Compositions and methods for targeting, editing or modifying human genes

Gill, Ryan T; Warnecke, Tanya; Baumgartner, Roland Franz

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 August 2021 (12.08.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/158918 A1

(51) International Patent Classification:
C12N 15/11 (2006.01) *C12N 5/0783* (2010.01)
C12N 9/22 (2006.01)

(21) International Application Number:
PCT/US2021/016823

(22) International Filing Date:
05 February 2021 (05.02.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/970,455 05 February 2020 (05.02.2020) US

(71) Applicant: **DANMARKS TEKNISKE UNIVERSITET** [DK/DK]; Anker Engelunds Vej 101A, 2800 Kgs. Lyngby (DK).

(72) Inventor; and

(71) Applicant *(for US only)*: **GILL, Ryan T.** [US/US]; 2025 Ash Street, Denver, Colorado 80207 (US).

(72) Inventors: **WARNECKE, Tanya**; 460 Oneida Street, Boulder, Colorado 80303 (US). **BAUMGARTNER, Roland Franz**; Jagerhausgasse 51, 1120 Vienna (AT).

(74) Agent: **XU, Lily**; Goodwin Procter LLP, 100 Northern Avenue, IP DOCKETING DEPT/7TH FL, Boston, Massachusetts 03310 (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,

(54) **Title:** COMPOSITIONS AND METHODS FOR TARGETING, EDITING OR MODIFYING HUMAN GENES

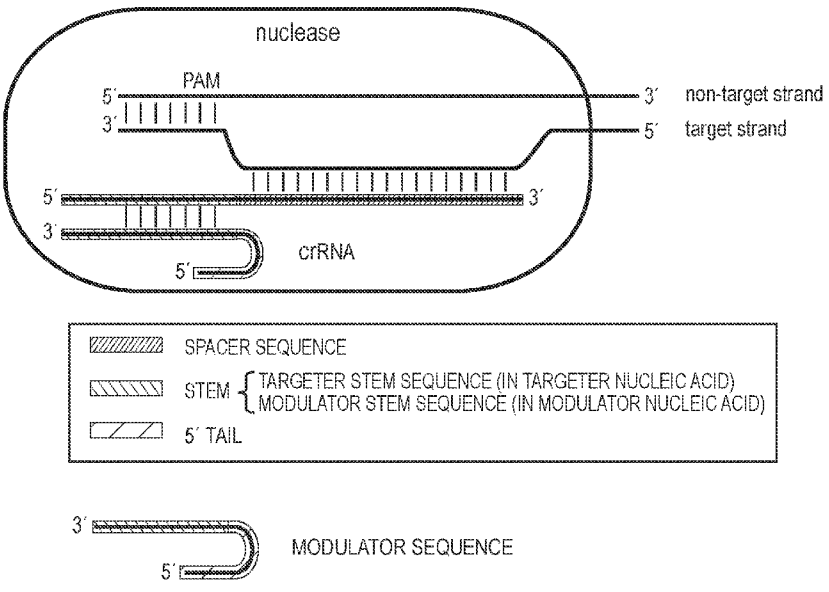

FIG. 1B

(57) **Abstract:** The present invention relates to engineered Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems and corresponding guide RNAs that target specific nucleotide sequences at certain gene loci in the human genome. Also provided are methods of targeting, editing, and/or modifying of the human genes using the engineered CRISPR systems, and compositions and cells comprising the engineered CRISPR systems.

[Continued on next page]

WO 2021/158918 A1

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).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

COMPOSITIONS AND METHODS FOR TARGETING, EDITING OR MODIFYING HUMAN GENES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of and priority to U.S. Provisional Patent Application No. 62/970,455, filed February 5, 2020, the disclosure of which is hereby incorporated by reference in its entirety for all purposes.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on January 28, 2021, is named ATS-002WO_SL.txt and is 333,008 bytes in size.

FIELD OF THE INVENTION

[0003] The present invention relates to engineered Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems and corresponding guide RNAs that target specific nucleotide sequences at certain gene loci in the human genome, methods of targeting, editing, and/or modifying human genes using the engineered CRISPR systems, and compositions and cells comprising the engineered CRISPR systems.

BACKGROUND OF THE INVENTION

[0004] Recent advances have been made in precise genome targeting technologies. For example, specific loci in genomic DNA can be targeted, edited, or otherwise modified by designer meganucleases, zinc finger nucleases, or transcription activator-like effectors (TALEs). Furthermore, the CRISPR-Cas systems of bacterial and archaeal adaptive immunity have been adapted for precise targeting of genomic DNA in eukaryotic cells. Compared to the earlier generations of genome editing tools, the CRISPR-Cas systems are easy to set up, scalable, and amenable to targeting multiple positions within the eukaryotic genome, thereby providing a major resource for new applications in genome engineering.

[0005] Two distinct classes of CRISPR-Cas systems have been identified. Class 1 CRISPR-Cas systems utilize multi-protein effector complexes, whereas class 2 CRISPR-Cas systems utilize single-protein effectors (see, Makarova *et al.* (2017) CELL, 168: 328). Among

the three types of class 2 CRISPR-Cas systems, type II and type V systems typically target DNA and type VI systems typically target RNA (*id.*). Naturally occurring type II effector complexes consist of Cas9, CRISPR RNA (crRNA), and trans-activating CRISPR RNA (tracrRNA), but the crRNA and tracrRNA can be fused as a single guide RNA in an engineered system for simplicity (see, Wang *et al.* (2016) ANNU. REV. BIOCHEM., 85: 227). Certain naturally occurring type V systems, such as type V-A, type V-C, and type V-D systems, do not require tracrRNA and use crRNA alone as the guide for cleavage of target DNA (see, Zetsche *et al.* (2015) CELL, 163: 759; Makarova *et al.* (2017) CELL, 168: 328).

[0006] The CRISPR-Cas systems have been engineered for various purposes, such as genomic DNA cleavage, base editing, epigenome editing, and genomic imaging (see, *e.g.*, Wang *et al.* (2016) ANNU. REV. BIOCHEM., 85: 227 and Rees *et al.* (2018) NAT. REV. GENET., 19: 770). Although significant developments have been made, there remains a need for new and useful CRISPR-Cas systems as powerful genome targeting tools.

SUMMARY OF THE INVENTION

[0007] The present invention is based, in part, upon the development of engineered CRISPR-Cas systems (*e.g.*, type V-A CRISPR-Cas systems) that can be used to target, edit, or otherwise modify specific target nucleotide sequences in human ADORA2A, B2M, CD52, CIITA, CTLA4, DCK, FAS, HAVCR2 (also called TIM3), LAG3, PDCD1 (also called PD-1), PTPN6, TIGIT, TRAC, TRBC1, TRBC2, CARD11, CD247, IL7R, LCK, or PLCG1 gene. In particular, guide nucleic acids, such as single guide nucleic acids and dual guide nucleic acids, can be designed to hybridize with the selected target nucleotide sequence and activate a Cas nuclease to edit the human genes. CRISPR-Cas systems comprising such guide nucleic acids are also useful for targeting or modifying the human genes.

[0008] A CRISPR-Cas system generally comprises a Cas protein and one or more guide nucleic acids (*e.g.*, RNAs). The Cas protein can be directed to a specific location in a double-stranded DNA target by recognizing a protospacer adjacent motif (PAM) in the non-target strand of the DNA, and the one or more guide nucleic acids can be directed to a specific location by hybridizing with a target nucleotide sequence in the target strand of the DNA. Both PAM recognition and target nucleotide sequence hybridization are required for stable binding of a CRISPR-Cas complex to the DNA target and, if the Cas protein has an effector function (*e.g.*, nuclease activity), activation of the effector function. As a result, when creating a CRISPR-Cas system, a guide nucleic acid can be designed to comprise a

nucleotide sequence called spacer sequence that hybridizes with a target nucleotide sequence, where target nucleotide sequence is located adjacent to a PAM in an orientation operable with the Cas protein. It has been observed that not all CRISPR-Cas systems designed by these criteria are equally effective. The present invention identifies target nucleotide sequences in particular human genes that can be efficiently edited, and provides CRISPR-Cas systems directed to these target nucleotide sequences.

[0009] Accordingly, in one aspect, the present invention provides a guide nucleic acid comprising a targeter stem sequence and a spacer sequence, wherein the spacer sequence comprises a nucleotide sequence listed in Table 1, 2, or 3.

[0010] In certain embodiments, the targeter stem sequence comprises a nucleotide sequence of GUAGA. In certain embodiments, the targeter stem sequence is 5' to the spacer sequence, optionally wherein the targeter stem sequence is linked to the spacer sequence by a linker consisting of 1, 2, 3, 4, or 5 nucleotides.

[0011] In certain embodiments, the guide nucleic acid is capable of activating a CRISPR Associated (Cas) nuclease in the absence of a tracrRNA (*e.g.*, the guide nucleic acid being a single guide nucleic acid). In certain embodiments, the guide nucleic acid comprises from 5' to 3' a modulator stem sequence, a loop sequence, a targeter stem sequence, and the spacer sequence.

[0012] In certain embodiments, the guide nucleic acid is a targeter nucleic acid that, in combination with a modulator nucleic acid, is capable of activating a Cas nuclease. In certain embodiments, the guide nucleic acid comprises from 5' to 3' a targeter stem sequence and the spacer sequence.

[0013] In certain embodiments, the Cas nuclease is a type V Cas nuclease. In certain embodiments, the Cas nuclease is a type V-A Cas nuclease. In certain embodiments, the Cas nuclease comprises an amino acid sequence at least 80% identical to SEQ ID NO: 1. In certain embodiments, the Cas nuclease is Cpf1. In certain embodiments, the Cas nuclease recognizes a protospacer adjacent motif (PAM) consisting of the nucleotide sequence of TTTN or CTTN.

[0014] In certain embodiments, the guide nucleic acid comprises a ribonucleic acid (RNA). In certain embodiments, the guide nucleic acid comprises a modified RNA. In certain embodiments, the guide nucleic acid comprises a combination of RNA and DNA. In certain embodiments, the guide nucleic acid comprises a chemical modification. In certain

embodiments, the chemical modification is present in one or more nucleotides at the 5' end of the guide nucleic acid. In certain embodiments, the chemical modification is present in one or more nucleotides at the 3' end of the guide nucleic acid. In certain embodiments, the chemical modification is selected from the group consisting of 2′-O-methyl, 2′-fluoro, 2′-O-methoxyethyl, phosphorothioate, phosphorodithioate, pseudouridine, and any combinations thereof.

[0015] The present invention also provides an engineered, non-naturally occurring system comprising a guide nucleic acid (*e.g.*, a single guide nucleic acid) disclosed herein. In certain embodiments, the engineered, non-naturally occurring system further comprising the Cas nuclease. In certain embodiments, the guide nucleic acid and the Cas nuclease are present in a ribonucleoprotein (RNP) complex.

[0016] The present invention also provides an engineered, non-naturally occurring system comprising the guide nucleic acid (*e.g.*, targeter nucleic acid) disclosed herein, wherein the engineered, non-naturally occurring system further comprises the modulator nucleic acid. In certain embodiments, the engineered, non-naturally occurring system, further comprises the Cas nuclease. In certain embodiments, the guide nucleic acid, the modulator nucleic acid, and the Cas nuclease are present in an RNP complex.

[0017] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 51 and 131-137, wherein the spacer sequence is capable of hybridizing with the human ADORA2A gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the ADORA2A gene locus is edited in at least 1.5% of the cells.

[0018] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 52, 64-66, 138-145, 622, 625-626, and 634-635, wherein the spacer sequence is capable of hybridizing with the human B2M gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the B2M gene locus is edited in at least 1.5% of the cells.

[0019] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 724, 726-727, 730-732, 735-738, 741-742, and 744-745, wherein the spacer

sequence is capable of hybridizing with the human CD247 gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the CD247 gene locus is edited in at least 1.5% of the cells.

[0020] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 53 and 146, wherein the spacer sequence is capable of hybridizing with the human CD52 gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the CD52 gene locus is edited in at least 1.5% of the cells.

[0021] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 54, 147-148, 636-640, 642, 644-648, 650-652, 655-656, 660-663, 666, 668, 670-671, 673-676, 678-679, and 682-685, wherein the spacer sequence is capable of hybridizing with the human CIITA gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the CIITA gene locus is edited in at least 1.5% of the cells.

[0022] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 55, 67-70, and 149-155, wherein the spacer sequence is capable of hybridizing with the human CTLA4 gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the CTLA4 gene locus is edited in at least 1.5% of the cells.

[0023] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 56, 71-74, and 156-159, wherein the spacer sequence is capable of hybridizing with the human DCK gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the DCK gene locus is edited in at least 1.5% of the cells.

[0024] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 57, 75-79, and 160-173, wherein the spacer sequence is capable of hybridizing with the human FAS gene. In certain embodiments, when the system is delivered into a

population of human cells *ex vivo*, the genomic sequence at the FAS gene locus is edited in at least 1.5% of the cells.

[0025] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 58, 80-86, and 174-187, wherein the spacer sequence is capable of hybridizing with the human HAVCR2 gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the HAVCR2 gene locus is edited in at least 1.5% of the cells.

[0026] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 748-749 and 753-754, wherein the spacer sequence is capable of hybridizing with the human IL7R gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the IL7R gene locus is edited in at least 1.5% of the cells.

[0027] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 59, 87, 88, and 188-198, wherein the spacer sequence is capable of hybridizing with the human LAG3 gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the LAG3 gene locus is edited in at least 1.5% of the cells.

[0028] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises the nucleotide sequence of SEQ ID NO: 757, wherein the spacer sequence is capable of hybridizing with the human LCK gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the LCK gene locus is edited in at least 1.5% of the cells.

[0029] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 60, 89-92, and 199-201, wherein the spacer sequence is capable of hybridizing with the human PDCD1 gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the PDCD1 gene locus is edited in at least 1.5% of the cells.

[0030] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of of SEQ ID NOs: 759 and 761-762, wherein the spacer sequence is capable of hybridizing with the human PLCG1 gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the PLCG1 gene locus is edited in at least 1.5% of the cells.

[0031] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 61, 93-104, and 202-213, wherein the spacer sequence is capable of hybridizing with the human PTPN6 gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the PTPN6 gene locus is edited in at least 1.5% of the cells.

[0032] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 62, 105, and 214-217, wherein the spacer sequence is capable of hybridizing with the human TIGIT gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the TIGIT gene locus is edited in at least 1.5% of the cells.

[0033] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63, 106-130, and 218-241, wherein the spacer sequence is capable of hybridizing with the human TRAC gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the TRAC gene locus is edited in at least 1.5% of the cells.

[0034] In certain embodiments of the engineered, non-naturally occurring system, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 705-706, 711-712, 714-715, 717, and 719-720, wherein the spacer sequence is capable of hybridizing with the human TRBC2 gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the TRBC2 gene locus is edited in at least 1.5% of the cells. In certain embodiments, the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 705-706, wherein the spacer sequence is capable of hybridizing with both the human

TRBC1 gene and the human TRBC2 gene. In certain embodiments, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the TRBC1 gene locus is edited in at least 1.5% of the cells.

[0035] In certain embodiments of the engineered, non-naturally occurring system, genomic mutations are detected in no more than 2% of the cells at any off-target loci by CIRCLE-Seq. In certain embodiments, genomic mutations are detected in no more than 1% of the cells at any off-target loci by CIRCLE-Seq.

[0036] In another aspect, the present invention provides a human cell comprising an engineered, non-naturally occurring system disclosed herein.

[0037] In another aspect, the present invention provides a composition comprising a guide nucleic acid, engineered, non-naturally occurring system, or human cell disclosed herein.

[0038] In another aspect, the present invention provides a method of cleaving a target DNA comprising the sequence of a preselected target gene or a portion thereof, the method comprising contacting the target DNA with an engineered, non-naturally occurring system disclosed herein, thereby resulting in cleavage of the target DNA. In certain embodiments, the contacting occurs *in vitro*. In certain embodiments, the contacting occurs in a cell *ex vivo*. In certain embodiments, the target DNA is genomic DNA of the cell.

[0039] In another aspect, the present invention provides a method of editing human genomic sequence at a preselected target gene locus, the method comprising delivering an engineered, non-naturally occurring system disclosed herein into a human cell, thereby resulting in editing of the genomic sequence at the target gene locus in the human cell. In certain embodiments, the cell is an immune cell. In certain embodiments, the immune cell is a T lymphocyte.

[0040] In certain embodiments, the method of editing human genomic sequence at a preselected target gene locus comprises delivering an engineered, non-naturally occurring system disclosed herein into a population of human cells, thereby resulting in editing of the genomic sequence at the target gene locus in at least a portion of the human cells. In certain embodiments, the population of human cells comprises human immune cells. In certain embodiments, the population of human cells is an isolated population of human immune cells. In certain embodiments, the immune cells are T lymphocytes.

[0041] In certain embodiments of the method of editing human genomic sequence at a preselected target gene locus, the engineered, non-naturally occurring system is delivered into the cell(s) as a pre-formed RNP complex. In certain embodiments, the pre-formed RNP complex is delivered into the cell(s) by electroporation.

[0042] In certain embodiments, the target gene is human ADORA2A gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 51 and 131-137. In certain embodiments, the genomic sequence at the ADORA2A gene locus is edited in at least 1.5% of the human cells.

[0043] In certain embodiments, the target gene is human B2M gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 52, 64-66, 138-145, 622, 625-626, and 634-635,. In certain embodiments, the genomic sequence at the B2M gene locus is edited in at least 1.5% of the human cells.

[0044] In certain embodiments, the target gene is human CD52 gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 53 and 146. In certain embodiments, the genomic sequence at the CD52 gene locus is edited in at least 1.5% of the human cells.

[0045] In certain embodiments, the target gene is human CD247 gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 724, 726-727, 730-732, 735-738, 741-742, and 744-745. In certain embodiments, the genomic sequence at the CD247 gene locus is edited in at least 1.5% of the human cells.

[0046] In certain embodiments, the target gene is human CIITA gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 54, 147-148, 636-640, 642, 644-648, 650-652, 655-656, 660-663, 666, 668, 670-671, 673-676, 678-679, and 682-685. In certain embodiments, the genomic sequence at the CIITA gene locus is edited in at least 1.5% of the human cells.

[0047] In certain embodiments, the target gene is human CTLA4 gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 55, 67-70, and 149-155. In certain embodiments, the genomic sequence at the CTLA4 gene locus is edited in at least 1.5% of the human cells.

[0048] In certain embodiments, the target gene is human DCK gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID

NOs: 56, 71-74, and 156-159. In certain embodiments, the genomic sequence at the DCK gene locus is edited in at least 1.5% of the human cells.

[0049] In certain embodiments, the target gene is human FAS gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 57, 75-79, and 160-173. In certain embodiments, the genomic sequence at the FAS gene locus is edited in at least 1.5% of the human cells.

[0050] In certain embodiments, the target gene is human HAVCR2 gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 58, 80-86, and 174-187. In certain embodiments, the genomic sequence at the HAVCR2 gene locus is edited in at least 1.5% of the human cells.

[0051] In certain embodiments, the target gene is human IL7R gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 748-749 and 753-754. In certain embodiments, the genomic sequence at the IL7R gene locus is edited in at least 1.5% of the human cells.

[0052] In certain embodiments, the target gene is human LAG3 gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 59, 87, 88, and 188-198. In certain embodiments, the genomic sequence at the LAG3 gene locus is edited in at least 1.5% of the human cells.

[0053] In certain embodiments, the target gene is human LCK gene, wherein the spacer sequence comprises the nucleotide sequence of SEQ ID NO: 757. In certain embodiments, the genomic sequence at the LCK gene locus is edited in at least 1.5% of the human cells.

[0054] In certain embodiments, the target gene is human PDCD1 gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 60, 89-92, and 199-201. In certain embodiments, the genomic sequence at the PDCD1 gene locus is edited in at least 1.5% of the human cells.

[0055] In certain embodiments, the target gene is human PLCG1 gene, wherein the spacer sequence comprises a sequence of SEQ ID NO: 759 and 761-762. In certain embodiments, the genomic sequence at the PLCG1 gene locus is edited in at least 1.5% of the human cells.

[0056] In certain embodiments, the target gene is human PTPN6 gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID

NOs: 61, 93-104, and 202-213. In certain embodiments, the genomic sequence at the PTPN6 gene locus is edited in at least 1.5% of the human cells.

[0057] In certain embodiments, the target gene is human TIGIT gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 62, 105, and 214-217. In certain embodiments, the genomic sequence at the TIGIT gene locus is edited in at least 1.5% of the human cells.

[0058] In certain embodiments, the target gene is human TRAC gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63, 106-130, and 218-241. In certain embodiments, the genomic sequence at the TRAC gene locus is edited in at least 1.5% of the human cells.

[0059] In certain embodiments, the target gene is human TRBC2 gene, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 705-706, 711-712, 714-715, 717, and 719-720. In certain embodiments, the genomic sequence at the TRBC2 gene locus is edited in at least 1.5% of the human cells. In certain embodiments, the method further results in editing of the genomic sequence at human TRBC1 gene locus in the human cell, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 705-706. In certain embodiments, the genomic sequence at the TRBC1 gene locus is edited in at least 1.5% of the human cells.

[0060] In certain embodiments, genomic mutations are detected in no more than 2% of the cells at any off-target loci by CIRCLE-Seq. In certain embodiments, genomic mutations are detected in no more than 1% of the cells at any off-target loci by CIRCLE-Seq.

BRIEF DESCRIPTION OF THE DRAWINGS

[0061] **Figure 1A** is a schematic representation showing the structure of an exemplary single guide type V-A CRISPR system. **Figure 1B** is a schematic representation showing the structure of an exemplary dual guide type V-A CRISPR system.

[0062] **Figures 2A-2C** are a series of schematic representation showing incorporation of a protecting group (*e.g.*, a protective nucleotide sequence or a chemical modification) (**Figure 2A**), a donor template-recruiting sequence (**Figure 2B**), and an editing enhancer (**Figure 2C**) into a type V-A CRISPR-Cas system. These additional elements are shown in the context of a dual guide type V-A CRISPR system, but it is understood that they can also

be present other CRISPR systems, including a single guide type V-A CRISPR system, a single guide type II CRISPR system, or a dual guide type II CRISPR system.

DETAILED DESCRIPTION OF THE INVENTION

[0063] The present invention is based, in part, upon the development of engineered CRISPR-Cas systems (*e.g.*, type V-A CRISPR-Cas systems) that can be used to target, edit, or otherwise modify specific target nucleotide sequences in human ADORA2A, B2M, CD52, CIITA, CTLA4, DCK, FAS, HAVCR2 (also called TIM3), LAG3, PDCD1 (also called PD-1), PTPN6, TIGIT, TRAC, TRBC1, TRBC2, CARD11, CD247, IL7R, LCK, or PLCG1 gene. In particular, guide nucleic acids, such as single guide nucleic acids and dual guide nucleic acids, can be designed to hybridize with the selected target nucleotide sequence and activate a Cas nuclease to edit the human genes. CRISPR-Cas systems comprising such guide nucleic acids are also useful for targeting or modifying the human genes.

[0064] A CRISPR-Cas system generally comprises a Cas protein and one or more guide nucleic acids (*e.g.*, RNAs). The Cas protein can be directed to a specific location in a double-stranded DNA target by recognizing a protospacer adjacent motif (PAM) in the non-target strand of the DNA, and the one or more guide nucleic acids can be directed to a specific location by hybridizing with a target nucleotide sequence in the target strand of the DNA. Both PAM recognition and target nucleotide sequence hybridization are required for stable binding of a CRISPR-Cas complex to the DNA target and, if the Cas protein has an effector function (*e.g.*, nuclease activity), activation of the effector function. As a result, when creating a CRISPR-Cas system, a guide nucleic acid can be designed to comprise a nucleotide sequence called spacer sequence that hybridizes with a target nucleotide sequence, where target nucleotide sequence is located adjacent to a PAM in an orientation operable with the Cas protein. It has been observed that not all CRISPR-Cas systems designed by these criteria are equally effective. The present invention identifies target nucleotide sequences in particular human genes that can be efficiently edited, and provides CRISPR-Cas systems directed to these target nucleotide sequences.

[0065] Naturally occurring Type V-A, type V-C, and type V-D CRISPR-Cas systems lack a tracrRNA and rely on a single crRNA to guide the CRISPR-Cas complex to the target DNA. Dual guide nucleic acids capable of activating type V-A, type V-C, or type V-D Cas nucleases have been developed, for example, by splitting the single crRNA into a targeter nucleic acid and a modulator nucleic acid (see, U.S. Provisional Patent Application No.

62/910,055). Naturally occurring type V-A Cas proteins comprise a RuvC-like nuclease domain but lack an HNH endonuclease domain, and recognize a 5' T-rich PAM located immediately upstream from the target nucleotide sequence, the orientation determined using the non-target strand (*i.e.*, the strand not hybridized with the spacer sequence) as the coordinate. The CRISPR-Cas systems cleave a double-stranded DNA to generate a staggered double-stranded break rather than a blunt end. The cleavage site is distant from the PAM site (*e.g.*, separated by at least 10, 11, 12, 13, 14, or 15 nucleotides downstream from the PAM on the non-target strand and/or separated by at least 15, 16, 17, 18, or 19 nucleotides upstream from the sequence complementary to PAM on the target strand).

[0066] Naturally occurring type II CRISPR-Cas systems (*e.g.*, CRISPR-Cas9 systems) generally comprise two guide nucleic acids, called crRNA and tracrRNA, which form a complex by nucleotide hybridization. Single guide nucleic acids capable of activating type II Cas nucleases have been developed, for example, by linking the crRNA and the tracrRNA (see, *e.g.*, U.S. Patent Application Publication Nos. 2014/0242664 and 2014/0068797). Naturally occurring type II Cas proteins comprise a RuvC-like nuclease domain and an HNH endonuclease domain, and recognize a 3' G-rich PAM located immediately downstream from the target nucleotide sequence, the orientation determined using the non-target strand (*i.e.*, the strand not hybridized with the spacer sequence) as the coordinate. The CRISPR-Cas systems cleave a double-stranded DNA to generate a blunt end. The cleavage site is generally 3-4 nucleotides upstream from the PAM on the non-target strand.

[0067] Elements in an exemplary single guide type V-A CRISPR-Cas system are shown in **Figure 1A**. The single guide nucleic acid is also called a "crRNA" where it is present in the form of an RNA. It comprises, from 5' to 3', an optional 5' tail, a modulator stem sequence, a loop, a targeter stem sequence complementary to the modulator stem sequence, and a spacer sequence that hybridizes with the target strand of the target DNA. Where a 5' tail is present, the sequence including the 5' tail and the modulator stem sequence is also called a "modulator sequence" herein. A fragment of the single guide nucleic acid from the optional 5' tail to the targeter stem sequence, also called a "scaffold sequence" herein, bind the Cas protein. In addition, the PAM in the non-target strand of the target DNA binds the Cas protein.

[0068] Elements in an exemplary dual guide type V-A CRISPR-Cas system are shown in **Figure 1B**. The first guide nucleic acid, called "modulator nucleic acid" herein, comprises, from 5' to 3', an optional 5' tail and a modulator stem sequence. Where a 5' tail is present,

the sequence including the 5' tail and the modulator stem sequence is also called a "modulator sequence" herein. The second guide nucleic acid, called "targeter nucleic acid" herein, comprises, from 5' to 3', a targeter stem sequence complementary to the modulator stem sequence and a spacer sequence that hybridizes with the target strand of the target DNA. The duplex between the modulator stem sequence and the targeter stem sequence, plus the optional 5' tail, constitute a structure that binds the Cas protein. In addition, the PAM in the non-target strand of the target DNA binds the Cas protein.

[0069] The terms "targeter stem sequence" and "modulator stem sequence," as used herein, refer to a pair of nucleotide sequences in one or more guide nucleic acids that hybridize with each other. When a targeter stem sequence and a modulator stem sequence are contained in a single guide nucleic acid, the targeter stem sequence is proximal to a spacer sequence designed to hybridize with a target nucleotide sequence, and the modulator stem sequence is proximal to the targeter stem sequence. When a targeter stem sequence and a modulator stem sequence are in separate nucleic acids, the targeter stem sequence is in the same nucleic acid as a spacer sequence designed to hybridize with a target nucleotide sequence. In a CRISPR-Cas system that naturally includes separate crRNA and tracrRNA (*e.g.*, a type II system), the duplex formed between the targeter stem sequence and the modulator stem sequence corresponds to the duplex formed between the crRNA and the tracrRNA. In a CRISPR-Cas system that naturally includes a single crRNA but no tracrRNA (*e.g.*, a type V-A system), the duplex formed between the targeter stem sequence and the modulator stem sequence corresponds to the stem portion of a stem-loop structure in the scaffold sequence (also called direct repeat sequence) of the crRNA. It is understood that 100% complementarity is not required between the targeter stem sequence and the modulator stem sequence. In a type V-A CRISPR-Cas system, however, the targeter stem sequence is typically 100% complementary to the modulator stem sequence.

[0070] The term "targeter nucleic acid," as used herein in the context of a dual guide CRISPR-Cas system, refers to a nucleic acid comprising (i) a spacer sequence designed to hybridize with a target nucleotide sequence; and (ii) a targeter stem sequence capable of hybridizing with an additional nucleic acid to form a complex, wherein the complex is capable of activating a Cas nuclease (*e.g.*, a type II or type V-A Cas nuclease) under suitable conditions, and wherein the targeter nucleic acid alone, in the absence of the additional nucleic acid, is not capable of activating the Cas nuclease under the same conditions.

[0071] The term "modulator nucleic acid," as used herein in connection with a given targeter nucleic acid and its corresponding Cas nuclease, refers to a nucleic acid capable of hybridizing with the targeter nucleic acid to form a complex, wherein the complex, but not the modulator nucleic acid alone, is capable of activating the type Cas nuclease under suitable conditions.

[0072] The term "suitable conditions," as used in connection with the definitions of "targeter nucleic acid" and "modulator nucleic acid," refers to the conditions under which a naturally occurring CRISPR-Cas system is operative, such as in a prokaryotic cell, in a eukaryotic (*e.g.*, mammalian or human) cell, or in an *in vitro* assay.

[0073] The features and uses of the guide nucleic acids and CRISPR-Cas systems are discussed in the following sections.

I. Guide Nucleic Acids and Engineered, Non-Naturally Occurring CRISPR-Cas Systems

[0074] The present invention provides a guide nucleic acid comprising a targeter stem sequence and a spacer sequence, wherein the spacer sequence comprises a nucleotide sequence listed Table 1, 2, or 3, or a portion thereof sufficient to hybridize with the corresponding target gene listed in the table. In particular, Table 1 lists the guide nucleic acid that showed the best editing efficiency for each target gene using the method described in Example 1. Table 2 lists the guide nucleic acids that showed at least 10% editing efficiency using the method described in Example 1. Table 3 lists the guide nucleic acids that showed at least 1.5% and lower than 10% editing efficiency using the method described in Example 1.

[0075] In certain embodiments, a guide nucleic acid of the present invention is capable of binding the genomic locus of the corresponding target gene in the human genome. In certain embodiments, a guide nucleic acid of the present invention, alone or in combination with a modulator nucleic acid, is capable of directing a Cas protein to the genomic locus of the corresponding target gene in the human genome. In certain embodiments, a guide nucleic acid of the present invention, alone or in combination with a modulator nucleic acid, is capable of directing a Cas nuclease to the genomic locus of the corresponding target gene in the human genome, thereby resulting in cleavage of the genomic DNA at the genomic locus.

Table 1. Selected Spacer Sequences Targeting Human Genes

Target Gene	crRNA	Spacer Sequence	SEQ ID NO
TRAC	gTRAC006	TGAGGGTGAAGGATAGACGCT	63
ADORA2A	gADORA2A_12	AGGATGTGGTCCCCATGAACT	51
B2M	gB2M_41	ATAGATCGAGACATGTAAGCA	635
CARD11	gCARD11_1	TAGTACCGCTCCTGGAAGGTT	721
CD247	gCD247_12	CTAGCAGAGAAGGAAGAACCC	735
CD52	gCD52_1	CTCTTCCTCCTACTCACCATC	53
CIITA	gCIITA_32	CCTTGGGGCTCTGACAGGTAG	636
CTLA4	gCTLA4_4	AGCGGCACAAGGCTCAGCTGA	55
DCK	gDCK_6	CGGAGGCTCCTTACCGATGTT	56
FAS	gFAS_36	GTGTTGCTGGTGAGTGTGCAT	57
HAVCR2	gTIM3_6	CTTGTAAGTAGTAGCAGCAGC	58
IL7R	gIL7R_3	CAGGGGAGATGGATCCTATCT	749
LAG3	gLAG3_6	GGGTGCATACCTGTCTGGCTG	59
LCK	gLCK1_3	ACCCATCAACCCGTAGGGATG	757
PDCD1	gPD_23	TCTGCAGGGACAATAGGAGCC	60
PLCG1	gPLCG1_2	CCTTTCTGCGCTTCGTGGTGT	759
PTPN6	gPTPN6_6	TATGACCTGTATGGAGGGGAG	61
TIGIT	gTIGIT_2	AGGCCTTACCTGAGGCGAGGG	62
TRBC1+2	gTRBC1+2_3	CGCTGTCAAGTCCAGTTCTAC	706
TRBC2	gTRBC2_12	CCGGAGGTGAAGCCACAGTCT	712

Table 2. Selected Spacer Sequences Targeting Human Genes

Target Gene	crRNA	Spacer sequence	SEQ ID NO
ADORA2A	gADORA2A_12	AGGATGTGGTCCCCATGAACT	51
B2M	gB2M_4	CTCACGTCATCCAGCAGAGAA	52
B2M	gB2M_7	ACTTTCCATTCTCTGCTGGAT	64
B2M	gB2M_2	TGGCCTGGAGGCTATCCAGCG	65
B2M	gB2M_17	TATCTCTTGTACTACACTGAA	66
B2M	gB2M_30	AGTGGGGGTGAATTCAGTGTA	625
B2M	gB2M_41	ATAGATCGAGACATGTAAGCA	635
CIITA	gCIITA_32	CCTTGGGGCTCTGACAGGTAG	636
CIITA	gCIITA_33	ACCTTGGGGCTCTGACAGGTA	637
CIITA	gCIITA_35	CTCCCAGAACCCGACACAGAC	639

Target Gene	crRNA	Spacer sequence	SEQ ID NO
CIITA	gCIITA_36	TGGGCTCAGGTGCTTCCTCAC	640
CIITA	gCIITA_38	CTTGTCTGGGCAGCGGAACTG	642
CIITA	gCIITA_40	TCAAAGTAGAGCACATAGGAC	644
CIITA	gCIITA_41	TGCCCAACTTCTGCTGGCATC	645
CIITA	gCIITA_43	TCTGCAGCCTTCCCAGAGGAG	647
CIITA	gCIITA_44	TCCAGGCGCATCTGGCCGGAG	648
CIITA	gCIITA_48	CTCGGGAGGTCAGGGCAGGTT	652
CIITA	gCIITA_57	CAGAAGAAGCTGCTCCGAGGT	660
CIITA	gCIITA_59	AGAGCTCAGGGATGACAGAGC	662
CIITA	gCIITA_60	TGCCGGGCAGTGTGCCAGCTC	663
CIITA	gCIITA_63	GCCACTCAGAGCCAGCCACAG	666
CIITA	gCIITA_65	GCAGCACGTGGTACAGGAGCT	668
CIITA	gCIITA_67	TGGGCACCCGCCTCACGCCTC	670
CIITA	gCIITA_70	CCAGGTCTTCCACATCCTTCA	673
CIITA	gCIITA_71	AAAGCCAAGTCCCTGAAGGAT	674
CIITA	gCIITA_72	GGTCCCGAACAGCAGGGAGCT	675
CIITA	gCIITA_73	TTTAGGTCCCGAACAGCAGGG	676
CIITA	gCIITA_76	GGGAAAGCCTGGGGGCCTGAG	679
CIITA	gCIITA_80	CAAGGACTTCAGCTGGGGGAA	682
CIITA	gCIITA_81	TAGGCACCCAGGTCAGTGATG	683
CIITA	gCIITA_82	CGACAGCTTGTACAATAACTG	684
CD247	gCD247_1	TGTGTTGCAGTTCAGCAGGAG	724
CD247	gCD247_3	CGGAGGGTCTACGGCGAGGCT	726
CD247	gCD247_4	TTATCTGTTATAGGAGCTCAA	727
CD247	gCD247_8	GACAAGAGACGTGGCCGGGAC	731
CD247	gCD247_12	CTAGCAGAGAAGGAAGAACCC	735
CD247	gCD247_15	ATCCCAATCTCACTGTAGGCC	738
CD247	gCD247_18	TCATTTCACTCCCAAACAACC	741
CD247	gCD247_19	ACTCCCAAACAACCAGCGCCG	742
CD52	gCD52_1	CTCTTCCTCCTACTCACCATC	53
CIITA	gCIITA_4	TAGGGGCCCCAACTCCATGGT	54
CTLA4	gCTLA4_4	AGCGGCACAAGGCTCAGCTGA	55

Target Gene	crRNA	Spacer sequence	SEQ ID NO
CTLA4	gCTLA4_14	CCTGGAGATGCATACTCACAC	67
CTLA4	gCTLA4_6	CAGAAGACAGGGATGAAGAGA	68
CTLA4	gCTLA4_19	CACTGGAGGTGCCCGTGCAGA	69
CTLA4	gCTLA4_13	TGTGTGAGTATGCATCTCCAG	70
DCK	gDCK_6	CGGAGGCTCCTTACCGATGTT	56
DCK	gDCK_2	TCAGCCAGCTCTGAGGGGACC	71
DCK	gDCK_8	CTCACAACAGCTGCAGGGAAG	72
DCK	gDCK_26	AGCTTGCCATTCAGAGAGGCA	73
DCK	gDCK_30	TACATACCTGTCACTATACAC	74
FAS	gFAS_36	GTGTTGCTGGTGAGTGTGCAT	57
FAS	gFAS_34	TTTTTCTAGATGTGAACATGG	75
FAS	gFAS_35	ATGATTCCATGTTCACATCTA	76
FAS	gFAS_12	GTGTAACATACCTGGAGGACA	77
FAS	gFAS_1	GGAGGATTGCTCAACAACCAT	78
FAS	gFAS_59	TAGGAAACAGTGGCAATAAAT	79
HAVCR2	gTIM3_6	CTTGTAAGTAGTAGCAGCAGC	58
HAVCR2	gTIM3_29	CAAGGATGCTTACCACCAGGG	80
HAVCR2	gTIM3_6	TAAGTAGTAGCAGCAGCAGCA	81
HAVCR2	gTIM3_32	TATCAGGGAGGCTCCCCAGTG	82
HAVCR2	gTIM3_30	CCACCAGGGGACATGGCCCAG	83
HAVCR2	gTIM3_12	AATGTGGCAACGTGGTGCTCA	84
HAVCR2	gTIM3_25	TGACATTAGCCAAGGTCACCC	85
HAVCR2	gTIM3_18	CGCAAAGGAGATGTGTCCCTG	86
IL7R	gIL7R_3	CAGGGGAGATGGATCCTATCT	749
IL7R	gIL7R_8	CATAACACACAGGCCAAGATG	754
LAG3	gLAG3_6	GGGTGCATACCTGTCTGGCTG	59
LAG3	gLAG3_38	TCAGGACCTTGGCTGGAGGCA	87
LAG3	gLAG3_33	GGTCACCTGGATCCCTGGGGA	88
LCK	gLCK1_3	ACCCATCAACCCGTAGGGATG	757
PDCD1	gPD_23	TCTGCAGGGACAATAGGAGCC	60
PDCD1	gPD_2	CCTTCCGCTCACCTCCGCCTG	89
PDCD1	gPD_8	GCACGAAGCTCTCCGATGTGT	90

Target Gene	crRNA	Spacer sequence	SEQ ID NO
PDCD1	gPD_29	CTAGCGGAATGGGCACCTCAT	91
PDCD1	gPD_27	CAGTGGCGAGAGAAGACCCCG	92
PTPN6	gPTPN6_6	TATGACCTGTATGGAGGGGAG	61
PTPN6	gPTPN6_46	ACTGCCCCCCACCCAGGCCTG	93
PTPN6	gPTPN6_7	CGACTCTGACAGAGCTGGTGG	94
PTPN6	gPTPN6_26	CAGAAGCAGGAGGTGAAGAAC	95
PTPN6	gPTPN6_1	ACCGAGACCTCAGTGGGCTGG	96
PTPN6	gPTPN6_37	TGGGCCCTACTCTGTGACCAA	97
PTPN6	gPTPN6_16	TGTGCTCAGTGACCAGCCCAA	98
PTPN6	gPTPN6_25	CCCACCCACATCTCAGAGTTT	99
PTPN6	gPTPN6_12	TTGTGCGTGAGAGCCTCAGCC	100
PTPN6	gPTPN6_22	AAGAAGACGGGGATTGAGGAG	101
PTPN6	gPTPN6_5	TCCCCTCCATACAGGTCATAG	102
PTPN6	gPTPN6_19	GCTCCCCCCAGGGTGGACGCT	103
PTPN6	gPTPN6_14	GGCTGGTCACTGAGCACAGAA	104
TIGIT	gTIGIT_2	AGGCCTTACCTGAGGCGAGGG	62
TIGIT	gTIGIT_18	GTCCTCCCTCTAGTGGCTGAG	105
TRAC	gTRAC006	TGAGGGTGAAGGATAGACGCT	63
TRAC	gTRAC073	GCAGACAGGGAGAAATAAGGA	106
TRAC	gTRAC017	CAGGTGAAATTCCTGAGATGT	107
TRAC	gTRAC059	GACATCATTGACCAGAGCTCT	108
TRAC	gTRAC078	CCAGCTCACTAAGTCAGTCTC	109
TRAC	gTRAC012	TATGGAGAAGCTCTCATTTCT	110
TRAC	gTRAC039	TAAGATGCTATTTCCCGTATA	111
TRAC	gTRAC067	CCGTGTCATTCTCTGGACTGC	112
TRAC	gTRAC079	ATTCCTCCACTTCAACACCTG	113
TRAC	gTRAC038	TACGGGAAATAGCATCTTAGA	114
TRAC	gTRAC061	GTGGCAATGGATAAGGCCGAG	115
TRAC	gTRAC058	CTTGCTTCAGGAATGGCCAGG	116
TRAC	gTRAC021	TAGTTCAAAACCTCTATCAAT	117
TRAC	gTRAC049	TCTGTGATATACACATCAGAA	118
TRAC	gTRAC074	GGCAGACAGGGAGAAATAAGG	119

Target Gene	crRNA	Spacer sequence	SEQ ID NO
TRAC	gTRAC018	CTCGATATAAGGCCTTGAGCA	120
TRAC	gTRAC043	GAGTCTCTCAGCTGGTACACG	121
TRAC	gTRAC075	TGGCAGACAGGGAGAAATAAG	122
TRAC	gTRAC082	CCAGCTGACAGATGGGCTCCC	123
TRAC	gTRAC040	CCGTATAAAGCATGAGACCGT	124
TRAC	gTRAC041	CCCCAACCCAGGCTGGAGTCC	125
TRAC	gTRAC076	TTGGCAGACAGGGAGAAATAA	126
TRAC	gTRAC014	TCAGAAGAGCCTGGCTAGGAA	127
TRAC	gTRAC029	CTCTGCCAGAGTTATATTGCT	128
TRAC	gTRAC028	CCATGCCTGCCTTTACTCTGC	129
TRAC	gTRAC050	GTCTGTGATATACACATCAGA	130
TRBC1+2	gTRBC1+2_1	AGCCATCAGAAGCAGAGATCT	705
TRBC1+2	gTRBC1+2_3	CGCTGTCAAGTCCAGTTCTAC	706
TRBC2	gTRBC2_11	AGACTGTGGCTTCACCTCCGG	711
TRBC2	gTRBC2_12	CCGGAGGTGAAGCCACAGTCT	712
TRBC2	gTRBC2_15	CTAGGGAAGGCCACCTTGTAT	715
TRBC2	gTRBC2_21	GAGCTAGCCTCTGGAATCCTT	720

Table 3. Selected Spacer Sequences Targeting Human Genes

Target Gene	crRNA	Spacer sequence	SEQ ID NO
ADORA2A	gADORA2A_16	CGGATCTTCCTGGCGGCGCGA	131
ADORA2A	gADORA2A_28	AAGGCAGCTGGCACCAGTGCC	132
ADORA2A	gADORA2A_2	TGGTGTCACTGGCGGCGGCCG	133
ADORA2A	gADORA2A_23	TTCTGCCCCGACTGCAGCCAC	134
ADORA2A	gADORA2A_7	GTGACCGGCACGAGGGCTAAG	135
ADORA2A	gADORA2A_8	CCATCGGCCTGACTCCCATGC	136
ADORA2A	gADORA2A_4	CCATCACCATCAGCACCGGGT	137
B2M	gB2M_21	TCACAGCCCAAGATAGTTAAG	138
B2M	gB2M_8	CTGAATTGCTATGTGTCTGGG	139
B2M	gB2M_11	CTGAAGAATGGAGAGAGAATT	140
B2M	gB2M_18	TCAGTGGGGGTGAATTCAGTG	141

Target Gene	crRNA	Spacer sequence	SEQ ID NO
B2M	gB2M_5	CATTCTCTGCTGGATGACGTG	142
B2M	gB2M_10	ATCCATCCGACATTGAAGTTG	143
B2M	gB2M_22	CCCCACTTAACTATCTTGGGC	144
B2M	gB2M_1	GCTGTGCTCGCGCTACTCTCT	145
B2M	gB2M_27	AATTCTCTCTCCATTCTTCAG	622
B2M	gB2M_31	CAGTGGGGGTGAATTCAGTGT	626
B2M	gB2M_40	CATAGATCGAGACATGTAAGC	634
CD247	gCD247_7	CCCCCATCTCAGGGTCCCGGC	730
CD247	gCD247_9	TCTCCCTCTAACGTCTTCCCG	732
CD247	gCD247_13	TGCAGTTCCTGCAGAAGAGGG	736
CD247	gCD247_14	TGCAGGAACTGCAGAAAGATA	737
CD247	gCD247_21	TGATTTGCTTTCACGCCAGGG	744
CD247	gCD247_22	CTTTCACGCCAGGGTCTCAGT	745
CD52	gCD52_4	GCTGGTGTCGTTTTGTCCTGA	146
CIITA	gCIITA_18	TGCTGGCATCTCCATACTCTC	147
CIITA	gCIITA_29	GTCTCTTGCAGTGCCTTTCTC	148
CIITA	gCIITA_34	CCGGCCTTTTTACCTTGGGGC	638
CIITA	gCIITA_42	TGACTTTTCTGCCCAACTTCT	646
CIITA	gCIITA_46	CCAGAGCCCATGGGGCAGAGT	650
CIITA	gCIITA_47	TCCCCACCATCTCCACTCTGC	651
CIITA	gCIITA_51	CAGAGCCGGTGGAGCAGTTCT	655
CIITA	gCIITA_52	CCCAGCACAGCAATCACTCGT	656
CIITA	gCIITA_55	AGCCACATCTTGAAGAGACCT	658
CIITA	gCIITA_58	AGCTGTCCGGCTTCTCCATGG	661
CIITA	gCIITA_68	CCCCTCTGGATTGGGGAGCCT	671
CIITA	gCIITA_75	CCTCCTAGGCTGGGCCCTGTC	678
CIITA	gCIITA_83	TCTTGCCAGCGTCCAGTACAA	685
CTLA4	gCTLA4_27	CTGTTGCAGATCCAGAACCGT	149
CTLA4	gCTLA4_36	ACAGCTAAAGAAAAGAAGCCC	150
CTLA4	gCTLA4_41	TCAATTGATGGGAATAAAATA	151
CTLA4	gCTLA4_28	CTCCTCTGGATCCTTGCAGCA	152
CTLA4	gCTLA4_37	CACATAGACCCCTGTTGTAAG	153

Target Gene	crRNA	Spacer sequence	SEQ ID NO
CTLA4	gCTLA4_18	CTAGATGATTCCATCTGCACG	154
CTLA4	gCTLA4_5	TTCTTCTCTTCATCCCTGTCT	155
DCK	gDCK_9	AGGATATTCACAAATGTTGAC	156
DCK	gDCK_22	GAAGGTAAAAGACCATCGTTC	157
DCK	gDCK_21	TCATACATCATCTGAAGAACA	158
DCK	gDCK_7	ATCTTTCCTCACAACAGCTGC	159
FAS	gFAS_47	AGTGAAGAGAAAGGAAGTACA	160
FAS	gFAS_45	TTTGTTCTTTCAGTGAAGAGA	161
FAS	gFAS_25	CTAGGCTTAGAAGTGGAAATA	162
FAS	gFAS_10	GAAGGCCTGCATCATGATGGC	163
FAS	gFAS_32	GTGCAAGGGTCACAGTGTTCA	164
FAS	gFAS_5	GGACGATAATCTAGCAACAGA	165
FAS	gFAS_14	TTCCTTGGGCAGGTGAAAGGA	166
FAS	gFAS_29	GTTTACATCTGCACTTGGTAT	167
FAS	gFAS_33	CTTGGTGCAAGGGTCACAGTG	168
FAS	gFAS_71	CTGTTCTGCTGTGTCTTGGAC	169
FAS	gFAS_38	CTCTTTGCACTTGGTGTTGCT	170
FAS	gFAS_70	TGTTCTGCTGTGTCTTGGACA	171
FAS	gFAS_4	ACAGGTTCTTACGTCTGTTGC	172
FAS	gFAS_15	GGCAGGTGAAAGGAAAGCTAG	173
HAVCR2	gTIM3_42	CTAGGGTATTCTCATAGCAAA	174
HAVCR2	gTIM3_10	CCCCAGCAGACGGGCACGAGG	175
HAVCR2	gTIM3_47	GCCAACCTCCCTCCCTCAGGA	176
HAVCR2	gTIM3_34	TGTTTCCATAGCAAATATCCA	177
HAVCR2	gTIM3_19	GATCCGGCAGCAGTAGATCCC	178
HAVCR2	gTIM3_48	CCAATCCTGAGGGAGGGAGGT	179
HAVCR2	gTIM3_36	CGGGACTCTGGAGCAACCATC	180
HAVCR2	gTIM3_15	GCCAGTATCTGGATGTCCAAT	181
HAVCR2	gTIM3_27	ACTGCAGCCTTTCCAAGGATG	182
HAVCR2	gTIM3_41	CCCCTTACTAGGGTATTCTCA	183
HAVCR2	gTIM3_23	ACCTGAAGTTGGTCATCAAAC	184
HAVCR2	gTIM3_28	CCAAGGATGCTTACCACCAGG	185

Target Gene	crRNA	Spacer sequence	SEQ ID NO
HAVCR2	gTIM3_40	GTTTCCCCCTTACTAGGGTAT	186
HAVCR2	gTIM3_13	ATCAGTCCTGAGCACCACGTT	187
IL7R	gIL7R_2	CCAGGGGAGATGGATCCTATC	748
IL7R	gIL7R_7	TCTGTCGCTCTGTTGGTCATC	753
LAG3	gLAG3_35	TGAGGTGACTCCAGTATCTGG	188
LAG3	gLAG3_41	CCAGCCTTGGCAATGCCAGCT	189
LAG3	gLAG3_37	TGTGGAGCTCTCTGGACACCC	190
LAG3	gLAG3_16	GGGCAGGAAGAGGAAGCTTTC	191
LAG3	gLAG3_46	TCCATAGGTGCCCAACGCTCT	192
LAG3	gLAG3_27	CCACCTGAGGCTGACCTGTGA	193
LAG3	gLAG3_31	CCCAGGGATCCAGGTGACCCA	194
LAG3	gLAG3_3	ACCTGGAGCCACCCAAAGCGG	195
LAG3	gLAG3_25	CCCTTCGACTAGAGGATGTGA	196
LAG3	gLAG3_13	CGCTAAGTGGTGATGGGGGGA	197
LAG3	gLAG3_22	GCAGTGAGGAAAGACCGGGTC	198
PDCD1	gPD_20	CAGAGAGAAGGGCAGAAGTGC	199
PDCD1	gPD_22	GAACTGGCCGGCTGGCCTGGG	200
PDCD1	gPD_18	GTGCCCTTCCAGAGAGAAGGG	201
PLCG1	gPLCG1_2	CCTTTCTGCGCTTCGTGGTGT	759
PLCG1	gPLCG1_4	TGCGCTTCGTGGTGTATGAGG	761
PLCG1	gPLCG1_5	GTGGTGTATGAGGAAGACATG	762
PTPN6	gPTPN6_20	GAGACCTTCGACAGCCTCACG	202
PTPN6	gPTPN6_41	CTGGACCAGATCAACCAGCGG	203
PTPN6	gPTPN6_53	CCCCCCTGCACCCGGCTGCAG	204
PTPN6	gPTPN6_28	CACCAGCGTCTGGAAGGGCAG	205
PTPN6	gPTPN6_42	CTGCCGCTGGTTGATCTGGTC	206
PTPN6	gPTPN6_32	TGGCAGATGGCGTGGCAGGAG	207
PTPN6	gPTPN6_4	CTGGCTCGGCCCAGTCGCAAG	208
PTPN6	gPTPN6_8	AGGTGGATGATGGTGCCGTCG	209
PTPN6	gPTPN6_40	GGGAGACCTGATTCGGGAGAT	210
PTPN6	gPTPN6_48	AATGAACTGGGCGATGGCCAC	211
PTPN6	gPTPN6_10	TCTAGGTGGTACCATGGCCAC	212

Target Gene	crRNA	Spacer sequence	SEQ ID NO
PTPN6	gPTPN6_39	CAGGTCTCCCCGCTGGACAAT	213
TIGIT	gTIGIT_11	GGGTGGCACATCTCCCCATCC	214
TIGIT	gTIGIT_7	TGCAGAGAAAGGTGGCTCTAT	215
TIGIT	gTIGIT_10	TAATGCTGACTTGGGGTGGCA	216
TIGIT	gTIGIT_27	CTCCTGAGGTCACCTTCCACA	217
TRAC	gTRAC066	CTAAGAAACAGTGAGCCTTGT	218
TRAC	gTRAC042	CCTCTTTGCCCCAACCCAGGC	219
TRAC	gTRAC035	AGGTTTCCTTGAGTGGCAGGC	220
TRAC	gTRAC044	AGAATCAAAATCGGTGAATAG	221
TRAC	gTRAC072	CCCCTTACTGCTCTTCTAGGC	222
TRAC	gTRAC062	GGTGGCAATGGATAAGGCCGA	223
TRAC	gTRAC020	GAACTATAAATCAGAACACCT	224
TRAC	gTRAC013	TTTCTCAGAAGAGCCTGGCTA	225
TRAC	gTRAC068	CCCGTGTCATTCTCTGGACTG	226
TRAC	gTRAC025	CTGGGCCTTTTTCCCATGCCT	227
TRAC	gTRAC019	AACTATAAATCAGAACACCTG	228
TRAC	gTRAC048	ATTCTCAAACAAATGTGTCAC	229
TRAC	gTRAC036	CTTGAGTGGCAGGCCAGGCCT	230
TRAC	gTRAC056	CATGTGCAAACGCCTTCAACA	231
TRAC	gTRAC064	TACTAAGAAACAGTGAGCCTT	232
TRAC	gTRAC071	CTCAGACTGTTTGCCCCTTAC	233
TRAC	gTRAC081	TAATTCCTCCACTTCAACACC	234
TRAC	gTRAC030	ATAGGATCTTCTTCAAAACCC	235
TRAC	gTRAC033	GAAGAAGATCCTATTAAATAA	236
TRAC	gTRAC001	TGTTTTTAATGTGACTCTCAT	237
TRAC	gTRAC009	GTACTTTACAGTTTATTAAAT	238
TRAC	gTRAC007	ATAAACTGTAAAGTACCAAAC	239
TRAC	gTRAC084	GACTTTTCCCAGCTGACAGAT	240
TRAC	gTRAC083	CCCAGCTGACAGATGGGCTCC	241
TRBC2	gTRBC2_14	CCAGCAAGGGGTCCTGTCTGC	714
TRBC2	gTRBC2_17	CCATGGCCATCAGCACGAGGG	717
TRBC2	gTRBC2_19	CACAGGTCAAGAGAAAGGATT	719

[0076] The spacer sequences provided in Tables 1-3 are designed based upon identification of target nucleotide sequences associated with a PAM in a given target gene locus, and are selected based upon the editing efficiency detected in human cells.

[0077] To provide sufficient targeting to the target nucleotide sequence, the spacer sequence is generally 16 or more nucleotides in length. In certain embodiments, the spacer sequence is at least 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides in length. In certain embodiments, the spacer sequence is shorter than or equal to 75, 50, 45, 40, 35, 30, 25, or 20 nucleotides in length. Shorter spacer sequence may be desirable for reducing off-target events. Accordingly, in certain embodiments, the spacer sequence is shorter than or equal to 21, 20, 19, 18, or 17 nucleotides. In certain embodiments, the spacer sequence is 17-30 nucleotides in length, *e.g.*, 17-21, 17-22, 17-23, 17-24, 17-25, 17-30, 20-21, 20-22, 20-23, 20-24, 20-25, or 20-30 nucleotides in length. In certain embodiments, the spacer sequence is about 20 nucleotides in length. In certain embodiments, the spacer sequence is about 21 nucleotides in length. In certain embodiments, the spacer sequence is 20 nucleotides in length.

[0078] In certain embodiments, the spacer sequence comprises a portion of a spacer sequence listed in Table 1, 2, or 3, wherein the portion is 16, 17, 18, 19, or 20 nucleotides in length. In certain embodiments, the spacer sequence comprises nucleotides 1-16, 1-17, 1-18, 1-19, or 1-20 of a spacer sequence listed in Table 1, 2, or 3. In specific embodiments, the spacer sequence consists of nucleotides 1-16, 1-17, 1-18, 1-19, or 1-20 of a spacer sequence listed in Table 1, 2, or 3.

[0079] In certain embodiments, the spacer sequence is 21 nucleotides in length. In certain embodiments, the spacer sequence consists of a spacer sequence shown in Table 1, 2, or 3.

[0080] In certain embodiments, the spacer sequence, where it is longer than 21 nucleotides in length, comprises a spacer sequence shown in Table 1, 2, or 3 and one or more nucleotides. In certain embodiments, the one or more nucleotides are 3' to the spacer sequence shown in Table 1, 2, or 3.

[0081] In certain embodiments, the spacer sequence is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% complementary to the target nucleotide sequence. In certain embodiments, the spacer sequence is 100% complementary to the target nucleotide

sequence in the seed region (about 5 base pairs proximal to the PAM). In certain embodiments, the spacer sequence is 100% complementary to the target nucleotide sequence. The spacer sequences listed in Tables 1-3 are designed to be 100% complementary to the wild-type sequence of the corresponding target gene. Accordingly, it is contemplated that a spacer sequence useful for targeting a gene listed in Table 1, 2, or 3 can be at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to a corresponding spacer sequence listed in Table 1, 2, or 3, or a portion thereof disclosed herein. In certain embodiments, the spacer sequence is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides different from a sequence listed in Table 1, 2, or 3. In certain embodiments, the spacer sequence is 100% identical to a sequence listed in Table 1, 2, or 3 in the seed region (about 5 base pairs proximal to the PAM). It has been reported that compared to DNA binding, DNA cleavage is less tolerant to mismatches between the spacer sequence and the target nucleotide sequence (see, Klein *et al.* (2018) CELL REPORTS, 22: 1413). Accordingly, in certain embodiments, a guide nucleic acid to be used with a Cas nuclease comprises a spacer sequence 100% complementary to the target nucleotide sequence. In certain embodiments, a guide nucleic acid to be used with a Cas nuclease comprises a spacer sequence listed in Table 1, 2, or 3, or a portion thereof disclosed herein.

[0082] The present invention also provides guide nucleic acids targeting human DHODH, PLK1, MVD, TUBB, or U6 gene comprising the spacer sequences provided below in Table 25. DHODH, PLK1, MVD, and TUBB are known to be essential genes. It is contemplated that the guide nucleic acids targeting these genes, particularly the ones that edit the respective genomic locus at hight efficiency (e.g., at least 50%, at least 60%, at least 70%, at least 80%, or at least 90%), can be used as positive controls for assessing transfection efficiency and other experimental processes. The spacer sequences targeting U6 in Table 25 are designed to hybridize with the promoter region of human U6 gene and can be used to assess expression of an inserted gene from the endogenous U6 promoter.

Cas Proteins

[0083] The guide nucleic acid of the present invention, either as a single guide nucleic acid alone or as a targeter nucleic acid used in combination with a cognate modulator nucleic acid, is capable of binding a CRISPR Associated (Cas) protein. In certain embodiments, the guide nucleic acid, either as a single guide nucleic acid alone or as a targeter nucleic acid

used in combination with a cognate modulator nucleic acid, is capable of activating a Cas nuclease.

[0084] The terms "CRISPR-Associated protein," "Cas protein," and "Cas," as used interchangeably herein, refer to a naturally occurring Cas protein or an engineered Cas protein. Non-limiting examples of Cas protein engineering includes but are not limited to mutations and modifications of the Cas protein that alter the activity of the Cas, alter the PAM specificity, broaden the range of recognized PAMs, and/or reduce the ability to modify one or more off-target loci as compared to a corresponding unmodified Cas. In certain embodiments, the altered activity of the engineered Cas comprises altered ability (*e.g.*, specificity or kinetics) to bind the naturally occurring crRNA or engineered dual guide nucleic acids, altered ability (*e.g.*, specificity or kinetics) to bind the target nucleotide sequence, altered processivity of nucleic acid scanning, and/or altered effector (*e.g.*, nuclease) activity. A Cas protein having the nuclease activity is referred to as a "CRISPR-Associated nuclease" or "Cas nuclease," as used interchangeably herein.

[0085] In certain embodiments, the Cas protein is a type V-A, type V-C, or type V-D Cas protein. In certain embodiments, the Cas protein is a type V-A Cas protein. In other embodiments, the Cas protein is a type II Cas protein, *e.g.*, a Cas9 protein.

[0086] In certain embodiments, the Cas nuclease is a type V-A, type V-C, or type V-D Cas nuclease. In certain embodiments, the Cas nuclease is a type V-A Cas nuclease. In other embodiments, the Cas protein is a type II Cas nuclease, *e.g.*, a Cas9 nuclease.

[0087] In certain embodiments, the type V-A Cas protein comprises Cpf1. Cpf1 proteins are known in the art and are described in U.S. Patent Nos. 9,790,490 and 10,113,179. Cpf1 orthologs can be found in various bacterial and archaeal genomes. For example, in certain embodiments, the Cpf1 protein is derived from *Francisella novicida U112* (Fn), *Acidaminococcus sp. BV3L6* (As), *Lachnospiraceae bacterium ND2006* (Lb), *Lachnospiraceae bacterium* MA2020 (Lb2), *Candidatus Methanoplasma termitum* (CMt), *Moraxella bovoculi 237* (Mb), *Porphyromonas crevioricanis* (Pc), *Prevotella disiens* (Pd), *Francisella tularensis 1*, *Francisella tularensis subsp. novicida*, *Prevotella albensis*, *Lachnospiraceae bacterium MC2017 1*, *Butyrivibrio proteoclasticus*, *Peregrinibacteria bacterium GW2011_GWA2_33_10*, *Parcubacteria bacterium GW2011_GWC2_44_17*, *Smithella sp. SCADC*, *Eubacterium eligens*, *Leptospira inadai*, *Porphyromonas macacae*, *Prevotella bryantii* (Pb), *Proteocatella sphenisci* (Ps), *Anaerovibrio* sp. RM50 (As2),

Moraxella caprae (Mc), *Lachnospiraceae bacterium* COE1 (Lb3), or *Eubacterium coprostanoligenes* (Ec).

[0088] In certain embodiments, the type V-A Cas protein comprises AsCpf1 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 3. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 3.

AsCpf1 (SEQ ID NO: 3)

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYK
TYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDN
LTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYE
NRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVS
TSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAH
IIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFN
ELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSL
KHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLD
SLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVE
KFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYD
LNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSS
QYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPN
LHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQ
KTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHV
PITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLN
TIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAV
VVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQL
TDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDF
LHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRI
VPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALI
RSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKG
QLLLNHLKESKDLKLQNGISNQDWLAYIQELRN

[0089] In certain embodiments, the type V-A Cas protein comprises LbCpf1 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 4. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 4.

LbCpf1 (SEQ ID NO: 4)

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDR
YYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGY
KSLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCI
NENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGID
VYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEG
YTSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFG
EWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLS
VVEKLKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFE
NYIKAFFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLY
FQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYE
KINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKL
IDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKL
VEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRA
SLKKEELVVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPK
NIFKINTEVRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGI
RIKTDYHSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIAL
EDLNSGFKNSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNK
FESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPE
EDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYK
ELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLISP
VKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKV
KIAISNKEWLEYAQTSVKH

[0090] In certain embodiments, the type V-A Cas protein comprises FnCpf1 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 5. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 5.

FnCpf1 (SEQ ID NO: 5)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE

NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDG
KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV
SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRI
KNNQEGKKLNLVIKNEEYFEFVQNRNN

[0091] In certain embodiments, the type V-A Cas protein comprises PbCpf1 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 6. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 6.

PbCpf1 (SEQ ID NO: 6)

MQINNLKIIYMKFTDFTGLYSLSKTLRFELKPIGKTLENIKKAGLLEQDQHRADSYKK
VKKIIDEYHKAFIEKSLSNFELKYQSEDKLDSLEEYLMYYSMKRIEKTEKDKFAKIQD
NLRKQIADHLKGDESYKTIFSKDLIRKNLPDFVKSDEERTLIKEFKDFTTYFKGFYEN
RENMYSAEDKSTAISHRIIHENLPKFVDNINAFSKIILIPELREKLNQIYQDFEEYLNVE
SIDEIFHLDYFSMVMTQKQIEVYNAIIGGKSTNDKKIQGLNEYINLYNQKHKDCKLPK
LKLLFKQILSDRIAISWLPDNFKDDQEALDSIDTCYKNLLNDGNVLGEGNLKLLLENI
DTYNLKGIFIRNDLQLTDISQKMYASWNVIQDAVILDLKKQVSRKKKESAEDYNDRL
KKLYTSQESFSIQYLNDCLRAYGKTENIQDYFAKLGAVNNEHEQTINLFAQVRNAYT
SVQAILTTPYPENANLAQDKETVALIKNLLDSLKRLQRFIKPLLGKGDESDKDERFYG
DFTPLWETLNQITPLYNMVRNYMTRKPYSQEKIKLNFENSTLLGGWDLNKEHDNTA
IILRKNGLYYLAIMKKSANKIFDKDKLDNSGDCYEKMVYKLLPGANKMLPKVFFSK
SRIDEFKPSENIIENYKKGTHKKGANFNLADCHNLIDFFKSSISKHEDWSKFNFHFSDT
SSYEDLSDFYREVEQQGYSISFCDVSVEYINKMVEKGDLYLFQIYNKDFSEFSKGTPN
MHTLYWNSLFSKENLNNIIYKLNGQAEIFFRKKSLNYKRPTHPAHQAIKNKNKCNEK
KESIFDYDLVKDKRYTVDKFQFHVPITMNFKSTGNTNINQQVIDYLRTEDDTHIIGID
RGERHLLYLVVIDSHGKIVEQFTLNEIVNEYGGNIYRTNYHDLLDTREQNREKARES
WQTIENIKELKEGYISQVIHKITDLMQKYHAVVVLEDLNMGFMRGRQKVEKQVYQK
FEEMLINKLNYLVNKKADQNSAGGLLHAYQLTSKFESFQKLGKQSGFLFYIPAWNTS
KIDPVTGFVNLFDTRYESIDKAKAFFGKFDSIRYNADKDWFEFAFDYNNFTTKAEGT
RTNWTICTYGSRIRTFRNQAKNSQWDNEEIDLTKAYKAFFAKHGINIYDNIKEAIAME
TEKSFFEDLLHLLKLTLQMRNSITGTTTDYLISPVHDSKGNFYDSRICDNSLPANADA
NGAYNIARKGLMLIQQIKDSTSSNRFKFSPITNKDWLIFAQEKPYLND

[0092] In certain embodiments, the type V-A Cas protein comprises PsCpf1 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at

least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 7. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 7.

PsCpf1 (SEQ ID NO: 7)

MENFKNLYPINKTLRFELRPYGKTLENFKKSGLLEKDAFKANSRRSMQAIIDEKFKET
IEERLKYTEFSECDLGNMTSKDKKITDKAATNLKKQVILSFDDEIFNNYLKPDKNIDA
LFKNDPSNPVISTFKGFTTYFVNFFEIRKHIFKGESSGSMAYRIIDENLTTYLNNIEKIK
KLPEELKSQLEGIDQIDKLNNYNEFITQSGITHYNEIIGGISKSENVKIQGINEGINLYCQ
KNKVKLPRLTPLYKMILSDRVSNSFVLDTIENDTELIEMISDLINKTEISQDVIMSDIQN
IFIKYKQLGNLPGISYSSIVNAICSDYDNNFGDGKRKKSYENDRKKHLETNVYSINYIS
ELLTDTDVSSNIKMRYKELEQNYQVCKENFNATNWMNIKNIKQSEKTNLIKDLLDIL
KSIQRFYDLFDIVDEDKNPSAEFYTWLSKNAEKLDFEFNSVYNKSRNYLTRKQYSDK
KIKLNFDSPTLAKGWDANKEIDNSTIIMRKFNNDRGDYDYFLGIWNKSTPANEKIIPL
EDNGLFEKMQYKLYPDPSKMLPKQFLSKIWKAKHPTTPEFDKKYKEGRHKKGPDFE
KEFLHELIDCFKHGLVNHDEKYQDVFGFNLRNTEDYNSYTEFLEDVERCNYNLSFNK
IADTSNLINDGKLYVFQIWSKDFSIDSKGTKNLNTIYFESLFSEENMIEKMFKLSGEAE
IFYRPASLNYCEDIIKKGHHHAELKDKFDYPIIKDKRYSQDKFFFHVPMVINYKSEKL
NSKSLNNRTNENLGQFTHIIGIDRGERHLIYLTVVDVSTGEIVEQKHLDEIINTDTKGV
EHKTHYLNKLEEKSKTRDNERKSWEAIETIKELKEGYISHVINEIQKLQEKYNALIVM
ENLNYGFKNSRIKVEKQVYQKFETALIKKFNYIIDKKDPETYIHGYQLTNPITTLDKIG
NQSGIVLYIPAWNTSKIDPVTGFVNLLYADDLKYKNQEQAKSFIQKIDNIYFENGEFK
FDIDFSKWNNRYSISKTKWTLTSYGTRIQTFRNPQKNNKWDSAEYDLTEEFKLILNID
GTLKSQDVETYKKFMSLFKLMLQLRNSVTGTDIDYMISPVTDKTGTHFDSRENIKNL
PADADANGAYNIARKGIMAIENIMNGISDPLKISNEDYLKYIQNQQE

[0093] In certain embodiments, the type V-A Cas protein comprises As2Cpf1 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 8. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 8.

As2Cpf1 (SEQ ID NO: 8)

MVAFIDEFVGQYPVSKTLRFEARPVPETKKWLESDQCSVLFNDQKRNEYYGVLKEL
LDDYYRAYIEDALTSFTLDKALLENAYDLYCNRDTNAFSSCCEKLRKDLVKAFGNL
KDYLLGSDQLKDLVKLKAKVDAPAGKGKKKIEVDSRLINWLNNNAKYSAEDREKYI
KAIESFEGFVTYLTNYKQARENMFSSEDKSTAIAFRVIDQNMVTYFGNIRIYEKIKAK
YPELYSALKGFEKFFSPTAYSEILSQSKIDEYNYQCIGRPIDDADFKGVNSLINEYRQK
NGIKARELPVMSMLYKQILSDRDNSFMSEVINRNEEAIECAKNGYKVSYALFNELLQ
LYKKIFTEDNYGNIYVKTQPLTELSQALFGDWSILRNALDNGKYDKDIINLAELEKYF
SEYCKVLDADDAAKIQDKFNLKDYFIQKNALDATLPDLDKITQYKPHLDAMLQAIR
KYKLFSMYNGRKKMDVPENGIDFSNEFNAIYDKLSEFSILYDRIRNFATKKPYSDEK

MKLSFNMPTMLAGWDYNNETANGCFLFIKDGKYFLGVADSKSKNIFDFKKNPHLLD
KYSSKDIYYKVKYKQVSGSAKMLPKVVFAGSNEKIFGHLISKRILEIREKKLYTAAA
GDRKAVAEWIDFMKSAIAIHPEWNEYFKFKFKNTAEYDNANKFYEDIDKQTYSLEK
VEIPTEYIDEMVSQHKLYLFQLYTKDFSDKKKKKGTDNLHTMYWHGVFSDENLKA
VTEGTQPIIKLNGEAEMFMRNPSIEFQVTHEHNKPIANKNPLNTKKESVFNYDLIKDK
RYTERKFYFHCPITLNFRADKPIKYNEKINRFVENNPDVCIIGIDRGERHLLYYTVINQ
TGDILEQGSLNKISGSYTNDKGEKVNKETDYHDLLDRKEKGKHVAQQAWETIENIKE
LKAGYLSQVVYKLTQLMLQYNAVIVLENLNVGFKRGRTKVEKQVYQKFEKAMIDK
LNYLVFKDRGYEMNGSYAKGLQLTDKFESFDKIGKQTGCIYYVIPSYTSHIDPKTGF
VNLLNAKLRYENITKAQDTIRKFDSISYNAKADYFEFAFDYRSFGVDMARNEWVVC
TCGDLRWEYSAKTRETKAYSVTDRLKELFKAHGIDYVGGENLVSHITEVADKHFLS
TLLFYLRLVLKMRYTVSGTENENDFILSPVEYAPGKFFDSREATSTEPMNADANGAY
HIALKGLMTIRGIEDGKLHNYGKGGENAAWFKFMQNQEYKNNG

[0094] In certain embodiments, the type V-A Cas protein comprises McCpf1 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 9. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 9.

McCpf1 (SEQ ID NO: 9)

MLFQDFTHLYPLSKTMRFELKPIGKTLEHIHAKNFLSQDETMADMYQKVKAILDDY
HRDFIADMMGEVKLTKLAEFYDVYLKFRKNPKDDGLQKQLKDLQAVLRKEIVKPIG
NGGKYKAGYDRLFGAKLFKDGKELGDLAKFVIAQEGESSPKLAHLAHFEKFSTYFT
GFHDNRKNMYSDEDKHTAITYRLIHENLPRFIDNLQILATIKQKHSALYDQIINELTAS
GLDVSLASHLDGYHKLLTQEGITAYNTLLGGISGEAGSRKIQGINELINSHHNQHCHK
SERIAKLRPLHKQILSDGMGVSFLPSKFADDSEMCQAVNEFYRHYADVFAKVQSLFD
GFDDHQKDGIYVEHKNLNELSKQAFGDFALLGRVLDGYYVDVVNPEFNERFAKAK
TDNAKAKLTKEKDKFIKGVHSLASLEQAIEHYTARHDDESVQAGKLGQYFKHGLAG
VDNPIQKIHNNHSTIKGFLERERPAGERALPKIKSGKNPEMTQLRQLKELLDNALNVA
HFAKLLTTKTTLDNQDGNFYGEFGALYDELAKIPTLYNKVRDYLSQKPFSTEKYKLN
FGNPTLLNGWDLNKEKDNFGIILQKDGCYYLALLDKAHKKVFDNAPNTGKNVYQK
MIYKLLPGPNKMLPKVFFAKSNLDYYNPSAELLDKYAQGTHKKGNNFNLKDCHALI
DFFKAGINKHPEWQHFGFKFSPTSSYQDLSDFYREVEPQGYQVKFVDINADYINELV
EQGQLYLFQIYNKDFSPKAHGKPNLHTLYFKALFSKDNLANPIYKLNGEAQIFYRKA
SLDMNETTIHRAGEVLENKNPDNPKKRQFVYDIIKDKRYTQDKFMLHVPITMNFGV
QGMTIKEFNKKVNQSIQQYDEVNVIGIDRGERHLLYLTVINSKGEILEQRSLNDITTAS
ANGTQMTTPYHKILDKREIERLNARVGWGEIETIKELKSGYLSHVVHQISQLMLKYN
AIVVLEDLNFGFKRGRFKVEKQIYQNFENALIKKLNHLVLKDEADDEIGSYKNALQL
TNNFTDLKSIGKQTGFLFYVPAWNTSKIDPETGFVDLLKPRYENIAQSQAFFGKFDKI
CYNADKDYFEFHIDYAKFTDKAKNSRQIWKICSHGDKRYVYDKTANQNKGATKGI
NVNDELKSLFARHHINDKQPNLVMDICQNNDKEFHKSLIYLLKTLLALRYSNASSDE
DFILSPVANDEGMFFNSALADDTQPQNADANGAYHIALKGLWVLEQIKNSDDLNKV
KLAIDNQTWLNFAQNR

[0095] In certain embodiments, the type V-A Cas protein comprises Lb3Cpf1 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 10. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 10.

Lb3Cpf1 (SEQ ID NO: 10)

MHENNGKIADNFIGIYPVSKTLRFELKPVGKTQEYIEKHGILDEDLKRAGDYKSVKKI
IDAYHKYFIDEALNGIQLDGLKNYYELYEKKRDNNEEKEFQKIQMSLRKQIVKRFSE
HPQYKYLFKKELIKNVLPEFTKDNAEEQTLVKSFQEFTTYFEGFHQNRKNMYSDEEK
STAIAYRVVHQNLPKYIDNMRIFSMILNTDIRSDLTELFNNLKTKMDITIVEEYFAIDG
FNKVVNQKGIDVYNTILGAFSTDDNTKIKGLNEYINLYNQKNKAKLPKLKPLFKQILS
DRDKISFIPEQFDSDTEVLEAVDMFYNRLLQFVIENEGQITISKLLTNFSAYDLNKIYV
KNDTTISAISNDLFDDWSYISKAVRENYDSENVDKNKRAAAYEEKKEKALSKIKMYS
IEELNFFVKKYSCNECHIEGYFERRILEILDKMRYAYESCKILHDKGLINNISLCQDRQ
AISELKDFLDSIKEVQWLLKPLMIGQEQADKEEAFYTELLRIWEELEPITLLYNKVRN
YVTKKPYTLEKVKLNFYKSTLLDGWDKNKEKDNLGIILLKDGQYYLGIMNRRNNKI
ADDAPLAKTDNVYRKMEYKLLTKVSANLPRIFLKDKYNPSEEMLEKYEKGTHLKGE
NFCIDDCRELIDFFKKGIKQYEDWGQFDFKFSDTESYDDISAFYKEVEHQGYKITFRDI
DETYIDSLVNEGKLYLFQIYNKDFSPYSKGTKNLHTLYWEMLFSQQNLQNIVYKLNG
NAEIFYRKASINQKDVVVHKADLPIKNKDPQNSKKESMFDYDIIKDKRFTCDKYQFH
VPITMNFKALGENHFNRKVNRLIHDAENMHIIGIDRGERNLIYLCMIDMKGNIVKQIS
LNEIISYDKNKLEHKRNYHQLLKTREDENKSARQSWQTIHTIKELKEGYLSQVIHVIT
DLMVEYNAIVVLEDLNFGFKQGRQKFERQVYQKFEKMLIDKLNYLVDKSKGMDED
GGLLHAYQLTDEFKSFKQLGKQSGFLYYIPAWNTSKLDPTTGFVNLFYTKYESVEKS
KEFINNFTSILYNQEREYFEFLFDYSAFTSKAEGSRLKWTVCSKGERVETYRNPKKNN
EWDTQKIDLTFELKKLFNDYSISLLDGDLREQMGKIDKADFYKKFMKLFALIVQMR
NSDEREDKLISPVLNKYGAFFETGKNERMPLDADANGAYNIARKGLWIIEKIKNTDV
EQLDKVKLTISNKEWLQYAQEHIL

[0096] In certain embodiments, the type V-A Cas protein comprises EcCpf1 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 11. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 11.

EcCpf1 (SEQ ID NO: 11)

MDFFKNDMYFLCINGIIVISKLFAYLFLMYKRGVVMIKDNFVNVYSLSKTIRMALIP
WGKTEDNFYKKFLLEEDEERAKNYIKVKGYMDEYHKNFIESALNSVVLNGVDEYCE

LYFKQNKSDSEVKKIESLEASMRKQISKAMKEYTVDGVKIYPLLSKKEFIRELLPEFL
TQDEEIETLEQFNDFSTYFQGFWENRKNIYTDEEKSTGVPYRCINDNLPKFLDNVKSF
EKVILALPQKAVDELNANFNGVYNVDVQDVFSVDYFNFVLSQSGIEKYNNIIGGYSN
SDASKVQGLNEKINLYNQQIAKSDKSKKLPLLKPLYKQILSDRSSLSFIPEKFKDDNE
VLNSINVLYDNIAESLEKANDLMSDIANYNTDNIFISSGVAVTDISKKVFGDWSLIRN
NWNDEYESTHKKGKNEEKFYEKEDKEFKKIKSFSVSELQRLANSDLSIVDYLVDESA
SLYADIKTAYNNAKDLLSNEYSHSKRLSKNDDAIELIKSFLDSIKNYEAFLKPLCGTG
KEESKDNAFYGAFLECFEEIRQVDAVYNKVRNHITQKPYSNDKIKLNFQNPQFLAGW
DKNKERAYRSVLLRNGEKYYLAIMEKGKSKLFEDFPEDESSPFEKIDYKLLPEPSKM
LPKVFFATSNKDLFNPSDEILNIRATGSFKKGDSFNLDDCHKFIDFYKASIENHPDWS
KFDFDFSETNDYEDISKFFKEVSDQGYSIGYRKISESYLEEMVDNGSLYMFQLYNKDF
SENRKSKGTPNLHTLYFKMLFDERNLEDVVYKLSGGAEMFYRKPSIDKNEMIVHPK
NQPIDNKNPNNVKKTSTFEYDIVKDMRYTKPQFQLHLPIVLNFKANSKGYINDDVRN
VLKNSEDTYVIGIDRGERNLVYACVVDGNGKLVEQVPLNVIEADNGYKTDYHKLLN
DREEKRNEARKSWKTIGNIKELKEGYISQVVHKICQLVVKYDAVIAMEDLNSGFVNS
RKKVEKQVYQKFERMLTQKLNYLVDKKLDPNEMGGLLNAYQLTNEATKVRNGRQ
DGIIFYIPAWLTSKIDPTTGFVNLLKPKYNSVSASKEFFSKFDEIRYNEKENYFEFSFNY
DNFPKCNADFKREWTVCTYGDRIRTFRDPENNNKFNSEVVVLNDEFKNLFVEFDIDY
TDNLKEQILAMDEKSFYKKLMGLLSLTLQMRNSISKNVDVDYLISPVKNSNGEFYDS
RNYDITSSLPCDADSNGAYNIARKGLWAINQIKQADDETKANISIKNSEWLQYAQNC
DEV

[0097] In certain embodiments, the type V-A Cas protein is not Cpf1. In certain embodiments, the type V-A Cas nuclease is not AsCpf1.

[0098] In certain embodiments, the type V-A Cas protein comprises MAD1, MAD2, MAD3, MAD4, MAD5, MAD6, MAD7, MAD8, MAD9, MAD10, MAD11, MAD12, MAD13, MAD14, MAD15, MAD16, MAD17, MAD18, MAD19, or MAD20, or variants thereof. MAD1-MAD20 are known in the art and are described in U.S. Patent No. 9,982,279.

[0099] In certain embodiments, the type V-A Cas protein comprises MAD7 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 1. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 1.

MAD7 (SEQ ID NO: 1)

MNNGTNNFQNFIGISSLQKTLRNALIPTETTQQFIVKNGIIKEDELRGENRQILKDIMD
DYYRGFISETLSSIDDIDWTSLFEKMEIQLKNGDNKDTLIKEQTEYRKAIHKKFANDD
RFKNMFSAKLISDILPEFVIHNNNYSASEKEEKTQVIKLFSRFATSFKDYFKNRANCFS
ADDISSSSCHRIVNDNAEIFFSNALVYRRIVKSLSNDDINKISGDMKDSLKEMSLEEIY
SYEKYGEFITQEGISFYNDICGKVNSFMNLYCQKNKENKNLYKLQKLHKQILCIADTS
YEVPYKFESDEEVYQSVNGFLDNISSKHIVERLRKIGDNYNGYNLDKIYIVSKFYESV

SQKTYRDWETINTALEIHYNNILPGNGKSKADKVKKAVKNDLQKSITEINELVSNYK
LCSDDNIKAETYIHEISHILNNFEAQELKYNPEIHLVESELKASELKNVLDVIMNAFH
WCSVFMTEELVDKDNNFYAELEEIYDEIYPVISLYNLVRNYVTQKPYSTKKIKLNFGI
PTLADGWSKSKEYSNNAIILMRDNLYYLGIFNAKNKPDKKIIEGNTSENKGDYKKMI
YNLLPGPNKMIPKVFLSSKTGVETYKPSAYILEGYKQNKHIKSSKDFDITFCHDLIDYF
KNCIAIHPEWKNFGFDFSDTSTYEDISGFYREVELQGYKIDWTYISEKDIDLLQEKGQ
LYLFQIYNKDFSKKSTGNDNLHTMYLKNLFSEENLKDIVLKLNGEAEIFFRKSSIKNPI
IHKKGSILVNRTYEAEEKDQFGNIQIVRKNIPENIYQELYKYFNDKSDKELSDEAAKL
KNVVGHHEAATNIVKDYRYTYDKYFLHMPITINFKANKTGFINDRILQYIAKEKDLH
VIGIDRGERNLIYVSVIDTCGNIVEQKSFNIVNGYDYQIKLKQQEGARQIARKEWKEI
GKIKEIKEGYLSLVIHEISKMVIKYNAIIAMEDLSYGFKKGRFKVERQVYQKFETMLI
NKLNYLVFKDISITENGGLLKGYQLTYIPDKLKNVGHQCGCIFYVPAAYTSKIDPTTG
FVNIFKFKDLTVDAKREFIKKFDSIRYDSEKNLFCFTFDYNNFITQNTVMSKSSWSVY
TYGVRIKRRFVNGRFSNESDTIDITKDMEKTLEMTDINWRDGHDLRQDIIDYEIVQHI
FEIFRLTVQMRNSLSELEDRDYDRLISPVLNENNIFYDSAKAGDALPKDADANGAYCI
ALKGLYEIKQITENWKEDGKFSRDKLKISNKDWFDFIQNKRYL

[0100] In certain embodiments, the type V-A Cas protein comprises MAD2 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 2. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 2.

MAD2 (SEQ ID NO: 2)

MSSLTKFTNKYSKQLTIKNELIPVGKTLENIKENGLIDGDEQLNENYQKAKIIVDDFLR
DFINKALNNTQIGNWRELADALNKEDEDNIEKLQDKIRGIIVSKFETFDLFSSYSIKKD
EKIIDDDNDVEEEELDLGKKTSSFKYIFKKNLFKLVLPSYLKTTNQDKLKIISSFDNFS
TYFRGFFENRKNIFTKKPISTSIAYRIVHDNFPKFLDNIRCFNVWQTECPQLIVKADNY
LKSKNVIAKDKSLANYFTVGAYDYFLSQNGIDFYNNIIGGLPAFAGHEKIQGLNEFIN
QECQKDSELKSKLKNRHAFKMAVLFKQILSDREKSFVIDEFESDAQVIDAVKNFYAE
QCKDNNVIFNLLNLIKNIAFLSDDELDGIFIEGKYLSSVSQKLYSDWSKLRNDIEDSAN
SKQGNKELAKKIKTNKGDVEKAISKYEFSLSELNSIVHDNTKFSDLLSCTLHKVASEK
LVKVNEGDWPKHLKNNEEKQKIKEPLDALLEIYNTLLIFNCKSFNKNGNFYVDYDR
CINELSSVVYLYNKTRNYCTKKPYNTDKFKLNFNSPQLGEGFSKSKENDCLTLLFKK
DDNYYVGIIRKGAKINFDDTQAIADNTDNCIFKMNYFLLKDAKKFIPKCSIQLKEVKA
HFKKSEDDYILSDKEKFASPLVIKKSTFLLATAHVKGKKGNIKKFQKEYSKENPTEYR
NSLNEWIAFCKEFLKTYKAATIFDITTLKKAEEYADIVEFYKDVDNLCYKLEFCPIKT
SFIENLIDNGDLYLFRINNKDFSSKSTGTKNLHTLYLQAIFDERNLNNPTIMLNGGAEL
FYRKESIEQKNRITHKAGSILVNKVCKDGTSLDDKIRNEIYQYENKFIDTLSDEAKKV
LPNVIKKEATHDITKDKRFTSDKFFFHCPLTINYKEGDTKQFNNEVLSFLRGNPDINII
GIDRGERNLIYVTVINQKGEILDSVSFNTVTNKSSKIEQTVDYEEKLAVREKERIEAKR
SWDSISKIATLKEGYLSAIVHEICLLMIKHNAIVVLENLNAGFKRIRGGLSEKSVYQKF
EKMLINKLNYFVSKKESDWNKPSGLLNGLQLSDQFESFEKLGIQSGFIFYVPAAYTSK
IDPTTGFANVLNLSKVRNVDAIKSFFSNFNEISYSKKEALFKFSFDLDSLSKKGFSSFV
KFSKSKWNVYTFGERIIKPKNKQGYREDKRINLTFEMKKLLNEYKVSFDLENNLIPN

LTSANLKDTFWKELFFIFKTTLQLRNSVTNGKEDVLISPVKNAKGEFFVSGTHNKTLP
QDCDANGAYHIALKGLMILERNNLVREEKDTKKIMAISNVDWFEYVQKRRGVL

[0101] In certain embodiments, the type V-A Cas protein comprises Csm1. Csm1 proteins are known in the art and are described in U.S. Patent No. 9,896,696. Csm1 orthologs can be found in various bacterial and archaeal genomes. For example, in certain embodiments, the Csm1 protein is derived from *Smithella* sp. SCADC (Sm), *Sulfuricurvum sp.* (Ss), or *Microgenomates (Roizmanbacteria) bacterium* (Mb).

[0102] In certain embodiments, the type V-A Cas protein comprises SmCsm1 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 12. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 12.

SmCsm1 (SEQ ID NO: 12)

MEKYKITKTIRFKLLPDKIQDISRQVAVLQNSTNAEKKNNLLRLVQRGQELPKLLNE
YIRYSDNHKLKSNVTVHFRWLRLFTKDLFYNWKKDNTEKKIKISDVVYLSHVFEAFL
KEWESTIERVNADCNKPEESKTRDAEIALSIRKLGIKHQLPFIKGFVDNSNDKNSEDT
KSKLTALLSEFEAVLKICEQNYLPSQSSGIAIAKASFNYYTINKKQKDFEAEIVALKKQ
LHARYGNKKYDQLLRELNLIPLKELPLKELPLIEFYSEIKKRKSTKKSEFLEAVSNGLV
FDDLKSKFPLFQTESNKYDEYLKLSNKITQKSTAKSLLSKDSPEAQKLQTEITKLKKN
RGEYFKKAFGKYVQLCELYKEIAGKRGKLKGQIKGIENERIDSQRLQYWALVLEDNL
KHSLILIPKEKTNELYRKVWGAKDDGASSSSSSTLYYFESMTYRALRKLCFGINGNTF
LPEIQKELPQYNQKEFGEFCFHKSNDDKEIDEPKLISFYQSVLKTDFVKNTLALPQSVF
NEVAIQSFETRQDFQIALEKCCYAKKQIISESLKKEILENYNTQIFKITSLDLQRSEQKN
LKGHTRIWNRFWTKQNEEINYNLRLNPEIAIVWRKAKKTRIEKYGERSVLYEPEKRN
RYLHEQYTLCTTVTDNALNNEITFAFEDTKKKGTEIVKYNEKINQTLKKEFNKNQLW
FYGIDAGEIELATLALMNKDKEPQLFTVYELKKLDFFKHGYIYNKERELVIREKPYK
AIQNLSYFLNEELYEKTFRDGKFNETYNELFKEKHVSAIDLTTAKVINGKIILNGDMIT
FLNLRILHAQRKIYEELIENPHAELKEKDYKLYFEIEGKDKDIYISRLDFEYIKPYQEIS
NYLFAYFASQQINEAREEEQINQTKRALAGNMIGVIYYLYQKYRGIISIEDLKQTKVE
SDRNKFEGNIERPLEWALYRKFQQEGYVPPISELIKLRELEKFPLKDVKQPKYENIQQ
FGIIKFVSPEETSTTCPKCLRRFKDYDKNKQEGFCKCQCGFDTRNDLKGFEGLNDPD
KVAAFNIAKRGFEDLQKYK

[0103] In certain embodiments, the type V-A Cas protein comprises SsCsm1 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid

sequence set forth in SEQ ID NO: 13. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 13.

SsCsm1 (SEQ ID NO: 13)

MLHAFTNQYQLSKTLRFGATLKEDEKKCKSHEELKGFVDISYENMKSSATIAESLNE
NELVKKCERCYSEIVKFHNAWEKIYYRTDQIAVYKDFYRQLSRKARFDAGKQNSQLI
TLASLCGMYQGAKLSRYITNYWKDNITRQKSFLKDFSQQLHQYTRALEKSDKAHTK
PNLINFNKTFMVLANLVNEIVIPLSNGAISFPNISKLEDGEESHLIEFALNDYSQLSELIG
ELKDAIATNGGYTPFAKVTLNHYTAEQKPHVFKNDIDAKIRELKLIGLVETLKGKSSE
QIEEYFSNLDKFSTYNDRNQSVIVRTQCFKYKPIPFLVKHQLAKYISEPNGWDEDAVA
KVLDAVGAIRSPAHDYANNQEGFDLNHYPIKVAFDYAWEQLANSLYTTVTFPQEMC
EKYLNSIYGCEVSKEPVFKFYADLLYIRKNLAVLEHKNNLPSNQEEFICKINNTFENIV
LPYKISQFETYKKDILAWINDGHDHKKYTDAKQQLGFIRGGLKGRIKAEEVSQKDKY
GKIKSYYENPYTKLTNEFKQISSTYGKTFAELRDKFKEKNEITKITHFGIIIEDKNRDRY
LLASELKHEQINHVSTILNKLDKSSEFITYQVKSLTSKTLIKLIKNHTTKKGAISPYADF
HTSKTGFNKNEIEKNWDNYKREQVLVEYVKDCLTDSTMAKNQNWAEFGWNFEKC
NSYEDIEHEIDQKSYLLQSDTISKQSIASLVEGGCLLLPIINQDITSKERKDKNQFSKD
WNHIFEGSKEFRLHPEFAVSYRTPIEGYPVQKRYGRLQFVCAFNAHIVPQNGEFINLK
KQIENFNDEDVQKRNVTEFNKKVNHALSDKEYVVIGIDRGLKQLATLCVLDKRGKIL
GDFEIYKKEFVRAEKRSESHWEHTQAETRHILDLSNLRVETTIEGKKVLVDQSLTLVK
KNRDTPDEEATEENKQKIKLKQLSYIRKLQHKMQTNEQDVLDLINNEPSDEEFKKRIE
GLISSFGEGQKYADLPINTMREMISDLQGVIARGNNQTEKNKIIELDAADNLKQGIVA
NMIGIVNYIFAKYSYKAYISLEDLSRAYGGAKSGYDGRYLPSTSQDEDVDFKEQQNQ
MLAGLGTYQFFEMQLLKKLQKIQSDNTVLRFVPAFRSADNYRNILRLEETKYKSKPF
GVVHFIDPKFTSKKCPVCSKTNVYRDKDDILVCKECGFRSDSQLKERENNIHYIHNG
DDNGAYHIALKSVENLIQMK

[0104] In certain embodiments, the type V-A Cas protein comprises MbCsm1 or a variant thereof. In certain embodiments, the type V-A Cas protein comprises an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in SEQ ID NO: 14. In certain embodiments, the type V-A Cas protein comprises the amino acid sequence set forth in SEQ ID NO: 14.

MbCsm1 (SEQ ID NO: 14)

MEIQELKNLYEVKKTVRFELKPSKKKIFEGGDVIKLQKDFEKVQKFFLDIFVYKNEHT
KLEFKKKREIKYTWLRTNTKNEFYNWRGKSDTGKNYALNKIGFLAEEILRWLNEWQ
ELTKSLKDLTQREEHKQERKSDIAFVLRNFLKRQNLPFIKDFFNAVIDIQGKQGKESD
DKIRKFREEIKEIEKNLNACSREYLPTQSNGVLLYKASFSYYTLNKTPKEYEDLKKEK
ESELSSVLLKEIYRRKRFNRTTNQKDTLFECTSDWLVKIKLGKDIYEWTLDEAYQKM
KIWKANQKSNFIEAVAGDKLTHQNFRKQFPLFDASDEDFETFYRLTKALDKNPENA
KKIAQKRGKFFNAPNETVQTKNYHELCELYKRIAVKRGKIIAEIKGIENEEVQSQLLT
HWAVIAEERDKKFIVLIPRKNGGKLENHKNAHAFLQEKDRKEPNDIKVYHFKSLTLR
SLEKLCFKEAKNTFAPEIKKETNPKIWFPTYKQEWNSTPERLIKFYKQVLQSNYAQTY
LDLVDFGNLNTFLETHFTTLEEFESDLEKTCYTKVPVYFAKKELETFADEFEAEVFEI

TTRSISTESKRKENAHAEIWRDFWSRENEEENHITRLNPEVSVLYRDEIKEKSNTSRK
NRKSNANNRFSDPRFTLATTITLNADKKKSNLAFKTVEDINIHIDNFNKKFSKNFSGE
WVYGIDRGLKELATLNVVKFSDVKNVFGVSQPKEFAKIPIYKLRDEKAILKDENGLS
LKNAKGEARKVIDNISDVLEEGKEPDSTLFEKREVSSIDLTRAKLIKGHIISNGDQKTY
LKLKETSAKRRIFELFSTAKIDKSSQFHVRKTIELSGTKIYWLCEWQRQDSWRTEKVS
LRNTLKGYLQNLDLKNRFENIETIEKINHLRDAITANMVGILSHLQNKLEMQGVIALE
NLDTVREQSNKKMIDEHFEQSNEHVSRRLEWALYCKFANTGEVPPQIKESIFLRDEF
KVCQIGILNFIDVKGTSSNCPNCDQESRKTGSHFICNFQNNCIFSSKENRNLLEQNLHN
SDDVAAFNIAKRGLEIVKV

[0105] More type V-A Cas proteins and their corresponding naturally occurring CRISPR-Cas systems can be identified by computational and experimental methods known in the art, *e.g.*, as described in U.S. Patent No. 9,790,490 and Shmakov *et al.* (2015) MOL. CELL, 60: 385. Exemplary computational methods include analysis of putative Cas proteins by homology modeling, structural BLAST, PSI-BLAST, or HHPred, and analysis of putative CRISPR loci by identification of CRISPR arrays. Exemplary experimental methods include *in vitro* cleavage assays and in-cell nuclease assays (*e.g.*, the Surveyor assay) as described in Zetsche *et al.* (2015) CELL, 163: 759.

[0106] In certain embodiments, the Cas protein is a Cas nuclease that directs cleavage of one or both strands at the target locus, such as the target strand (*i.e.*, the strand having the target nucleotide sequence that hybridizes with a single guide nucleic acid or dual guide nucleic acids) and/or the non-target strand. In certain embodiments, the Cas 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 nucleotides from the first or last nucleotide of the target nucleotide sequence or its complementary sequence. In certain embodiments, the cleavage is staggered, *i.e.* generating sticky ends. In certain embodiments, the cleavage generates a staggered cut with a 5' overhang. In certain embodiments, the cleavage generates a staggered cut with a 5' overhang of 1 to 5 nucleotides, *e.g.*, of 4 or 5 nucleotides. In certain embodiments, the cleavage site is distant from the PAM, *e.g.*, the cleavage occurs after the 18th nucleotide on the non-target strand and after the 23rd nucleotide on the target strand.

[0107] In certain embodiments, the Cas protein lacks substantially all DNA cleavage activity. Such a Cas protein can be generated by introducing one or more mutations to an active Cas nuclease (*e.g.*, a naturally occurring Cas nuclease). A mutated Cas protein is considered to substantially lack all DNA cleavage activity when the DNA cleavage activity of the protein has no more than about 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the DNA cleavage activity of the corresponding non-mutated form, for example, nil or negligible as

compared with the non-mutated form. Thus, the Cas protein may comprise one or more mutations (*e.g.*, a mutation in the RuvC domain of a type V-A Cas protein) and be used as a generic DNA binding protein with or without fusion to an effector domain. Exemplary mutations include D908A, E993A, and D1263A with reference to the amino acid positions in AsCpf1; D832A, E925A, and D1180A with reference to the amino acid positions in LbCpf1; and D917A, E1006A, and D1255A with reference to the amino acid position numbering of the FnCpf1. More mutations can be designed and generated according to the crystal structure described in Yamano *et al.* (2016) CELL, 165: 949.

[0108] It is understood that the Cas protein, rather than losing nuclease activity to cleave all DNA, may lose the ability to cleave only the target strand or only the non-target strand of a double-stranded DNA, thereby being functional as a nickase (see, Gao *et al.* (2016) CELL RES., 26: 901). Accordingly, in certain embodiments, the Cas nuclease is a Cas nickase. In certain embodiments, the Cas nuclease has the activity to cleave the non-target strand but substantially lacks the activity to cleave the target strand, *e.g.*, by a mutation in the Nuc domain. In certain embodiments, the Cas nuclease has the cleavage activity to cleave the target strand but substantially lacks the activity to cleave the non-target strand.

[0109] In other embodiments, the Cas nuclease has the activity to cleave a double-stranded DNA and result in a double-strand break.

[0110] Cas proteins that lack substantially all DNA cleavage activity or have the ability to cleave only one strand may also be identified from naturally occurring systems. For example, certain naturally occurring CRISPR-Cas systems may retain the ability to bind the target nucleotide sequence but lose entire or partial DNA cleavage activity in eukaryotic (*e.g.*, mammalian or human) cells. Such type V-A proteins are disclosed, for example, in Kim *et al.* (2017) ACS SYNTH. BIOL. 6(7): 1273-82 and Zhang *et al.* (2017) CELL DISCOV. 3:17018.

[0111] The activity of the Cas protein (*e.g.*, Cas nuclease) can be altered, thereby creating an engineered Cas protein. In certain embodiments, the altered activity of the engineered Cas protein comprises increased targeting efficiency and/or decreased off-target binding. While not wishing to be bound by theory, it is hypothesized that off-target binding can be recognized by the Cas protein, for example, by the presence of one or more mismatches between the spacer sequence and the target nucleotide sequence, which may affect the stability and/or conformation of the CRISPR-Cas complex. In certain embodiments, the altered activity comprises modified binding, *e.g.*, increased binding to the target locus (*e.g.*,

the target strand or the non-target strand) and/or decreased binding to off-target loci. In certain embodiments, the altered activity comprises altered charge in a region of the protein that associates with a single guide nucleic acid or dual guide nucleic acids. In certain embodiments, the altered activity of the engineered Cas protein comprises altered charge in a region of the protein that associates with the target strand and/or the non-target strand. In certain embodiments, the altered activity of the engineered Cas protein comprises altered charge in a region of the protein that associates with an off-target locus. The altered charge can include decreased positive charge, decreased negative charge, increased positive charge, and increased negative charge. For example, decreased negative charge and increased positive charge may generally strengthen the binding to the nucleic acid(s) whereas decreased positive charge and increased negative charge may weaken the binding to the nucleic acid(s). In certain embodiments, the altered activity comprises increased or decreased steric hindrance between the protein and a single guide nucleic acid or dual guide nucleic acids. In certain embodiments, the altered activity comprises increased or decreased steric hindrance between the protein and the target strand and/or the non-target strand. In certain embodiments, the altered activity comprises increased or decreased steric hindrance between the protein and an off-target locus. In certain embodiments, the modification or mutation comprises a substitution of Lys, His, Arg, Glu, Asp, Ser, Gly, or Thr. In certain embodiments, the modification or mutation comprises a substitution with Gly, Ala, Ile, Glu, or Asp. In certain embodiments, the modification or mutation comprises an amino acid substitution in the groove between the WED and RuvC domain of the Cas protein (*e.g.*, a type V-A Cas protein).

[0112] In certain embodiments, the altered activity of the engineered Cas protein comprises increased nuclease activity to cleave the target locus. In certain embodiments, the altered activity of the engineered Cas protein comprises decreased nuclease activity to cleave an off-target locus. In certain embodiments, the altered activity of the engineered Cas protein comprises altered helicase kinetics. In certain embodiments, the engineered Cas protein comprises a modification that alters formation of the CRISPR complex.

[0113] In certain embodiments, a protospacer adjacent motif (PAM) or PAM-like motif directs binding of the Cas protein complex to the target locus. Many Cas proteins have PAM specificity. The precise sequence and length requirements for the PAM differ depending on the Cas protein used. PAM sequences are typically 2-5 base pairs in length and are adjacent to (but located on a different strand of target DNA from) the target nucleotide sequence.

PAM sequences can be identified using a method known in the art, such as testing cleavage, targeting, or modification of oligonucleotides having the target nucleotide sequence and different PAM sequences.

[0114] Exemplary PAM sequences are provided in Tables 4 and 5. In one embodiment, the Cas protein is MAD7 and the PAM is TTTN, wherein N is A, C, G, or T. In another embodiment, the Cas protein is MAD7 and the PAM is CTTN, wherein N is A, C, G, or T. In another embodiment, the Cas protein is AsCpf1 and the PAM is TTTN, wherein N is A, C, G, or T. In another embodiment, the Cas protein is FnCpf1 and the PAM is 5' TTN, wherein N is A, C, G, or T. PAM sequences for certain other type V-A Cas proteins are disclosed in Zetsche *et al.* (2015) CELL, 163: 759 and U.S. Patent No. 9,982,279. Further, engineering of the PAM Interacting (PI) domain of a Cas protein may allow programing of PAM specificity, improve target site recognition fidelity, and increase the versatility of the engineered, non-naturally occurring system. Exemplary approaches to alter the PAM specificity of Cpf1 is described in Gao *et al.* (2017) NAT. BIOTECHNOL., 35: 789.

[0115] In certain embodiments, the engineered Cas protein comprises a modification that alters the Cas protein specificity in concert with modification to targeting range. Cas mutants can be designed to have increased target specificity as well as accommodating modifications in PAM recognition, for example by choosing mutations that alter PAM specificity (*e.g.*, in the PI domain) and combining those mutations with groove mutations that increase (or if desired, decrease) specificity for the on-target locus versus off-target loci. The Cas modifications described herein can be used to counter loss of specificity resulting from alteration of PAM recognition, enhance gain of specificity resulting from alteration of PAM recognition, counter gain of specificity resulting from alteration of PAM recognition, or enhance loss of specificity resulting from alteration of PAM recognition.

[0116] In certain embodiments, the engineered Cas protein comprises one or more nuclear localization signal (NLS) motifs. In certain embodiments, the engineered Cas protein comprises at least 2 (*e.g.*, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10) NLS motifs. Non-limiting examples of NLS motifs include: the NLS of SV40 large T-antigen, having the amino acid sequence of PKKKRKV (SEQ ID NO: 35); the NLS from nucleoplasmin, *e.g.*, the nucleoplasmin bipartite NLS having the amino acid sequence of KRPAATKKAGQAKKKK (SEQ ID NO: 36); the c-myc NLS, having the amino acid sequence of PAAKRVKLD (SEQ ID NO: 37) or RQRRNELKRSP (SEQ ID NO: 38); the hRNPA1 M9 NLS, having the amino acid sequence of

NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 39); the importin-α IBB domain NLS, having the amino acid sequence of RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: 40); the myoma T protein NLS, having the amino acid sequence of VSRKRPRP (SEQ ID NO: 41) or PPKKARED (SEQ ID NO: 42); the human p53 NLS, having the amino acid sequence of PQPKKKPL (SEQ ID NO: 43); the mouse c-abl IV NLS, having the amino acid sequence of SALIKKKKKMAP (SEQ ID NO: 44); the influenza virus NS1 NLS, having the amino acid sequence of DRLRR (SEQ ID NO: 45) or PKQKKRK (SEQ ID NO: 46); the hepatitis virus δ antigen NLS, having the amino acid sequence of RKLKKKIKKL (SEQ ID NO: 47); the mouse Mx1 protein NLS, having the amino acid sequence of REKKKFLKRR (SEQ ID NO: 48); the human poly(ADP-ribose) polymerase NLS, having the amino acid sequence of KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 49); the human glucocorticoid receptor NLS, having the amino acid sequence of RKCLQAGMNLEARKTKK (SEQ ID NO: 33), and synthetic NLS motifs such as PAAKKKKLD (SEQ ID NO: 34).

[0117] In general, the one or more NLS motifs are of sufficient strength to drive accumulation of the Cas protein in a detectable amount in the nucleus of a eukaryotic cell. The strength of nuclear localization activity may derive from the number of NLS motif(s) in the Cas protein, the particular NLS motif(s) used, the position(s) of the NLS motif(s), or a combination of these factors. In certain embodiments, the engineered Cas protein comprises at least 1 (*e.g.*, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10) NLS motif(s) at or near the N-terminus (*e.g.*, within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N-terminus). In certain embodiments, the engineered Cas protein comprises at least 1 (*e.g.*, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10) NLS motif(s) at or near the C-terminus (*e.g.*, within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the C-terminus). In certain embodiments, the engineered Cas protein comprises at least 1 (*e.g.*, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10) NLS motif(s) at or near the C-terminus and at least 1 (*e.g.*, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10) NLS motif(s) at or near the N-terminus. In certain embodiments, the engineered Cas protein comprises one, two, or three NLS motifs at or near the C-terminus. In certain embodiments, the engineered Cas protein comprises one NLS motif at or near the N-terminus and one, two, or three NLS motifs at or near the C-terminus.

In certain embodiments, the engineered Cas protein comprises a nucleoplasmin NLS at or near the C-terminus.

[0118] Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the nucleic acid-targeting protein, such that location within a cell may be visualized. Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting the protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay that detects the effect of the nuclear import of a Cas protein complex (*e.g.*, assay for DNA cleavage or mutation at the target locus, or assay for altered gene expression activity) as compared to a control not exposed to the Cas protein or exposed to a Cas protein lacking one or more of the NLS motifs.

[0119] In certain embodiments, the Cas protein is a chimeric Cas protein, *e.g.*, a Cas protein having enhanced function by being a chimera. Chimeric Cas proteins may be new Cas proteins containing fragments from more than one naturally occurring Cas proteins or variants thereof. For example, fragments of multiple type V-A Cas homologs (*e.g.*, orthologs) may be fused to form a chimeric Cas protein. In certain embodiments, the chimeric Cas protein comprises fragments of Cpf1 orthologs from multiple species and/or strains.

[0120] In certain embodiments, the Cas protein comprises one or more effector domains. The one or more effector domains may be located at or near the N-terminus of the Cas protein and/or at or near the C-terminus of the Cas protein. In certain embodiments, an effector domain comprised in the Cas protein is a transcriptional activation domain (*e.g.*, VP64), a transcriptional repression domain (*e.g.*, a KRAB domain or an SID domain), an exogenous nuclease domain (*e.g.*, FokI), a deaminase domain (*e.g.*, cytidine deaminase or adenine deaminase), or a reverse transcriptase domain (*e.g.*, a high fidelity reverse transcriptase domain). Other activities of effector domains include but are not limited to methylase activity, demethylase activity, transcription release factor activity, translational initiation activity, translational activation activity, translational repression activity, histone modification (*e.g.*, acetylation or demethylation) activity, single-stranded RNA cleavage activity, double-strand RNA cleavage activity, single-strand DNA cleavage activity, double-strand DNA cleavage activity, and nucleic acid binding activity.

[0121] In certain embodiments, the Cas protein comprises one or more protein domains that enhance homology-directed repair (HDR) and/or inhibit non-homologous end joining (NHEJ). Exemplary protein domains having such functions are described in Jayavaradhan *et al.* (2019) NAT. COMMUN. 10(1): 2866 and Janssen *et al.* (2019) MOL. THER. NUCLEIC ACIDS 16: 141-54. In certain embodiments, the Cas protein comprises a dominant negative version of p53-binding protein 1 (53BP1), for example, a fragment of 53BP1 comprising a minimum focus forming region (*e.g.*, amino acids 1231-1644 of human 53BP1). In certain embodiments, the Cas protein comprises a motif that is targeted by APC-Cdh1, such as amino acids 1-110 of human Geminin, thereby resulting in degradation of the fusion protein during the HDR non-permissive G1 phase of the cell cycle.

[0122] In certain embodiments, the Cas protein comprises an inducible or controllable domain. Non-limiting examples of inducers or controllers include light, hormones, and small molecule drugs. In certain embodiments, the Cas protein comprises a light inducible or controllable domain. In certain embodiments, the Cas protein comprises a chemically inducible or controllable domain.

[0123] In certain embodiments, the Cas protein comprises a tag protein or peptide for ease of tracking or purification. Non-limiting examples of tag proteins and peptides include fluorescent proteins (*e.g.*, green fluorescent protein (GFP), YFP, RFP, CFP, mCherry, tdTomato), HIS tags (*e.g.*, 6×His tag, (SEQ ID NO: 789)), hemagglutinin (HA) tag, FLAG tag, and Myc tag.

[0124] In certain embodiments, the Cas protein is conjugated to a non-protein moiety, such as a fluorophore useful for genomic imaging. In certain embodiments, the Cas protein is covalently conjugated to the non-protein moiety. The terms "CRISPR-Associated protein," "Cas protein," "Cas," "CRISPR-Associated nuclease," and "Cas nuclease" are used herein to include such conjugates despite the presence of one or more non-protein moieties.

Guide Nucleic Acids

[0125] In certain embodiments, the guide nucleic acid of the present invention is a guide nucleic acid that is capable of binding a Cas protein alone (*e.g.*, in the absence of a tracrRNA). Such guide nucleic acid is also called a single guide nucleic acid. In certain embodiments, the single guide nucleic acid is capable of activating a Cas nuclease alone (*e.g.*, in the absence of a tracrRNA). The present invention also provides an engineered, non-naturally occurring system comprising the single guide nucleic acid. In certain embodiments,

the system further comprises the Cas protein that the single guide nucleic acid is capable of binding or the Cas nuclease that the single guide nucleic acid is capable of activating.

[0126] In other embodiments, the guide nucleic acid of the present invention is a targeter nucleic acid that, in combination with a modulator nucleic acid, is capable of binding a Cas protein. In certain embodiments, the guide nucleic acid is a targeter nucleic acid that, in combination with a modulator nucleic acid, is capable of activating a Cas nuclease. The present invention also provides an engineered, non-naturally occurring system comprising the targeter nucleic acid and the cognate modulator nucleic acid. In certain embodiments, the system further comprises the Cas protein that the targeter nucleic acid and the modulator nucleic acid are capable of binding or the Cas nuclease that the targeter nucleic acid and the modulator nucleic acid are capable of activating.

[0127] It is contemplated that the single or dual guide nucleic acids need to be the compatible with a Cas protein (*e.g.*, Cas nuclease) to provide an operative CRISPR system. For example, the targeter stem sequence and the modulator stem sequence can be derived from a naturally occurring crRNA capable of activating a Cas nuclease in the absence of a tracrRNA. Alternatively, the targeter stem sequence and the modulator stem sequence can be derived from a naturally occurring set of crRNA and tracrRNA, respectively, that are capable of activating a Cas nuclease. In certain embodiments, the nucleotide sequences of the targeter stem sequence and the modulator stem sequence are identical to the corresponding stem sequences of a stem-loop structure in such naturally occurring crRNA.

[0128] Guide nucleic acid sequences that are operative with a type II or type V Cas protein are known in the art and are disclosed, for example, in U.S. Patent Nos. 9,790,490, 9,896,696, and 10,113,179, and U.S. Patent Application Publication Nos. 2014/0242664 and 2014/0068797. Exemplary single guide and dual guide sequences that are operative with certain type V-A Cas proteins are provided in Tables 4 and 5, respectively. It is understood that these sequences are merely illustrative, and other guide nucleic acid sequences may also be used with these Cas proteins.

Table 4. Type V-A Cas Protein and Corresponding Single Guide Nucleic Acid Sequences

Cas Protein	Scaffold Sequence[1]	PAM[2]
MAD7 (SEQ ID NO: 1)	UAAUUUCUACUCUUGUAGA (SEQ ID NO: 15), AUCUACAACAGUAGA (SEQ ID NO: 16), AUCUACAAAAGUAGA (SEQ ID NO: 17), GGAAUUUCUACUCUUGUAGA (SEQ ID NO: 18), UAAUUCCCACUCUUGUGGG (SEQ ID NO: 19)	5' TTTN or 5' CTTN
MAD2 (SEQ ID NO: 2)	AUCUACAAGAGUAGA (SEQ ID NO: 20), AUCUACAACAGUAGA (SEQ ID NO: 16), AUCUACAAAAGUAGA (SEQ ID NO: 17), AUCUACACUAGUAGA (SEQ ID NO: 21)	5' TTTN
AsCpf1 (SEQ ID NO: 3)	UAAUUUCUACUCUUGUAGA (SEQ ID NO: 15)	5' TTTN
LbCpf1 (SEQ ID NO: 4)	UAAUUUCUACUAAGUGUAGA (SEQ ID NO: 22)	5' TTTN
FnCpf1 (SEQ ID NO: 5)	UAAUUUUCUACUUGUUGUAGA (SEQ ID NO: 23)	5' TTN
PbCpf1 (SEQ ID NO: 6)	AAUUUCUACUGUUGUAGA (SEQ ID NO: 24)	5' TTTC
PsCpf1 (SEQ ID NO: 7)	AAUUUCUACUGUUGUAGA (SEQ ID NO: 24)	5' TTTC
As2Cpf1 (SEQ ID NO: 8)	AAUUUCUACUGUUGUAGA (SEQ ID NO: 24)	5' TTTC
McCpf1 (SEQ ID NO: 9)	GAAUUUCUACUGUUGUAGA (SEQ ID NO: 25)	5' TTTC
Lb3Cpf1 (SEQ ID NO: 10)	GAAUUUCUACUGUUGUAGA (SEQ ID NO: 25)	5' TTTC
EcCpf1 (SEQ ID NO: 11)	GAAUUUCUACUGUUGUAGA (SEQ ID NO: 25)	5' TTTC
SmCsm1 (SEQ ID NO: 12)	GAAUUUCUACUGUUGUAGA (SEQ ID NO: 25)	5' TTTC
SsCsm1 (SEQ ID NO: 13)	GAAUUUCUACUGUUGUAGA (SEQ ID NO: 25)	5' TTTC
MbCsm1 (SEQ ID NO: 14)	GAAUUUCUACUGUUGUAGA (SEQ ID NO: 25)	5' TTTC

[1] The modulator sequence in the scaffold sequence is underlined; the targeter stem sequence in the scaffold sequence is bold-underlined. It is understood that a "scaffold sequence" listed herein constitutes a portion of a single guide nucleic acid. Additional nucleotide sequences, other than the spacer sequence, can be comprised in the single guide nucleic acid.

[2] In the consensus PAM sequences, N represents A, C, G, or T. Where the PAM sequence is preceded by "5'," it means that the PAM is located immediately upstream of the target nucleotide sequence when using the non-target strand (*i.e.*, the strand not hybridized with the spacer sequence) as the coordinate.

Table 5. Type V-A Cas Protein and Corresponding Dual Guide Nucleic Acid Sequences

Cas Protein	Modulator Sequence[1]	Targeter Stem Sequence	PAM[2]
MAD7 (SEQ ID NO: 1)	UAAUUUCUAC (SEQ ID NO: 26)	GUAGA	5' TTTN or 5' CTTN
	AUCUAC (SEQ ID NO: 27)	GUAGA	
	GGAAUUUCUAC (SEQ ID NO: 28)	GUAGA	
	UAAUUCCCAC (SEQ ID NO: 29)	GUGGG	
MAD2 (SEQ ID NO: 2)	AUCUAC (SEQ ID NO: 27)	GUAGA	5' TTTN
AsCpf1 (SEQ ID NO: 3)	UAAUUUCUAC (SEQ ID NO: 26)	GUAGA	5' TTTN
LbCpf1 (SEQ ID NO: 4)	UAAUUUCUAC (SEQ ID NO: 26)	GUAGA	5' TTTN
FnCpf1 (SEQ ID NO: 5)	UAAUUUUCUACU (SEQ ID NO: 30)	GUAGA	5' TTN
PbCpf1 (SEQ ID NO: 6)	AAUUUCUAC (SEQ ID NO: 31)	GUAGA	5' TTTC
PsCpf1 (SEQ ID NO: 7)	AAUUUCUAC (SEQ ID NO: 31)	GUAGA	5' TTTC
As2Cpf1 (SEQ ID NO: 8)	AAUUUCUAC (SEQ ID NO: 31)	GUAGA	5' TTTC
McCpf1 (SEQ ID NO: 9)	GAAUUUCUAC (SEQ ID NO: 32)	GUAGA	5' TTTC
Lb3Cpf1 (SEQ ID NO: 10)	GAAUUUCUAC (SEQ ID NO: 32)	GUAGA	5' TTTC
EcCpf1 (SEQ ID NO: 11)	GAAUUUCUAC (SEQ ID NO: 32)	GUAGA	5' TTTC
SmCsm1 (SEQ ID NO: 12)	GAAUUUCUAC (SEQ ID NO: 32)	GUAGA	5' TTTC
SsCsm1 (SEQ ID NO: 13)	GAAUUUCUAC (SEQ ID NO: 32)	GUAGA	5' TTTC
MbCsm1 (SEQ ID NO: 14)	GAAUUUCUAC (SEQ ID NO: 32)	GUAGA	5' TTTC

[1] It is understood that a "modulator sequence" listed herein may constitute the nucleotide sequence of a modulator nucleic acid. Alternatively, additional nucleotide sequences can be comprised in the modulator nucleic acid 5' and/or 3' to a "modulator sequence" listed herein.
[2] In the consensus PAM sequences, N represents A, C, G, or T. Where the PAM sequence is preceded by "5'," it means that the PAM is located immediately upstream of the target nucleotide sequence when using the non-target strand (*i.e.*, the strand not hybridized with the spacer sequence) as the coordinate.

[0129] In certain embodiments, the guide nucleic acid of the present invention, in the context of a type V-A CRISPR-Cas system, comprises a targeter stem sequence listed in Table 5. The same targeter stem sequences, as a portion of scaffold sequences, are bold-underlined in Table 4.

[0130] In certain embodiments, the guide nucleic acid is a single guide nucleic acid that comprises, from 5' to 3', a modulator stem sequence, a loop sequence, a targeter stem sequence, and a spacer sequence disclosed herein. In certain embodiments, the targeter stem sequence in the single guide nucleic acid is listed in Table 4 as a bold-underlined portion of scaffold sequence, and the modulator stem sequence is complementary (*e.g.*, 100% complementary) to the targeter stem sequence. In certain embodiments, the single guide nucleic acid comprises, from 5' to 3', a modulator sequence listed in Table 4 as an underlined portion of a scaffold sequence, a loop sequence, a targeter stem sequence a bold-underlined portion of the same scaffold sequence, and a spacer sequence disclosed herein. In certain embodiments, an engineered, non-naturally occurring system of the present invention comprises the single guide nucleic acid comprising a scaffold sequence listed in Table 4. In certain embodiments, the system further comprises a Cas protein (*e.g.*, Cas nuclease) comprising an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the amino acid sequence set forth in the SEQ ID NO listed in the same line of Table 4. In certain embodiments, the system further comprises a Cas protein (*e.g.*, Cas nuclease) comprising the amino acid sequence set forth in the SEQ ID NO listed in the same line of Table 4. In certain embodiments, the system is useful for targeting, editing, or modifying a nucleic acid comprising a target nucleotide sequence close or adjacent to (*e.g.*, immediately downstream of) a PAM listed in the same line of Table 4 when using the non-target strand (*i.e.*, the strand not hybridized with the spacer sequence) as the coordinate.

[0131] In certain embodiments, the guide nucleic acid is a targeter guide nucleic acid that comprises, from 5' to 3', a targeter stem sequence and a spacer sequence disclosed herein. In certain embodiments, the targeter stem sequence in the targeter nucleic acid is listed in Table 5. In certain embodiments, an engineered, non-naturally occurring system of the present invention comprises the targeter nucleic acid and a modulator stem sequence complementary (*e.g.*, 100% complementary) to the targeter stem sequence. In certain embodiments, the modulator nucleic acid comprises a modulator sequence listed in the same line of Table 5. In certain embodiments, the system further comprises a Cas protein (*e.g.*, Cas nuclease) comprising an amino acid sequence at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%

identical to the amino acid sequence set forth in the SEQ ID NO listed in the same line of Table 5. In certain embodiments, the system further comprises a Cas protein (*e.g.*, Cas nuclease) comprising the amino acid sequence set forth in the SEQ ID NO listed in the same line of Table 5. In certain embodiments, the system is useful for targeting, editing, or modifying a nucleic acid comprising a target nucleotide sequence close or adjacent to (*e.g.*, immediately downstream of) a PAM listed in the same line of Table 5 when using the non-target strand (*i.e.*, the strand not hybridized with the spacer sequence) as the coordinate.

[0132] The single guide nucleic acid, the targeter nucleic acid, and/or the modulator nucleic acid can be synthesized chemically or produced in a biological process (*e.g.*, catalyzed by an RNA polymerase in an *in vitro* reaction). Such reaction or process may limit the lengths of the single guide nucleic acid, targeter nucleic acid, and modulator nucleic acid. In certain embodiments, the single guide nucleic acid is no more than 100, 90, 80, 70, 60, 50, 40, 30, or 25 nucleotides in length. In certain embodiments, the single guide nucleic acid is at least 20, 25, 30, 40, 50, 60, 70, 80, or 90 nucleotides in length. In certain embodiments, the single guide nucleic acid is 20-100, 20-90, 20-80, 20-70, 20-60, 20-50, 20-40, 20-30, 20-25, 25-100, 25-90, 25-80, 25-70, 25-60, 25-50, 25-40, 25-30, 30-100, 30-90, 30-80, 30-70, 30-60, 30-50, 30-40, 40-100, 40-90, 40-80, 40-70, 40-60, 40-50, 50-100, 50-90, 50-80, 50-70, 50-60, 60-100, 60-90, 60-80, 60-70, 70-100, 70-90, 70-80, 80-100, 80-90, or 90-100 nucleotides in length. In certain embodiments, the targeter nucleic acid is no more than 100, 90, 80, 70, 60, 50, 40, 30, or 25 nucleotides in length. In certain embodiments, the targeter nucleic acid is at least 20, 25, 30, 40, 50, 60, 70, 80, or 90 nucleotides in length. In certain embodiments, the targeter nucleic acid is 20-100, 20-90, 20-80, 20-70, 20-60, 20-50, 20-40, 20-30, 20-25, 25-100, 25-90, 25-80, 25-70, 25-60, 25-50, 25-40, 25-30, 30-100, 30-90, 30-80, 30-70, 30-60, 30-50, 30-40, 40-100, 40-90, 40-80, 40-70, 40-60, 40-50, 50-100, 50-90, 50-80, 50-70, 50-60, 60-100, 60-90, 60-80, 60-70, 70-100, 70-90, 70-80, 80-100, 80-90, or 90-100 nucleotides in length. In certain embodiments, the modulator nucleic acid is no more than 100, 90, 80, 70, 60, 50, 40, 30, or 20 nucleotides in length. In certain embodiments, the modulator nucleic acid is at least 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, or 90 nucleotides in length. In certain embodiments, the modulator nucleic acid is 10-100, 10-90, 10-80, 10-70, 10-60, 10-50, 10-40, 10-30, 10-20, 15-100, 15-90, 15-80, 15-70, 15-60, 15-50, 15-40, 15-30, 15-20, 20-100, 20-90, 20-80, 20-70, 20-60, 20-50, 20-40, 20-30, 25-100, 25-90, 25-80, 25-70, 25-60, 25-50, 25-40, 25-30, 30-100, 30-90, 30-80, 30-70, 30-60, 30-50, 30-40, 40-100, 40-90,

40-80, 40-70, 40-60, 40-50, 50-100, 50-90, 50-80, 50-70, 50-60, 60-100, 60-90, 60-80, 60-70, 70-100, 70-90, 70-80, 80-100, 80-90, or 90-100 nucleotides in length.

[0133] It is contemplated that the length of the duplex formed within the single guide nuclei acid or formed between the targeter nucleic acid and the modulator nucleic acid may be a factor in providing an operative CRISPR system. In certain embodiments, the targeter stem sequence and the modulator stem sequence each consist of 4-10 nucleotides that base pair with each other. In certain embodiments, the targeter stem sequence and the modulator stem sequence each consist of 4-9, 4-8, 4-7, 4-6, 4-5, 5-10, 5-9, 5-8, 5-7, or 5-6 nucleotides that base pair with each other. In certain embodiments, the targeter stem sequence and the modulator stem sequence each consist of 4, 5, 6, 7, 8, 9, or 10 nucleotides. It is understood that the composition of the nucleotides in each sequence affects the stability of the duplex, and a C-G base pair confers greater stability than an A-U base pair. In certain embodiments, 20%-80%, 20%-70%, 20%-60%, 20%-50%, 20%-40%, 20%-30%, 30%-80%, 30%-70%, 30%-60%, 30%-50%, 30%-40%, 40%-80%, 40%-70%, 40%-60%, 40%-50%, 50%-80%, 50%-70%, 50%-60%, 60%-80%, 60%-70%, or 70%-80% of the base pairs are C-G base pairs.

[0134] In certain embodiments, the targeter stem sequence and the modulator stem sequence each consist of 5 nucleotides. As such, the targeter stem sequence and the modulator stem sequence form a duplex of 5 base pairs. In certain embodiments, 0-4, 0-3, 0-2, 0-1, 1-5, 1-4, 1-3, 1-2, 2-5, 2-4, 2-3, 3-5, 3-4, or 4-5 out of the 5 base pairs are C-G base pairs. In certain embodiments, 0, 1, 2, 3, 4, or 5 out of the 5 base pairs are C-G base pairs. In certain embodiments, the targeter stem sequence consists of 5'-GUAGA-3' and the modulator stem sequence consists of 5'-UCUAC-3'. In certain embodiments, the targeter stem sequence consists of 5'-GUGGG-3' and the modulator stem sequence consists of 5'-CCCAC-3'.

[0135] In certain embodiments, in a type V-A system, the 3' end of the targeter stem sequence is linked by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides to the 5' end of the spacer sequence. In certain embodiments, the targeter stem sequence and the spacer sequence are adjacent to each other, directly linked by an internucleotide bond. In certain embodiments, the targeter stem sequence and the spacer sequence are linked by one nucleotide, *e.g.*, a uridine. In certain embodiments, the targeter stem sequence and the spacer sequence are linked by two or more nucleotides. In certain embodiments, the targeter stem sequence and the spacer sequence are linked by 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides.

[0136] In certain embodiments, the targeter nucleic acid further comprises an additional nucleotide sequence 5' to the targeter stem sequence. In certain embodiments, the additional nucleotide sequence comprises at least 1 (*e.g.*, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, or at least 50) nucleotides. In certain embodiments, the additional nucleotide sequence consists of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides. In certain embodiments, the additional nucleotide sequence consists of 2 nucleotides. In certain embodiments, the additional nucleotide sequence is reminiscent to the loop or a fragment thereof (*e.g.*, one, two, three, or four nucleotides at the 3' end of the loop) in a crRNA of a corresponding single guide CRISPR-Cas system. It is understood that an additional nucleotide sequence 5' to the targeter stem sequence is dispensable. Accordingly, in certain embodiments, the targeter nucleic acid does not comprise any additional nucleotide 5' to the targeter stem sequence.

[0137] In certain embodiments, the targeter nucleic acid or the single guide nucleic acid further comprises an additional nucleotide sequence containing one or more nucleotides at the 3' end that does not hybridize with the target nucleotide sequence. The additional nucleotide sequence may protect the targeter nucleic acid from degradation by 3'-5' exonuclease. In certain embodiments, the additional nucleotide sequence is no more than 100 nucleotides in length. In certain embodiments, the additional nucleotide sequence is no more than 90, 80, 70, 60, 50, 40, 30, 20, or 10 nucleotides in length. In certain embodiments, the additional nucleotide sequence is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides in length. In certain embodiments, the additional nucleotide sequence is 5-100, 5-50, 5-40, 5-30, 5-25, 5-20, 5-15, 5-10, 10-100, 10-50, 10-40, 10-30, 10-25, 10-20, 10-15, 15-100, 15-50, 15-40, 15-30, 15-25, 15-20, 20-100, 20-50, 20-40, 20-30, 20-25, 25-100, 25-50, 25-40, 25-30, 30-100, 30-50, 30-40, 40-100, 40-50, or 50-100 nucleotides in length.

[0138] In certain embodiments, the additional nucleotide sequence forms a hairpin with the spacer sequence. Such secondary structure may increase the specificity of guide nucleic acid or the engineered, non-naturally occurring system (see, Kocak *et al.* (2019) NAT. BIOTECH. 37: 657-66). In certain embodiments, the free energy change during the hairpin formation is greater than or equal to -20 kcal/mol, -15 kcal/mol, -14 kcal/mol, -13 kcal/mol, -12 kcal/mol, -11 kcal/mol, or -10 kcal/mol. In certain embodiments, the free energy change during the hairpin formation is greater than or equal to -5 kcal/mol, -6 kcal/mol, -7 kcal/mol, -8 kcal/mol, -9 kcal/mol, -10 kcal/mol, -11 kcal/mol, -12 kcal/mol, -13 kcal/mol, -14

kcal/mol, or -15 kcal/mol. In certain embodiments, the free energy change during the hairpin formation is in the range of -20 to -10 kcal/mol, -20 to -11 kcal/mol, -20 to -12 kcal/mol, -20 to -13 kcal/mol, -20 to -14 kcal/mol, -20 to -15 kcal/mol, -15 to -10 kcal/mol, -15 to -11 kcal/mol, -15 to -12 kcal/mol, -15 to -13 kcal/mol, -15 to -14 kcal/mol, -14 to -10 kcal/mol, -14 to -11 kcal/mol, -14 to -12 kcal/mol, -14 to -13 kcal/mol, -13 to -10 kcal/mol, -13 to -11 kcal/mol, -13 to -12 kcal/mol, -12 to -10 kcal/mol, -12 to -11 kcal/mol, or -11 to -10 kcal/mol. In other embodiments, the targeter nucleic acid or the single guide nucleic acid does not comprise any nucleotide 3' to the spacer sequence.

[0139] In certain embodiments, the modulator nucleic acid further comprises an additional nucleotide sequence 3' to the modulator stem sequence. In certain embodiments, the additional nucleotide sequence comprises at least 1 (*e.g.*, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, or at least 50) nucleotides. In certain embodiments, the additional nucleotide sequence consists of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, or 50 nucleotides. In certain embodiments, the additional nucleotide sequence consists of 1 nucleotide (*e.g.*, uridine). In certain embodiments, the additional nucleotide sequence consists of 2 nucleotides. In certain embodiments, the additional nucleotide sequence is reminiscent to the loop or a fragment thereof (*e.g.*, one, two, three, or four nucleotides at the 5' end of the loop) in a crRNA of a corresponding single guide CRISPR-Cas system. It is understood that an additional nucleotide sequence 3' to the modulator stem sequence is dispensable. Accordingly, in certain embodiments, the modulator nucleic acid does not comprise any additional nucleotide 3' to the modulator stem sequence.

[0140] It is understood that the additional nucleotide sequence 5' to the targeter stem sequence and the additional nucleotide sequence 3' to the modulator stem sequence, if present, may interact with each other. For example, although the nucleotide immediately 5' to the targeter stem sequence and the nucleotide immediately 3' to the modulator stem sequence do not form a Watson-Crick base pair (otherwise they would constitute part of the targeter stem sequence and part of the modulator stem sequence, respectively), other nucleotides in the additional nucleotide sequence 5' to the targeter stem sequence and the additional nucleotide sequence 3' to the modulator stem sequence may form one, two, three, or more base pairs (*e.g.*, Watson-Crick base pairs). Such interaction may affect the stability of the complex comprising the targeter nucleic acid and the modulator nucleic acid.

[0141] The stability of a complex comprising a targeter nucleic acid and a modulator nucleic acid can be assessed by the Gibbs free energy change (ΔG) during the formation of the complex, either calculated or actually measured. Where all the predicted base pairing in the complex occurs between a base in the targeter nucleic acid and a base in the modulator nucleic acid, *i.e.*, there is no intra-strand secondary structure, the ΔG during the formation of the complex correlates generally with the ΔG during the formation of a secondary structure within the corresponding single guide nucleic acid. Methods of calculating or measuring the ΔG are known in the art. An exemplary method is RNAfold (rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) as disclosed in Gruber *et al.* (2008) NUCLEIC ACIDS RES., 36(Web Server issue): W70–W74. Unless indicated otherwise, the ΔG values in the present disclosure are calculated by RNAfold for the formation of a secondary structure within a corresponding single guide nucleic acid. In certain embodiments, the ΔG is lower than or equal to -1 kcal/mol, *e.g.*, lower than or equal to -2 kcal/mol, lower than or equal to -3 kcal/mol, lower than or equal to -4 kcal/mol, lower than or equal to -5 kcal/mol, lower than or equal to -6 kcal/mol, lower than or equal to -7 kcal/mol, lower than or equal to -7.5 kcal/mol, or lower than or equal to -8 kcal/mol. In certain embodiments, the ΔG is greater than or equal to -10 kcal/mol, *e.g.*, greater than or equal to -9 kcal/mol, greater than or equal to -8.5 kcal/mol, or greater than or equal to -8 kcal/mol. In certain embodiments, the ΔG is in the range of -10 to -4 kcal/mol. In certain embodiments, the ΔG is in the range of -8 to -4 kcal/mol, -7 to -4 kcal/mol, -6 to -4 kcal/mol, -5 to -4 kcal/mol, -8 to -4.5 kcal/mol, -7 to -4.5 kcal/mol, -6 to -4.5 kcal/mol, or -5 to -4.5 kcal/mol. In certain embodiments, the ΔG is about -8 kcal/mol, -7 kcal/mol, -6 kcal/mol, -5 kcal/mol, -4.9 kcal/mol, -4.8 kcal/mol, -4.7 kcal/mol, -4.6 kcal/mol, -4.5 kcal/mol, -4.4 kcal/mol, -4.3 kcal/mol, -4.2 kcal/mol, -4.1 kcal/mol, or -4 kcal/mol.

[0142] It is understood that the ΔG may be affected by a sequence in the targeter nucleic acid that is not within the targeter stem sequence, and/or a sequence in the modulator nucleic acid that is not within the modulator stem sequence. For example, one or more base pairs (*e.g.*, Watson-Crick base pair) between an additional sequence 5' to the targeter stem sequence and an additional sequence 3' to the modulator stem sequence may reduce the ΔG, *i.e.*, stabilize the nucleic acid complex. In certain embodiments, the nucleotide immediately 5' to the targeter stem sequence comprises a uracil or is a uridine, and the nucleotide immediately 3' to the modulator stem sequence comprises a uracil or is a uridine, thereby forming a nonconventional U-U base pair.

[0143] In certain embodiments, the modulator nucleic acid or the single guide nucleic acid comprises a nucleotide sequence referred to herein as a "5' tail" positioned 5' to the modulator stem sequence. In a naturally occurring type V-A CRISPR-Cas system, the 5' tail is a nucleotide sequence positioned 5' to the stem-loop structure of the crRNA. A 5' tail in an engineered type V-A CRISPR-Cas system, whether single guide or dual guide, can be reminiscent to the 5' tail in a corresponding naturally occurring type V-A CRISPR-Cas system.

[0144] Without being bound by theory, it is contemplated that the 5' tail may participate in the formation of the CRISPR-Cas complex. For example, in certain embodiments, the 5' tail forms a pseudoknot structure with the modulator stem sequence, which is recognized by the Cas protein (see, Yamano *et al.* (2016) CELL, 165: 949). In certain embodiments, the 5' tail is at least 3 (*e.g.*, at least 4 or at least 5) nucleotides in length. In certain embodiments, the 5' tail is 3, 4, or 5 nucleotides in length. In certain embodiments, the nucleotide at the 3' end of the 5' tail comprises a uracil or is a uridine. In certain embodiments, the second nucleotide in the 5' tail, the position counted from the 3' end, comprises a uracil or is a uridine. In certain embodiments, the third nucleotide in the 5' tail, the position counted from the 3' end, comprises an adenine or is an adenosine. This third nucleotide may form a base pair (*e.g.*, a Watson-Crick base pair) with a nucleotide 5' to the modulator stem sequence. Accordingly, in certain embodiments, the modulator nucleic acid comprises a uridine or a uracil-containing nucleotide 5' to the modulator stem sequence. In certain embodiments, the 5' tail comprises the nucleotide sequence of 5'-AUU-3'. In certain embodiments, the 5' tail comprises the nucleotide sequence of 5'-AAUU-3'. In certain embodiments, the 5' tail comprises the nucleotide sequence of 5'-UAAUU-3'. In certain embodiments, the 5' tail is positioned immediately 5' to the modulator stem sequence.

[0145] In certain embodiments, the single guide nucleic acid, the targeter nucleic acid, and/or the modulator nucleic acid are designed to reduce the degree of secondary structure other than the hybridization between the targeter stem sequence and the modulator stem sequence. In certain embodiments, no more than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the single guide nucleic acid other than the targeter stem sequence and the modulator stem sequence participate in self-complementary base pairing when optimally folded. In certain embodiments, no more than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the targeter nucleic acid and/or the modulator nucleic acid participate in self-complementary base pairing when

optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see *e.g.*, A. R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[0146] The targeter nucleic acid is directed to a specific target nucleotide sequence, and a donor template can be designed to modify the target nucleotide sequence or a sequence nearby. It is understood, therefore, that association of the single guide nucleic acid, the targeter nucleic acid, or the modulator nucleic acid with a donor template can increase editing efficiency and reduce off-targeting. Accordingly, in certain embodiments, the single guide nucleic acid or the modulator nucleic acid further comprises a donor template-recruiting sequence capable of hybridizing with a donor template (see **Figure 2B**). Donor templates are described in the "Donor Templates" subsection of section II *infra*. The donor template and donor template-recruiting sequence can be designed such that they bear sequence complementarity. In certain embodiments, the donor template-recruiting sequence is at least 90% (*e.g.*, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%) complementary to at least a portion of the donor template. In certain embodiments, the donor template-recruiting sequence is 100% complementary to at least a portion of the donor template. In certain embodiments, where the donor template comprises an engineered sequence not homologous to the sequence to be repaired, the donor template-recruiting sequence is capable of hybridizing with the engineered sequence in the donor template. In certain embodiments, the donor template-recruiting sequence is at least 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides in length. In certain embodiments, the donor template-recruiting sequence is positioned at or near the 5' end of the single guide nucleic acid or at or near the 5' end of the modulator nucleic acid. In certain embodiments, the donor template-recruiting sequence is linked to the 5' tail, if present, or to the modulator stem sequence, of the single guide nucleic acid or the modulator nucleic acid through an internucleotide bond or a nucleotide linker.

[0147] In certain embodiments, the single guide nucleic acid or the modulator nucleic acid further comprises an editing enhancer sequence, which increases the efficiency of gene editing and/or homology-directed repair (HDR) (see **Figure 2C**). Exemplary editing

enhancer sequences are described in Park *et al.* (2018) NAT. COMMUN. 9: 3313. In certain embodiments, the editing enhancer sequence is positioned 5' to the 5' tail, if present, or 5' to the single guide nucleic acid or the modulator stem sequence. In certain embodiments, the editing enhancer sequence is 1-50, 4-50, 9-50, 15-50, 25-50, 1-25, 4-25, 9-25, 15-25, 1-15, 4-15, 9-15, 1-9, 4-9, or 1-4 nucleotides in length. In certain embodiments, the editing enhancer sequence is about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, or 55 nucleotides in length. The editing enhancer sequence is designed to minimize homology to the target nucleotide sequence or any other sequence that the engineered, non-naturally occurring system may be contacted to, *e.g.*, the genome sequence of a cell into which the engineered, non-naturally occurring system is delivered. In certain embodiments, the editing enhancer is designed to minimize the presence of hairpin structure. The editing enhancer can comprise one or more of the chemical modifications disclosed herein.

[0148] The single guide nucleic acid, the modulator nucleic acid, and/or the targeter nucleic acid can further comprise a protective nucleotide sequence that prevents or reduces nucleic acid degradation. In certain embodiments, the protective nucleotide sequence is at least 5 (*e.g.*, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, or at least 50) nucleotides in length. The length of the protective nucleotide sequence increases the time for an exonuclease to reach the 5' tail, modulator stem sequence, targeter stem sequence, and/or spacer sequence, thereby protecting these portions of the single guide nucleic acid, the modulator nucleic acid, and/or the targeter nucleic acid from degradation by an exonuclease. In certain embodiments, the protective nucleotide sequence forms a secondary structure, such as a hairpin or a tRNA structure, to reduce the speed of degradation by an exonuclease (see, for example, Wu *et al.* (2018) CELL. MOL. LIFE SCI., 75(19): 3593-3607). Secondary structures can be predicted by methods known in the art, such as the online webserver RNAfold developed at University of Vienna using the centroid structure prediction algorithm (see, Gruber *et al.* (2008) NUCLEIC ACIDS RES., 36: W70). Certain chemical modifications, which may be present in the protective nucleotide sequence, can also prevent or reduce nucleic acid degradation, as disclosed in the "RNA Modifications" subsection *infra*.

[0149] A protective nucleotide sequence is typically located at the 5' or 3' end of the single guide nucleic acid, the modulator nucleic acid, and/or the targeter nucleic acid. In certain embodiments, the single guide nucleic acid comprises a protective nucleotide sequence at the 5' end, at the 3' end, or at both ends, optionally through a nucleotide linker.

In certain embodiments, the modulator nucleic acid comprises a protective nucleotide sequence at the 5' end, at the 3' end, or at both ends, optionally through a nucleotide linker. In particular embodiments, the modulator nucleic acid comprises a protective nucleotide sequence at the 5' end (see **Figure 2A**). In certain embodiments, the targeter nucleic acid comprises a protective nucleotide sequence at the 5' end, at the 3' end, or at both ends, optionally through a nucleotide linker.

[0150] As described above, various nucleotide sequences can be present in the 5' portion of a single nucleic acid or a modulator nucleic acid, including but not limited to a donor template-recruiting sequence, an editing enhancer sequence, a protective nucleotide sequence, and a linker connecting such sequence to the 5' tail, if present, or to the modulator stem sequence. It is understood that the functions of donor template recruitment, editing enhancement, protection against degradation, and linkage are not exclusive to each other, and one nucleotide sequence can have one or more of such functions. For example, in certain embodiments, the single guide nucleic acid or the modulator nucleic acid comprises a nucleotide sequence that is both a donor template-recruiting sequence and an editing enhancer sequence. In certain embodiments, the single guide nucleic acid or the modulator nucleic acid comprises a nucleotide sequence that is both a donor template-recruiting sequence and a protective sequence. In certain embodiments, the single guide nucleic acid or the modulator nucleic acid comprises a nucleotide sequence that is both an editing enhancer sequence and a protective sequence. In certain embodiments, the single guide nucleic acid or the modulator nucleic acid comprises a nucleotide sequence that is a donor template-recruiting sequence, an editing enhancer sequence, and a protective sequence. In certain embodiments, the nucleotide sequence 5' to the 5' tail, if present, or 5' to the modulator stem sequence is 1-90, 1-80, 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10, 10-90, 10-80, 10-70, 10-60, 10-50, 10-40, 10-30, 10-20, 20-90, 20-80, 20-70, 20-60, 20-50, 20-40, 20-30, 30-90, 30-80, 30-70, 30-60, 30-50, 30-40, 40-90, 40-80, 40-70, 40-60, 40-50, 50-90, 50-80, 50-70, 50-60, 60-90, 60-80, 60-70, 70-90, 70-80, or 80-90 nucleotides in length.

[0151] In certain embodiments, the engineered, non-naturally occurring system further comprises one or more compounds (*e.g.*, small molecule compounds) that enhance HDR and/or inhibit NHEJ. Exemplary compounds having such functions are described in Maruyama *et al.* (2015) NAT BIOTECHNOL. 33(5): 538-42; Chu *et al.* (2015) NAT BIOTECHNOL. 33(5): 543-48; Yu *et al.* (2015) CELL STEM CELL 16(2): 142-47; Pinder *et al.* (2015) NUCLEIC ACIDS RES. 43(19): 9379-92; and Yagiz *et al.* (2019) COMMUN. BIOL. 2: 198.

In certain embodiments, the engineered, non-naturally occurring system further comprises one or more compounds selected from the group consisting of DNA ligase IV antagonists (*e.g.*, SCR7 compound, Ad4 E1B55K protein, and Ad4 E4orf6 protein), RAD51 agonists (*e.g.*, RS-1), DNA-dependent protein kinase (DNA-PK) antagonists (*e.g.*, NU7441 and KU0060648), β3-adrenergic receptor agonists (*e.g.*, L755507), inhibitors of intracellular protein transport from the ER to the Golgi apparatus (*e.g.*, brefeldin A), and any combinations thereof.

[0152] In certain embodiments, the engineered, non-naturally occurring system comprising a targeter nucleic acid and a modulator nucleic acid is tunable or inducible. For example, in certain embodiments, the targeter nucleic acid, the modulator nucleic acid, and/or the Cas protein can be introduced to the target nucleotide sequence at different times, the system becoming active only when all components are present. In certain embodiments, the amounts of the targeter nucleic acid, the modulator nucleic acid, and/or the Cas protein can be titrated to achieve desired efficiency and specificity. In certain embodiments, excess amount of a nucleic acid comprising the targeter stem sequence or the modulator stem sequence can be added to the system, thereby dissociating the complex of the targeter nucleic and modulator nucleic acid and turning off the system.

RNA Modifications

[0153] The guide nucleic acids disclosed herein, including a single guide nucleic acid, a targeter nucleic acid, and/or a modulator nucleic acid, may comprise a DNA (*e.g.*, modified DNA), an RNA (*e.g.*, modified RNA), or a combination thereof. In certain embodiments, the single guide nucleic acid comprises a DNA (*e.g.*, modified DNA), an RNA (*e.g.*, modified RNA), or a combination thereof. In certain embodiments, the targeter nucleic acid comprises a DNA (*e.g.*, modified DNA), an RNA (*e.g.*, modified RNA), or a combination thereof. In certain embodiments, the modulator nucleic acid comprises a DNA (*e.g.*, modified DNA), an RNA (*e.g.*, modified RNA), or a combination thereof. The spacer sequences disclosed herein are presented as DNA sequences by including thymidines (T) rather than uridines (U). It is understood that corresponding RNA sequences and DNA/RNA chimeric sequences are also contemplated. For example, where the spacer sequence is an RNA, its sequence can be derived from a DNA sequence disclosed herein by replacing each T with U. As a result, for the purpose of describing a nucleotide sequence, T and U are used interchangeably herein.

[0154] In certain embodiments, the single guide nucleic acid is an RNA. A single guide nucleic acid in the form of an RNA is also called a single guide RNA. In certain embodiments, the targeter nucleic acid is an RNA and the modulator nucleic acid is an RNA. A targeter nucleic acid in the form of an RNA is also called targeter RNA, and a modulator nucleic acid in the form of an RNA is also called modulator RNA.

[0155] In certain embodiments, the single guide nucleic acid, the targeter nucleic acid, and/or the modulator nucleic acid are RNAs with one or more modifications in a ribose group, one or more modifications in a phosphate group, one or more modifications in a nucleobase, one or more terminal modifications, or a combination thereof. Exemplary modifications are disclosed in U.S. Patent Application Publication Nos. 2016/0289675, 2017/0355985, 2018/0119140, Watts *et al.* (2008) Drug Discov. Today 13: 842-55, and Hendel *et al.* (2015) NAT. BIOTECHNOL. 33: 985.

[0156] Modifications in a ribose group include but are not limited to modifications at the 2′ position or modifications at the 4′ position. For example, in certain embodiments, the ribose comprises 2′-O-C1-4alkyl, such as 2′-O-methyl (2′-OMe). In certain embodiments, the ribose comprises 2′-O-C1-3alkyl-O-C1-3alkyl, such as 2′-methoxyethoxy (2′-O—$CH_2CH_2OCH_3$) also known as 2′-O-(2-methoxyethyl) or 2′-MOE. In certain embodiments, the ribose comprises 2′-O-allyl. In certain embodiments, the ribose comprises 2′-O-2,4-Dinitrophenol (DNP). In certain embodiments, the ribose comprises 2′-halo, such as 2′-F, 2′-Br, 2′-Cl, or 2′-I. In certain embodiments, the ribose comprises 2′-NH_2. In certain embodiments, the ribose comprises 2′-H (*e.g.*, a deoxynucleotide). In certain embodiments, the ribose comprises 2′-arabino or 2′-F-arabino. In certain embodiments, the ribose comprises 2′-LNA or 2′-ULNA. In certain embodiments, the ribose comprises a 4′-thioribosyl.

[0157] Modifications in a phosphate group include but are not limited to a phosphorothioate internucleotide linkage, a chiral phosphorothioate internucleotide linkage, a phosphorodithioate internucleotide linkage, a boranophosphonate internucleotide linkage, a C_{1-4}alkyl phosphonate internucleotide linkage such as a methylphosphonate internucleotide linkage, a boranophosphonate internucleotide linkage, a phosphonocarboxylate internucleotide linkage such as a phosphonoacetate internucleotide linkage, a phosphonocarboxylate ester internucleotide linkage such as a phosphonoacetate ester internucleotide linkage, an amide linkage, a thiophosphonocarboxylate internucleotide linkage such as a thiophosphonoacetate internucleotide linkage, a thiophosphonocarboxylate

ester internucleotide linkage such as a thiophosphonoacetate ester internucleotide linkage, and a 2′,5′-linkage having a phosphodiester linker or any of the linkers above. Various salts, mixed salts and free acid forms are also included.

[0158] Modifications in a nucleobase include but are not limited to 2-thiouracil, 2-thiocytosine, 4-thiouracil, 6-thioguanine, 2-aminoadenine, 2-aminopurine, pseudouracil, hypoxanthine, 7-deazaguanine, 7-deaza-8-azaguanine, 7-deazaadenine, 7-deaza-8-azaadenine, 5-methylcytosine, 5-methyluracil, 5-hydroxymethylcytosine, 5-hydroxymethyluracil, 5,6-dehydrouracil, 5-propynylcytosine, 5-propynyluracil, 5-ethynylcytosine, 5-ethynyluracil, 5-allyluracil, 5-allylcytosine, 5-aminoallyluracil, 5-aminoallyl-cytosine, 5-bromouracil, 5-iodouracil, diaminopurine, difluorotoluene, dihydrouracil, an abasic nucleotide, Z base, P base, Unstructured Nucleic Acid, isoguanine, isocytosine (see, Piccirilli *et al.* (1990) NATURE, 343: 33), 5-methyl-2-pyrimidine (see, Rappaport (1993) BIOCHEMISTRY, 32: 3047), x(A,G,C,T), and y(A,G,C,T).

[0159] Terminal modifications include but are not limited to polyethyleneglycol (PEG), hydrocarbon linkers (such as heteroatom (O,S,N)-substituted hydrocarbon spacers; halo-substituted hydrocarbon spacers; keto-, carboxyl-, amido-, thionyl-, carbamoyl-, thionocarbamaoyl-containing hydrocarbon spacers), spermine linkers, dyes such as fluorescent dyes (for example, fluoresceins, rhodamines, cyanines), quenchers (for example, dabcyl, BHQ), and other labels (for example biotin, digoxigenin, acridine, streptavidin, avidin, peptides and/or proteins). In certain embodiments, a terminal modification comprises a conjugation (or ligation) of the RNA to another molecule comprising an oligonucleotide (such as deoxyribonucleotides and/or ribonucleotides), a peptide, a protein, a sugar, an oligosaccharide, a steroid, a lipid, a folic acid, a vitamin and/or other molecule. In certain embodiments, a terminal modification incorporated into the RNA is located internally in the RNA sequence via a linker such as 2-(4-butylamidofluorescein)propane-1,3-diol bis(phosphodiester) linker, which is incorporated as a phosphodiester linkage and can be incorporated anywhere between two nucleotides in the RNA.

[0160] The modifications disclosed above can be combined in the single guide RNA, the targeter RNA, and/or the modulator RNA. In certain embodiments, the modification in the RNA is selected from the group consisting of incorporation of 2'-O-methyl-3'phosphorothioate, 2′-O-methyl-3′-phosphonoacetate, 2′-O-methyl-3′-thiophosphonoacetate, 2′-halo-3′-phosphorothioate (*e.g.*, 2′-fluoro-3′-phosphorothioate), 2′-halo-3′-phosphonoacetate

(*e.g.*, 2′-fluoro-3′-phosphonoacetate), and 2′-halo-3′-thiophosphonoacetate (*e.g.*, 2′-fluoro-3′-thiophosphonoacetate).

[0161] In certain embodiments, the modification alters the stability of the RNA. In certain embodiments, the modification enhances the stability of the RNA, *e.g.*, by increasing nuclease resistance of the RNA relative to a corresponding RNA without the modification. Stability-enhancing modifications include but are not limited to incorporation of 2′-O-methyl, a 2′-O-C_{1-4}alkyl, 2′-halo (*e.g.*, 2′-F, 2′-Br, 2′-Cl, or 2′-I), 2′MOE, a 2′-O-C_{1-3}alkyl-O-C_{1-3}alkyl, 2′-NH_2, 2′-H (or 2′-deoxy), 2′-arabino, 2′-F-arabino, 4′-thioribosyl sugar moiety, 3′-phosphorothioate, 3′-phosphonoacetate, 3′-thiophosphonoacetate, 3′-methylphosphonate, 3′-boranophosphate, 3′-phosphorodithioate, locked nucleic acid ("LNA") nucleotide which comprises a methylene bridge between the 2' and 4' carbons of the ribose ring, and unlocked nucleic acid ("ULNA") nucleotide. Such modifications are suitable for use as a protecting group to prevent or reduce degradation of the 5' tail, modulator stem sequence, targeter stem sequence, and/or spacer sequence (see, the "Guide Nucleic Acids" subsection *supra*).

[0162] In certain embodiments, the modification alters the specificity of the engineered, non-naturally occurring system. In certain embodiments, the modification enhances the specification of the engineered, non-naturally occurring system, *e.g.*, by enhancing on-target binding and/or cleavage, or reducing off-target binding and/or cleavage, or a combination thereof. Specificity-enhancing modifications include but are not limited to 2-thiouracil, 2-thiocytosine, 4-thiouracil, 6-thioguanine, 2-aminoadenine, and pseudouracil.

[0163] In certain embodiments, the modification alters the immunostimulatory effect of the RNA relative to a corresponding RNA without the modification. For example, in certain embodiments, the modification reduces the ability of the RNA to activate TLR7, TLR8, TLR9, TLR3, RIG-I, and/or MDA5.

[0164] In certain embodiments, the single guide nucleic acid, the targeter nucleic acid, and/or the modulator nucleic acid comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 modified nucleotides. The modification can be made at one or more positions in the single guide nucleic acid, the targeter nucleic acid, and/or the modulator nucleic acid such that these nucleic acids retain functionality. For example, the modified nucleic acids can still direct the Cas protein to the target nucleotide sequence and allow the Cas protein to exert its effector function. It is understood that the particular modification(s) at a position may be

selected based on the functionality of the nucleotide at the position. For example, a specificity-enhancing modification may be suitable for a nucleotide in the spacer sequence, the targeter stem sequence, or the modulator stem sequence. A stability-enhancing modification may be suitable for one or more terminal nucleotides in the single guide nucleic acid, the targeter nucleic acid, and/or the modulator nucleic acid. In certain embodiments, at least 1 (*e.g.*, at least 2, at least 3, at least 4, or at least 5) terminal nucleotides at the 5' end and/or at least 1 (*e.g.*, at least 2, at least 3, at least 4, or at least 5) terminal nucleotides at the 3' end of the single guide nucleic acid are modified nucleotides. In certain embodiments, 5 or fewer (*e.g.*, 1 or fewer, 2 or fewer, 3 or fewer, or 4 or fewer) terminal nucleotides at the 5' end and/or 5 or fewer (*e.g.*, 1 or fewer, 2 or fewer, 3 or fewer, or 4 or fewer) terminal nucleotides at the 3' end of the single guide nucleic acid are modified nucleotides. In certain embodiments, at least 1 (*e.g.*, at least 2, at least 3, at least 4, or at least 5) terminal nucleotides at the 5' end and/or at least 1 (*e.g.*, at least 2, at least 3, at least 4, or at least 5) terminal nucleotides at the 3' end of the targeter nucleic acid are modified nucleotides. In certain embodiments, 5 or fewer (*e.g.*, 1 or fewer, 2 or fewer, 3 or fewer, or 4 or fewer) terminal nucleotides at the 5' end and/or 5 or fewer (*e.g.*, 1 or fewer, 2 or fewer, 3 or fewer, or 4 or fewer) terminal nucleotides at the 3' end of the targeter nucleic acid are modified nucleotides. In certain embodiments, at least 1 (*e.g.*, at least 2, at least 3, at least 4, or at least 5) terminal nucleotides at the 5' end and/or at least 1 (*e.g.*, at least 2, at least 3, at least 4, or at least 5) terminal nucleotides at the 3' end of the modulator nucleic acid are modified nucleotides. In certain embodiments, 5 or fewer (*e.g.*, 1 or fewer, 2 or fewer, 3 or fewer, or 4 or fewer) terminal nucleotides at the 5' end and/or 5 or fewer (*e.g.*, 1 or fewer, 2 or fewer, 3 or fewer, or 4 or fewer) terminal nucleotides at the 3' end of the modulator nucleic acid are modified nucleotides. Selection of positions for modifications is described in U.S. Patent Application Publication Nos. 2016/0289675 and 2017/0355985. As used in this paragraph, where the targeter or modulator nucleic acid is a combination of DNA and RNA, the nucleic acid as a whole is considered as an RNA, and the DNA nucleotide(s) are considered as modification(s) of the RNA, including a 2'-H modification of the ribose and optionally a modification of the nucleobase.

[0165] It is understood that the targeter nucleic acid and the modulator nucleic acid, while not in the same nucleic acids, *i.e.*, not linked end-to-end through a traditional internucleotide bond, can be covalently conjugated to each other through one or more chemical modifications

introduced into these nucleic acids, thereby increasing the stability of the double-stranded complex and/or improving other characteristics of the system.

II. Methods of Targeting, Editing, and/or Modifying Genomic DNA

[0166] The engineered, non-naturally occurring system disclosed herein are useful for targeting, editing, and/or modifying a target nucleic acid, such as a DNA (*e.g.*, genomic DNA) in a cell or organism. For example, in certain embodiments, with respect to a given target gene listed in Table 1, 2, or 3, an engineered, non-naturally occurring system disclosed herein that comprises a guide nucleic acid comprising a corresponding spacer sequence, when delivered into a population of human cells (*e.g.*, Jurkat cells) *ex vivo*, edits the genomic sequence at the locus of the target gene in at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0167] The present invention provides a method of cleaving a target nucleic acid (*e.g.*, DNA) comprising the sequence of a preselected target gene or a portion thereof, the method comprising contacting the target DNA with an engineered, non-naturally occurring system disclosed herein, thereby resulting in cleavage of the target DNA.

[0168] In addition, the present invention provides a method of binding a target nucleic acid (*e.g.*, DNA) comprising the sequence of a preselected target gene or a portion thereof, the method comprising contacting the target DNA with an engineered, non-naturally occurring system disclosed herein, thereby resulting in binding of the system to the target DNA. This method is useful for detecting the presence and/or location of the preselected target gene, for example, if a component of the system (*e.g.*, the Cas protein) comprises a detectable marker.

[0169] In addition, the present invention provides a method of modifying a target nucleic acid (*e.g.*, DNA) comprising the sequence of a preselected target gene or a portion thereof, or a structure (*e.g.*, protein) associated with the target DNA (*e.g.*, a histone protein in a chromosome), the method comprising contacting the target DNA with an engineered, non-naturally occurring system disclosed herein, wherein the Cas protein comprises an effector domain or is associated with an effector protein, thereby resulting in modification of the target DNA or the structure associated with the target DNA. The modification corresponds to

the function of the effector domain or effector protein. Exemplary functions described in the "Cas Proteins" subsection in Section I *supra* are applicable hereto.

[0170] The engineered, non-naturally occurring system can be contacted with the target nucleic acid as a complex. Accordingly, in certain embodiments, the method comprises contacting the target nucleic acid with a CRISPR-Cas complex comprising a targeter nucleic acid, a modulator nucleic acid, and a Cas protein disclosed herein. In certain embodiments, the Cas protein is a type V-A, type V-C, or type V-D Cas protein (*e.g.*, Cas nuclease). In certain embodiments, the Cas protein is a type V-A Cas protein (*e.g.*, Cas nuclease).

[0171] The preselected target genes include human ADORA2A, B2M, CD52, CIITA, CTLA4, DCK, FAS, HAVCR2, LAG3, PDCD1, PTPN6, TIGIT, TRAC, TRBC1, TRBC2, CARD11, CD247, IL7R, LCK, and PLCG1 genes. Accordingly, the present invention also provides a method of editing a human genomic sequence at one of these preselected target gene loci, the method comprising delivering the engineered, non-naturally occurring system disclosed herein into a human cell, thereby resulting in editing of the genomic sequence at the target gene locus in the human cell. In addition, the present invention provides a method of detecting a human genomic sequence at one of these preselected target gene loci, the method comprising delivering the engineered, non-naturally occurring system disclosed herein into a human cell, wherein a component of the system (*e.g.*, the Cas protein) comprises a detectable marker, thereby detecting the target gene locus in the human cell. In addition, the present invention provides a method of modifying a human chromosome at one of these preselected target gene loci, the method comprising delivering the engineered, non-naturally occurring system disclosed herein into a human cell, wherein the Cas protein comprises an effector domain or is associated with an effector protein, thereby resulting in modification of the chromosome at the target gene locus in the human cell.

[0172] The CRISPR-Cas complex may be delivered to a cell by introducing a pre-formed ribonucleoprotein (RNP) complex into the cell. Alternatively, one or more components of the CRISPR-Cas complex may be expressed in the cell. Exemplary methods of delivery are known in the art and described in, for example, U.S. Patent Nos. 10,113,167 and 8,697,359 and U.S. Patent Application Publication Nos. 2015/0344912, 2018/0044700, 2018/0003696, 2018/0119140, 2017/0107539, 2018/0282763, and 2018/0363009.

[0173] It is understood that contacting a DNA (*e.g.*, genomic DNA) in a cell with a CRISPR-Cas complex does not require delivery of all components of the complex into the

cell. For examples, one or more of the components may be pre-existing in the cell. In certain embodiments, the cell (or a parental/ancestral cell thereof) has been engineered to express the Cas protein, and the single guide nucleic acid (or a nucleic acid comprising a regulatory element operably linked to a nucleotide sequence encoding the single guide nucleic acid), the targeter nucleic acid (or a nucleic acid comprising a regulatory element operably linked to a nucleotide sequence encoding the targeter nucleic acid), and/or the modulator nucleic acid (or a nucleic acid comprising a regulatory element operably linked to a nucleotide sequence encoding the modulator nucleic acid) are delivered into the cell. In certain embodiments, the cell (or a parental/ancestral cell thereof) has been engineered to express the modulator nucleic acid, and the Cas protein (or a nucleic acid comprising a regulatory element operably linked to a nucleotide sequence encoding the Cas protein) and the targeter nucleic acid (or a nucleic acid comprising a regulatory element operably linked to a nucleotide sequence encoding the targeter nucleic acid) are delivered into the cell. In certain embodiments, the cell (or a parental/ancestral cell thereof) has been engineered to express the Cas protein and the modulator nucleic acid, and the targeter nucleic acid (or a nucleic acid comprising a regulatory element operably linked to a nucleotide sequence encoding the targeter nucleic acid) is delivered into the cell.

[0174] In certain embodiments, the target DNA is in the genome of a target cell. Accordingly, the present invention also provides a cell comprising the non-naturally occurring system or a CRISPR expression system described herein. In addition, the present invention provides a cell whose genome has been modified by the CRISPR-Cas system or complex disclosed herein.

[0175] The target cells can be mitotic or post-mitotic cells from any organism, such as a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a plant cell, an algal cell, *e.g.*, Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa, Sargassum patens C. Agardh, and the like, a fungal cell (*e.g.*, a yeast cell), an animal cell, a cell from an invertebrate animal (*e.g.* fruit fly, enidarian, echinoderm, nematode, *etc.*), a cell from a vertebrate animal (*e.g.*, fish, amphibian, reptile, bird, mammal), a cell from a mammal, a cell from a rodent, or a cell from a human. The types of target cells include but are not limited to a stem cell (*e.g.*, an embryonic stem (ES) cell, an induced pluripotent stem (iPS) cell, a germ cell), a somatic cell (*e.g.*, a fibroblast, a hematopoietic cell, a T lymphocyte (*e.g.*, $CD8^+$ T lymphocyte), an NK cell, a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell), an in vitro or in vivo embryonic cell of an embryo at any

stage (*e.g.*, a 1-cell, 2-cell, 4-cell, 8-cell; stage zebrafish embryo). Cells may be from established cell lines or may be primary cells (*i.e.*, cells and cells cultures that have been derived from a subject and allowed to grow in vitro for a limited number of passages of the culture). For example, primary cultures are cultures that may have been passaged within 0 times, 1 time, 2 times, 4 times, 5 times, 10 times, or 15 times, but not enough times to go through the crisis stage. Typically, the primary cell lines of the present invention are maintained for fewer than 10 passages in vitro. If the cells are primary cells, they may be harvest from an individual by any suitable method. For example, leukocytes may be harvested by apheresis, leukocytapheresis, or density gradient separation, while cells from tissues such as skin, muscle, bone marrow, spleen, liver, pancreas, lung, intestine, or stomach can be harvested by biopsy. The harvested cells may be used immediately, or may be stored under frozen conditions with a cryopreservative and thawed at a later time in a manner as commonly known in the art.

Ribonucleoprotein (RNP) Delivery and "Cas RNA" Delivery

[0176] The engineered, non-naturally occurring system disclosed herein can be delivered into a cell by suitable methods known in the art, including but not limited to ribonucleoprotein (RNP) delivery and "Cas RNA" delivery described below.

[0177] In certain embodiments, a CRISPR-Cas system including a single guide nucleic acid and a Cas protein, or a CRISPR-Cas system including a targeter nucleic acid, a modulator nucleic acid, and a Cas protein, can be combined into a RNP complex and then delivered into the cell as a pre-formed complex. This method is suitable for active modification of the genetic or epigenetic information in a cell during a limited time period. For example, where the Cas protein has nuclease activity to modify the genomic DNA of the cell, the nuclease activity only needs to be retained for a period of time to allow DNA cleavage, and prolonged nuclease activity may increase off-targeting. Similarly, certain epigenetic modifications can be maintained in a cell once established and can be inherited by daughter cells.

[0178] A "ribonucleoprotein" or "RNP," as used herein, refers to a complex comprising a nucleoprotein and a ribonucleic acid. A "nucleoprotein" as provided herein refers to a protein capable of binding a nucleic acid (*e.g.*, RNA, DNA). Where the nucleoprotein binds a ribonucleic acid it is referred to as "ribonucleoprotein." The interaction between the ribonucleoprotein and the ribonucleic acid may be direct, *e.g.*, by covalent bond, or indirect,

e.g., by non-covalent bond (*e.g.* electrostatic interactions (*e.g.* ionic bond, hydrogen bond, halogen bond), van der Waals interactions (*e.g.* dipole-dipole, dipole-induced dipole, London dispersion), ring stacking (pi effects), hydrophobic interactions, and the like). In certain embodiments, the ribonucleoprotein includes an RNA-binding motif non-covalently bound to the ribonucleic acid. For example, positively charged aromatic amino acid residues (*e.g.*, lysine residues) in the RNA-binding motif may form electrostatic interactions with the negative nucleic acid phosphate backbones of the RNA.

[0179] To ensure efficient loading of the Cas protein, the single guide nucleic acid, or the combination of the targeter nucleic acid and the modulator nucleic acid, can be provided in excess molar amount (*e.g.*, about 2 fold, about 3 fold, about 4 fold, or about 5 fold) relative to the Cas protein. In certain embodiments, the targeter nucleic acid and the modulator nucleic acid are annealed under suitable conditions prior to complexing with the Cas protein. In other embodiments, the targeter nucleic acid, the modulator nucleic acid, and the Cas protein are directly mixed together to form an RNP.

[0180] A variety of delivery methods can be used to introduce an RNP disclosed herein into a cell. Exemplary delivery methods or vehicles include but are not limited to microinjection, liposomes (see, *e.g.*, U.S. Patent Publication No. 2017/0107539) such as molecular trojan horses liposomes that delivers molecules across the blood brain barrier (see, Pardridge *et al.* (2010) COLD SPRING HARB. PROTOC., doi:10.1101/pdb.prot5407), immunoliposomes, virosomes, microvesicles (*e.g.*, exosomes and ARMMs), polycations, lipid:nucleic acid conjugates, electroporation, cell permeable peptides (see, U.S. Patent Publication No. 2018/0363009), nanoparticles, nanowires (see, Shalek *et al.* (2012) NANO LETTERS, 12: 6498), exosomes, and perturbation of cell membrane (*e.g.*, by passing cells through a constriction in a microfluidic system, see, U.S. Patent Publication No. 2018/0003696). Where the target cell is a proliferating cell, the efficiency of RNP delivery can be enhanced by cell cycle synchronization (see, U.S. Patent Publication No. 2018/0044700).

[0181] In other embodiments, the dual guide CRISPR-Cas system is delivered into a cell in a "Cas RNA" approach, *i.e.*, delivering (a) a single guide nucleic acid, or a combination of a targeter nucleic acid and a modulator nucleic acid, and (b) an RNA (*e.g.*, messenger RNA (mRNA)) encoding a Cas protein. The RNA encoding the Cas protein can be translated in the cell and form a complex with the single guide nucleic acid or combination of the targeter nucleic acid and the modulator nucleic acid intracellularly. Similar to the RNP approach,

RNAs have limited half-lives in cells, even though stability-increasing modification(s) can be made in one or more of the RNAs. Accordingly, the "Cas RNA" approach is suitable for active modification of the genetic or epigenetic information in a cell during a limited time period, such as DNA cleavage, and has the advantage of reducing off-targeting.

[0182] The mRNA can be produced by transcription of a DNA comprising a regulatory element operably linked to a Cas coding sequence. Given that multiple copies of Cas protein can be generated from one mRNA, the targeter nucleic acid and the modulator nucleic acid are generally provided in excess molar amount (*e.g.*, at least 5 fold, at least 10 fold, at least 20 fold, at least 30 fold, at least 50 fold, or at least 100 fold) relative to the mRNA. In certain embodiments, the targeter nucleic acid and the modulator nucleic acid are annealed under suitable conditions prior to delivery into the cells. In other embodiments, the targeter nucleic acid and the modulator nucleic acid are delivered into the cells without annealing *in vitro*.

[0183] A variety of delivery systems can be used to introduce an "Cas RNA" system into a cell. Non-limiting examples of delivery methods or vehicles include microinjection, biolistic particles, liposomes (see, *e.g.*, U.S. Patent Publication No. 2017/0107539) such as molecular trojan horses liposomes that delivers molecules across the blood brain barrier (see, Pardridge *et al.* (2010) COLD SPRING HARB. PROTOC., doi:10.1101/pdb.prot5407), immunoliposomes, virosomes, polycations, lipid:nucleic acid conjugates, electroporation, nanoparticles, nanowires (see, Shalek *et al.* (2012) NANO LETTERS, 12: 6498), exosomes, and perturbation of cell membrane (*e.g.*, by passing cells through a constriction in a microfluidic system, see, U.S. Patent Publication No. 2018/0003696). Specific examples of the "nucleic acid only" approach by electroporation are described in International (PCT) Publication No. WO2016/164356.

[0184] In other embodiments, the CRISPR-Cas system is delivered into a cell in the form of (a) a single guide nucleic acid or a combination of a targeter nucleic acid and a modulator nucleic acid, and (b) a DNA comprising a regulatory element operably linked to a Cas coding sequence. The DNA can be provided in a plasmid, viral vector, or any other form described in the "CRISPR Expression Systems" subsection. Such delivery method may result in constitutive expression of Cas protein in the target cell (*e.g.*, if the DNA is maintained in the cell in an episomal vector or is integrated into the genome), and may increase the risk of off-targeting which is undesirable when the Cas protein has nuclease activity. Notwithstanding, this approach is useful when the Cas protein comprises a non-nuclease effector (*e.g.*, a

transcriptional activator or repressor). It is also useful for research purposes and for genome editing of plants.

CRISPR Expression Systems

[0185] The present invention also provides a nucleic acid comprising a regulatory element operably linked to a nucleotide sequence encoding a guide nucleic acid disclosed herein. In certain embodiments, the nucleic acid comprises a regulatory element operably linked to a nucleotide sequence encoding a single guide nucleic acid disclosed herein; this nucleic acid alone can constitute a CRISPR expression system. In certain embodiments, the nucleic acid comprises a regulatory element operably linked to a nucleotide sequence encoding a targeter nucleic acid disclosed herein. In certain embodiments, the nucleic acid further comprises a nucleotide sequence encoding a modulator nucleic acid disclosed herein, wherein the nucleotide sequence encoding the modulator nucleic acid is operably linked to the same regulatory element as the nucleotide sequence encoding the targeter nucleic acid or a different regulatory element; this nucleic acid alone can constitute a CRISPR expression system.

[0186] In addition, the present invention provides a CRISPR expression system comprising: (a) a nucleic acid comprising a first regulatory element operably linked to a nucleotide sequence encoding a targeter nucleic acid disclosed herein and (b) a nucleic acid comprising a second regulatory element operably linked to a nucleotide sequence encoding a modulator nucleic acid disclosed herein.

[0187] In certain embodiments, the CRISPR expression system disclosed herein further comprises a nucleic acid comprising a third regulatory element operably linked to a nucleotide sequence encoding a Cas protein disclosed herein. In certain embodiments, the Cas protein is a type V-A, type V-C, or type V-D Cas protein (*e.g.*, Cas nuclease). In certain embodiments, the Cas protein is a type V-A Cas protein (*e.g.*, Cas nuclease).

[0188] As used in this context, the term "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

[0189] The nucleic acids of the CRISPR expression system described above may be independently selected from various nucleic acids such as DNA (*e.g.*, modified DNA) and RNA (*e.g.*, modified RNA). In certain embodiments, the nucleic acids comprising a

regulatory element operably linked to one or more nucleotide sequences encoding the guide nucleic acids are in the form of DNA. In certain embodiments, the nucleic acid comprising a third regulatory element operably linked to a nucleotide sequence encoding the Cas protein is in the form of DNA. The third regulatory element can be a constitutive or inducible promoter that drives the expression of the Cas protein. In other embodiments, the nucleic acid comprising a third regulatory element operably linked to a nucleotide sequence encoding the Cas protein is in the form of RNA (*e.g.*, mRNA).

[0190] The nucleic acids of the CRISPR expression system can be provided in one or more vectors. The term "vector," as used herein, refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in cells, such as prokaryotic cells, eukaryotic cells, mammalian cells, or target tissues. 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. Gene therapy procedures are known in the art and disclosed in Van Brunt (1988) BIOTECHNOLOGY, 6: 1149; Anderson (1992) SCIENCE, 256: 808; Nabel & Feigner (1993) TIBTECH, 11: 211; Mitani & Caskey (1993) TIBTECH, 11: 162; Dillon (1993) TIBTECH, 11: 167; Miller (1992) NATURE, 357: 455; Vigne,(1995) RESTORATIVE NEUROLOGY AND NEUROSCIENCE, 8: 35; Kremer & Perricaudet (1995) BRITISH MEDICAL BULLETIN, 51: 31; Haddada *et al.* (1995) CURRENT TOPICS IN MICROBIOLOGY AND IMMUNOLOGY, 199: 297; Yu *et al.* (1994) GENE THERAPY, 1: 13; and Doerfler and Bohm (Eds.) (2012) The Molecular Repertoire of Adenoviruses II: Molecular Biology of Virus-Cell Interactions. In certain embodiments, at least one of the vectors is a DNA plasmid. In certain embodiments, at least one of the vectors is a viral vector (*e.g.*, retrovirus, adenovirus, or adeno-associated virus).

[0191] Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors and replication defective viral vectors) do not autonomously replicate in the host cell. Certain vectors, however, may be integrated into the genome of the host cell and thereby are replicated along with the host genome. A skilled person in the art will appreciate that

different vectors may be suitable for different delivery methods and have different host tropism, and will be able to select one or more vectors suitable for the use.

[0192] The term "regulatory element," as used herein, refers to a transcriptional and/or translational control sequence, such as a promoter, enhancer, transcription termination signal (*e.g.*, polyadenylation signal), internal ribosomal entry sites (IRES), protein degradation signal, and the like, that provide for and/or regulate transcription of a non-coding sequence (*e.g.*, a targeter nucleic acid or a modulator nucleic acid) or a coding sequence (*e.g.*, a Cas protein) and/or regulate translation of an encoded polypeptide. Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY, 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (*e.g.*, liver, pancreas), or particular cell types (*e.g.*, lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In certain embodiments, a vector comprises one or more pol III promoter (*e.g.*, 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (*e.g.*, 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (*e.g.*, 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer), the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter. Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I (see, Takebe *et al.* (1988) MOL. CELL. BIOL., 8: 466); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β-globin (see, O'Hare *et al.* (1981) PROC. NATL. ACAD. SCI. USA., 78: 1527). It will be appreciated by those skilled in the art that the design of the expression vector can depend on factors such as the choice of the host cell to be transformed, the level of expression desired, *etc.* A vector can be introduced into host cells to produce transcripts, proteins, or peptides, including fusion proteins or

peptides, encoded by nucleic acids as described herein (*e.g.*, CRISPR transcripts, proteins, enzymes, mutant forms thereof, or fusion proteins thereof).

[0193] In certain embodiments, the nucleotide sequence encoding the Cas protein is codon optimized for expression in a eukaryotic host cell, *e.g.*, a yeast cell, a mammalian cell (*e.g.*, a mouse cell, a rat cell, or a human cell), or a plant cell. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, 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" available at kazusa.or.jp/codon/ and these tables can be adapted in a number of ways (see, Nakamura *et al.* (2000) NUCL. ACIDS RES., 28: 292). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell, such as Gene Forge (Aptagen; Jacobus, Pa.), are also available. In certain embodiments, the codon optimization facilitates or improves expression of the Cas protein in the host cell.

Donor Templates

[0194] Cleavage of a target nucleotide sequence in the genome of a cell by the CRISPR-Cas system or complex disclosed herein can activate the DNA damage pathways, which may rejoin the cleaved DNA fragments by NHEJ or HDR. HDR requires a repair template, either endogenous or exogenous, to transfer the sequence information from the repair template to the target.

[0195] In certain embodiments, the engineered, non-naturally occurring system or CRISPR expression system further comprises a donor template. As used herein, the term "donor template" refers to a nucleic acid designed to serve as a repair template at or near the target nucleotide sequence upon introduction into a cell or organism. In certain embodiments, the donor template is complementary to a polynucleotide comprising the target nucleotide sequence or a portion thereof. When optimally aligned, a donor template may overlap with one or more nucleotides of a target nucleotide sequences (*e.g.* about or more than about 1, 5, 10, 15, 20, 25, 30, 35, 40, or more nucleotides). The nucleotide sequence of

the donor template is typically not identical to the genomic sequence that it replaces. Rather, the donor template may contain one or more substitutions, insertions, deletions, inversions or rearrangements with respect to the genomic sequence, so long as sufficient homology is present to support homology-directed repair. In certain embodiments, the donor template comprises a non-homologous sequence flanked by two regions of homology (*i.e.*, homology arms), such that homology-directed repair between the target DNA region and the two flanking sequences results in insertion of the non-homologous sequence at the target region. In certain embodiments, the donor template comprises a non-homologous sequence 10-100 nucleotides, 50-500 nucleotides, 100-1,000 nucleotides, 200-2,000 nucleotides, or 500-5,000 nucleotides in length positioned between two homology arms.

[0196] Generally, the homologous region(s) of a donor template has at least 50% sequence identity to a genomic sequence with which recombination is desired. The homology arms are designed or selected such that they are capable of recombining with the nucleotide sequences flanking the target nucleotide sequence under intracellular conditions. In certain embodiments, where HDR of the non-target strand is desired, the donor template comprises a first homology arm homologous to a sequence 5' to the target nucleotide sequence and a second homology arm homologous to a sequence 3' to the target nucleotide sequence. In certain embodiments, the first homology arm is at least 50% (*e.g.*, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to a sequence 5' to the target nucleotide sequence. In certain embodiments, the second homology arm is at least 50% (*e.g.*, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to a sequence 3' to the target nucleotide sequence. In certain embodiments, when the donor template sequence and a polynucleotide comprising a target nucleotide sequence are optimally aligned, the nearest nucleotide of the donor template is within about 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000, or more nucleotides from the target nucleotide sequence.

[0197] In certain embodiments, the donor template futher comprises an engineered sequence not homologous to the sequence to be repaired. Such engineered sequence can harbor a barcode and/or a sequence capable of hybridizing with a donor template-recruiting sequence disclosed herein.

[0198] In certain embodiments, the donor template further comprises one or more mutations relative to the genomic sequence, wherein the one or more mutations reduce or prevent cleavage, by the same CRISPR-Cas system, of the donor template or of a modified genomic sequence with at least a portion of the donor template sequence incorporated. In certain embodiments, in the donor template, the PAM adjacent to the target nucleotide sequence and recognized by the Cas nuclease is mutated to a sequence not recognized by the same Cas nuclease. In certain embodiments, in the donor template, the target nucleotide sequence (*e.g.*, the seed region) is mutated. In certain embodiments, the one or more mutations are silent with respect to the reading frame of a protein-coding sequence encompassing the mutated sites.

[0199] The donor template can be provided to the cell as single-stranded DNA, single-stranded RNA, double-stranded DNA, or double-stranded RNA. It is understood that the CRISPR-Cas system disclosed herein may possess nuclease activity to cleave the target strand, the non-target strand, or both. When HDR of the target strand is desired, a donor template having a nucleic acid sequence complementary to the target strand is also contemplated.

[0200] The donor template can be introduced into a cell in linear or circular form. If introduced in linear form, the ends of the donor template may be protected (*e.g.*, from exonucleolytic degradation) by methods known to those of skill in the art. For example, one or more dideoxynucleotide residues are added to the 3′ terminus of a linear molecule and/or self-complementary oligonucleotides are ligated to one or both ends (see, for example, Chang *et al.* (1987) PROC. NATL. ACAD SCI USA, 84: 4959; Nehls *et al.* (1996) SCIENCE, 272: 886; see also the chemical modifications for increasing stability and/or specificity of RNA disclosed *supra*). Additional methods for protecting exogenous polynucleotides from degradation include, but are not limited to, addition of terminal amino group(s) and the use of modified internucleotide linkages such as, for example, phosphorothioates, phosphoramidates, and O-methyl ribose or deoxyribose residues. As an alternative to protecting the termini of a linear donor template, additional lengths of sequence may be included outside of the regions of homology that can be degraded without impacting recombination.

[0201] A donor 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 certain embodiments, the donor template is a

DNA. In certain embodiments, a donor template is in the same nucleic acid as a sequence encoding the single guide nucleic acid, a sequence encoding the targeter nucleic acid, a sequence encoding the modulator nucleic acid, and/or a sequence encoding the Cas protein, where applicable. In certain embodiments, a donor template is provided in a separate nucleic acid. A donor template polynucleotide may be of any suitable length, such as about or at least about 50, 75, 100, 150, 200, 500, 1000, 2000, 3000, 4000, or more nucleotides in length.

[0202] A donor template can be introduced into a cell as an isolated nucleic acid. Alternatively, a donor template can be introduced into a cell as part of a vector (*e.g.*, a plasmid) having additional sequences such as, for example, replication origins, promoters and genes encoding antibiotic resistance, that are not intended for insertion into the DNA region of interest. Alternatively, a donor template can be delivered by viruses (*e.g.*, adenovirus, adeno-associated virus (AAV)). In certain embodiments, the donor template is introduced as an AAV, *e.g.*, a pseudotyped AAV. The capsid proteins of the AAV can be selected by a person skilled in the art based upon the tropism of the AAV and the target cell type. For example, in certain embodiments, the donor template is introduced into a hepatocyte as AAV8 or AAV9. In certain embodiments, the donor template is introduced into a hematopoietic stem cell, a hematopoietic progenitor cell, or a T lymphocyte (*e.g.*, $CD8^+$ T lymphocyte) as AAV6 or an AAVHSC (see, U.S. Patent No. 9,890,396). It is understood that the sequence of a capsid protein (VP1, VP2, or VP3) may be modified from a wild-type AAV capsid protein, for example, having at least 50% (*e.g.*, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%) sequence identity to a wild-type AAV capsid sequence.

[0203] The donor template can be delivered to a cell (*e.g.*, a primary cell) by various delivery methods, such as a viral or non-viral method disclosed herein. In certain embodiments, a non-viral donor template is introduced into the target cell as a naked nucleic acid or in complex with a liposome or poloxamer. In certain embodiments, a non-viral donor template is introduced into the target cell by electroporation. In other embodiments, a viral donor template is introduced into the target cell by infection. The engineered, non-naturally occurring system can be delivered before, after, or simultaneously with the donor template (see, International (PCT) Application Publication No. WO2017/053729). A skilled person in the art will be able to choose proper timing based upon the form of delivery (consider, for example, the time needed for transcription and translation of RNA and protein components)

and the half-life of the molecule(s) in the cell. In particular embodiments, where the CRISPR-Cas system including the Cas protein is delivered by electroporation (*e.g.*, as an RNP), the donor template (*e.g.*, as an AAV) is introduced into the cell within 4 hours (*e.g.*, within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 90, 120, 150, 180, 210, or 240 minutes) after the introduction of the engineered, non-naturally occurring system.

[0204] In certain embodiments, the donor template is conjugated covalently to the modulator nucleic acid. Covalent linkages suitable for this conjugation are known in the art and are described, for example, in U.S. Patent No. 9,982,278 and Savic *et al.* (2018) eLife 7:e33761. In certain embodiments, the donor template is covalently linked to the modulator nucleic acid (*e.g.*, the 5' end of the modulator nucleic acid) through an internucleotide bond. In certain embodiments, the donor template is covalently linked to the modulator nucleic acid (*e.g.*, the 5' end of the modulator nucleic acid) through a linker.

Efficiency and Specificity

[0205] The engineered, non-naturally occurring system of the present invention has the advantage of high efficiency and/or high specificity in nucleic acid targeting, cleavage, or modification.

[0206] In certain embodiments, the engineered, non-naturally occurring system has high efficiency. For example, in certain embodiments, at least 1%, at least 1.5%, at least 2%, at least 2.5%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of a population of nucleic acids having the target nucleotide sequence and a cognate PAM, when contacted with the engineered, non-naturally occurring system, is targeted, cleaved, or modified. In certain embodiments, the genomes of at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of a population of cells, when the engineered, non-naturally occurring system is delivered into the cells, are targeted, cleaved, or modified.

[0207] In certain embodiments, where the engineered, non-naturally occurring system comprises a guide nucleic acid comprising a spacer sequence listed in Table 2 or a portion thereof, the genomes of at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of a population of human cells are targeted, cleaved, edited, or modified when the engineered, non-naturally occurring system is delivered into the cells. In certain embodiments, where the engineered, non-naturally occurring system comprises a guide nucleic acid comprising a spacer sequence listed in Table 2 or a portion thereof, the genomes of at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of a population of human cells are edited when the engineered, non-naturally occurring system is delivered into the cells.

[0208] In certain embodiments, where the engineered, non-naturally occurring system comprises a guide nucleic acid comprising a spacer sequence listed in Table 3 or a portion thereof, the genomes of at least 1%, at least 1.5%, at least 2%, at least 2.5%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of a population of human cells are targeted, cleaved, edited, or modified when the engineered, non-naturally occurring system is delivered into the cells. In certain embodiments, where the engineered, non-naturally occurring system comprises a guide nucleic acid comprising a spacer sequence listed in Table 3 or a portion thereof, the genomes of at least 1%, at least 1.5%, at least 2%, at least 2.5%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of a population of human cells are edited when the engineered, non-naturally occurring system is delivered into the cells.

[0209] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 51 is delivered into a population of human cells *ex vivo*, the genome sequence at the ADORA2A

gene locus is edited in at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0210] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 52 is delivered into a population of human cells *ex vivo*, the genome sequence at the B2M gene locus is edited in at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0211] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 53 is delivered into a population of human cells *ex vivo*, the genome sequence at the CD52 gene locus is edited in at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0212] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 54 is delivered into a population of human cells *ex vivo*, the genome sequence at the CIITA gene locus is edited in at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0213] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 55, 67, 68, or 69 is delivered into a population of human cells *ex vivo*, the genome sequence at the CTLA4 gene locus is edited in at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0214] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 56, 71, or 72 is delivered into a population of human cells *ex vivo*, the genome sequence at the DCK gene locus is edited in at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0215] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 57, 75, 76, 77, or 78 is delivered into a population of human cells *ex vivo*, the genome sequence at the FAS gene locus is edited in at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0216] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 58, 80, or 81 is delivered into a population of human cells *ex vivo*, the genome sequence at the HAVCR2 gene locus is edited in at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0217] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 59 is delivered into a population of human cells *ex vivo*, the genome sequence at the LAG3 gene locus is edited in at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0218] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 60, 89, 90, 91, or 92 is delivered into a population of human cells *ex vivo*, the genome sequence at the PDCD1 gene locus is edited in at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at

least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0219] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 61, 93, 94, 95, 96, 97, 98, or 99 is delivered into a population of human cells *ex vivo*, the genome sequence at the PTPN6 gene locus is edited in at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0220] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 62 or 105 is delivered into a population of human cells *ex vivo*, the genome sequence at the TIGIT gene locus is edited in at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0221] In certain embodiments, when an engineered, non-naturally occurring system comprising a guide nucleic acid comprising a spacer sequence set forth in SEQ ID NO: 63, 106, 107, 108, 109, 110, 111, 112, 113, 114, or 115 is delivered into a population of human cells *ex vivo*, the genome sequence at the TRAC gene locus is edited in at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the cells.

[0222] It has been observed that for a given spacer sequence, the occurrence of on-target events and the occurrence of off-target events are generally correlated. For certain therapeutic purposes, lower on-target efficiency can be tolerated and low off-target frequency is more desirable. For example, when editing or modifying a proliferating cell that will be delivered to a subject and proliferate *in vivo*, tolerance to off-target events is low. Prior to delivery, it is possible to assess the on-target and off-target events, thereby selecting one or more colonies that have the desired edit or modification and lack any undesired edit or modification. Notwithstanding, the on-target efficiency needs to meet a certain standard to be suitable for therapeutic use. The high editing efficiency observed with the spacer sequences disclosed herein in a standard CRISPR-Cas system allows tuning of the system, for

example, by reducing the binding of the guide nucleic acids to the Cas protein, without losing therapeutic applicability.

[0223] In certain embodiments, when a population of nucleic acids having the target nucleotide sequence and a cognate PAM is contacted with the engineered, non-naturally occurring system disclosed herein, the frequency of off-target events (*e.g.*, targeting, cleavage, or modification, depending on the function of the CRISPR-Cas system) is reduced. Methods of assessing off-target events were summarized in Lazzarotto *et al.* (2018) NAT PROTOC. 13(11): 2615-42, and include discovery of in *situ* Cas off-targets and verification by sequencing (DISCOVER-seq) as disclosed in Wienert *et al.* (2019) SCIENCE 364(6437): 286-89; genome-wide unbiased identification of double-stranded breaks (DSBs) enabled by sequencing (GUIDE-seq) as disclosed in Kleinstiver *et al.* (2016) NAT. BIOTECH. 34: 869-74; circularization for *in vitro* reporting of cleavage effects by sequencing (CIRCLE-seq) as described in Kocak *et al.* (2019) NAT. BIOTECH. 37: 657-66. In certain embodiments, the off-target events include targeting, cleavage, or modification at a given off-target locus (*e.g.*, the locus with the highest occurrence of off-target events detected). In certain embodiments, the off-target events include targeting, cleavage, or modification at all the loci with detectable off-target events, collectively.

[0224] In certain embodiments, genomic mutations are detected in no more than 0.0001%, 0.0002%, 0.0003%, 0.0004%, 0.0005%, 0.0006%, 0.0007%, 0.0008%, 0.0009%, 0.001%, 0.002%, 0.003%, 0.004%, 0.005%, 0.006%, 0.007%, 0.008%, 0.009%, 0.01%, 0.02%, 0.03%, 0.04%, 0.05%, 0.06%, 0.07%, 0.08%, 0.09%, 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 2%, 3%, 4%, or 5% of the cells at any off-target loci (in aggregate). In certain embodiments, the ratio of the percentage of cells having an on-target event to the percentage of cells having any off-target event (*e.g.*, the ratio of the percentage of cells having an on-target editing event to the percentage of cells having a mutation at any off-target loci) is at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000. It is understood that genetic variation may be present in a population of cells, for example, by spontaneous mutations, and such mutations are not included as off-target events.

Multiplex Methods

[0225] The method of targeting, editing, and/or modifying a genomic DNA disclosed herein can be conducted in multiplicity. For example, a library of targeter nucleic acids can

be used to target multiple genomic loci; a library of donor templates can also be used to generate multiple insertions, deletions, and/or substitutions. The multiplex assay can be conducted in a screening method wherein each separate cell culture (*e.g.*, in a well of a 96-well plate or a 384-well plate) is exposed to a different guide nucleic acid having a different targeter stem sequence and/or a different donor template. The multiplex assay can also be conducted in a selection method wherein a cell culture is exposed to a mixed population of different guide nucleic acids and/or donor templates, and the cells with desired characteristics (*e.g.*, functionality) are enriched or selected by advantageous survival or growth, resistance to a certain agent, expression of a detectable protein (*e.g.*, a fluorescent protein that is detectable by flow cytometry), *etc.*

[0226] In certain embodiments, the plurality of guide nucleic acids and/or the plurality of donor templates are designed for saturation editing. For example, in certain embodiments, each nucleotide position in a sequence of interest is systematically modified with each of all four traditional bases, A, T, G and C. In other embodiments, at least one sequence in each gene from a pool of genes of interest is modified, for example, according to a CRISPR design algorithm. In certain embodiments, each sequence from a pool of exogenous elements of interest (*e.g.*, protein coding sequences, non-protein coding genes, regulatory elements) is inserted into one or more given loci of the genome.

[0227] It is understood that the multiplex methods suitable for the purpose of carrying out a screening or selection method, which is typically conducted for research purposes, may be different from the methods suitable for therapeutic purposes. For example, constitutive expression of certain elements (*e.g.*, a Cas nuclease and/or a guide nucleic acid) may be undesirable for therapeutic purposes due to the potential of increased off-targeting. Conversely, for research purposes, constitutive expression of a Cas nuclease and/or a guide nucleic acid may be desirable. For example, the constitutive expression provides a large window during which other elements can be introduced. When a stable cell line is established for the constitutive expression, the number of exogenous elements that need to be co-delivered into a single cell is also reduced. Therefore, constitutive expression of certain elements can increase the efficiency and reduce the complexity of a screening or selection process. Inducible expression of certain elements of the system disclosed herein may also be used for research purposes given similar advantages. Expression may be induced by an exogenous agent (*e.g.*, a small molecule) or by an endogenous molecule or complex present in a particular cell type (*e.g.*, at a particular stage of differentiation). Methods known in the

art, such as those described in the "CRISPR Expression Systems" subsection *supra*, can be used for constitutively or inducibly expressing one or more elements.

[0228] It is further understood that despite the need to introduce multiple elements—the single guide nucleic acid and the Cas protein; or the targeter nucleic acid, the modulator nucleic acid, and the Cas protein—these elements can be delivered into the cell as a single complex of pre-formed RNP. Therefore, the efficiency of the screening or selection process can also be achieved by pre-assembling a plurality of RNP complexes in a multiplex manner.

[0229] In certain embodiments, the method disclosed herein further comprises a step of identifying a guide nucleic acid, a Cas protein, a donor template, or a combination of two or more of these elements from the screening or selection process. A set of barcodes may be used, for example, in the donor template between two homology arms, to facilitate the identification. In specific embodiments, the method further comprises harvesting the population of cells; selectively amplifying a genomic DNA or RNA sample including the target nucleotide sequence(s) and/or the barcodes; and/or sequencing the genomic DNA or RNA sample and/or the barcodes that has been selectively amplified.

[0230] In addition, the present invention provides a library comprising a plurality of guide nucleic acids disclosed herein. In another aspect, the present invention provides a library comprising a plurality of nucleic acids each comprising a regulatory element operably linked to a different guide nucleic acid disclosed herein. These libraries can be used in combination with one or more Cas proteins or Cas-coding nucleic acids disclosed herein, and/or one or more donor templates as disclosed herein for a screening or selection method.

III. Pharmaceutical Compositions

[0231] The present invention provides a composition (*e.g.*, pharmaceutical composition) comprising a guide nucleic acid, an engineered, non-naturally occurring system, or a eukaryotic cell disclosed herein. In certain embodiments, the composition comprises an RNP comprising a guide nucleic acid disclosed herein and a Cas protein (*e.g.*, Cas nuclease). In certain embodiments, the composition comprises a complex of a targeter nucleic acid and a modulator nucleic acid disclosed herein. In certain embodiments, the composition comprises an RNP comprising the targeter nucleic acid, the modulator nucleic acid, and a Cas protein (*e.g.*, Cas nuclease).

[0232] In addition, the present invention provides a method of producing a composition, the method comprising incubating a single guide nucleic acid disclosed herein with a Cas

protein, thereby producing a complex of the single guide nucleic acid and the Cas protein (*e.g.*, an RNP). In certain embodiments, the method further comprises purifying the complex (*e.g.*, the RNP).

[0233] In addition, the present invention provides a method of producing a composition, the method comprising incubating a targeter nucleic acid and a modulator nucleic acid disclosed herein under suitable conditions, thereby producing a composition (*e.g.*, pharmaceutical composition) comprising a complex of the targeter nucleic acid and the modulator nucleic acid. In certain embodiments, the method further comprises incubating the targeter nucleic acid and the modulator nucleic acid with a Cas protein (*e.g.*, the Cas nuclease that the targeter nucleic acid and the modulator nucleic acid are capable of activating or a related Cas protein), thereby producing a complex of the targeter nucleic acid, the modulator nucleic acid, and the Cas protein (*e.g.*, an RNP). In certain embodiments, the method further comprises purifying the complex (*e.g.*, the RNP).

[0234] For therapeutic use, a guide nucleic acid, an engineered, non-naturally occurring system, a CRISPR expression system, or a cell comprising such system or modified by such system disclosed herein is combined with a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable" as used herein refers to those compounds, materials, compositions, and/or dosage forms which are, within the scope of sound medical judgment, suitable for use in contact with the tissues of human beings and animals without excessive toxicity, irritation, allergic response, or other problem or complication, commensurate with a reasonable benefit-to-risk ratio.

[0235] The term "pharmaceutically acceptable carrier" as used herein refers to buffers, carriers, and excipients suitable for use in contact with the tissues of human beings and animals without excessive toxicity, irritation, allergic response, or other problem or complication, commensurate with a reasonable benefit/risk ratio. Pharmaceutically acceptable carriers include any of the standard pharmaceutical carriers, such as a phosphate buffered saline solution, water, emulsions (*e.g.*, such as an oil/water or water/oil emulsions), and various types of wetting agents. The compositions also can include stabilizers and preservatives. For examples of carriers, stabilizers and adjuvants, see, *e.g.*, Martin, Remington's Pharmaceutical Sciences, 15th Ed., Mack Publ. Co., Easton, PA (1975). Pharmaceutically acceptable carriers include buffers, solvents, dispersion media, coatings, isotonic and absorption delaying agents, and the like, that are compatible with pharmaceutical

administration. The use of such media and agents for pharmaceutically active substances is known in the art.

[0236] In certain embodiments, a pharmaceutical composition disclosed herein comprises a salt, e.g., NaCl, $MgCl_2$, KCl, $MgSO_4$, *etc.*; a buffering agent, e.g., a Tris buffer, N-(2-Hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid) (HEPES), 2-(N-Morpholino)ethanesulfonic acid (MES), MES sodium salt, 3-(N-Morpholino)propanesulfonic acid (MOPS), N-tris[Hydroxymethyl]methyl-3-aminopropanesulfonic acid (TAPS), *etc.*; a solubilizing agent; a detergent, *e.g.*, a non-ionic detergent such as Tween-20, *etc.*; a nuclease inhibitor; and the like. For example, in certain embodiments, a subject composition comprises a subject DNA-targeting RNA and a buffer for stabilizing nucleic acids.

[0237] In certain embodiments, a pharmaceutical composition may contain formulation materials for modifying, maintaining or preserving, for example, the pH, osmolarity, viscosity, clarity, color, isotonicity, odor, sterility, stability, rate of dissolution or release, adsorption or penetration of the composition. In such embodiments, suitable formulation materials include, but are not limited to, amino acids (such as glycine, glutamine, asparagine, arginine or lysine); antimicrobials; antioxidants (such as ascorbic acid, sodium sulfite or sodium hydrogen-sulfite); buffers (such as borate, bicarbonate, Tris-HCl, citrates, phosphates or other organic acids); bulking agents (such as mannitol or glycine); chelating agents (such as ethylenediamine tetraacetic acid (EDTA)); complexing agents (such as caffeine, polyvinylpyrrolidone, beta-cyclodextrin or hydroxypropyl-beta-cyclodextrin); fillers; monosaccharides; disaccharides; and other carbohydrates (such as glucose, mannose or dextrins); proteins (such as serum albumin, gelatin or immunoglobulins); coloring, flavoring and diluting agents; emulsifying agents; hydrophilic polymers (such as polyvinylpyrrolidone); low molecular weight polypeptides; salt-forming counterions (such as sodium); preservatives (such as benzalkonium chloride, benzoic acid, salicylic acid, thimerosal, phenethyl alcohol, methylparaben, propylparaben, chlorhexidine, sorbic acid or hydrogen peroxide); solvents (such as glycerin, propylene glycol or polyethylene glycol); sugar alcohols (such as mannitol or sorbitol); suspending agents; surfactants or wetting agents (such as pluronics, PEG, sorbitan esters, polysorbates such as polysorbate 20, polysorbate, triton, tromethamine, lecithin, cholesterol, tyloxapal); stability enhancing agents (such as sucrose or sorbitol); tonicity enhancing agents (such as alkali metal halides, preferably sodium or potassium chloride, mannitol sorbitol); delivery vehicles; diluents; excipients

and/or pharmaceutical adjuvants (see, *Remington's Pharmaceutical Sciences*, 18th ed. (Mack Publishing Company, 1990).

[0238] In certain embodiments, a pharmaceutical composition may contain nanoparticles, *e.g.*, polymeric nanoparticles, liposomes, or micelles (See Anselmo *et al.* (2016) BIOENG. TRANSL. MED. 1: 10-29). In certain embodiment, the pharmaceutical composition comprises an inorganic nanoparticle. Exemplary inorganic nanoparticles include, *e.g.*, magnetic nanoparticles (*e.g.*, Fe_3MnO_2) or silica. The outer surface of the nanoparticle can be conjugated with a positively charged polymer (*e.g.*, polyethylenimine, polylysine, polyserine) which allows for attachment (*e.g.*, conjugation or entrapment) of payload. In certain embodiment, the pharmaceutical composition comprises an organic nanoparticle (*e.g.*, entrapment of the payload inside the nanoparticle). Exemplary organic nanoparticles include, *e.g.*, SNALP liposomes that contain cationic lipids together with neutral helper lipids which are coated with polyethylene glycol (PEG) and protamine and nucleic acid complex coated with lipid coating. In certain embodiment, the pharmaceutical composition comprises a liposome, for example, a liposome disclosed in International Application Publication No. WO 2015/148863.

[0239] In certain embodiments, the pharmaceutical composition comprises a targeting moiety to increase target cell binding or update of nanoparticles and liposomes. Exemplary targeting moieties include cell specific antigens, monoclonal antibodies, single chain antibodies, aptamers, polymers, sugars, and cell penetrating peptides. In certain embodiments, the pharmaceutical composition comprises a fusogenic or endosome-destabilizing peptide or polymer.

[0240] In certain embodiments, a pharmaceutical composition may contain a sustained- or controlled-delivery formulation. Techniques for formulating sustained- or controlled-delivery means, such as liposome carriers, bio-erodible microparticles or porous beads and depot injections, are also known to those skilled in the art. Sustained-release preparations may include, *e.g.*, porous polymeric microparticles or semipermeable polymer matrices in the form of shaped articles, *e.g.*, films, or microcapsules. Sustained release matrices may include polyesters, hydrogels, polylactides, copolymers of L-glutamic acid and gamma ethyl-L-glutamate, poly (2-hydroxyethyl-inethacrylate), ethylene vinyl acetate, or poly-D(−)-3-hydroxybutyric acid. Sustained release compositions may also include liposomes that can be prepared by any of several methods known in the art.

[0241] A pharmaceutical composition of the invention can be administered by a variety of methods known in the art. The route and/or mode of administration vary depending upon the desired results. Administration can be intravenous, intramuscular, intraperitoneal, or subcutaneous, or administered proximal to the site of the target. The pharmaceutically acceptable carrier should be suitable for intravenous, intramuscular, subcutaneous, parenteral, spinal or epidermal administration (*e.g.*, by injection or infusion). Depending on the route of administration, the active compound (*e.g.*, the guide nucleic acid, engineered, non-naturally occurring system, or CRISPR expression system of the invention) may be coated in a material to protect the compound from the action of acids and other natural conditions that may inactivate the compound.

[0242] Formulation components suitable for parenteral administration include a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerin, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as EDTA; buffers such as acetates, citrates or phosphates; and agents for the adjustment of tonicity such as sodium chloride or dextrose.

[0243] For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). The carrier should be stable under the conditions of manufacture and storage, and should be preserved against microorganisms. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol), and suitable mixtures thereof.

[0244] Pharmaceutical formulations preferably are sterile. Sterilization can be accomplished by any suitable method, *e.g.*, filtration through sterile filtration membranes. Where the composition is lyophilized, filter sterilization can be conducted prior to or following lyophilization and reconstitution. In certain embodiments, the pharmaceutical composition is lyophilized, and then reconstituted in buffered saline, at the time of administration.

[0245] Pharmaceutical compositions of the invention can be prepared in accordance with methods well known and routinely practiced in the art. See, *e.g.*, Remington: The Science and Practice of Pharmacy, Mack Publishing Co., 20th ed., 2000; and Sustained and Controlled Release Drug Delivery Systems, J. R. Robinson, ed., Marcel Dekker, Inc., New York, 1978.

Pharmaceutical compositions are preferably manufactured under GMP conditions. Typically, a therapeutically effective dose or efficacious dose of the guide nucleic acid, engineered, non-naturally occurring system, or CRISPR expression system of the invention is employed in the pharmaceutical compositions of the invention. The multispecific antibodies of the invention are formulated into pharmaceutically acceptable dosage forms by conventional methods known to those of skill in the art. Dosage regimens are adjusted to provide the optimum desired response (*e.g.*, a therapeutic response). For example, a single bolus may be administered, several divided doses may be administered over time or the dose may be proportionally reduced or increased as indicated by the exigencies of the therapeutic situation. It is especially advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subjects to be treated; each unit contains a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[0246] Actual dosage levels of the active ingredients in the pharmaceutical compositions of the invention can be varied so as to obtain an amount of the active ingredient which is effective to achieve the desired therapeutic response for a particular patient, composition, and mode of administration, without being toxic to the patient. The selected dosage level depends upon a variety of pharmacokinetic factors including the activity of the particular compositions of the present invention employed, or the ester, salt or amide thereof, the route of administration, the time of administration, the rate of excretion of the particular compound being employed, the duration of the treatment, other drugs, compounds and/or materials used in combination with the particular compositions employed, the age, sex, weight, condition, general health and prior medical history of the patient being treated, and like factors.

IV. Therapeutic Uses

[0247] The guide nucleic acids, the engineered, non-naturally occurring systems, and the CRISPR expression systems disclosed herein are useful for targeting, editing, and/or modifying the genomic DNA in a cell or organism. These guide nucleic acids and systems, as well as a cell comprising one of the systems or a cell whose genome has been modified by one of the systems, can be used to treat a disease or disorder in which modification of genetic or epigenetic information is desirable. Accordingly, the present invention provides a method of treating a disease or disorder, the method comprising administering to a subject in need

thereof a guide nucleic acid, a non-naturally occurring system, a CRISPR expression system, or a cell disclosed herein.

[0248] The term "subject" includes human and non-human animals. Non-human animals include all vertebrates, *e.g.*, mammals and non-mammals, such as non-human primates, sheep, dog, cow, chickens, amphibians, and reptiles. Except when noted, the terms "patient" or "subject" are used herein interchangeably.

[0249] The terms "treatment", "treating", "treat", "treated", and the like, as used herein, refer to obtaining a desired pharmacologic and/or physiologic effect. The effect may be therapeutic in terms of a partial or complete cure for a disease and/or adverse effect attributable to the disease or delaying the disease progression. "Treatment", as used herein, covers any treatment of a disease in a mammal, *e.g.*, in a human, and includes: (a) inhibiting the disease, *i.e.*, arresting its development; and (b) relieving the disease, *i.e.*, causing regression of the disease. It is understood that a disease or disorder may be identified by genetic methods and treated prior to manifestation of any medical symptom.

[0250] For minimization of toxicity and off-target effect, it is important to control the concentration of the CRISPR-Cas system delivered. Optimal concentrations can be determined by testing different concentrations in a cellular, tissue, or non-human eukaryote animal model and using deep sequencing to analyze the extent of modification at potential off-target genomic loci. The concentration that gives the highest level of on-target modification while minimizing the level of off-target modification should be selected for *ex vivo* or *in vivo* delivery.

[0251] It is understood that the guide nucleic acid, the engineered, non-naturally occurring system, and the CRISPR expression system disclosed herein can be used to treat any disease or disorder that can be improved by editing or modifying human ADORA2A, B2M, CD52, CIITA, CTLA4, DCK, FAS, HAVCR2, LAG3, PDCD1, PTPN6, TIGIT, TRAC, TRBC1, TRBC2, CARD11, CD247, IL7R, LCK, or PLCG1 gene in a cell. In certain embodiments, the guide nucleic acid, the engineered, non-naturally occurring system, and the CRISPR expression system disclosed herein can be used to engineer an immune cell. Immune cells include but are not limited to lymphocytes (*e.g.*, B lymphocytes or B cells, T lymphocytes or T cells, and natural killer cells), myeloid cells (*e.g.*, monocytes, macrophages, eosinophils, mast cells, basophils, and granulocytes), and the stem and progenitor cells that can differentiate into these cell types (*e.g.*, hematopoietic stem cells,

hematopoietic progenitor cells, and lymphoid progenitor cells). The cells can include autologous cells derived from a subject to be treated, or alternatively allogenic cells derived from a donor.

[0252] In certain embodiments, the immune cell is a T cell, which can be, for example, a cultured T cell, a primary T cell, a T cell from a cultured T cell line (*e.g.*, Jurkat, SupTi), or a T cell obtained from a mammal, for example, from a subject to be treated. If obtained from a mammal, the T cell can be obtained from numerous sources, including but not limited to blood, bone marrow, lymph node, the thymus, or other tissues or fluids. T cells can also be enriched or purified. The T cell can be any type of T cell and can be of any developmental stage, including but not limited to, $CD4^+/CD8^+$ double positive T cells, $CD4^+$ helper T cells (*e.g.*, Th1 and Th2 cells), $CD8^+$ T cells (*e.g.*, cytotoxic T cells), tumor infiltrating lymphocytes (TILs), memory T cells (*e.g.*, central memory T cells and effector memory T cells), regulatory T cells, naive T cells, and the like.

[0253] In certain embodiments, an immune cell, *e.g.*, a T cell, is engineered to express an exogenous gene. For example, in certain embodiments, the guide nucleic acid, the engineered, non-naturally occurring system, and the CRISPR expression system disclosed herein may be used to engineer an immune cell to express an exogenous gene at the locus of a human ADORA2A, B2M, CD52, CIITA, CTLA4, DCK, FAS, HAVCR2, LAG3, PDCD1, PTPN6, TIGIT, TRAC, TRBC1, TRBC2, CARD11, CD247, IL7R, LCK, or PLCG1 gene. For example, in certain embodiments, an engineered CRISPR system disclosed herein may catalyze DNA cleavage at the gene locus, allowing for site-specific integration of the exogenous gene at the gene locus by HDR.

[0254] In certain embodiments, an immune cell, *e.g.*, a T cell, is engineered to express a chimeric antigen receptor (CAR), *i.e.*, the T cell comprises an exogenous nucleotide sequence encoding a CAR. As used herein, the term "chimeric antigen receptor" or "CAR" refers to any artificial receptor including an antigen-specific binding moiety and one or more signaling chains derived from an immune receptor. CARs can comprise a single chain fragment variable (scFv) of an antibody specific for an antigen coupled via hinge and transmembrane regions to cytoplasmic domains of T cell signaling molecules, *e.g.* a T cell costimulatory domain (*e.g.*, from CD28, CD137, OX40, ICOS, or CD27) in tandem with a T cell triggering domain (*e.g.* from CD3ζ). A T cell expressing a chimeric antigen receptor is referred to as a CAR T cell. Exemplary CAR T cells include CD19 targeted CTL019 cells (see, Grupp *et al.* (2015) BLOOD, 126: 4983), 19-28z cells (see, Park *et al.* (2015) J. CLIN. ONCOL., 33: 7010),

and KTE-C19 cells (see, Locke *et al.* (2015) BLOOD, 126: 3991). Additional exemplary CAR T cells are described in U.S. Patent Nos. 8,399,645, 8,906,682, 7,446,190, 9,181,527, 9,272,002, and 9,266,960, U.S. Patent Publication Nos. 2016/0362472, 2016/0200824, and 2016/0311917, and International (PCT) Publication Nos. WO2013/142034, WO2015/120180, WO2015/188141, WO2016/120220, and WO2017/040945. Exemplary approaches to express CARs using CRISPR systems are described in Hale *et al.* (2017) MOL THER METHODS CLIN DEV., 4: 192, MacLeod *et al.* (2017) MOL THER, 25: 949, and Eyquem *et al.* (2017) NATURE, 543: 113.

[0255] In certain embodiments, an immune cell, *e.g.*, a T cell, binds an antigen, *e.g.*, a cancer antigen, through an endogenous T cell receptor (TCR). In certain embodiments, an immune cell, *e.g.*, a T cell, is engineered to express an exogenous TCR, *e.g.*, an exogenous naturally occurring TCR or an exogenous engineered TCR. T cell receptors comprise two chains referred to as the α- and β-chains, that combine on the surface of a T cell to form a heterodimeric receptor that can recognize MHC-restricted antigens. Each of α- and β- chain comprises a constant region and a variable region. Each variable region of the α- and β-chains defines three loops, referred to as complementary determining regions (CDRs) known as CDR_1, CDR_2, and CDR_3 that confer the T cell receptor with antigen binding activity and binding specificity.

[0256] In certain embodiments, a CAR or TCR binds a cancer antigen selected from B-cell maturation antigen (BCMA), mesothelin, prostate specific membrane antigen (PSMA), prostate stem cell antigen (PCSA), carbonic anhydrase IX (CAIX), carcinoembryonic antigen (CEA), CD5, CD7, CD10, CD19, CD20, CD22, CD30, CD33, CD34, CD38, CD41, CD44, CD49f, CD56, CD70, CD74, CD123, CD133, CD138, epithelial glycoprotein2 (EGP 2), epithelial glycoprotein-40 (EGP-40), epithelial cell adhesion molecule (EpCAM), receptor-type tyrosine-protein kinase (FLT3), folate-binding protein (FBP), fetal acetylcholine receptor (AChR), folate receptor-a and β (FRa and β), Ganglioside G2 (GD2), Ganglioside G3 (GD3), epidermal growth factor receptor 2 (HER-2/ERB2), epidermal growth factor receptor vIII (EGFRvIII), ERB3, ERB4, human telom erase reverse transcriptase (hTERT), Interleukin-13 receptor subunit alpha-2 (IL- 13Ra2), K-light chain, kinase insert domain receptor (KDR), Lewis A (CA19.9), Lewis Y (LeY), LI cell adhesion molecule (LICAM), melanoma-associated antigen 1 (melanoma antigen family Al, MAGE-A1), Mucin 16 (MUC-16), Mucin 1 (MUC-1; *e.g.*, a truncated MUC-1), KG2D ligands, cancer-testis antigen NY-ESO-1, oncofetal antigen (h5T4), tumor-associated glycoprotein 72 (TAG-72), vascular

endothelial growth factor R2 (VEGF-R2), Wilms tumor protein (WT-1), type 1 tyrosine-protein kinase transmembrane receptor (ROR1), B7-H3 (CD276), B7-H6 (Nkp30), Chondroitin sulfate proteoglycan-4 (CSPG4), DNAX Accessory Molecule (DNAM-1), Ephrin type A Receptor 2 (EpHA2), Fibroblast Associated Protein (FAP), Gpl00/HLA-A2, Glypican 3 (GPC3), HA-IH, HERK-V, IL-1 IRa, Latent Membrane Protein 1 (LMP1), Neural cell-adhesion molecule (N-CAM/CD56), and Trail Receptor (TRAIL-R).

[0257] Genetic loci suitable for insertion of a CAR- or exogenous TCR-encoding sequence include but are not limited to TCR subunit loci (*e.g.*, the TCRα constant (TRAC) locus, the TCRβ constant 1 (TRBC1) locus, and the TCRβ constant 2 (TRBC2) locus). It is understood that insertion in the TRAC locus reduces tonic CAR signaling and enhances T cell potency (see, Eyquem *et al.* (2017) NATURE, 543: 113). Furthermore, inactivation of the endogenous TRAC, TRBC1, or TRBC2 gene may reduce a graft-versus-host disease (GVHD) response, thereby allowing use of allogeneic T cells as starting materials for preparation of CAR-T cells. Accordingly, in certain embodiments, an immune cell, *e.g.*, a T cell, is engineered to have reduced expression of an endogenous TCR or TCR subunit, *e.g.*, TRAC, TRBC1, and/or TRBC2. The cell may be engineered to have partially reduced or no expression of the endogenous TCR or TCR subunit. For example, in certain embodiments, the immune cell, *e.g.*, a T cell, is engineered to have less than 80% (*e.g.*, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, or less than 5%) of the expression of the endogenous TCR or TCR subunit relative to a corresponding unmodified or parental cell. In certain embodiments, the immune cell, *e.g.*, a T cell, is engineered to have no detectable expression of the endogenous TCR or TCR subunit. Exemplary approaches to reduce expression of TCRs using CRISPR systems are described in U.S. Patent No. 9,181,527, Liu *et al.* (2017) CELL RES, 27: 154, Ren *et al.* (2017) CLIN CANCER RES, 23: 2255, Cooper *et al.* (2018) LEUKEMIA, 32: 1970, and Ren *et al.* (2017) ONCOTARGET, 8: 17002.

[0258] It is understood that certain immune cells, such as T cells, also express major histocompatibility complex (MHC) or human leukocyte antigen (HLA) genes, and inactivation of these endogenous gene may reduce a GVHD response, thereby allowing use of allogeneic T cells as starting materials for preparation of CAR-T cells. Accordingly, in certain embodiments, an immune cell, *e.g.*, a T-cell, is engineered to have reduced expression of one or more endogenous class I or class II MHCs or HLAs (*e.g.*, beta 2-microglobulin (B2M), class II major histocompatibility complex transactivator (CIITA), HLA-E, and/or

HLA-G). The cell may be engineered to have partially reduced or no expression of an endogenous MHC or HLA. For example, in certain embodiments, the immune cell, *e.g.*, a T-cell, is engineered to have less than less than 80% (*e.g.*, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, or less than 5%) of the expression of endogenous MHC (*e.g.*, B2M, CIITA, HLA-E, or HLA-G) relative to a corresponding unmodified or parental cell. In certain embodiments, the immune cell, *e.g.*, a T cell, is engineered to have no detectable expression of an endogenous MHC (*e.g.*, B2M, CIITA, HLA-E, or HLA-G). Exemplary approaches to reduce expression of MHCs using CRISPR systems are described in Liu *et al.* (2017) CELL RES, 27: 154, Ren *et al.* (2017) CLIN CANCER RES, 23: 2255, and Ren *et al.* (2017) ONCOTARGET, 8: 17002.

[0259] Other genes that may be inactivated to reduce a GVHD response include but are not limited to CD3, CD52, and deoxycytidine kinase (DCK). For example, inactivation of DCK may render the immune cells (*e.g.*, T cells) resistant to purine nucleotide analogue (PNA) compounds, which are often used to compromise the host immune system in order to reduce a GVHD response during an immune cell therapy. In certain embodiments, the immune cell, *e.g.*, a T-cell, is engineered to have less than less than 80% (*e.g.*, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, or less than 5%) of the expression of endogenous CD52 or DCK relative to a corresponding unmodified or parental cell.

[0260] It is understood that the activity of an immune cell (*e.g.*, T cell) may be enhanced by inactivating or reducing the expression of an immune suppressor such as an immune checkpoint protein. Accordingly, in certain embodiments, an immune cell, *e.g.*, a T cell, is engineered to have reduced expression of an immune checkpoint protein. Exemplary immune checkpoint proteins expressed by wild-type T cells include but are not limited to PDCD1 (PD-1), CTLA4, ADORA2A (A2AR), B7-H3, B7-H4, BTLA, KIR, LAG3, HAVCR2 (TIM3), TIGIT, VISTA, PTPN6 (SHP-1), and FAS. The cell may be modified to have partially reduced or no expression of the immune checkpoint protein. For example, in certain embodiments, the immune cell, *e.g.*, a T cell, is engineered to have less than 80% (*e.g.*, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, or less than 5%) of the expression of the immune checkpoint protein relative to a corresponding unmodified or parental cell. In certain embodiments, the immune cell, *e.g.*, a T cell, is engineered to have no detectable expression of the immune checkpoint protein. Exemplary approaches to reduce expression of immune checkpoint proteins using

CRISPR systems are described in International (PCT) Publication No. WO2017/017184, Cooper *et al.* (2018) LEUKEMIA, 32: 1970, Su *et al.* (2016) ONCOIMMUNOLOGY, 6: e1249558, and Zhang *et al.* (2017) FRONT MED, 11: 554.

[0261] The immune cell can be engineered to have reduced expression of an endogenous gene, *e.g.*, an endogenous genes described above, by gene editing or modification. For example, in certain embodiments, an engineered CRISPR system disclosed herein may result in DNA cleavage at a gene locus, thereby inactivating the targeted gene. In other embodiments, an engineered CRISPR system disclosed herein may be fused to an effector domain (*e.g.*, a transcriptional repressor or histone methylase) to reduce the expression of the target gene.

[0262] The immune cell can also be engineered to express an exogenous protein (besides an antigen-binding protein described above) at the locus of a human ADORA2A, B2M, CD52, CIITA, CTLA4, DCK, FAS, HAVCR2, LAG3, PDCD1, PTPN6, TIGIT, TRAC, TRBC1, TRBC2, CARD11, CD247, IL7R, LCK, or PLCG1 gene.

[0263] In certain embodiments, an immune cell, *e.g.*, a T cell, is modified to express a dominant-negative form of an immune checkpoint protein. In certain embodiments, the dominant-negative form of the checkpoint inhibitor can act as a decoy receptor to bind or otherwise sequester the natural ligand that would otherwise bind and activate the wild-type immune checkpoint protein. Examples of engineered immune cells, for example, T cells containing dominant-negative forms of an immune suppressor are described, for example, in International (PCT) Publication No. WO2017/040945.

[0264] In certain embodiments, an immune cell, *e.g.*, a T cell, is modified to express a gene (*e.g.*, a transcription factor, a cytokine, or an enzyme) that regulates the survival, proliferation, activity, or differentiation (*e.g.*, into a memory cell) of the immune cell. In certain embodiments, the immune cell is modified to express TET2, FOXO1, IL-12, IL-15, IL-18, IL-21, IL-7, GLUT1, GLUT3, HK1, HK2, GAPDH, LDHA, PDK1, PKM2, PFKFB3, PGK1, ENO1, GYS1, and/or ALDOA. In certain embodiments, the modification is an insertion of a nucleotide sequence encoding the protein operably linked to a regulatory element. In certain embodiments, the modification is a substitution of a single nucleotide polymorphism (SNP) site in the endogenous gene. In certain embodiments, an immune cell, *e.g.*, a T cell, is modified to express a variant of a gene, for example, a variant that has greater activity than the respective wild-type gene. In certain embodiments, the immune cell is

modified to express a variant of CARD11, CD247, IL7R, LCK, or PLCG1. For example, certain gain-of-function variants of IL7R were disclosed in Zenatti *et al.*, (2011) NAT. GENET. 43(10):932-39. The variant can be expressed from the native locus of the respective wild-type gene by delivering an engineered system described herein for targeting the native locus in combination with a donor template that carries the variant or a portion thereof.

[0265] In certain embodiments, an immune cell, *e.g.*, a T cell, is modified to express a protein (*e.g.*, a cytokine or an enzyme) that regulates the microenvironment that the immune cell is designed to migrate to (*e.g.*, a tumor microenvironment). In certain embodiments, the immune cell is modified to express CA9, CA12, a V-ATPase subunit, NHE1, and/or MCT-1.

V. Kits

[0266] It is understood that the guide nucleic acid, the engineered, non-naturally occurring system, the CRISPR expression system, and the library disclosed herein can be packaged in a kit suitable for use by a medical provider. Accordingly, in another aspect, the invention provides kits containing any one or more of the elements disclosed in the above systems, libraries, methods, and compositions. In certain embodiments, the kit comprises an engineered, non-naturally occurring system as disclosed herein and instructions for using the kit. The instructions may be specific to the applications and methods described herein. In certain embodiments, one or more of the elements of the system are provided in a solution. In certain embodiments, one or more of the elements of the system are provided in lyophilized form, and the kit further comprises a diluent. Elements may be provided individually or in combinations, and may be provided in any suitable container, such as a vial, a bottle, a tube, or immobilized on the surface of a solid base (*e.g.*, chip or microarray). In certain embodiments, the kit comprises one or more of the nucleic acids and/or proteins described herein. In certain embodiments, the kit provides all elements of the systems of the invention.

[0267] In certain embodiments of a kit comprising the engineered, non-naturally occurring dual guide system, the targeter nucleic acid and the modulator nucleic acid are provided in separate containers. In other embodiments, the targeter nucleic acid and the modulator nucleic acid are pre-complexed, and the complex is provided in a single container.

[0268] In certain embodiments, the kit comprises a Cas protein or a nucleic acid comprising a regulatory element operably linked to a nucleic acid encoding a Cas protein provided in a separate container. In other embodiments, the kit comprises a Cas protein pre-

complexed with the single guide nucleic acid or a combination of the targeter nucleic acid and the modulator nucleic acid, and the complex is provided in a single container.

[0269] In certain embodiments, the kit further comprises one or more donor templates provided in one or more separate containers. In certain embodiments, the kit comprises a plurality of donor templates as disclosed herein (*e.g.*, in separate tubes or immobilized on the surface of a solid base such as a chip or a microarray), one or more guide nucleic acids disclosed herein, and optionally a Cas protein or a regulatory element operably linked to a nucleic acid encoding a Cas protein as disclosed herein. Such kits are useful for identifying a donor template that introduces optimal genetic modification in a multiplex assay. The CRISPR expression systems as disclosed herein are also suitable for use in a kit.

[0270] In certain embodiments, a kit further comprises one or more reagents and/or buffers for use in a process utilizing one or more of the elements described herein. Reagents may be provided in any suitable container and may 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 may be a reaction or storage 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 certain embodiments, the buffer has a pH from about 7 to about 10. In certain embodiments, the kit further comprises a pharmaceutically acceptable carrier. In certain embodiments, the kit further comprises one or more devices or other materials for administration to a subject.

[0271] Throughout the description, where compositions are described as having, including, or comprising specific components, or where processes and methods are described as having, including, or comprising specific steps, it is contemplated that, additionally, there are compositions of the present invention that consist essentially of, or consist of, the recited components, and that there are processes and methods according to the present invention that consist essentially of, or consist of, the recited processing steps.

[0272] In the application, where an element or component is said to be included in and/or selected from a list of recited elements or components, it should be understood that the element or component can be any one of the recited elements or components, or the element or component can be selected from a group consisting of two or more of the recited elements or components.

[0273] Further, it should be understood that elements and/or features of a composition or a method described herein can be combined in a variety of ways without departing from the spirit and scope of the present invention, whether explicit or implicit herein. For example, where reference is made to a particular compound, that compound can be used in various embodiments of compositions of the present invention and/or in methods of the present invention, unless otherwise understood from the context. In other words, within this application, embodiments have been described and depicted in a way that enables a clear and concise application to be written and drawn, but it is intended and will be appreciated that embodiments may be variously combined or separated without parting from the present teachings and invention(s). For example, it will be appreciated that all features described and depicted herein can be applicable to all aspects of the invention(s) described and depicted herein.

[0274] The terms "a" and "an" and "the" and similar references in the context of describing the invention (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. For example, the term "a cell" includes a plurality of cells, including mixtures thereof. Where the plural form is used for compounds, salts, and the like, this is taken to mean also a single compound, salt, or the like.

[0275] It should be understood that the expression "at least one of" includes individually each of the recited objects after the expression and the various combinations of two or more of the recited objects unless otherwise understood from the context and use. The expression "and/or" in connection with three or more recited objects should be understood to have the same meaning unless otherwise understood from the context.

[0276] The use of the term "include," "includes," "including," "have," "has," "having," "contain," "contains," or "containing," including grammatical equivalents thereof, should be understood generally as open-ended and non-limiting, for example, not excluding additional unrecited elements or steps, unless otherwise specifically stated or understood from the context.

[0277] Where the use of the term "about" is before a quantitative value, the present invention also includes the specific quantitative value itself, unless specifically stated otherwise. As used herein, the term "about" refers to a ±10% variation from the nominal value unless otherwise indicated or inferred.

[0278] It should be understood that the order of steps or order for performing certain actions is immaterial so long as the present invention remain operable. Moreover, two or more steps or actions may be conducted simultaneously.

[0279] The use of any and all examples, or exemplary language herein, for example, "such as" or "including," is intended merely to illustrate better the present invention and does not pose a limitation on the scope of the invention unless claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the present invention.

EXAMPLES

[0280] The following Examples are merely illustrative and are not intended to limit the scope or content of the invention in any way.

Example 1. Cleavage of Genomic DNA by Single Guide MAD7 CRISPR-Cas Systems

[0281] MAD7 is a type V-A Cas protein that has endonuclease activity when complexed with a single guide RNA, also known as a crRNA in a type V-A system (see, U.S. Patent No. 9,982,279). This example describes cleavage of the genomic DNA of Jurkat cells using MAD7 in complex with single guide nucleic acids targeting human ADORA2A, B2M, CARD11, CD247, CD52, CIITA, CTLA4, DCK, DHODH, FAS, HAVCR2, IL7R, LAG3, LCK, MDV, PDCD1, PLCG1, PLK1, PTPN6, TIGIT, TRAC, TRBC1, TRBC2, TUBB, or U6 gene.

[0282] Briefly, Jurkat cells were grown in RPMI 1640 medium (Thermo Fisher Scientific, A1049101) supplemented with 10% fetus bovine serum at 37° C in a 5% CO_2 environment, and split every 2-3 days to a density of 100,000 cells/mL. MAD7 protein, which contained a nucleoplasmin NLS at the C-terminus, was expressed in *E. Coli* and purified by fast protein liquid chromatography (FPLC). RNP complexes were prepared by incubating 66 pmol MAD7 protein with 100 pmol chemically synthesized single guide RNA for 10 minutes at room temperature. The RNPs were mixed with 200,000 Jurkat cells in a final volume of 25 µL. Electroporation was carried out on a 4D-Nucleofector (Lonza) using program CL-120. Following electroporation, the cells were cultured for three days.

[0283] Genomic DNA of the cells was extracted using the Quick Extract DNA extraction solution 1.0 (Epicentre). The genes were amplified from the genomic DNA samples in a PCR reaction with primers with or without overhang adaptors and processed using the

Nextera XT Index Kit v2 Set A (Illumina, FC-131-2001) or the KAPA HyperPlus kit (Roche, cat. no. KK8514), respectively. The final PCR products were analyzed by next-generation sequencing, and the data were analyzed with the AmpliCan package (see, Labun et al. (2019), Accurate analysis of genuine CRISPR editing events with ampliCan, Genome Res., electronically published in advance). Editing efficiency was determined by the number of edited reads relative to the total number of reads obtained under each condition.

[0284] The nucleotide sequence of each single guide RNA used in this example consisted of, from 5' to 3', UAAUUUCUACUCUUGUAGAU (SEQ ID NO: 50) and a spacer sequence. In SEQ ID NO: 50, the modulator stem sequence (UCUAC) and the targeter stem sequence (GUAGA) are underlined. The editing efficiency of each single guide RNA was measured as the percentage of cells having one or more insertion or deletion at the target site (% indel). The spacer sequences tested for targeting human ADORA2A, B2M, CARD11, CD247, CD52, CIITA, CTLA4, DCK, DHODH, FAS, HAVCR2, IL7R, LAG3, LCK, MVD, PDCD1, PLCG1, PLK1, PTPN6, TIGIT, TRAC, TRBC1, TRBC2, TUBB, or U6 gene and the editing efficiency of each single guide RNA are shown in Tables 6-25 and illustrated in Figures 3-15, respectively. In Tables 6-25, N.D. means not determined.

Table 6. Tested crRNAs Targeting Human ADORA2A Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gADORA2A_1	GTGGTGTCACTGGCGGCGGCC	242	0.3
gADORA2A_2	TGGTGTCACTGGCGGCGGCCG	133	3.9
gADORA2A_3	GCCATCACCATCAGCACCGGG	243	0.5
gADORA2A_4	CCATCACCATCAGCACCGGGT	137	2.1
gADORA2A_5	GTCCTGGTCCTCACGCAGAGC	244	0.1
gADORA2A_6	GCCCTCGTGCCGGTCACCAAG	245	0.9
gADORA2A_7	GTGACCGGCACGAGGGCTAAG	135	2.8
gADORA2A_8	CCATCGGCCTGACTCCCATGC	136	2.2
gADORA2A_9	GCTGACCGCAGTTGTTCCAAC	246	1.1
gADORA2A_10	GGCTGACCGCAGTTGTTCCAA	247	0.5
gADORA2A_11	GCCCTCCCCGCAGCCCTGGGA	248	1.3
gADORA2A_12	AGGATGTGGTCCCCATGAACT	51	18.2
gADORA2A_13	AACTTCTTTGCCTGTGTGCTG	249	0.1
gADORA2A_14	TTTGCCTGTGTGCTGGTGCCC	250	0.2

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gADORA2A_15	CCTGTGTGCTGGTGCCCCTGC	251	1.1
gADORA2A_16	CGGATCTTCCTGGCGGCGCGA	131	7.8
gADORA2A_17	AGCTGTCGTCGCGCCGCCAGG	252	0.1
gADORA2A_18	TGCAGTGTGGACCGTGCCCGC	253	0.2
gADORA2A_19	GCAGCATGGACCTCCTTCTGC	254	0.4
gADORA2A_20	CCCTCTGCTGGCTGCCCCTAC	255	0.6
gADORA2A_21	ACTTTCTTCTGCCCCGACTGC	256	0.6
gADORA2A_22	CTTCTGCCCCGACTGCAGCCA	257	1.0
gADORA2A_23	TTCTGCCCCGACTGCAGCCAC	134	2.8
gADORA2A_24	ATCTACGCCTACCGTATCCGC	258	0.0
gADORA2A_25	CGCAAGATCATTCGCAGCCAC	259	0.1
gADORA2A_26	AAAGGTTCTTGCTGCCTCAGG	260	0.1
gADORA2A_27	CAAGGCAGCTGGCACCAGTGC	261	0.1
gADORA2A_28	AAGGCAGCTGGCACCAGTGCC	132	5.8
gADORA2A_29	AGCTCATGGCTAAGGAGCTCC	262	0.2
gADORA2A_30	GCCATGAGCTCAAGGGAGTGT	263	0.5

Table 7. Tested crRNAs Targeting Human B2M Gene

crRNA Name	Spacer Sequence	SEQ ID NO	% Indel
gB2M_1	GCTGTGCTCGCGCTACTCTCT	145	1.8
gB2M_2	TGGCCTGGAGGCTATCCAGCG	65	17.4
gB2M_3	CCCGATATTCCTCAGGTACTC	264	0.1
gB2M_4	CTCACGTCATCCAGCAGAGAA	52	74.1
gB2M_5	CATTCTCTGCTGGATGACGTG	142	2.2
gB2M_6	CCATTCTCTGCTGGATGACGT	265	1.0
gB2M_7	ACTTTCCATTCTCTGCTGGAT	64	17.9
gB2M_8	CTGAATTGCTATGTGTCTGGG	139	3.5
gB2M_9	AATGTCGGATGGATGAAACCC	266	0.5
gB2M_10	ATCCATCCGACATTGAAGTTG	143	2.0
gB2M_11	CTGAAGAATGGAGAGAGAATT	140	3.4
gB2M_12	TCAATTCTCTCTCCATTCTTC	267	0.7
gB2M_13	TTCAATTCTCTCTCCATTCTT	268	0.7
gB2M_14	CTGAAAGACAAGTCTGAATGC	269	0.4

crRNA Name	Spacer Sequence	SEQ ID NO	% Indel
gB2M_15	TCTTTCAGCAAGGACTGGTCT	270	0.9
gB2M_16	AGCAAGGACTGGTCTTTCTAT	271	0.3
gB2M_17	TATCTCTTGTACTACACTGAA	66	15.3
gB2M_18	TCAGTGGGGGTGAATTCAGTG	141	3.0
gB2M_19	ACTATCTTGGGCTGTGACAAA	272	0.1
gB2M_20	GTCACAGCCCAAGATAGTTAA	273	0.8
gB2M_21	TCACAGCCCAAGATAGTTAAG	138	5.3
gB2M_22	CCCCACTTAACTATCTTGGGC	144	2.0
gB2M_23	CTGGCCTGGAGGCTATCCAGC	618	0.77
gB2M_24	TCCCGATATTCCTCAGGTACT	619	0.54
gB2M_25	CCGATATTCCTCAGGTACTCC	620	0.14
gB2M_26	AGTAAGTCAACTTCAATGTCG	621	0.11
gB2M_27	AATTCTCTCTCCATTCTTCAG	622	2.70
gB2M_28	CAATTCTCTCTCCATTCTTCA	623	0.26
gB2M_29	CAGCAAGGACTGGTCTTTCTA	624	0.19
gB2M_30	AGTGGGGGTGAATTCAGTGTA	625	91.96
gB2M_31	CAGTGGGGGTGAATTCAGTGT	626	8.10
gB2M_33	CTATCTCTTGTACTACACTGA	627	0.21
gB2M_34	TACTACACTGAATTCACCCCC	628	0.80
gB2M_35	GGCTGTGACAAAGTCACATGG	629	0.18
gB2M_36	CAAAAGAATGTAAGACTTACC	630	0.13
gB2M_37	CCTCCATGATGCTGCTTACAT	631	0.81
gB2M_38	TTCATAGATCGAGACATGTAA	632	0.18
gB2M_39	TCATAGATCGAGACATGTAAG	633	0.20
gB2M_40	CATAGATCGAGACATGTAAGC	634	4.25
gB2M_41	ATAGATCGAGACATGTAAGCA	635	93.92

Table 8. Tested crRNAs Targeting Human CD52 Gene

crRNA Name	Spacer Sequence	SEQ ID NO	% Indel
gCD52_1	CTCTTCCTCCTACTCACCATC	53	28.4
gCD52_2	TCCTCCTACAGATACAAACTG	274	N.D.
gCD52_3	GTCCTGAGAGTCCAGTTTGTA	275	N.D.
gCD52_4	GCTGGTGTCGTTTTGTCCTGA	146	4.1

crRNA Name	Spacer Sequence	SEQ ID NO	% Indel
gCD52_5	TGTTGCTGGATGCTGAGGGGC	276	1.1
gCD52_6	CCTTTTCTTCGTGGCCAATGC	277	0.2
gCD52_7	TCTTCGTGGCCAATGCCATAA	278	0.2
gCD52_8	CTTCGTGGCCAATGCCATAAT	279	0.15

Table 9. Tested crRNAs Targeting Human CIITA Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gCIITA_1	GGGCTCTGACAGGTAGGACCC	280	0.5
gCIITA_2	TACCTTGGGGCTCTGACAGGT	281	0.0
gCIITA_3	TTACCTTGGGGCTCTGACAGG	282	0.0
gCIITA_4	TAGGGGCCCCAACTCCATGGT	54	13.5
gCIITA_5	TTAACAGCGATGCTGACCCCC	284	0.1
gCIITA_6	TATGACCAGATGGACCTGGCT	285	0.2
gCIITA_7	TCCTCCCAGAACCCGACACAG	286	0.1
gCIITA_8	CCTCCCAGAACCCGACACAGA	287	0.1
gCIITA_9	CATGTCACACAACAGCCTGCT	288	0.1
gCIITA_10	CTCACCGATATTGGCATAAGC	289	0.1
gCIITA_11	TCCTTGTCTGGGCAGCGGAAC	290	0.1
gCIITA_12	CCTTGTCTGGGCAGCGGAACT	291	0.4
gCIITA_13	TCTGGGCAGCGGAACTGGACC	292	0.1
gCIITA_14	CTCAGGCCCTCCAGCTGGGAG	293	0.2
gCIITA_15	CTGAAAATGTCCTTGCTCAGG	294	0.2
gCIITA_16	TCTCAAAGTAGAGCACATAGG	295	0.1
gCIITA_17	ATCTGGTCCTATGTGCTCTAC	296	0.2
gCIITA_18	TGCTGGCATCTCCATACTCTC	147	4.8
gCIITA_19	CTGCCCAACTTCTGCTGGCAT	297	0.5
gCIITA_20	TCTGCCCAACTTCTGCTGGCA	298	0.1
gCIITA_21	CTGACTTTTCTGCCCAACTTC	299	0.1
gCIITA_22	CTCTGCAGCCTTCCCAGAGGA	300	0.6
gCIITA_23	CCAGAGGAGCTTCCGGCAGAC	301	0.9
gCIITA_24	AGGTCTGCCGGAAGCTCCTCT	302	0.1
gCIITA_25	CAGTGCTTCAGGTCTGCCGGA	303	0.2
gCIITA_26	CGGCAGACCTGAAGCACTGGA	304	0.3

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gCIITA_27	CTCACAGCTGAGCCCCCCACT	305	0.4
gCIITA_28	CTCCAGGCGCATCTGGCCGGA	306	0.7
gCIITA_29	GTCTCTTGCAGTGCCTTTCTC	148	2.4
gCIITA_30	TCTCTTGCAGTGCCTTTCTCC	307	0.1
gCIITA_31	CTCCAGTTCCTCGTTGAGCTG	308	0.1
gCIITA_32	CCTTGGGGCTCTGACAGGTAG	636	93.85
gCIITA_33	ACCTTGGGGCTCTGACAGGTA	637	11.83
gCIITA_34	CCGGCCTTTTTACCTTGGGGC	638	2.26
gCIITA_35	CTCCCAGAACCCGACACAGAC	639	48.70
gCIITA_36	TGGGCTCAGGTGCTTCCTCAC	640	85.46
gCIITA_37	CTGGGCTCAGGTGCTTCCTCA	641	0.45
gCIITA_38	CTTGTCTGGGCAGCGGAACTG	642	38.38
gCIITA_39	CTCAAAGTAGAGCACATAGGA	643	0.25
gCIITA_40	TCAAAGTAGAGCACATAGGAC	644	15.68
gCIITA_41	TGCCCAACTTCTGCTGGCATC	645	46.21
gCIITA_42	TGACTTTTCTGCCCAACTTCT	646	2.72
gCIITA_43	TCTGCAGCCTTCCCAGAGGAG	647	55.09
gCIITA_44	TCCAGGCGCATCTGGCCGGAG	648	39.16
gCIITA_45	TCCAGTTCCTCGTTGAGCTGC	649	0.22
gCIITA_46	CCAGAGCCCATGGGGCAGAGT	650	1.51
gCIITA_47	TCCCCACCATCTCCACTCTGC	651	2.05
gCIITA_48	CTCGGGAGGTCAGGGCAGGTT	652	61.63
gCIITA_49	GAAGCTTGTTGGAGACCTCTC	653	0.67
gCIITA_50	GGAAGCTTGTTGGAGACCTCT	654	0.57
gCIITA_51	CAGAGCCGGTGGAGCAGTTCT	655	8.94
gCIITA_52	CCCAGCACAGCAATCACTCGT	656	2.63
gCIITA_53	TCTTCTCTGTCCCCTGCCATT	657	0.28
gCIITA_55	AGCCACATCTTGAAGAGACCT	658	5.71
gCIITA_56	CCAGAAGAAGCTGCTCCGAGG	659	0.52
gCIITA_57	CAGAAGAAGCTGCTCCGAGGT	660	12.02
gCIITA_58	AGCTGTCCGGCTTCTCCATGG	661	3.25
gCIITA_59	AGAGCTCAGGGATGACAGAGC	662	16.35

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gCIITA_60	TGCCGGGCAGTGTGCCAGCTC	663	11.98
gCIITA_61	ATGTCTGCGGCCCAGCTCCCA	664	1.25
gCIITA_62	GCCATCGCCCAGGTCCTCACG	665	1.29
gCIITA_63	GCCACTCAGAGCCAGCCACAG	666	35.47
gCIITA_64	TGGCTGGGCTGATCTTCCAGC	667	0.50
gCIITA_65	GCAGCACGTGGTACAGGAGCT	668	70.73
gCIITA_66	CTGGGCACCCGCCTCACGCCT	669	0.31
gCIITA_67	TGGGCACCCGCCTCACGCCTC	670	12.57
gCIITA_68	CCCCTCTGGATTGGGGAGCCT	671	4.61
gCIITA_69	AAAGGCTCGATGGTGAACTTC	672	1.17
gCIITA_70	CCAGGTCTTCCACATCCTTCA	673	38.98
gCIITA_71	AAAGCCAAGTCCCTGAAGGAT	674	39.50
gCIITA_72	GGTCCCGAACAGCAGGGAGCT	675	89.25
gCIITA_73	TTTAGGTCCCGAACAGCAGGG	676	10.88
gCIITA_74	CTTACGCAAACTCCAGTTTCT	677	0.79
gCIITA_75	CCTCCTAGGCTGGGCCCTGTC	678	2.78
gCIITA_76	GGGAAAGCCTGGGGGCCTGAG	679	68.93
gCIITA_77	CCCAAACTGGTGCGGATCCTC	680	0.57
gCIITA_79	CTCCCTGCAGCATCTGGAGTG	681	1.12
gCIITA_80	CAAGGACTTCAGCTGGGGGAA	682	87.87
gCIITA_81	TAGGCACCCAGGTCAGTGATG	683	44.56
gCIITA_82	CGACAGCTTGTACAATAACTG	684	34.37
gCIITA_83	TCTTGCCAGCGTCCAGTACAA	685	5.62
gCIITA_84	CCCGGCCTTTTTACCTTGGGG	686	0.38
gCIITA_85	CCTCCCAGGCAGCTCACAGTG	687	0.74
gCIITA_87	TCCAGCCAGGTCCATCTGGTC	688	0.15
gCIITA_88	TTCTCCAGCCAGGTCCATCTG	689	0.21
gCIITA_89	ATCACCTTCCATGTCACACAA	690	0.31
gCIITA_90	TCTGGGCTCAGGTGCTTCCTC	691	0.25
gCIITA_91	TGCCAATATCGGTGAGGAAGC	692	0.17
gCIITA_92	CAGGACTCCCAGCTGGAGGGC	693	0.61
gCIITA_93	TCTGACTTTTCTGCCCAACTT	694	0.21

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gCIITA_94	CAGTGCCTTTCTCCAGTTCCT	695	0.25
gCIITA_95	GCTGGCCTGGGGCACCTCACC	696	0.59
gCIITA_96	GCTCCATCAGCCACTGACCTG	697	0.29
gCIITA_97	CCTGTCATGTTTGCTCGGGAG	698	0.27
gCIITA_98	TCCATCTCCAGAGCACAAGAC	699	0.23
gCIITA_99	TTGGAGACCTCTCCAGCTGCC	700	0.99
gCIITA_100	GCAGAGCCGGTGGAGCAGTTC	701	0.46
gCIITA_101	CTGCTGCTCCTCTCCAGCCTG	702	0.23
gCIITA_103	GCAGCCAACAGCACCTCAGCC	703	0.22
gCIITA_104	GCCCAGCACAGCAATCACTCG	704	0.07

Table 10. Tested crRNAs Targeting Human CTLA4 Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gCTLA4_1	TGCCGCTGAAATCCAAGGCAA	309	1.3
gCTLA4_2	CCTTGGATTTCAGCGGCACAA	310	0.8
gCTLA4_3	GATTTCAGCGGCACAAGGCTC	311	0.6
gCTLA4_4	AGCGGCACAAGGCTCAGCTGA	55	58.4
gCTLA4_5	TTCTTCTCTTCATCCCTGTCT	155	1.7
gCTLA4_6	CAGAAGACAGGGATGAAGAGA	68	44.6
gCTLA4_7	GCAGAAGACAGGGATGAAGAG	312	0.2
gCTLA4_8	GGCTTTTCCATGCTAGCAATG	313	0.1
gCTLA4_9	GCTTTTCCATGCTAGCAATGC	314	0.2
gCTLA4_10	TCCATGCTAGCAATGCACGTG	315	0.1
gCTLA4_11	CCATGCTAGCAATGCACGTGG	316	0.1
gCTLA4_12	GTGTGTGAGTATGCATCTCCA	317	0.8
gCTLA4_13	TGTGTGAGTATGCATCTCCAG	70	12.6
gCTLA4_14	CCTGGAGATGCATACTCACAC	67	47.4
gCTLA4_15	GCCTGGAGATGCATACTCACA	318	0.2
gCTLA4_16	GGCAGGCTGACAGCCAGGTGA	319	1.2
gCTLA4_17	AGTCACCTGGCTGTCAGCCTG	320	0.4
gCTLA4_18	CTAGATGATTCCATCTGCACG	154	2.0
gCTLA4_19	CACTGGAGGTGCCCGTGCAGA	69	42.5
gCTLA4_20	ATTTCCACTGGAGGTGCCCGT	321	0.1

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gCTLA4_21	GATAGTGAGGTTCACTTGATT	322	0.6
gCTLA4_22	CAGATGTAGAGTCCCGTGTCC	323	0.6
gCTLA4_23	CTCACCAATTACATAAATCTG	324	0.8
gCTLA4_24	GCTCACCAATTACATAAATCT	325	1.0
gCTLA4_25	GTTTTCTGTTGCAGATCCAGA	326	0.1
gCTLA4_26	TTTTCTGTTGCAGATCCAGAA	327	0.1
gCTLA4_27	CTGTTGCAGATCCAGAACCGT	149	5.0
gCTLA4_28	CTCCTCTGGATCCTTGCAGCA	152	3.0
gCTLA4_29	CAGCAGTTAGTTCGGGGTTGT	328	0.7
gCTLA4_30	TTTATAGCTTTCTCCTCACAG	329	0.6
gCTLA4_31	CTCCTCACAGCTGTTTCTTTG	330	1.0
gCTLA4_32	TCCTCACAGCTGTTTCTTTGA	331	0.7
gCTLA4_33	GCTCAAAGAAACAGCTGTGAG	332	0.8
gCTLA4_34	TTTTTGTGTTTGACAGCTAAA	333	0.5
gCTLA4_35	TGTGTTTGACAGCTAAAGAAA	334	0.1
gCTLA4_36	ACAGCTAAAGAAAAGAAGCCC	150	3.9
gCTLA4_37	CACATAGACCCCTGTTGTAAG	153	2.9
gCTLA4_38	CACATTCTGGCTCTGTTGGGG	335	0.2
gCTLA4_39	TCACATTCTGGCTCTGTTGGG	336	0.3
gCTLA4_40	AGCCTTATTTTATTCCCATCA	337	0.3
gCTLA4_41	TCAATTGATGGGAATAAAATA	151	3.0

Table 11. Tested crRNAs Targeting Human DCK Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gDCK_1	TCTTGGGCGGGGTGGCCATTC	338	0.1
gDCK_2	TCAGCCAGCTCTGAGGGGACC	71	50.4
gDCK_3	CTTGATGCGGGTCCCCTCAGA	339	0.3
gDCK_4	GATGGAGATTTTCTTGATGCG	340	0.3
gDCK_5	CCGATGTTCCCTTCGATGGAG	341	0.5
gDCK_6	CGGAGGCTCCTTACCGATGTT	56	85.1
gDCK_7	ATCTTTCCTCACAACAGCTGC	159	1.5
gDCK_8	CTCACAACAGCTGCAGGGAAG	72	31.7
gDCK_9	AGGATATTCACAAATGTTGAC	156	8.1

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gDCK_10	TGAATATCCTTAAACAATTGT	342	1.0
gDCK_11	CCAATCTTCACACAATTGTTT	343	0.1
gDCK_12	AACAATTGTGTGAAGATTGGG	344	0.8
gDCK_13	AACATTGCACCATCTGGCAAC	345	1.2
gDCK_14	GAACATTGCACCATCTGGCAA	346	0.6
gDCK_15	CATACCTCAAATTCATCTTGA	347	0.3
gDCK_16	ATTTTCATACCTCAAATTCAT	348	0.1
gDCK_17	AATTTTATTTTCATACCTCAA	349	0.0
gDCK_18	TGCACATTCAAAATAGGAACT	350	0.4
gDCK_19	TCTGAGACATTGTAAGTTCCT	351	0.7
gDCK_20	CAATGTCTCAGAAAAATGGTG	352	0.6
gDCK_21	TCATACATCATCTGAAGAACA	158	3.6
gDCK_22	GAAGGTAAAAGACCATCGTTC	157	5.6
gDCK_23	ACCTTCCAAACATATGCCTGT	353	1.2
gDCK_24	CAAACATATGCCTGTCTCAGT	354	1.1
gDCK_25	CCATTCAGAGAGGCAAGCTGA	355	0.9
gDCK_26	AGCTTGCCATTCAGAGAGGCA	73	13.3
gDCK_27	CCTCTCTGAATGGCAAGCTCA	356	1.1
gDCK_28	TCTGCATCTTTGAGCTTGCCA	357	0.1
gDCK_29	TTGAACGATCTGTGTATAGTG	358	0.2
gDCK_30	TACATACCTGTCACTATACAC	74	12.8
gDCK_31	AGGTATATTTTTGCATCTAAT	359	0.05

Table 12. Tested crRNAs Targeting Human FAS Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gFAS_1	GGAGGATTGCTCAACAACCAT	78	22.6
gFAS_2	TATTTTACAGGTTCTTACGTC	360	0.1
gFAS_3	ATTTTACAGGTTCTTACGTCT	361	0.7
gFAS_4	ACAGGTTCTTACGTCTGTTGC	172	1.5
gFAS_5	GGACGATAATCTAGCAACAGA	165	1.9
gFAS_6	TGGACGATAATCTAGCAACAG	362	0.0
gFAS_7	GGCATTAACACTTTTGGACGA	363	0.1
gFAS_8	GAGTTGATGTCAGTCACTTGG	364	0.1

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gFAS_9	CAAGTTCTGAGTCTCAACTGT	365	0.1
gFAS_10	GAAGGCCTGCATCATGATGGC	163	2.4
gFAS_11	TGGCAGAATTGGCCATCATGA	366	0.8
gFAS_12	GTGTAACATACCTGGAGGACA	77	29.9
gFAS_13	TTTCCTTGGGCAGGTGAAAGG	367	1.1
gFAS_14	TTCCTTGGGCAGGTGAAAGGA	166	1.7
gFAS_15	GGCAGGTGAAAGGAAAGCTAG	173	1.5
gFAS_16	TTGGCAGGGCACGCAGTCTGG	368	0.7
gFAS_17	CCTTCTTGGCAGGGCACGCAG	369	0.8
gFAS_18	TCTGTGTACTCCTTCCCTTCT	370	1.0
gFAS_19	GTCTGTGTACTCCTTCCCTTC	371	0.6
gFAS_20	GAAGAAAAATGGGCTTTGTCT	372	0.7
gFAS_21	TCTTCCAAATGCAGAAGATGT	373	0.7
gFAS_22	ATCACACAATCTACATCTTCT	374	0.5
gFAS_23	AAGACTCTTACCATGTCCTTC	375	0.6
gFAS_24	CAAACTGATTTTCTAGGCTTA	376	0.1
gFAS_25	CTAGGCTTAGAAGTGGAAATA	162	3.5
gFAS_26	GAAGTGGAAATAAACTGCACC	377	0.3
gFAS_27	GTATTCTGGGTCCGGGTGCAG	378	1.3
gFAS_28	CATCTGCACTTGGTATTCTGG	379	1.2
gFAS_29	GTTTACATCTGCACTTGGTAT	167	1.6
gFAS_30	TTTTGTAACTCTACTGTATGT	380	0.8
gFAS_31	TTTGTAACTCTACTGTATGTG	381	1.4
gFAS_32	GTGCAAGGGTCACAGTGTTCA	164	2.4
gFAS_33	CTTGGTGCAAGGGTCACAGTG	168	1.6
gFAS_34	TTTTTCTAGATGTGAACATGG	75	59.1
gFAS_35	ATGATTCCATGTTCACATCTA	76	58.5
gFAS_36	GTGTTGCTGGTGAGTGTGCAT	57	61.9
gFAS_37	CACTTGGTGTTGCTGGTGAGT	382	1.3
gFAS_38	CTCTTTGCACTTGGTGTTGCT	170	1.5
gFAS_39	GGGTGGCTTTGTCTTCTTCTT	383	0.1
gFAS_40	GTCTTCTTCTTTTGCCAATTC	384	0.6

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gFAS_41	TCTTCTTCTTTTGCCAATTCC	385	0.1
gFAS_42	GCCAATTCCACTAATTGTTTG	386	0.4
gFAS_43	CCCCAAACAATTAGTGGAATT	387	0.4
gFAS_44	AACAAAGCAAGAACTTACCCC	388	0.3
gFAS_45	TTTGTTCTTTCAGTGAAGAGA	161	6.0
gFAS_46	TTCTTTCAGTGAAGAGAAAGG	389	0.9
gFAS_47	AGTGAAGAGAAAGGAAGTACA	160	9.8
gFAS_48	CTGTACTTCCTTTCTCTTCAC	390	0.8
gFAS_49	TGCATGTTTTCTGTACTTCCT	391	0.6
gFAS_50	CTGCATGTTTTCTGTACTTCC	392	0.4
gFAS_51	TGTGCTTTCTGCATGTTTTCT	393	0.3
gFAS_52	CTGTGCTTTCTGCATGTTTTC	394	0.3
gFAS_53	CCTTTCTGTGCTTTCTGCATG	395	0.3
gFAS_54	GTTTTCCTTTCTGTGCTTTCT	396	0.4
gFAS_55	AAGTTGGAGATTCATGAGAAC	397	0.4
gFAS_56	AATACCTACAGGATTTAAAGT	398	0.3
gFAS_57	TTGCTTTCTAGGAAACAGTGG	399	1.1
gFAS_58	CTAGGAAACAGTGGCAATAAA	400	1.3
gFAS_59	TAGGAAACAGTGGCAATAAAT	79	11.0
gFAS_60	CCAGATAAATTTATTGCCACT	401	0.7
gFAS_61	CTATTTTTCAGATGTTGACTT	402	0.1
gFAS_62	TCAGATGTTGACTTGAGTAAA	403	0.6
gFAS_63	AGTAAATATATCACCACTATT	404	0.8
gFAS_64	AACTTGACTTAGTGTCATGAC	405	0.4
gFAS_65	GAACAAAGCCTTTAACTTGAC	406	0.5
gFAS_66	GTTCGAAAGAATGGTGTCAAT	407	0.9
gFAS_67	ATTGACACCATTCTTTCGAAC	408	0.5
gFAS_68	TTCGAAAGAATGGTGTCAATG	409	0.7
gFAS_69	GGCTTCATTGACACCATTCTT	410	0.4
gFAS_70	TGTTCTGCTGTGTCTTGGACA	171	1.5
gFAS_71	CTGTTCTGCTGTGTCTTGGAC	169	1.5
gFAS_72	GTAATTGGCATCAACTTCATG	411	0.3

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gFAS_73	CATGAAGTTGATGCCAATTAC	412	0.8
gFAS_74	TTTCCATGAAGTTGATGCCAA	413	0.4
gFAS_75	TTTCTTTCCATGAAGTTGATG	414	0.5
gFAS_76	ATGGAAAGAAAGAAGCGTATG	415	1.3
gFAS_77	ATCAATGTGTCATACGCTTCT	416	0.8
gFAS_78	TTGAGATCTTTAATCAATGTG	417	1.0
gFAS_79	TTTGAGATCTTTAATCAATGT	418	0.9
gFAS_80	CTCTGCAAGAGTACAAAGATT	419	0.2
gFAS_81	TACTCTTGCAGAGAAAATTCA	420	0.2
gFAS_82	AGGATGATAGTCTGAATTTTC	421	0.4
gFAS_83	CTGAGTCACTAGTAATGTCCT	422	0.7
gFAS_84	AATTTTCTGAGTCACTAGTAA	423	0.6
gFAS_85	TGAAGTTTGAATTTTCTGAGT	424	0.4
gFAS_86	ATTTCTGAAGTTTGAATTTTC	425	0.3
gFAS_87	GATTTCATTTCTGAAGTTTGA	426	0.5
gFAS_88	GGATTTCATTTCTGAAGTTTG	427	0.5
gFAS_89	AGAAATGAAATCCAAAGCTTG	428	0.5
gFAS_90	TCACTCTAGACCAAGCTTTGG	429	0.5
gFAS_91	TTGTTTTTCACTCTAGACCAA	430	0.7
gFAS_92	GTCTAGAGTGAAAAACAACAA	431	0.5

Table 13. Tested crRNAs Targeting Human HAVCR2 Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gTIM3_1	TCTTCTGCAAGCTCCATGTTT	432	0.1
gTIM3_2	TCTTCTGCAAGCTCCATGTTT	433	0.07
gTIM3_3	CTTCTGCAAGCTCCATGTTTT	434	0.1
gTIM3_4	CACATCTTCCCTTTGACTGTG	435	0.8
gTIM3_5	GACTGTGTCCTGCTGCTGCTG	436	0.8
gTIM3_6	TAAGTAGTAGCAGCAGCAGCA	81	53.7
gTIM3_7	CTTGTAAGTAGTAGCAGCAGC	58	64.4
gTIM3_8	TCTCTCTATGCAGGGTCCTCA	437	0.1
gTIM3_9	TACACCCCAGCCGCCCCAGGG	438	1.0
gTIM3_10	CCCCAGCAGACGGGCACGAGG	175	7.3

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gTIM3_11	GCCCCAGCAGACGGGCACGAG	439	0.6
gTIM3_12	AATGTGGCAACGTGGTGCTCA	84	21.9
gTIM3_13	ATCAGTCCTGAGCACCACGTT	187	1.5
gTIM3_14	CATCAGTCCTGAGCACCACGT	440	0.1
gTIM3_15	GCCAGTATCTGGATGTCCAAT	181	2.9
gTIM3_16	CGGAAATCCCCATTTAGCCAG	441	0.4
gTIM3_17	GCGGAAATCCCCATTTAGCCA	442	0.1
gTIM3_18	CGCAAAGGAGATGTGTCCCTG	86	14.4
gTIM3_19	GATCCGGCAGCAGTAGATCCC	178	5.1
gTIM3_20	TCATCATTCATTATGCCTGGG	443	0.1
gTIM3_21	AGGTTAAATTTTTCATCATTC	444	0.1
gTIM3_22	ATGACCAACTTCAGGTTAAAT	445	0.1
gTIM3_23	ACCTGAAGTTGGTCATCAAAC	184	2.2
gTIM3_24	TGTTGTTTCTGACATTAGCCA	446	0.7
gTIM3_25	TGACATTAGCCAAGGTCACCC	85	15.7
gTIM3_26	GAAAGGCTGCAGTGAAGTCTC	447	0.1
gTIM3_27	ACTGCAGCCTTTCCAAGGATG	182	2.6
gTIM3_28	CCAAGGATGCTTACCACCAGG	185	1.9
gTIM3_29	CAAGGATGCTTACCACCAGGG	80	59.8
gTIM3_30	CCACCAGGGGACATGGCCCAG	83	22.1
gTIM3_31	TATAGCAGAGACACAGACACT	448	0.3
gTIM3_32	TATCAGGGAGGCTCCCCAGTG	82	22.4
gTIM3_33	CTGTTAGATTTATATCAGGGA	449	1.4
gTIM3_34	TGTTTCCATAGCAAATATCCA	177	5.6
gTIM3_35	CATAGCAAATATCCACATTGG	450	1.0
gTIM3_36	CGGGACTCTGGAGCAACCATC	180	3.3
gTIM3_37	AAAATTAAAGCGCCGAAGATA	451	0.2
gTIM3_38	CATTTGAAAATTAAAGCGCCG	452	0.1
gTIM3_39	TGTTTCCCCCTTACTAGGGTA	453	0.7
gTIM3_40	GTTTCCCCCTTACTAGGGTAT	186	1.7
gTIM3_41	CCCCTTACTAGGGTATTCTCA	183	2.2
gTIM3_42	CTAGGGTATTCTCATAGCAAA	174	8.5

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gTIM3_43	AATTCTGTATCTTCTCTTTGC	454	0.7
gTIM3_44	ATTTCCACAGCCTCATCTCTT	455	0.4
gTIM3_45	TTTCCACAGCCTCATCTCTTT	456	1.0
gTIM3_46	CACAGCCTCATCTCTTTGGCC	457	0.5
gTIM3_47	GCCAACCTCCCTCCCTCAGGA	176	6.0
gTIM3_48	CCAATCCTGAGGGAGGGAGGT	179	4.5
gTIM3_49	CTTCTGAGCGAATTCCCTCTG	458	0.7
gTIM3_50	ATATACGTTCTCTTCAATGGT	459	0.5
gTIM3_51	GGGTTGTCGCTTTGCAATGCC	460	0.5

Table 14. Tested crRNAs Targeting Human LAG3 Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gLAG3_1	CTGTTTCTGCAGCCGCTTTGG	461	0.2
gLAG3_2	TGCAGCCGCTTTGGGTGGCTC	462	0.2
gLAG3_3	ACCTGGAGCCACCCAAAGCGG	195	3.1
gLAG3_4	GCTCACCTAGTGAAGCCTCTC	463	1.3
gLAG3_5	TGCGAAGAGCAGGGGTCACTT	464	0.8
gLAG3_6	GGGTGCATACCTGTCTGGCTG	59	52.4
gLAG3_7	CCGCCCAGTGGCCCGCCCGCT	465	N.D.
gLAG3_8	TCGCTATGGCTGCGCCCAGCC	466	0.1
gLAG3_9	TCCTTGCACAGTGACTGCCAG	467	N.D.
gLAG3_10	CACAGTGACTGCCAGCCCCCC	468	N.D.
gLAG3_11	GAACTGCTCCTTCAGCCGCCC	469	0.1
gLAG3_12	AGCCGCCCTGACCGCCCAGCC	470	0.1
gLAG3_13	CGCTAAGTGGTGATGGGGGGA	197	2.3
gLAG3_14	CCGCTAAGTGGTGATGGGGGG	471	0.3
gLAG3_15	GCGGAAAGCTTCCTCTTCCTG	472	1.0
gLAG3_16	GGGCAGGAAGAGGAAGCTTTC	191	6.4
gLAG3_17	CTCTTCCTGCCCCAAGTCAGC	473	1.3
gLAG3_18	AACGTCTCCATCATGTATAAC	474	1.1
gLAG3_19	CTTTTCTCTTCAGGTCTGGAG	475	0.2
gLAG3_20	CTCTTCAGGTCTGGAGCCCCC	476	0.2
gLAG3_21	ACAGTGTACGCTGGAGCAGGT	477	0.1

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gLAG3_22	GCAGTGAGGAAAGACCGGGTC	198	2.1
gLAG3_23	CTCACTGCCAAGTGGACTCCT	478	0.4
gLAG3_24	ACCCTTCGACTAGAGGATGTG	479	0.8
gLAG3_25	CCCTTCGACTAGAGGATGTGA	196	2.7
gLAG3_26	GACTAGAGGATGTGAGCCAGG	480	1.0
gLAG3_27	CCACCTGAGGCTGACCTGTGA	193	3.4
gLAG3_28	CCCACCTGAGGCTGACCTGTG	481	0.8
gLAG3_29	TACTCTTTTCAGTGACTCCCA	482	0.3
gLAG3_30	CAGTGACTCCCAAATCCTTTG	483	0.1
gLAG3_31	CCCAGGGATCCAGGTGACCCA	194	3.1
gLAG3_32	GGGTCACCTGGATCCCTGGGG	484	0.2
gLAG3_33	GGTCACCTGGATCCCTGGGGA	88	17.1
gLAG3_34	GTGAGGTGACTCCAGTATCTG	485	0.7
gLAG3_35	TGAGGTGACTCCAGTATCTGG	188	9.3
gLAG3_36	GTGTGGAGCTCTCTGGACACC	486	0.9
gLAG3_37	TGTGGAGCTCTCTGGACACCC	190	6.9
gLAG3_38	TCAGGACCTTGGCTGGAGGCA	87	17.7
gLAG3_39	GCTGGAGGCACAGGAGGCCCA	487	0.3
gLAG3_40	CCCAGCCTTGGCAATGCCAGC	488	0.8
gLAG3_41	CCAGCCTTGGCAATGCCAGCT	189	8.3
gLAG3_42	GCAATGCCAGCTGTACCAGGG	489	0.6
gLAG3_43	TTGGAGCAGCAGTGTACTTCA	490	0.8
gLAG3_44	ACAGAGCTGTCTAGCCCAGGT	491	0.4
gLAG3_45	CTCCATAGGTGCCCAACGCTC	492	1.3
gLAG3_46	TCCATAGGTGCCCAACGCTCT	192	4.0
gLAG3_47	TCATCCTTGGTGTCCTTTCTC	493	0.4
gLAG3_48	GTGTCCTTTCTCTGCTCCTTT	494	0.1
gLAG3_49	CTCTGCTCCTTTTGGTGACTG	495	0.2
gLAG3_50	TCTGCTCCTTTTGGTGACTGG	496	0.1
gLAG3_51	TGGTGACTGGAGCCTTTGGCT	497	0.6
gLAG3_52	GGTGACTGGAGCCTTTGGCTT	498	0.2
gLAG3_53	GGCTTTCACCTTTGGAGAAGA	499	0.1

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gLAG3_54	GCTTTCACCTTTGGAGAAGAC	500	0.2
gLAG3_55	CTCTAAGGCAGAAAATCGTCT	501	0.1
gLAG3_56	CTGCCTTAGAGCAAGGGATTC	502	0.1
gLAG3_57	GAGCAAGGGATTCACCCTCCG	503	0.2

Table 15. Tested crRNAs Targeting Human PDCD1 Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gPD_1	AACCTGACCTGGGACAGTTTC	504	0.2
gPD_2	CCTTCCGCTCACCTCCGCCTG	89	46.9
gPD_3	CGCTCACCTCCGCCTGAGCAG	505	1.0
gPD_4	TCCACTGCTCAGGCGGAGGTG	506	0.6
gPD_5	TCCCCAGCCCTGCTCGTGGTG	507	1.2
gPD_6	GGTCACCACGAGCAGGGCTGG	508	0.7
gPD_7	ACCTGCAGCTTCTCCAACACA	509	0.2
gPD_8	GCACGAAGCTCTCCGATGTGT	90	41.7
gPD_9	TCCAACACATCGGAGAGCTTC	510	0.2
gPD_10	GTGCTAAACTGGTACCGCATG	511	0.2
gPD_11	TCCGTCTGGTTGCTGGGGCTC	512	0.1
gPD_12	CCCGAGGACCGCAGCCAGCCC	513	0.4
gPD_13	CGTGTCACACAACTGCCCAAC	514	0.5
gPD_14	CACATGAGCGTGGTCAGGGCC	515	0.1
gPD_15	GATCTGCGCCTTGGGGGCCAG	516	0.1
gPD_16	ATCTGCGCCTTGGGGGCCAGG	517	1.2
gPD_17	GGGGCCAGGGAGATGGCCCCA	518	0.6
gPD_18	GTGCCCTTCCAGAGAGAAGGG	201	1.7
gPD_19	TGCCCTTCCAGAGAGAAGGGC	519	0.9
gPD_20	CAGAGAGAAGGGCAGAAGTGC	199	2.5
gPD_21	TGCCCTTCTCTCTGGAAGGGC	520	1.4
gPD_22	GAACTGGCCGGCTGGCCTGGG	200	1.7
gPD_23	TCTGCAGGGACAATAGGAGCC	60	57.6
gPD_24	CTCCTCAAAGAAGGAGGACCC	521	0.1
gPD_25	TCCTCAAAGAAGGAGGACCCC	522	0.5
gPD_26	TCTCGCCACTGGAAATCCAGC	523	0.2

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gPD_27	CAGTGGCGAGAGAAGACCCCG	92	23.7
gPD_28	CCTAGCGGAATGGGCACCTCA	524	0.1
gPD_29	CTAGCGGAATGGGCACCTCAT	91	30.3
gPD_30	GCCCCTCTGACCGGCTTCCTT	525	0.3

Table 16. Tested crRNAs Targeting Human PTPN6 Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gPTPN6_1	ACCGAGACCTCAGTGGGCTGG	96	58.2
gPTPN6_2	AGCAGGGTCTCTGCATCCAGC	526	0.3
gPTPN6_4	CTGGCTCGGCCCAGTCGCAAG	208	4.3
gPTPN6_5	TCCCCTCCATACAGGTCATAG	102	14.8
gPTPN6_6	TATGACCTGTATGGAGGGGAG	61	83.4
gPTPN6_7	CGACTCTGACAGAGCTGGTGG	94	78.1
gPTPN6_8	AGGTGGATGATGGTGCCGTCG	209	3.5
gPTPN6_9	CCTGACGCTGCCTTCTCTAGG	527	0.8
gPTPN6_10	TCTAGGTGGTACCATGGCCAC	212	2.4
gPTPN6_11	GCCTGCAGCAGCGTCTCTGCC	528	0.2
gPTPN6_12	TTGTGCGTGAGAGCCTCAGCC	100	29.4
gPTPN6_13	GTGCTTTCTGTGCTCAGTGAC	529	0.8
gPTPN6_14	GGCTGGTCACTGAGCACAGAA	104	10.4
gPTPN6_15	CTGTGCTCAGTGACCAGCCCA	530	0.5
gPTPN6_16	TGTGCTCAGTGACCAGCCCAA	98	37.5
gPTPN6_17	ATGTGGGTGACCCTGAGCGGG	531	0.9
gPTPN6_18	CCTCGCACATGACCTTGATGT	532	1.4
gPTPN6_19	GCTCCCCCCAGGGTGGACGCT	103	13.5
gPTPN6_20	GAGACCTTCGACAGCCTCACG	202	9.7
gPTPN6_21	GACAGCCTCACGGACCTGGTG	533	0.5
gPTPN6_22	AAGAAGACGGGGATTGAGGAG	101	22.3
gPTPN6_23	TTGTTCAGTTCCAACACTCGG	534	0.1
gPTPN6_24	GCTGTATCCTCGGACTCCTGC	535	0.4
gPTPN6_25	CCCACCCACATCTCAGAGTTT	99	34.8
gPTPN6_26	CAGAAGCAGGAGGTGAAGAAC	95	77.5
gPTPN6_27	CAGACGCTGGTGCAAGTTCTT	536	0.3

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gPTPN6_28	CACCAGCGTCTGGAAGGGCAG	205	5.4
gPTPN6_29	TTCTCTGGCCGCTGCCCTTCC	537	0.1
gPTPN6_30	ATGTAGTTGGCATTGATGTAG	538	0.2
gPTPN6_31	CGTCCAGAACCAGCTGCTAGG	539	0.3
gPTPN6_32	TGGCAGATGGCGTGGCAGGAG	207	4.4
gPTPN6_33	TCCACCTCTCGGGTGGTCATG	540	0.7
gPTPN6_34	CTCCACCTCTCGGGTGGTCAT	541	1.2
gPTPN6_35	CCAGAACAAATGCGTCCCATA	542	0.2
gPTPN6_36	CAGAACAAATGCGTCCCATAC	543	0.5
gPTPN6_37	TGGGCCCTACTCTGTGACCAA	97	51.3
gPTPN6_38	TATTCGGTTGTGTCATGCTCC	544	0.1
gPTPN6_39	CAGGTCTCCCCGCTGGACAAT	213	1.6
gPTPN6_40	GGGAGACCTGATTCGGGAGAT	210	3.4
gPTPN6_41	CTGGACCAGATCAACCAGCGG	203	8.4
gPTPN6_42	CTGCCGCTGGTTGATCTGGTC	206	5.3
gPTPN6_43	CCTGCCGCTGGTTGATCTGGT	545	0.3
gPTPN6_44	CCCAGCGCCGGCATCGGCCGC	546	N.D.
gPTPN6_45	GTGGAGATGTTCTCCATGAGC	547	N.D.
gPTPN6_46	ACTGCCCCCCACCCAGGCCTG	93	80.3
gPTPN6_47	TACTGCGCCTCCGTCTGCACC	548	0.1
gPTPN6_48	AATGAACTGGGCGATGGCCAC	211	3.3
gPTPN6_49	TTCTTAGTGGTTTCAATGAAC	549	0.1
gPTPN6_50	GCATGGGCATTCTTCATGGCT	550	N.D.
gPTPN6_51	GACGAGGTGCGGGAGGCCTTG	551	N.D.
gPTPN6_52	GAGTCTAGTGCAGGGACCGTG	552	0.1
gPTPN6_53	CCCCCCTGCACCCGGCTGCAG	204	7.0
gPTPN6_54	TGTCTGCAGCCGGGTGCAGGG	553	0.9
gPTPN6_55	TCCTCCCTCTTGTTCTTAGTG	554	0.0
gPTPN6_56	CTCCTCCCTCTTGTTCTTAGT	555	0.1
gPTPN6_57	TTCACTTTCTCCTCCCTCTTG	556	0.2

Table 17. Tested crRNAs Targeting Human TIGIT Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gTIGIT_1	CCTGAGGCGAGGGGAGCCTGC	557	0.2
gTIGIT_2	AGGCCTTACCTGAGGCGAGGG	62	81.7
gTIGIT_3	GTCCTCTTCCCTAGGAATGAT	558	1.3
gTIGIT_4	TATTGTGCCTGTCATCATTCC	559	1.0
gTIGIT_5	TCTGCAGAAATGTTCCCCGTT	560	1.1
gTIGIT_6	CTCTGCAGAAATGTTCCCCGT	561	0.1
gTIGIT_7	TGCAGAGAAAGGTGGCTCTAT	215	6.0
gTIGIT_8	TGCCGTGGTGGAGGAGAGGTG	562	0.3
gTIGIT_9	TGGCCATTTGTAATGCTGACT	563	0.8
gTIGIT_10	TAATGCTGACTTGGGGTGGCA	216	1.6
gTIGIT_11	GGGTGGCACATCTCCCCATCC	214	9.7
gTIGIT_12	AAGGATGGGGAGATGTGCCAC	564	0.4
gTIGIT_13	AAGGATCGAGTGGCCCCAGGT	565	0.2
gTIGIT_14	TGCATCTATCACACCTACCCT	566	1.4
gTIGIT_15	TAGGACCTCCAGGAAGATTCT	567	0.4
gTIGIT_16	CTAGGACCTCCAGGAAGATTC	568	0.5
gTIGIT_17	CTCCAGCAGGAATACCTGAGC	569	0.8
gTIGIT_18	GTCCTCCCTCTAGTGGCTGAG	105	72.4
gTIGIT_19	GAGCCATGGCCGCGACGCTGG	570	0.9
gTIGIT_20	TAGTCAACGCGACCACCACGA	571	0.1
gTIGIT_21	CTAGTCAACGCGACCACCACG	572	0.1
gTIGIT_22	TAGTTTGTTTGTTTTTAGAAG	573	0.6
gTIGIT_23	TTTGTTTTTAGAAGAAAGCCC	574	1.0
gTIGIT_24	TTTTTAGAAGAAAGCCCTCAG	575	0.4
gTIGIT_25	TAGAAGAAAGCCCTCAGAATC	576	1.2
gTIGIT_26	CACAGAATGGATTCTGAGGGC	577	0.3
gTIGIT_27	CTCCTGAGGTCACCTTCCACA	217	1.6
gTIGIT_28	CTGGGGGTGAGGGAGCACTGG	578	0.5
gTIGIT_29	TGCCTGGACACAGCTTCCTGG	579	0.3
gTIGIT_30	TGTAACTCAGGACATTGAAGT	580	0.5
gTIGIT_31	AATGTCCTGAGTTACAGAAGC	581	0.5

Table 18. Tested crRNAs Targeting Human TRAC Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gTRAC001	TGTTTTTAATGTGACTCTCAT	237	1.8
gTRAC002	GTGTTTTTAATGTGACTCTCA	582	0.4
gTRAC003	CGTAGGATTTTGTGTTTTTAA	583	0.1
gTRAC004	CTTAGTGCTGAGACTCATTCT	584	0.7
gTRAC005	CCTTAGTGCTGAGACTCATTC	585	0.6
gTRAC006	TGAGGGTGAAGGATAGACGCT	63	81.8
gTRAC007	ATAAACTGTAAAGTACCAAAC	239	1.7
gTRAC008	TTTGGTACTTTACAGTTTATT	586	0.2
gTRAC009	GTACTTTACAGTTTATTAAAT	238	1.7
gTRAC010	CAGTTTATTAAATAGATGTTT	587	0.5
gTRAC011	TTAAATAGATGTTTATATGGA	588	0.0
gTRAC012	TATGGAGAAGCTCTCATTTCT	110	46.7
gTRAC013	TTTCTCAGAAGAGCCTGGCTA	225	5.8
gTRAC014	TCAGAAGAGCCTGGCTAGGAA	127	16.6
gTRAC015	ACCTGCAAAATGAATATGGTG	589	0.0
gTRAC016	GCAGGTGAAATTCCTGAGATG	590	0.2
gTRAC017	CAGGTGAAATTCCTGAGATGT	107	63.6
gTRAC018	CTCGATATAAGGCCTTGAGCA	120	26.0
gTRAC019	AACTATAAATCAGAACACCTG	228	4.5
gTRAC020	GAACTATAAATCAGAACACCT	224	6.4
gTRAC021	TAGTTCAAAACCTCTATCAAT	117	27.7
gTRAC022	TGGTATGTTGGCATTAAGTTG	591	1.0
gTRAC023	CCAACTTAATGCCAACATACC	592	1.4
gTRAC024	CTTTGCTGGGCCTTTTTCCCA	593	1.0
gTRAC025	CTGGGCCTTTTTCCCATGCCT	227	4.6
gTRAC026	TCCCATGCCTGCCTTTACTCT	594	0.6
gTRAC027	CCCATGCCTGCCTTTACTCTG	595	0.7
gTRAC028	CCATGCCTGCCTTTACTCTGC	129	15.3
gTRAC029	CTCTGCCAGAGTTATATTGCT	128	15.8
gTRAC030	ATAGGATCTTCTTCAAAACCC	235	2.2
gTRAC031	TTTAATAGGATCTTCTTCAAA	596	0.3

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gTRAC032	ATTTAATAGGATCTTCTTCAA	597	0.1
gTRAC033	GAAGAAGATCCTATTAAATAA	236	2.0
gTRAC034	AAGAAGATCCTATTAAATAAA	598	0.1
gTRAC035	AGGTTTCCTTGAGTGGCAGGC	220	7.5
gTRAC036	CTTGAGTGGCAGGCCAGGCCT	230	4.4
gTRAC037	AGTGAACGTTCACGGCCAGGC	599	0.7
gTRAC038	TACGGGAAATAGCATCTTAGA	114	40.7
gTRAC039	TAAGATGCTATTTCCCGTATA	111	45.8
gTRAC040	CCGTATAAAGCATGAGACCGT	124	21.5
gTRAC041	CCCCAACCCAGGCTGGAGTCC	125	18.7
gTRAC042	CCTCTTTGCCCCAACCCAGGC	219	7.6
gTRAC043	GAGTCTCTCAGCTGGTACACG	121	25.9
gTRAC044	AGAATCAAAATCGGTGAATAG	221	7.4
gTRAC045	TTTGAGAATCAAAATCGGTGA	600	1.3
gTRAC046	TGACACATTTGTTTGAGAATC	601	0.2
gTRAC047	GATTCTCAAACAAATGTGTCA	602	0.1
gTRAC048	ATTCTCAAACAAATGTGTCAC	229	4.5
gTRAC049	TCTGTGATATACACATCAGAA	118	27.6
gTRAC050	GTCTGTGATATACACATCAGA	130	11.4
gTRAC055	CACATGCAAAGTCAGATTTGT	603	1.0
gTRAC056	CATGTGCAAACGCCTTCAACA	231	3.9
gTRAC057	GTGCCTTCGCAGGCTGTTTCC	604	0.9
gTRAC058	CTTGCTTCAGGAATGGCCAGG	116	27.8
gTRAC059	GACATCATTGACCAGAGCTCT	108	50.1
gTRAC060	AGACATCATTGACCAGAGCTC	605	1.3
gTRAC061	GTGGCAATGGATAAGGCCGAG	115	38.8
gTRAC062	GGTGGCAATGGATAAGGCCGA	223	6.5
gTRAC063	TTAGTAAAAAGAGGGTTTTGG	606	1.4
gTRAC064	TACTAAGAAACAGTGAGCCTT	232	3.5
gTRAC065	ACTAAGAAACAGTGAGCCTTG	607	0.2
gTRAC066	CTAAGAAACAGTGAGCCTTGT	218	9.5
gTRAC067	CCGTGTCATTCTCTGGACTGC	112	45.4

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gTRAC068	CCCGTGTCATTCTCTGGACTG	226	5.3
gTRAC069	TCCCGTGTCATTCTCTGGACT	608	1.0
gTRAC070	TTCCCGTGTCATTCTCTGGAC	609	0.3
gTRAC071	CTCAGACTGTTTGCCCCTTAC	233	3.4
gTRAC072	CCCCTTACTGCTCTTCTAGGC	222	6.9
gTRAC073	GCAGACAGGGAGAAATAAGGA	106	66.9
gTRAC074	GGCAGACAGGGAGAAATAAGG	119	27.1
gTRAC075	TGGCAGACAGGGAGAAATAAG	122	25.2
gTRAC076	TTGGCAGACAGGGAGAAATAA	126	16.7
gTRAC077	TCCCTGTCTGCCAAAAAATCT	610	1.1
gTRAC078	CCAGCTCACTAAGTCAGTCTC	109	47.4
gTRAC079	ATTCCTCCACTTCAACACCTG	113	45.4
gTRAC080	AATTCCTCCACTTCAACACCT	611	0.5
gTRAC081	TAATTCCTCCACTTCAACACC	234	2.3
gTRAC082	CCAGCTGACAGATGGGCTCCC	123	21.5
gTRAC083	CCCAGCTGACAGATGGGCTCC	241	1.6
gTRAC084	GACTTTTCCCAGCTGACAGAT	240	1.6
gTRAC085	TCAACCCTGAGTTAAAACACA	612	0.5
gTRAC086	CTCAACCCTGAGTTAAAACAC	613	0.2
gTRAC087	TCCTGAAGGTAGCTGTTTTCT	614	0.2
gTRAC088	GTCCTGAAGGTAGCTGTTTTC	615	0.1
gTRAC089	AACTCAGGGTTGAGAAAACAG	616	0.7
gTRAC090	ACTCAGGGTTGAGAAAACAGC	617	0.1

Table 19. Tested crRNAs Targeting Human TRBC1/TRBC2 Genes

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gTRBC1+2_1	AGCCATCAGAAGCAGAGATCT	705	66.40 (TRBC1); 74.7 (TRBC2)
gTRBC1+2_3	CGCTGTCAAGTCCAGTTCTAC	706	71.28 (TRBC1)
gTRBC2_7	CCCTGTTTTCTTTCAGACTGT	707	0.09
gTRBC2_8	CTTTCAGACTGTGGCTTCACC	708	0.24

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gTRBC2_9	TTTCAGACTGTGGCTTCACCT	709	0.24
gTRBC2_10	CAGACTGTGGCTTCACCTCCG	710	0.16
gTRBC2_11	AGACTGTGGCTTCACCTCCGG	711	19.97
gTRBC2_12	CCGGAGGTGAAGCCACAGTCT	712	33.14
gTRBC2_13	TCAACAGAGTCTTACCAGCAA	713	1.20
gTRBC2_14	CCAGCAAGGGGTCCTGTCTGC	714	6.69
gTRBC2_15	CTAGGGAAGGCCACCTTGTAT	715	21.74
gTRBC2_16	TATGCCGTGCTGGTCAGTGCC	716	0.20
gTRBC2_17	CCATGGCCATCAGCACGAGGG	717	1.75
gTRBC2_18	CCTAGCAAGATCTCATAGAGG	718	0.37
gTRBC2_19	CACAGGTCAAGAGAAAGGATT	719	1.58
gTRBC2_21	GAGCTAGCCTCTGGAATCCTT	720	11.89

Table 20. Tested crRNAs Targeting Human CARD11 Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gCARD11_1	TAGTACCGCTCCTGGAAGGTT	721	1.37
gCARD11_2	ATCTTGTAGTACCGCTCCTGG	722	0.07
gCARD11_3	CTTCATCTTGTAGTACCGCTC	723	0.08

Table 21. Tested crRNAs Targeting Human CD247 gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gCD247_1	TGTGTTGCAGTTCAGCAGGAG	724	55.77
gCD247_2	CGTTATAGAGCTGGTTCTGGC	725	0.20
gCD247_3	CGGAGGGTCTACGGCGAGGCT	726	20.79
gCD247_4	TTATCTGTTATAGGAGCTCAA	727	12.31
gCD247_5	TCTGTTATAGGAGCTCAATCT	728	0.24
gCD247_6	TCCAAAACATCGTACTCCTCT	729	0.34
gCD247_7	CCCCCATCTCAGGGTCCCGGC	730	6.43
gCD247_8	GACAAGAGACGTGGCCGGGAC	731	40.95
gCD247_9	TCTCCCTCTAACGTCTTCCCG	732	4.13
gCD247_10	CTGAGGGTTCTTCCTTCTCTG	733	0.05
gCD247_11	CCGTTGTCTTTCCTAGCAGAG	734	1.18
gCD247_12	CTAGCAGAGAAGGAAGAACCC	735	70.64

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gCD247_13	TGCAGTTCCTGCAGAAGAGGG	736	4.93
gCD247_14	TGCAGGAACTGCAGAAAGATA	737	2.91
gCD247_15	ATCCCAATCTCACTGTAGGCC	738	31.12
gCD247_16	CATCCCAATCTCACTGTAGGC	739	0.10
gCD247_17	CTCATTTCACTCCCAAACAAC	740	0.30
gCD247_18	TCATTTCACTCCCAAACAACC	741	44.34
gCD247_19	ACTCCCAAACAACCAGCGCCG	742	43.17
gCD247_20	TTTTCTGATTTGCTTTCACGC	743	0.10
gCD247_21	TGATTTGCTTTCACGCCAGGG	744	5.23
gCD247_22	CTTTCACGCCAGGGTCTCAGT	745	8.24
gCD247_23	ACGCCAGGGTCTCAGTACAGC	746	0.30

Table 22. Tested crRNAs Targeting Human IL7R Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gIL7R_1	CTTTCCAGGGGAGATGGATCC	747	0.25
gIL7R_2	CCAGGGGAGATGGATCCTATC	748	8.35
gIL7R_3	CAGGGGAGATGGATCCTATCT	749	87.87
gIL7R_4	CTAACCATCAGCATTTTGAGT	750	0.11
gIL7R_5	GAGTTTTTTCTCTGTCGCTCT	751	0.07
gIL7R_6	AGTTTTTTCTCTGTCGCTCTG	752	0.06
gIL7R_7	TCTGTCGCTCTGTTGGTCATC	753	2.61
gIL7R_8	CATAACACACAGGCCAAGATG	754	25.83

Table 23. Tested crRNAs Targeting Human LCK Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gLCK1_1	ATGTCCTTTCACCCATCAACC	755	0.06
gLCK1_2	CACCCATCAACCCGTAGGGAT	756	0.17
gLCK1_3	ACCCATCAACCCGTAGGGATG	757	16.21

Table 24. Tested crRNAs Targeting Human PLCG1 Gene

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gPLCG1_1	CTCATACACCACGAAGCGCAG	758	0.09
gPLCG1_2	CCTTTCTGCGCTTCGTGGTGT	759	5.14

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gPLCG1_3	CTGCGCTTCGTGGTGTATGAG	760	0.05
gPLCG1_4	TGCGCTTCGTGGTGTATGAGG	761	1.91
gPLCG1_5	GTGGTGTATGAGGAAGACATG	762	3.53

Table 25. Tested crRNAs Targeting Certain Other Human Genes

crRNA	Spacer Sequence	SEQ ID NO	% Indel
gDHODH_1	TTGCAGAAGCGGGCCCAGGAT	770	0.60
gDHODH_2	TTGCAGAAGCGGGCCCAGGAT	771	0.59
gDHODH_3	TATGCTGAACACCTGATGCCG	772	74.94
gPLK1_1	CCAGGGTCGGCCGGTGCCCGT	773	29.06
gPLK1_2	GCCGGTGGAGCCGCCGCCGGA	774	2.01
gPLK1_3	TGGGCAAGGGCGGCTTTGCCA	775	2.26
gPLK1_4	GGGCAAGGGCGGCTTTGCCAA	776	28.24
gPLK1_5	GGCAAGGGCGGCTTTGCCAAG	777	28.41
gPLK1_6	CCAAGTGCTTCGAGATCTCGG	778	2.07
gPLK1_7	CATGGACATCTTCTCCCTCTG	779	90.07
gPLK1_8	TCGAGGACAACGACTTCGTGT	780	0.16
gPLK1_9	CGAGGACAACGACTTCGTGTT	781	6.84
gPLK1_10	GAGGACAACGACTTCGTGTTC	782	8.52
gMVD_1	CAGTTAAAAACCACCACAACA	783	1.42
gMVD_2	GCTGAATGGCCGGGAGGAGGA	784	14.06
gMVD_3	TGGAGTGGCAGATGGGAGAGC	785	63.22
gTUBB_1	AACCATGAGGGAAATCGTGCA	786	2.61
gTUBB_2	ACCATGAGGGAAATCGTGCAC	787	68.40
gTUBB_3	TTCTCTGTAGGTGGCAAATAT	788	18.67
gU6_1	GTCCTTTCCACAAGATATATA	763	68.1
gU6_2	GATTTCTTGGCTTTATATATC	764	0.71
gU6_3	TTGGCTTTATATATCTTGTGG	765	2.83
gU6_4	GCTTTATATATCTTGTGGAAA	766	0.37
gU6_5	ATATATCTTGTGGAAAGGACG	767	0.39
gU6_6	TATATCTTGTGGAAAGGACGA	768	0.39
gU6_7	TGGAAAGGACGAAACACCGTG	769	0.24

INCORPORATION BY REFERENCE

[0285] The entire disclosure of each of the patent and scientific documents referred to herein is incorporated by reference for all purposes.

EQUIVALENTS

[0286] The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The foregoing embodiments are therefore to be considered in all respects illustrative rather than limiting on the invention described herein. Scope of the invention is thus indicated by the appended claims rather than by the foregoing description, and all changes that come within the meaning and range of equivalency of the claims are intended to be embraced therein.

CLAIMS

1. A guide nucleic acid comprising a targeter stem sequence and a spacer sequence, wherein the spacer sequence comprises a nucleotide sequence listed in Table 1, 2, or 3.

2. The guide nucleic acid of claim 1, wherein the targeter stem sequence comprises a nucleotide sequence of GUAGA.

3. The guide nucleic acid of claim 1 or 2, wherein the targeter stem sequence is 5' to the spacer sequence, optionally wherein the targeter stem sequence is linked to the spacer sequence by a linker consisting of 1, 2, 3, 4, or 5 nucleotides.

4. The guide nucleic acid of any one of claims 1-3, wherein the guide nucleic acid is capable of activating a CRISPR Associated (Cas) nuclease in the absence of a tracrRNA.

5. The guide nucleic acid of claim 4, wherein the guide nucleic acid comprises from 5' to 3' a modulator stem sequence, a loop sequence, a targeter stem sequence, and the spacer sequence.

6. The guide nucleic acid of any one of claims 1-3, wherein the guide nucleic acid is a targeter nucleic acid that, in combination with a modulator nucleic acid, is capable of activating a Cas nuclease.

7. The guide nucleic acid of claim 6, wherein the guide nucleic acid comprises from 5' to 3' a targeter stem sequence and the spacer sequence.

8. The guide nucleic acid of any one of claims 4-7, wherein the Cas nuclease is a type V Cas nuclease.

9. The guide nucleic acid of claim 8, wherein the Cas nuclease is a type V-A Cas nuclease.

10. The guide nucleic acid of claim 9, wherein the Cas nuclease comprises an amino acid sequence at least 80% identical to SEQ ID NO: 1.

11. The guide nucleic acid of claim 9, wherein the Cas nuclease is Cpf1.

12. The guide nucleic acid of any one of claims 4-11, wherein the Cas nuclease recognizes a protospacer adjacent motif (PAM) consisting of the nucleotide sequence of TTTN or CTTN.

13. The guide nucleic acid of any one of the proceeding claims, wherein the guide nucleic acid comprises a ribonucleic acid (RNA).

14. The guide nucleic acid of claim 13, wherein the guide nucleic acid comprises a modified RNA.

15. The guide nucleic acid of claim 13 or 14, wherein the guide nucleic acid comprises a combination of RNA and DNA.

16. The guide nucleic acid of any one of claims 13-15, wherein the guide nucleic acid comprises a chemical modification.

17. The guide nucleic acid of claim 16, wherein the chemical modification is present in one or more nucleotides at the 5' end of the guide nucleic acid.

18. The guide nucleic acid of claim 16 or 17, wherein the chemical modification is present in one or more nucleotides at the 3' end of the guide nucleic acid.

19. The guide nucleic acid of any one of claims 16-18, wherein the chemical modification is selected from the group consisting of 2′-O-methyl, 2′-fluoro, 2′-O-methoxyethyl, phosphorothioate, phosphorodithioate, pseudouridine, and any combinations thereof.

20. An engineered, non-naturally occurring system comprising the guide nucleic acid of any one of claims 4-5 and 8-19.

21. The engineered, non-naturally occurring system of claim 20, further comprising the Cas nuclease.

22. The engineered, non-naturally occurring system of claim 21, wherein the guide nucleic acid and the Cas nuclease are present in a ribonucleoprotein (RNP) complex.

23. An engineered, non-naturally occurring system comprising the guide nucleic acid of any one of claims 6-19, further comprising the modulator nucleic acid.

24. The engineered, non-naturally occurring system of claim 23, further comprising the Cas nuclease.

25. The engineered, non-naturally occurring system of claim 24, wherein the guide nucleic acid, the modulator nucleic acid, and the Cas nuclease are present in an RNP complex.

26. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63, 106-130, and 218-241, and wherein the spacer sequence is capable of hybridizing with the human TRAC gene.

27. The engineered, non-naturally occurring system of claim 26, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the TRAC gene locus is edited in at least 1.5% of the cells.

28. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 51 and 131-137, and wherein the spacer sequence is capable of hybridizing with the human ADORA2A gene.

29. The engineered, non-naturally occurring system of claim 28, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the ADORA2A gene locus is edited in at least 1.5% of the cells.

30. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of

SEQ ID NOs: 52, 64-66, 138-145, 622, 625-626, and 634-635, and wherein the spacer sequence is capable of hybridizing with the human B2M gene.

31. The engineered, non-naturally occurring system of claim 30, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the B2M gene locus is edited in at least 1.5% of the cells.

32. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 724, 726-727, 730-732, 735-738, 741-742, and 744-745, and wherein the spacer sequence is capable of hybridizing with the human CD247 gene.

33. The engineered, non-naturally occurring system of claim 32, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the CD247 gene locus is edited in at least 1.5% of the cells.

34. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 53 and 146, and wherein the spacer sequence is capable of hybridizing with the human CD52 gene.

35. The engineered, non-naturally occurring system of claim 34, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the CD52 gene locus is edited in at least 1.5% of the cells.

36. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 54, 147-148, 636-640, 642, 644-648, 650-652, 655-656, 660-663, 666, 668, 670-671, 673-676, 678-679, and 682-685 and wherein the spacer sequence is capable of hybridizing with the human CIITA gene.

37. The engineered, non-naturally occurring system of claim 36, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the CIITA gene locus is edited in at least 1.5% of the cells.

38. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 55, 67-70, and 149-155, and wherein the spacer sequence is capable of hybridizing with the human CTLA4 gene.

39. The engineered, non-naturally occurring system of claim 38, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the CTLA4 gene locus is edited in at least 1.5% of the cells.

40. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 56, 71-74, and 156-159, and wherein the spacer sequence is capable of hybridizing with the human DCK gene.

41. The engineered, non-naturally occurring system of claim 40, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the DCK gene locus is edited in at least 1.5% of the cells.

42. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 57, 75-79, and 160-173, and wherein the spacer sequence is capable of hybridizing with the human FAS gene.

43. The engineered, non-naturally occurring system of claim 42, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the FAS gene locus is edited in at least 1.5% of the cells.

44. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 58, 80-86, and 174-187, and wherein the spacer sequence is capable of hybridizing with the human HAVCR2 gene.

45. The engineered, non-naturally occurring system of claim 44, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the HAVCR2 gene locus is edited in at least 1.5% of the cells.

46. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 748-749 and 753-754, and wherein the spacer sequence is capable of hybridizing with the human IL7R gene.

47. The engineered, non-naturally occurring system of claim 46, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the IL7R gene locus is edited in at least 1.5% of the cells.

48. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 59, 87, 88, and 188-198, and wherein the spacer sequence is capable of hybridizing with the human LAG3 gene.

49. The engineered, non-naturally occurring system of claim 48, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the LAG3 gene locus is edited in at least 1.5% of the cells.

50. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises the nucleotide sequence of SEQ ID NO: 757, and wherein the spacer sequence is capable of hybridizing with the human LCK gene.

51. The engineered, non-naturally occurring system of claim 50, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the LCK gene locus is edited in at least 1.5% of the cells.

52. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 60, 89-92, and 199-201, and wherein the spacer sequence is capable of hybridizing with the human PDCD1 gene.

53. The engineered, non-naturally occurring system of claim 52, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the PDCD1 gene locus is edited in at least 1.5% of the cells.

54. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of of SEQ ID NOs: 759 and 761-762, and wherein the spacer sequence is capable of hybridizing with the human PLCG1 gene.

55. The engineered, non-naturally occurring system of claim 54, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the PLCG1 gene locus is edited in at least 1.5% of the cells.

56. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 61, 93-104, and 202-213, and wherein the spacer sequence is capable of hybridizing with the human PTPN6 gene.

57. The engineered, non-naturally occurring system of claim 56, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the PTPN6 gene locus is edited in at least 1.5% of the cells.

58. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 62, 105, and 214-217, and wherein the spacer sequence is capable of hybridizing with the human TIGIT gene.

59. The engineered, non-naturally occurring system of claim 58, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the TIGIT gene locus is edited in at least 1.5% of the cells.

60. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of

SEQ ID NOs: 705-706, 711-712, 714-715, 717, and 719-720, and wherein the spacer sequence is capable of hybridizing with the human TRBC2 gene.

61. The engineered, non-naturally occurring system of claim 60, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the TRBC2 gene locus is edited in at least 1.5% of the cells.

62. The engineered, non-naturally occurring system of any one of claims 1-25, wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 705-706, and wherein the spacer sequence is capable of hybridizing with both the human TRBC1 gene and the human TRBC2 gene.

63. The engineered, non-naturally occurring system of claim 62, wherein, when the system is delivered into a population of human cells *ex vivo*, the genomic sequence at the TRBC1 gene locus is edited in at least 1.5% of the cells.

64. The engineered, non-naturally occurring system of any one of claims 20-63, wherein genomic mutations are detected in no more than 2% of the cells at any off-target loci by CIRCLE-Seq.

65. The engineered, non-naturally occurring system of claim 64, wherein genomic mutations are detected in no more than 1% of the cells at any off-target loci by CIRCLE-Seq.

66. A human cell comprising the engineered, non-naturally occurring system of any one of claims 20-65.

67. A composition comprising the guide nucleic acid of any one of claims 1-19, the engineered, non-naturally occurring system of any one of claims 20-65, or the human cell of claim 66.

68. A method of cleaving a target DNA comprising the sequence of a preselected target gene or a portion thereof, the method comprising contacting the target DNA with the engineered, non-naturally occurring system of any one of claims 20-65, thereby resulting in cleavage of the target DNA.

69. The method of claim 68, wherein the contacting occurs *in vitro*.

70. The method of claim 68, wherein the contacting occurs in a cell *ex vivo*.

71. The method of claim 70, wherein the target DNA is genomic DNA of the cell.

72. A method of editing human genomic sequence at a preselected target gene locus, the method comprising delivering the engineered, non-naturally occurring system of any one of claims 20-65 into a human cell, thereby resulting in editing of the genomic sequence at the target gene locus in the human cell.

73. The method of any one of claims 70-72, wherein the cell is an immune cell.

74. The method of claim 73, wherein the immune cell is a T lymphocyte.

75. The method of claim 72, the method comprising delivering the engineered, non-naturally occurring system of any one of claims 20-65 into a population of human cells, thereby resulting in editing of the genomic sequence at the target gene locus in at least a portion of the human cells.

76. The method of claim 75, wherein the population of human cells comprises human immune cells.

77. The method of claim 75 or 76, wherein the population of human cells is an isolated population of human immune cells.

78. The method of claim 76 or 77, wherein the immune cells are T lymphocytes.

79. The method of any one of claims 72-78, wherein the engineered, non-naturally occurring system is delivered into the cell(s) as a pre-formed RNP complex.

80. The method of claim 79, wherein the pre-formed RNP complex is delivered into the cell(s) by electroporation.

81. The method of any one of claims 72-80, wherein the target gene is human TRAC gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63, 106-130, and 218-241.

82. The method of any one of claims 75-81, wherein the genomic sequence at the TRAC gene locus is edited in at least 1.5% of the human cells.

83. The method of any one of claims 72-80, wherein the target gene is human ADORA2A gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 51 and 131-137.

84. The method of any one of claims 75-80 and 83, wherein the genomic sequence at the ADORA2A gene locus is edited in at least 1.5% of the human cells.

85. The method of any one of claims 72-80, wherein the target gene is human B2M gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 52, 64-66, 138-145, 622, 625-626, and 634-635,.

86. The method of any one of claims 75-80 and 85, wherein the genomic sequence at the B2M gene locus is edited in at least 1.5% of the human cells.

87. The method of any one of claims 72-80, wherein the target gene is human CD52 gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 53 and 146.

88. The method of any one of claims 75-80 and 87, wherein the genomic sequence at the CD52 gene locus is edited in at least 1.5% of the human cells.

89. The method of any one of claims 72-80, wherein the target gene is human CD247 gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 724, 726-727, 730-732, 735-738, 741-742, and 744-745.

90. The method of any one of claims 75-80 and 89, wherein the genomic sequence at the CD247 gene locus is edited in at least 1.5% of the human cells.

91. The method of any one of claims 72-80, wherein the target gene is human CIITA gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 54, 147-148, 636-640, 642, 644-648, 650-652, 655-656, 660-663, 666, 668, 670-671, 673-676, 678-679, and 682-685.

92. The method of any one of claims 75-80 and 91, wherein the genomic sequence at the CIITA gene locus is edited in at least 1.5% of the human cells.

93. The method of any one of claims 72-80, wherein the target gene is human CTLA4 gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 55, 67-70, and 149-155.

94. The method of any one of claims 75-80 and 93, wherein the genomic sequence at the CTLA4 gene locus is edited in at least 1.5% of the human cells.

95. The method of any one of claims 72-80, wherein the target gene is human DCK gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 56, 71-74, and 156-159.

96. The method of any one of claims 75-80 and 95, wherein the genomic sequence at the DCK gene locus is edited in at least 1.5% of the human cells.

97. The method of any one of claims 72-80, wherein the target gene is human FAS gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 57, 75-79, and 160-173.

98. The method of any one of claims 75-80 and 97, wherein the genomic sequence at the FAS gene locus is edited in at least 1.5% of the human cells.

99. The method of any one of claims 72-80, wherein the target gene is human HAVCR2 gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 58, 80-86, and 174-187.

100. The method of any one of claims 75-80 and 99, wherein the genomic sequence at the HAVCR2 gene locus is edited in at least 1.5% of the human cells.

101. The method of any one of claims 72-80, wherein the target gene is human IL7R gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 748-749 and 753-754.

102. The method of any one of claims 75-80 and 101, wherein the genomic sequence at the IL7R gene locus is edited in at least 1.5% of the human cells.

103. The method of any one of claims 72-80, wherein the target gene is human LAG3 gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 59, 87, 88, and 188-198.

104. The method of any one of claims 75-80 and 103, wherein the genomic sequence at the LAG3 gene locus is edited in at least 1.5% of the human cells.

105. The method of any one of claims 72-80, wherein the target gene is human LCK gene, and wherein the spacer sequence comprises the nucleotide sequence of SEQ ID NO: 757.

106. The method of any one of claims 75-80 and 105, wherein the genomic sequence at the LCK gene locus is edited in at least 1.5% of the human cells.

107. The method of any one of claims 72-80, wherein the target gene is human PDCD1 gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 60, 89-92, and 199-201.

108. The method of any one of claims 75-80 and 107, wherein the genomic sequence at the PDCD1 gene locus is edited in at least 1.5% of the human cells.

109. The method of any one of claims 69-77, wherein the target gene is human PLCG1 gene, and wherein the spacer sequence comprises a sequence of SEQ ID NO: 759 and 761-762.

110. The method of any one of claims 75-80 and 109, wherein the genomic sequence at the PLCG1 gene locus is edited in at least 1.5% of the human cells.

111. The method of any one of claims 72-80, wherein the target gene is human PTPN6 gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 61, 93-104, and 202-213.

112. The method of any one of claims 75-80 and 111, wherein the genomic sequence at the PTPN6 gene locus is edited in at least 1.5% of the human cells.

113. The method of any one of claims 72-80, wherein the target gene is human TIGIT gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 62, 105, and 214-217.

114. The method of any one of claims 75-80 and 113, wherein the genomic sequence at the TIGIT gene locus is edited in at least 1.5% of the human cells.

115. The method of any one of claims 72-80, wherein the target gene is human TRBC2 gene, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 705-706, 711-712, 714-715, 717, and 719-720.

116. The method of any one of claims 75-80 and 115, wherein the genomic sequence at the TRBC2 gene locus is edited in at least 1.5% of the human cells.

117. The method of claim 115 or 116, wherein the method further results in editing of the genomic sequence at human TRBC1 gene locus in the human cell, and wherein the spacer sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 705-706.

118. The method of claim 117, wherein the genomic sequence at the TRBC1 gene locus is edited in at least 1.5% of the human cells.

119. The method of any one of claims 75-118, wherein genomic mutations are detected in no more than 2% of the cells at any off-target loci by CIRCLE-Seq.

120. The method of any one of claims 75-119, wherein genomic mutations are detected in no more than 1% of the cells at any off-target loci by CIRCLE-Seq.

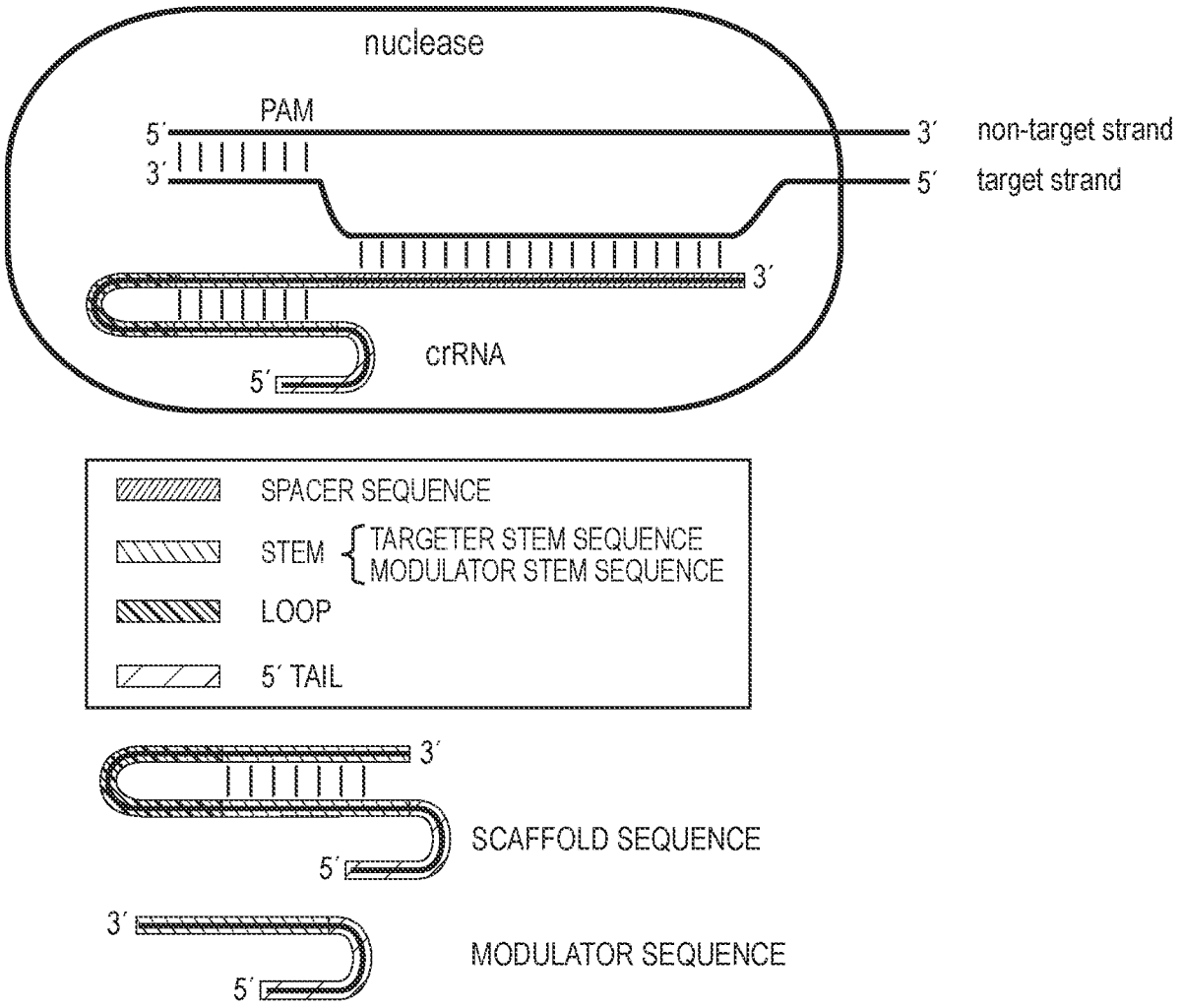
nuclease
PAM
5′
3′
3′
non-target strand
5′
target strand
3′
crRNA
5′
SPACER SEQUENCE
STEM
TARGETER STEM SEQUENCE
MODULATOR STEM SEQUENCE
LOOP
5′ TAIL
3′
5′
SCAFFOLD SEQUENCE
3′
5′
MODULATOR SEQUENCE

FIG. 1A

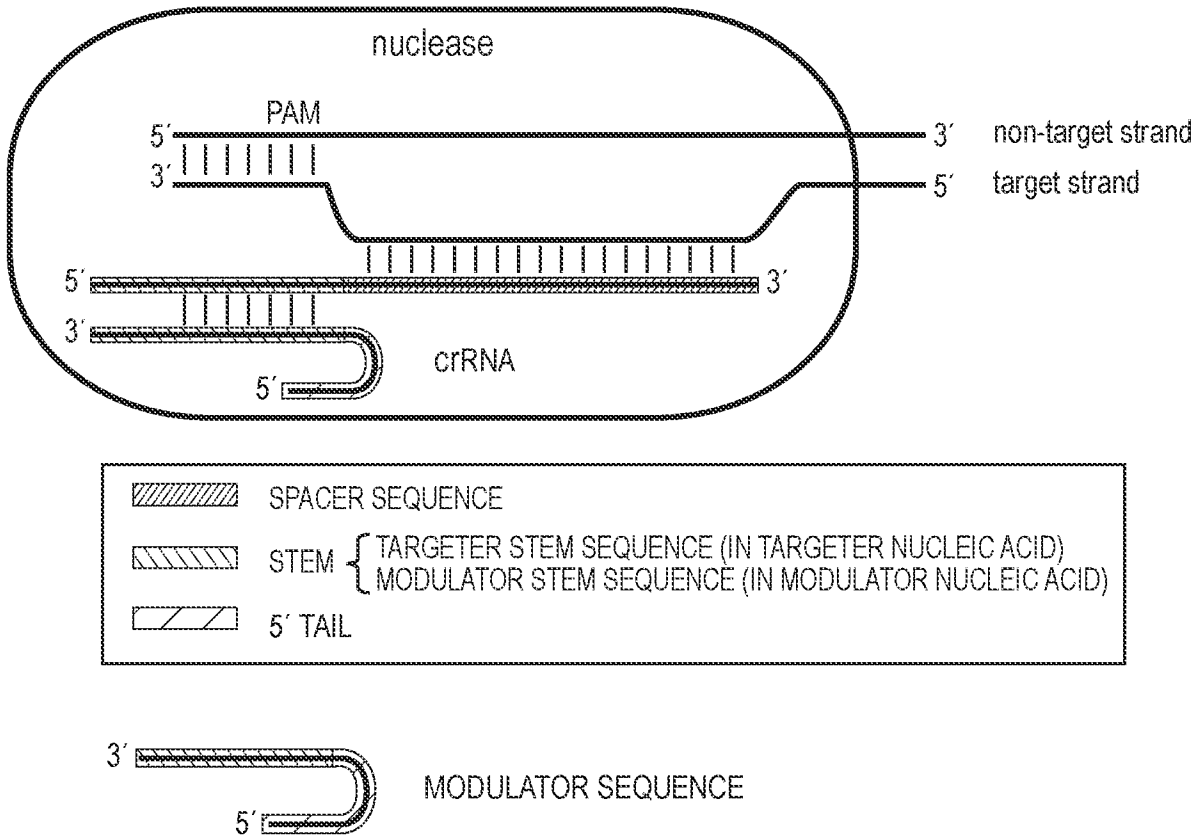
nuclease
PAM
5′
3′
3′
non-target strand
5′
target strand
5′
3′
3′
crRNA
5′
SPACER SEQUENCE
STEM
TARGETER STEM SEQUENCE (IN TARGETER NUCLEIC ACID)
MODULATOR STEM SEQUENCE (IN MODULATOR NUCLEIC ACID)
5′ TAIL
3′
5′
MODULATOR SEQUENCE

FIG. 1B

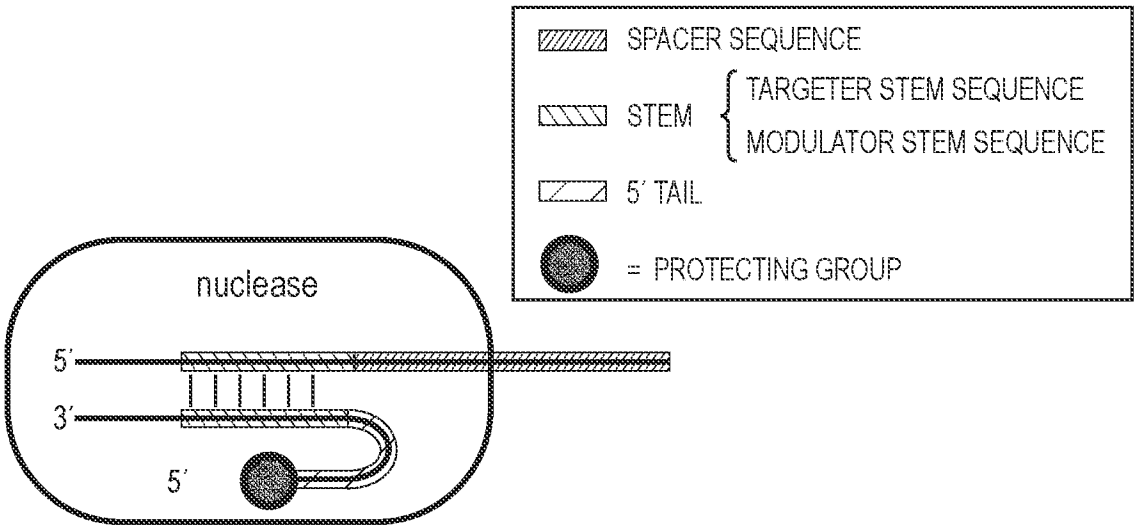

FIG. 2A

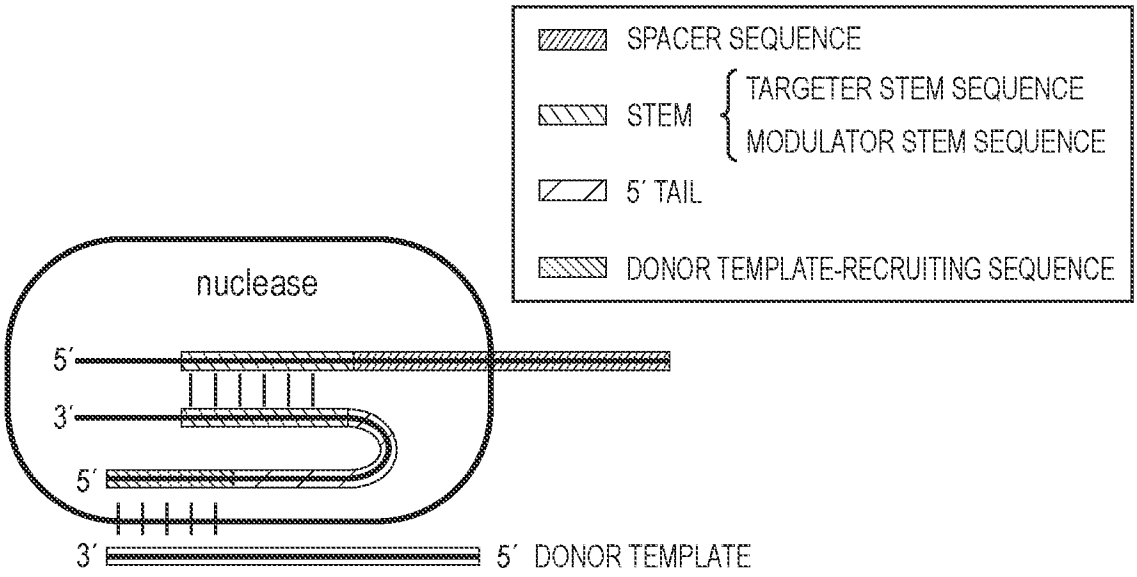

FIG. 2B

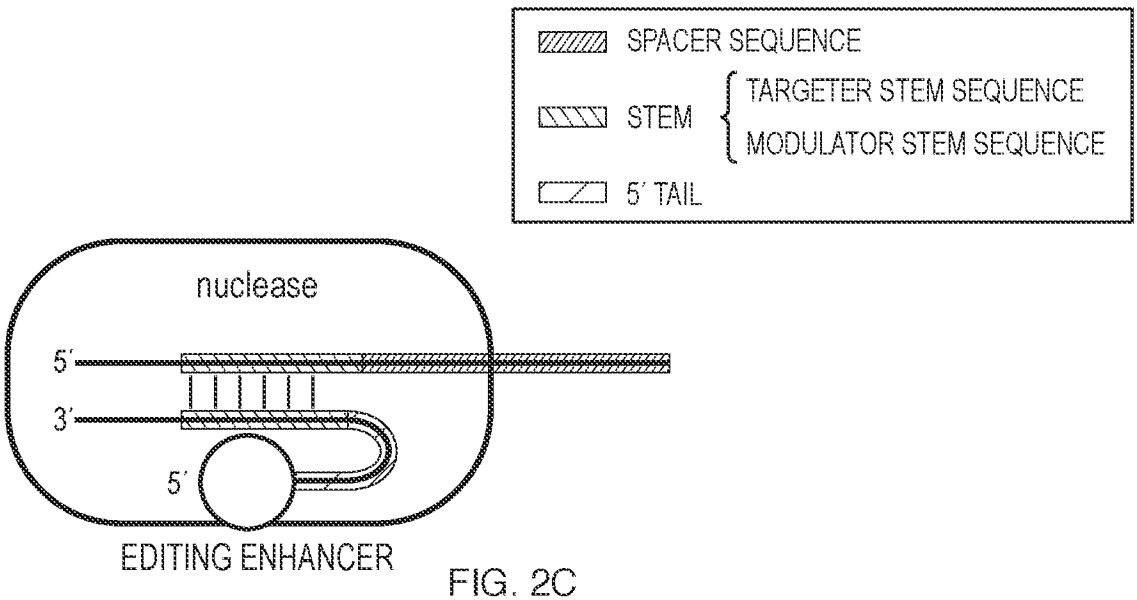

FIG. 2C

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/016823

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/11 (2006.01) C12N 9/22 (2006.01) C12N 5/0783 (2010.01)

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)

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

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

HCAPLUS, EMBASE, BIOSIS, MEDLINE, PATENW & keywords: TRAC, CRISPR, guide nucleic acid, CPF1, MAD7 and like terms; CPC: C12N2310/20
GENOMEQUEST: SEQ ID NOs: 1, 63, 106-130, 218-241
REGISTRY: SEQ ID NOs: 63, 107, 110, 120, 127, 225, 237, 238, 239
PUBMED, ESPACENET, Internal Databases: Inventor and Applicant names

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	

[X] Further documents are listed in the continuation of Box C [X] 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
16 April 2021	16 April 2021

Name and mailing address of the ISA/AU	**Authorised officer**
AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA Email address: pct@ipaustralia.gov.au	Richard Filmer AUSTRALIAN PATENT OFFICE (ISO 9001 Quality Certified Service) Telephone No. +61262832735

INTERNATIONAL SEARCH REPORT C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		International application No. **PCT/US2021/016823**
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2017/152015 A1 (EDITAS MEDICINE, INC.) 08 September 2017 Fig. 1; page 32 lines 11-16; page 32 line 29 - page 33 line 18; Page 37 lines 1-4; page 38 lines 14-23; page 42 lines 2-3; page 44 lines 20-31; page 90 lines 21-27; Example 1; SEQ ID NOs: 3564, 3515, 3536, 3431, 3452 and 3473	1-9, 11-27, 64-82
Y	page 39 line 32 - page 40 line 25	10, 119, 120
X	WO 2016/160721 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 06 October 2016 Abstract; page 2 lines 1 - 17; pages 97-106; Fig. 19; Fig. 25; SEQ ID NOs: 9751-9797	1-9, 11-27, 64-82
Y	page 66 lines 19-26	10, 119, 120
Y	US 9982279 B1 (INSCRIPTA, INC.) 29 May 2018 Example 6; SEQ ID NO: 7	10
Y	TSAI, S. Q. et al., "CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets", Nature Methods, 2017, vol. 14, no. 6, pages 607-614 Abstract; Methods	119, 120

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/016823

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. [X] forming part of the international application as filed:

      [X] 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 13ter.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 13ter.1(a)).

      [ ] on paper or in the form of an image file (Rule 13ter.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:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/016823

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:
the subject matter listed in Rule 39 on which, under Article 17(2)(a)(i), an international search is not required to be carried out, including

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:
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:

See Supplemental Box for Details

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.:
26, 27, 81 and 82 (completely) and 1-25, 64-80, 119 and 120 (in part)

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.

Supplemental Box

Continuation of: **Box III**

This International Application does not comply with the requirements of unity of invention because it does not relate to one invention or to a group of inventions so linked as to form a single general inventive concept.

This Authority has found that there are different inventions based on the following features that separate the claims into distinct groups:

• Invention 1 is defined by claims 26, 27, 81 and 82 (completely) and claims 1-25, 64-80, 119 and 120 (in part). The feature of a guide nucleic acid comprising a spacer sequence that targets the TRAC gene (SEQ ID NOs: 63, 106-130, 218-241) is specific to this group of claims.

• Inventions 2-20 are defined by claims 1-25, 28-80 and 83-120 (each in part). Each guide nucleic acid comprising a spacer sequence that targets one of ADORA2A, B2M, CD52, CIITA, CTLA4, DCK, FAS, HAVCR2, LAG3, PDCD1, PTPN6, TIGIT, TRBC1, TRBC2, CARD11, CD247, IL7R, LCK and PLCG1 is each directed to a different invention.

PCT Rule 13.2, first sentence, states that unity of invention is only fulfilled when there is a technical relationship among the claimed inventions involving one or more of the same or corresponding special technical features. PCT Rule 13.2, second sentence, defines a special technical feature as a feature which makes a contribution over the prior art.

When there is no special technical feature common to all the claimed inventions there is no unity of invention.

In the above groups of claims, the identified features may have the potential to make a contribution over the prior art but are not common to all the claimed inventions and therefore cannot provide the required technical relationship. The only feature common to all of the claimed inventions and which provides a technical relationship among them is a guide nucleic acid comprising a targeter stem sequence and a spacer sequence that targets a mammalian gene. However this feature does not make a contribution over the prior art because it is disclosed in:

WO 2017/152015 (EDITAS MEDICINE, INC.) 8 September 2017

This document discloses guide nucleic acids, each comprising a targeter stem sequence and a spacer sequence that hybridises to a sequence of one of the TRAC, FAS, CTLA4, PDCD1, PTPN6, B2M and TRBC genes. The guide nucleic acids form a complex with a Cpf1 nuclease (a type V-A Cas nuclease).

Therefore in the light of this document this common feature cannot be a special technical feature. Therefore there is no special technical feature common to all the claimed inventions and the requirements for unity of invention are consequently not satisfied *a posteriori*.

Additionally, as guide RNA molecules that target each of the genes of Tables 1-3 appear to be disclosed in the prior art, it may be considered that separate spacer sequence described in these tables defines a separate invention.

A search has conducted with respect to all sequences of Invention 1 as defined above.

INTERNATIONAL SEARCH REPORT Information on patent family members	International application No. **PCT/US2021/016823**

This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document/s Cited in Search Report		**Patent Family Member/s**	
Publication Number	**Publication Date**	**Publication Number**	**Publication Date**
WO 2017/152015 A1	08 September 2017	WO 2017152015 A1	08 Sep 2017
		AU 2017226172 A1	27 Sep 2018
		AU 2017226172 A2	04 Oct 2018
		CA 3016331 A1	08 Sep 2017
		CN 109312340 A	05 Feb 2019
		EP 3423580 A1	09 Jan 2019
		JP 2019507599 A	22 Mar 2019
		KR 20180118204 A	30 Oct 2018
		MX 2018010484 A	10 Jan 2019
		SG 11201807538P A	27 Sep 2018
		US 2019062735 A1	28 Feb 2019
WO 2016/160721 A1	06 October 2016	WO 2016160721 A1	06 Oct 2016
		AU 2016243583 A1	16 Nov 2017
		BR 112017020750 A2	26 Jun 2018
		CA 2984237 A1	06 Oct 2016
		CN 107847524 A	27 Mar 2018
		EP 3273976 A1	31 Jan 2018
		HK 1250481 A1	21 Dec 2018
		JP 2018510657 A	19 Apr 2018
		KR 20180020125 A	27 Feb 2018
		MX 2017012407 A	07 Mar 2018
		US 2016348073 A1	01 Dec 2016
US 9982279 B1	29 May 2018	US 9982279 B1	29 May 2018
		AU 2018289077 A1	30 Jan 2020
		CA 3067951 A1	27 Dec 2018
		CN 111511906 A	07 Aug 2020
		EP 3642334 A1	29 Apr 2020
		JP 2020530264 A	22 Oct 2020
		KR 20200020903 A	26 Feb 2020
		US 10011849 B1	03 Jul 2018
		US 2018371498 A1	27 Dec 2018
		US 10337028 B2	02 Jul 2019
		US 2018371497 A1	27 Dec 2018
		US 10435714 B2	08 Oct 2019
		US 2019390226 A1	26 Dec 2019

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.

INTERNATIONAL SEARCH REPORT Information on patent family members	International application No. **PCT/US2021/016823**

This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document/s Cited in Search Report		Patent Family Member/s	
Publication Number	**Publication Date**	**Publication Number**	**Publication Date**
		US 10626416 B2	21 Apr 2020
		US 2019360001 A1	28 Nov 2019
		US 2020231987 A1	23 Jul 2020
		WO 2018236548 A1	27 Dec 2018

End of Annex

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.