9. editing positive control #1
10. T7 treated editing positive control #1
11. editing positive control #2
12. T7 treated editing positive control #2
13. ABW1 + control gRNA
14. T7 treated ABW1 + control gRNA
15. ABW2 + control gRNA
16. T7 treated ABW2 + control gRNA
17. (ABW3 + control gRNA
18. T7 treated ABW3 + control gRNA
19. ABW4 + control gRNA
20. T7 treated ABW4 + control gRNA
21. ABW5 + control gRNA
22. T7 treated ABW5 + control gRNA
23. ABW8 + control gRNA
24. T7 treated ABW8 + control gRNA

FIG. 10

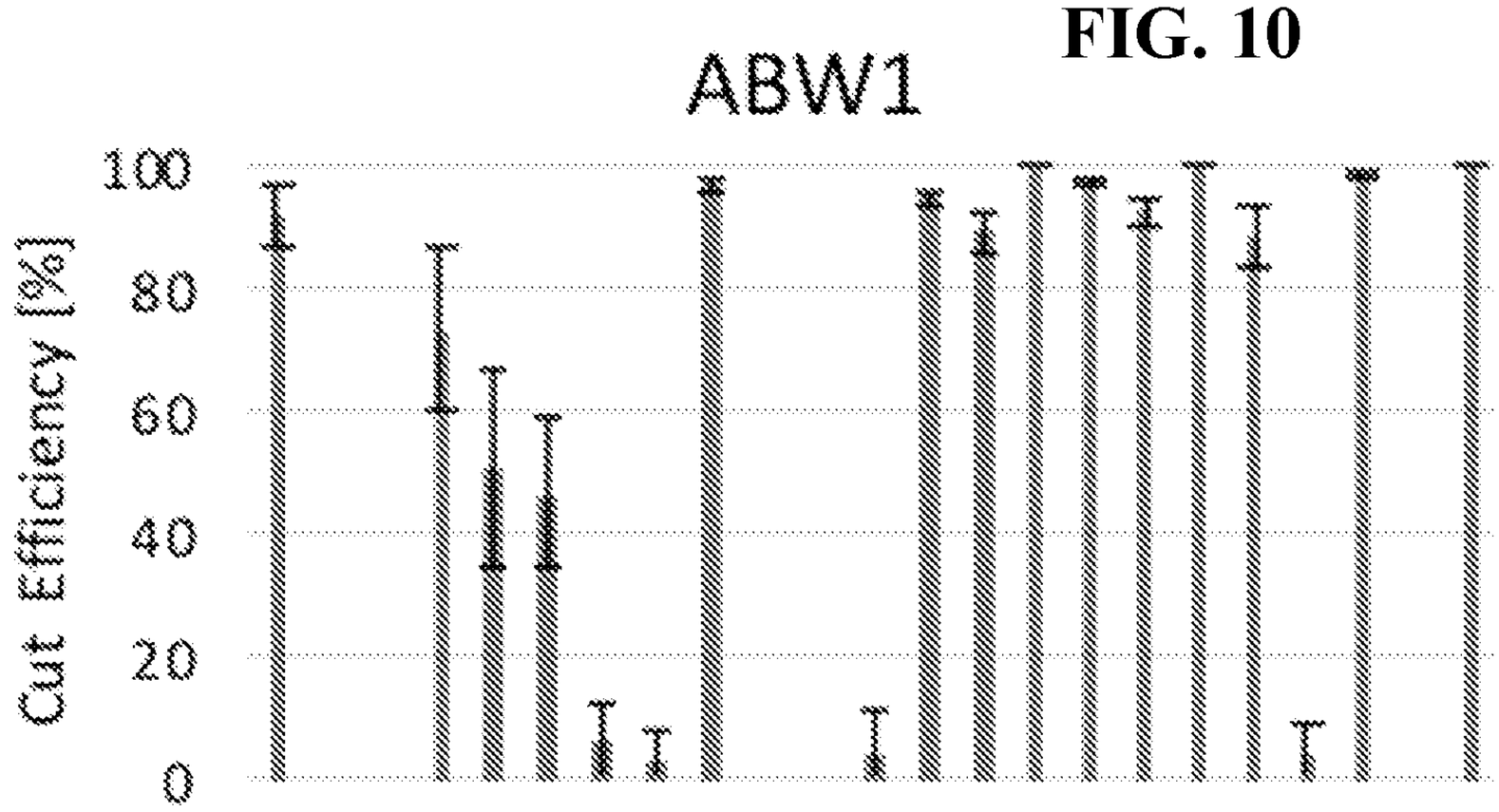


FIG. 11

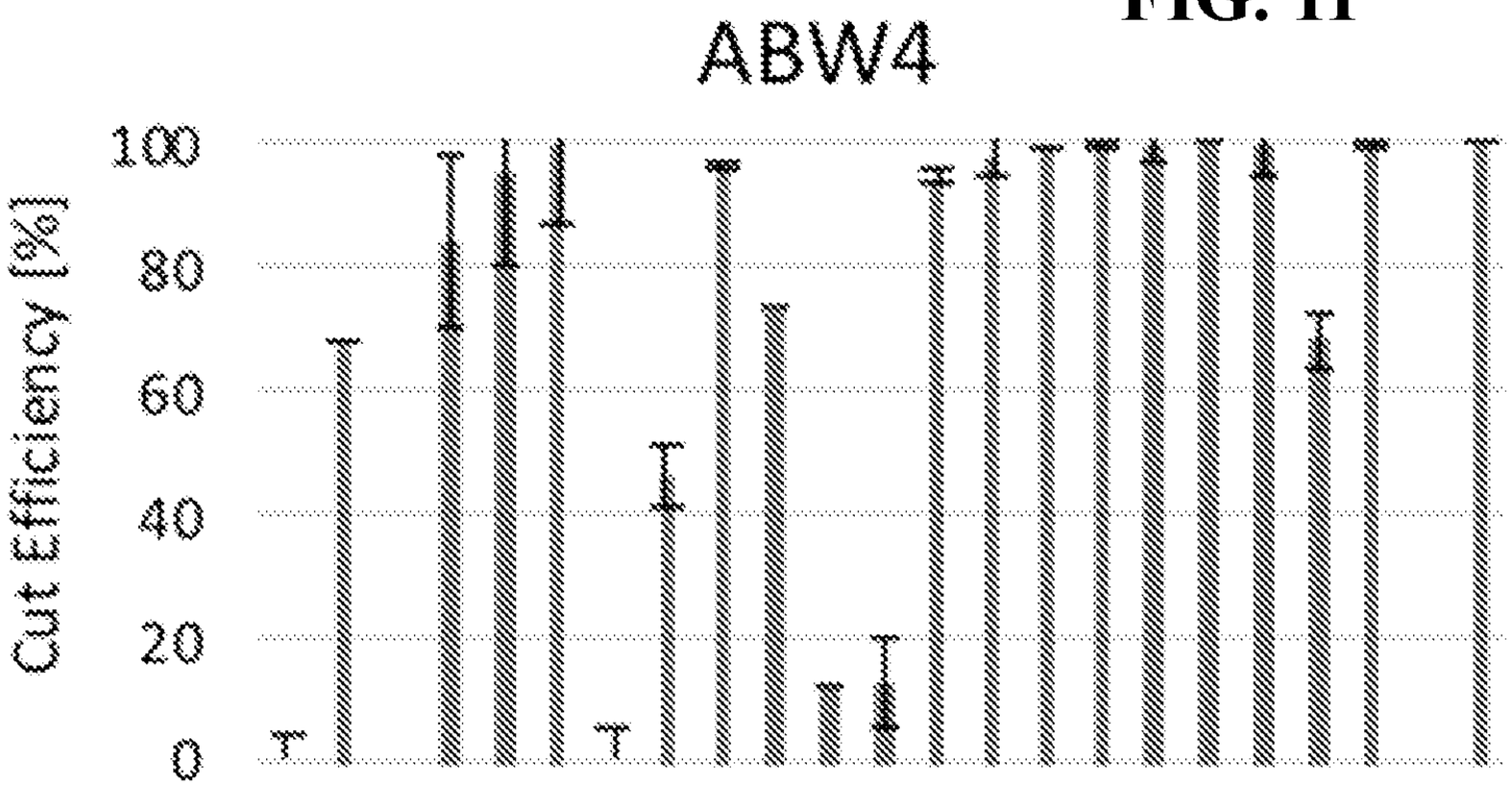


FIG. 12

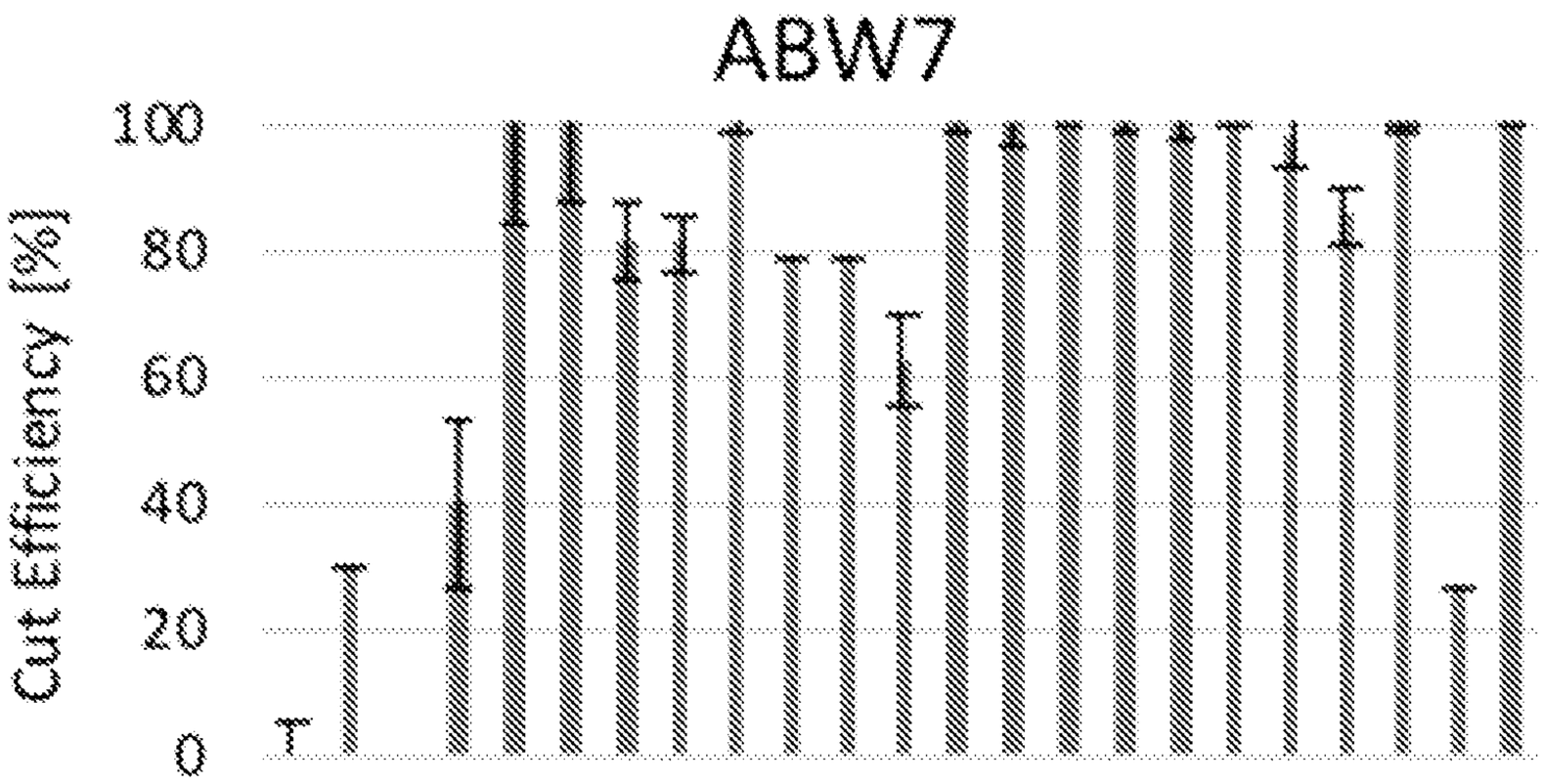


FIG. 13

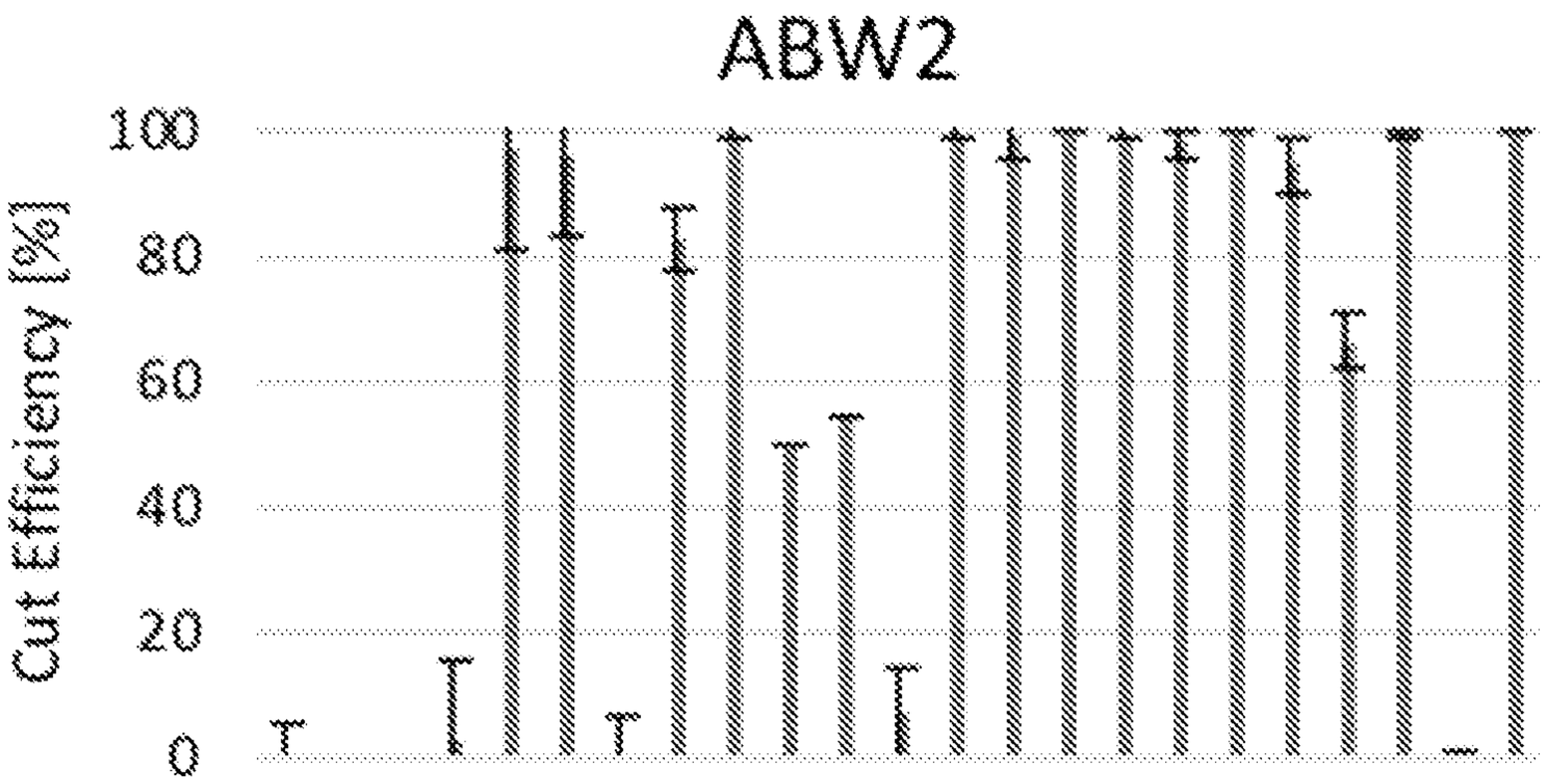


FIG. 14

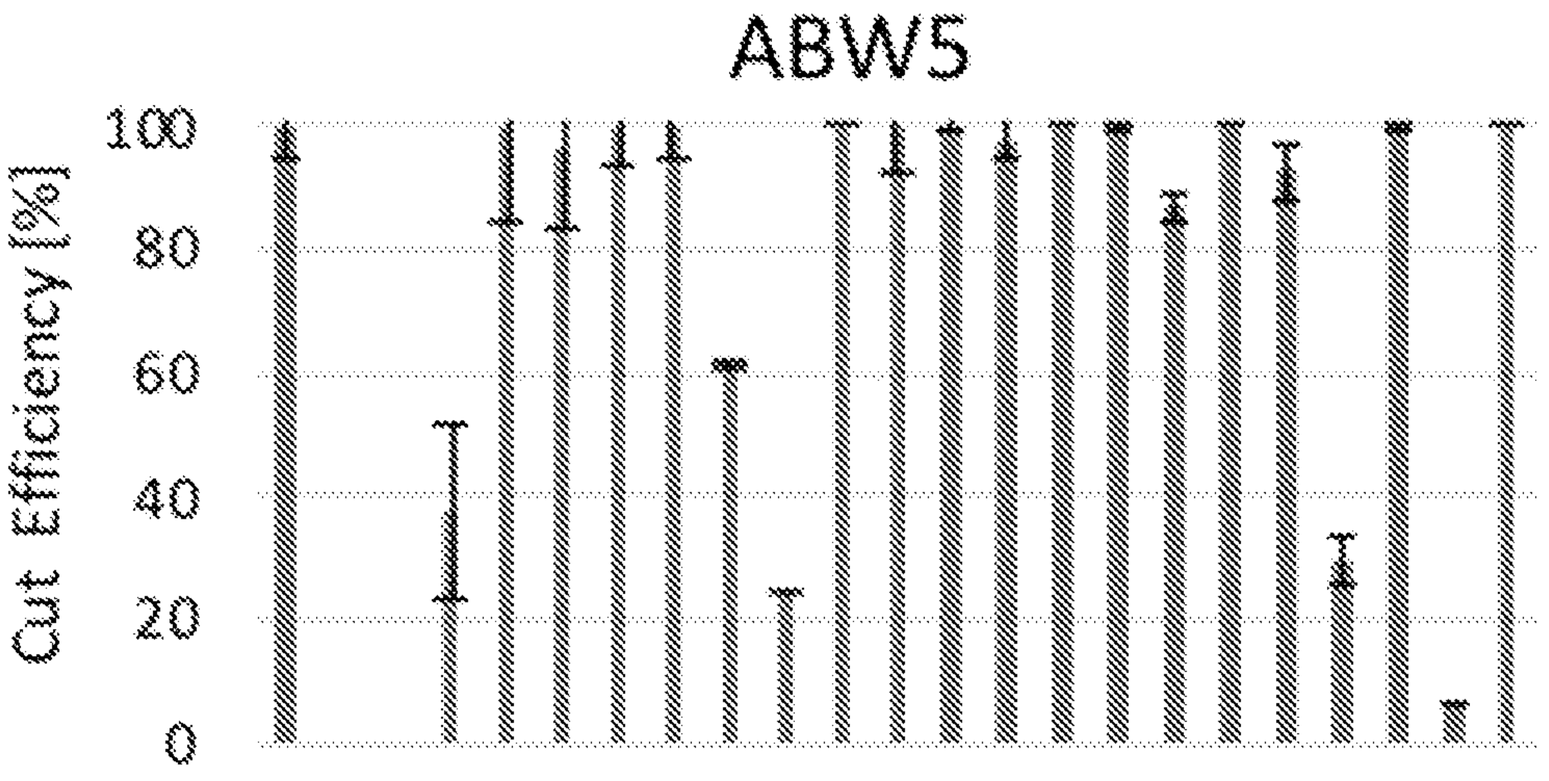


FIG. 15

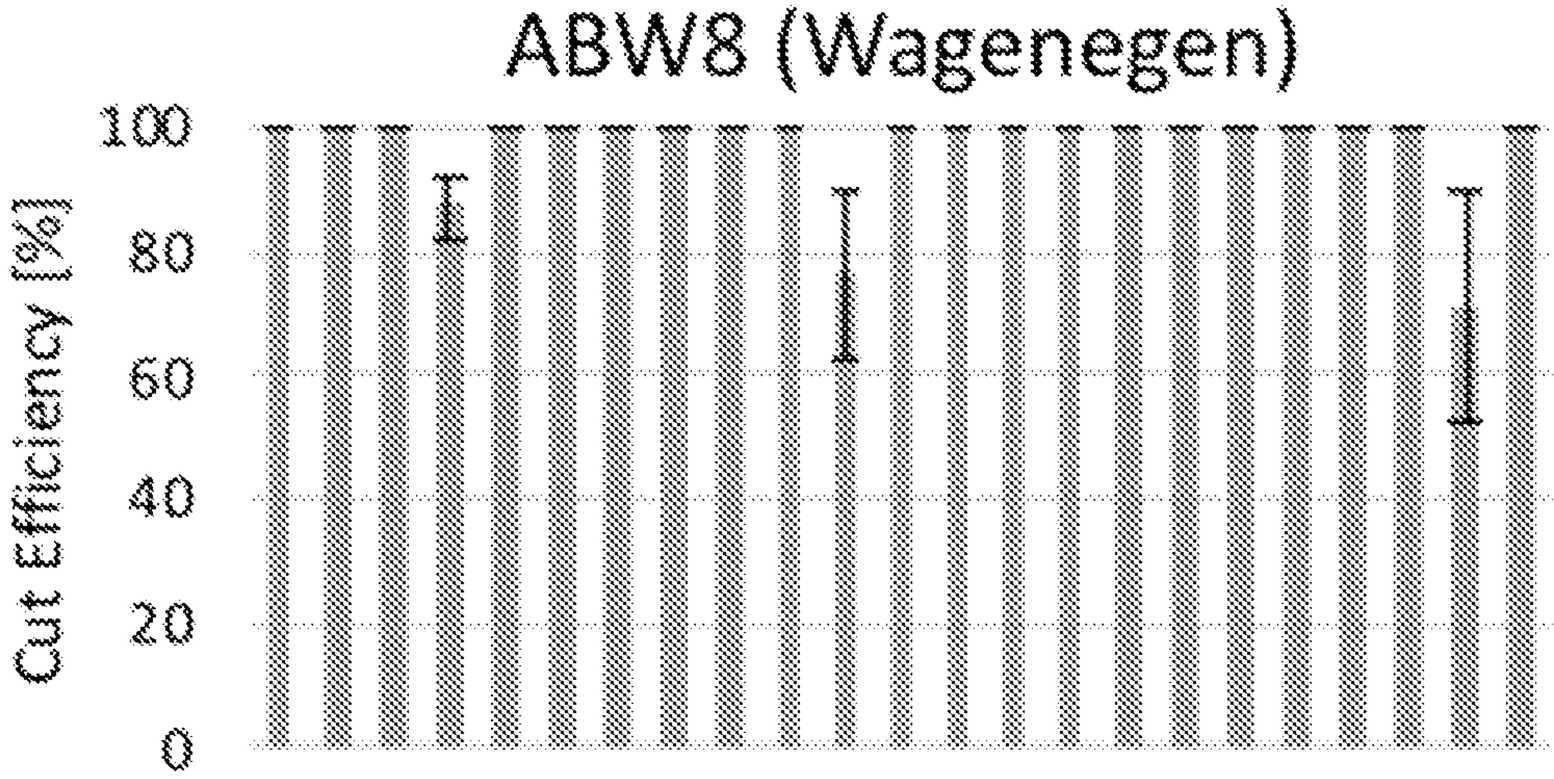


FIG. 16

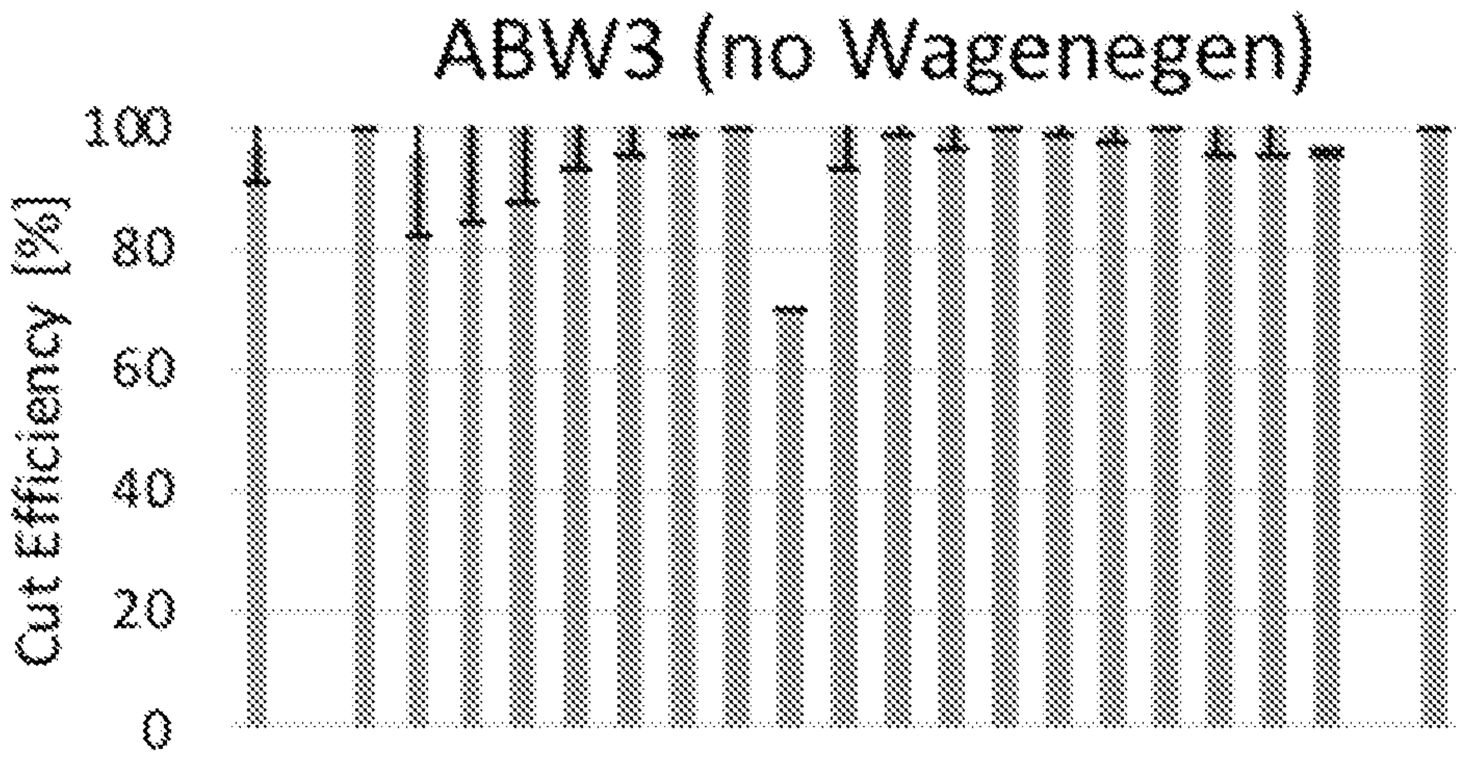


FIG. 17

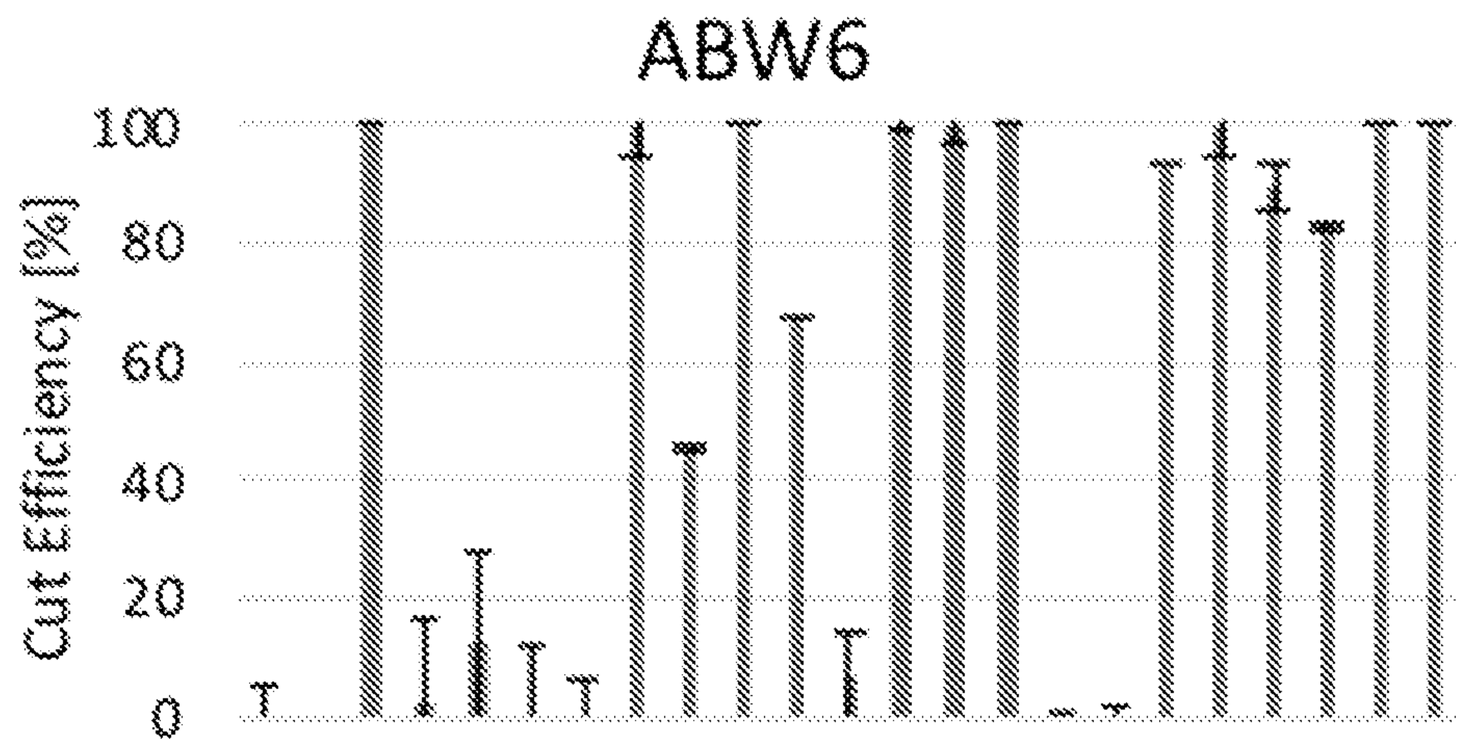
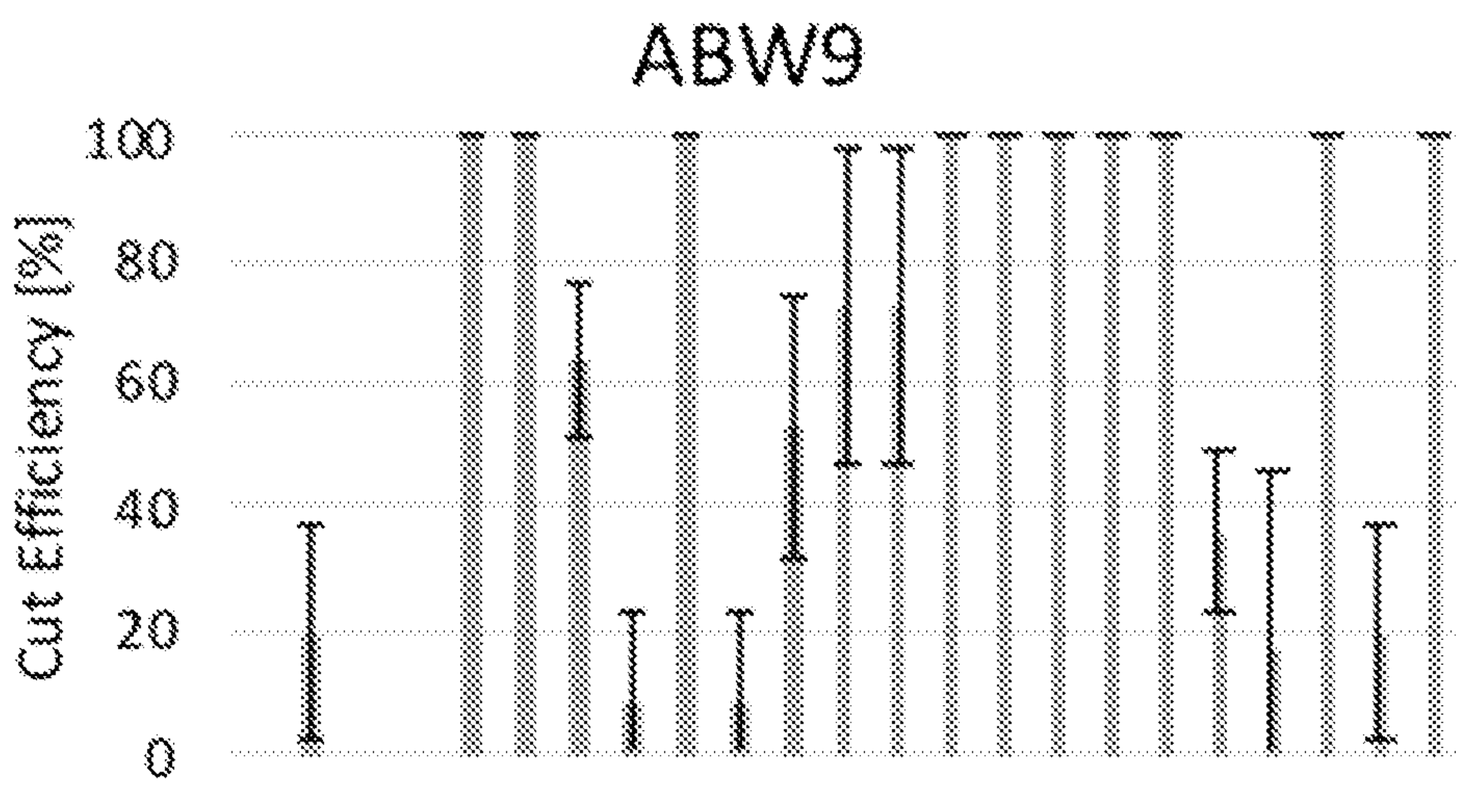


FIG. 18



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/061850

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 9/22; C12N 15/10; C12N 15/113; C12N 15/90 (2021.01)
CPC - C12N 9/22; C12N 15/102; C12N 15/907; C12N 2310/20 (2021.02)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
see Search History document

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2018/0155716 A1 (THE BROAD INSTITUTE, INC. et al) 07 June 2018 (07.06.2018) entire document	1-20, 30-32
A	WO 2017/106657 A1 (THE BROAD INSTITUTE INC. et al) 22 June 2017 (22.06.2017) entire document	1-20, 30-32
A	WO 2016/115179 A1 (MASSACHUSETTS INSTITUTE OF TECHNOLOGY et al) 21 July 2016 (21.07.2016) entire document	1-20, 30-32
A	WO 2018/226972 A2 (VILMORIN & CIE et al) 13 December 2018 (13.12.2018) entire document	1-20, 30-32
A	WO 2018/013990 A1 (ZYMERGEN INC.) 18 January 2018 (18.01.2018) entire document	1-20, 30-32

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
08 February 2021	MAR 10 2021
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/061850

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

SEQ ID NOs: 3, 16, 29, 42, 55, 68, 81, 94, 107, and 118-128 were searched.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/061850

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 21-29
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.