Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3064601 A1 2019/01/03

(21) 3 064 601

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2018/06/26
(87) Date publication PCT/PCT Publication Date: 2019/01/03
(85) Entrée phase nationale/National Entry: 2019/11/21
(86) N° demande PCT/PCT Application No.: US 2018/039616
(87) N° publication PCT/PCT Publication No.: 2019/005884
(30) Priorités/Priorities: 2017/06/26 (US62/525,181); 2017/07/03 (US62/528,391); 2017/07/18 (US62/534,016); 2017/09/21 (US62/561,638); 2017/10/04 (US62/568,304); 2017/10/18 (US62/574,158); 2017/11/27 (US62/591,187); 2017/12/22 (US62/610,105)

(51) Cl.Int./Int.Cl. *A61K 48/00* (2006.01), *C07H 21/04* (2006.01), *C12N 15/00* (2006.01), *C12N 15/82* (2006.01)

(71) Demandeurs/Applicants:
THE BROAD INSTITUTE, INC., US;
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, US;
PRESIDENT AND FELLOWS OF HARVARD COLLEGE, US

(72) Inventeurs/Inventors:
ZHANG, FENG, US;
GOOTENBERG, JONATHAN, US;
COX, DAVID BENJAMIN TURITZ, US;

(54) Titre : COMPOSITIONS A BASE DE CRISPR/CAS-ADENINE DESAMINASE, SYSTEMES ET PROCEDES D'EDITION CIBLEE D'ACIDES NUCLEIQUES

(54) Title: CRISPR/CAS-ADENINE DEAMINASE BASED COMPOSITIONS, SYSTEMS, AND METHODS FOR TARGETED NUCLEIC ACID EDITING

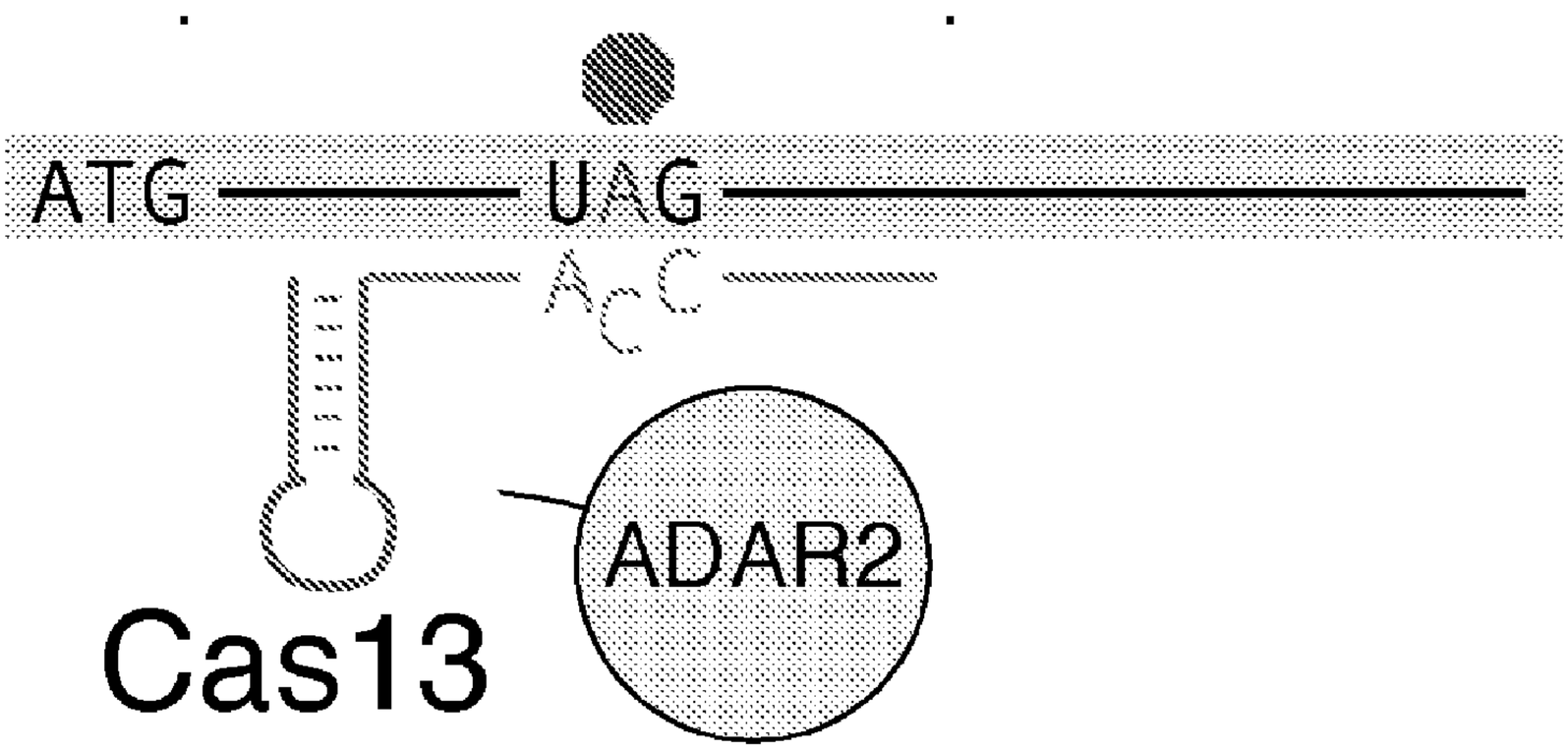


FIG. 1

(57) Abrégé/Abstract:
The invention provides for systems, methods, and compositions for targeting and editing nucleic acids. In particular, the invention provides non-naturally occurring or engineered RNA-targeting systems comprising a RNA-targeting Cas13 protein, at least one guide molecule, and at least one adenosine deaminase protein or catalytic domain thereof.

(21) 3 064 601

(13) A1

(72) Inventeurs(suite)/Inventors(continued): ABUDAYYEH, OMAR, US; KANNAN, SOUMYA, US

(74) Agent: GOWLING WLG (CANADA) LLP

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
03 January 2019 (03.01.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/005884 A1

(51) International Patent Classification:
A61K 48/00 (2006.01) *C12N 15/82* (2006.01)
C12N 15/00 (2006.01) *C07H 21/04* (2006.01)

(21) International Application Number:
PCT/US2018/039616

(22) International Filing Date:
26 June 2018 (26.06.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/525,181	26 June 2017 (26.06.2017)	US
62/528,391	03 July 2017 (03.07.2017)	US
62/534,016	18 July 2017 (18.07.2017)	US
62/561,638	21 September 2017 (21.09.2017)	US
62/568,304	04 October 2017 (04.10.2017)	US
62/574,158	18 October 2017 (18.10.2017)	US
62/591,187	27 November 2017 (27.11.2017)	US
62/610,105	22 December 2017 (22.12.2017)	US

(71) Applicants: **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, Massachusetts 02142 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US). **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, Massachusetts 02138 (US).

(72) Inventors: **ZHANG, Feng**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **GOOTENBERG, Jonathan**; c/o 17 Quincy Street, Cambridge, Massachusetts 02138 (US). **COX, David Benjamin Turitz**; c/o 77 Massachusetts Ave., Cambridge, Massachusetts 02139 (US). **ABUDAYYEH, Omar**; c/o 77 Massachusetts Ave, Cambridge, Massachusetts 02139 (US). **KANNAN, Soumya**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US).

(74) Agent: **NIX, F. Brent** et al.; Johnson, Marcou & Isaacs, LLC, P.O. Box 691, Hoschton, Georgia 30548 (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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH,

(54) Title: CRISPR/CAS-ADENINE DEAMINASE BASED COMPOSITIONS, SYSTEMS, AND METHODS FOR TARGETED NUCLEIC ACID EDITING

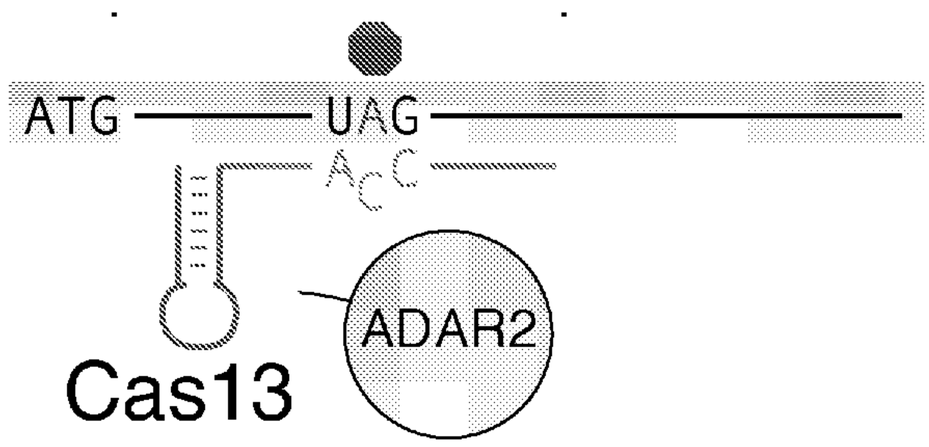


FIG. 1

(57) Abstract: The invention provides for systems, methods, and compositions for targeting and editing nucleic acids. In particular, the invention provides non-naturally occurring or engineered RNA-targeting systems comprising a RNA-targeting Cas13 protein, at least one guide molecule, and at least one adenosine deaminase protein or catalytic domain thereof.

[Continued on next page]

WO 2019/005884 A1

WO 2019/005884 A1

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

Published:

— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME	1	DE	3
CONTENANT LES PAGES	1	À	246

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME	1	OF	3
CONTAINING PAGES	1	TO	246

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

CRISPR/CAS-ADENINE DEAMINASE BASED COMPOSITIONS, SYSTEMS, AND METHODS FOR TARGETED NUCLEIC ACID EDITING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/525,181, filed June 26, 2017, U.S. Provisional Application No. 62/528,391, filed July 3, 2017, U.S. Provisional Application No. 62/534,016, filed July 18, 2017, U.S. Provisional Application No. 62/561,638, filed September 21, 2017, U.S. Provisional Application No. 62/568,304, filed October 4, 2017, U.S. Provisional Application No. 62/574,158, filed October 18, 2017, U.S. Provisional Application No. 62/591,187, filed November 27, 2017, and U.S. Provisional Application No. 62/610,105, filed December 22, 2017. The entire contents of the above-identified applications are hereby fully incorporated herein by reference.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under grant numbers MH100706, MH110049, and HL141201 awarded by the National Institutes of Health. The government has certain rights in the invention

REFERENCE TO DOCUMENTS CO-FILED IN COMPUTER READABLE FORMAT

[0003] An ASCII compliant text file entitled "Clin_var_pathogenic_SNPS_TC.txt" created on July 3, 2017 and 891,043 bytes in size is filed herewith via EFS-WEB, the contents of which are hereby incorporated herein by reference.

FIELD OF THE INVENTION

[0004] The present invention generally relates to systems, methods, and compositions for targeting and editing nucleic acids, in particular for programmable deamination of adenine at a target locus of interest.

BACKGROUND

[0005] Recent advances in genome sequencing techniques and analysis methods have significantly accelerated the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. Precise genome targeting technologies are needed

to enable systematic reverse engineering of causal genetic variations by allowing selective perturbation of individual genetic elements, as well as to advance synthetic biology, biotechnological, and medical applications. Although genome-editing techniques such as designer zinc fingers, transcription activator-like effectors (TALEs), or homing meganucleases are available for producing targeted genome perturbations, there remains a need for new genome engineering technologies that employ novel strategies and molecular mechanisms and are affordable, easy to set up, scalable, and amenable to targeting multiple positions within the eukaryotic genome. This would provide a major resource for new applications in genome engineering and biotechnology.

[0006] Programmable deamination of cytosine has been reported and may be used for correction of A→G and T→C point mutations. For example, Komor et al., Nature (2016) 533:420–424 reports targeted deamination of cytosine by APOBEC1 cytidine deaminase in a non-targeted DNA stranded displaced by the binding of a Cas9-guide RNA complex to a targeted DNA strand, which results in conversion of cytosine to uracil. See also Kim et al., Nature Biotechnology (2017) 35:371-376; Shimatani et al., Nature Biotechnology (2017) doi:10.1038/nbt.3833; Zong et al., Nature Biotechnology (2017) doi:10.1038/nbt.3811; Yang Nature Communication (2016) doi:10.1038/ncomms13330.

SUMMARY OF THE INVENTION

[0007] The present application relates to modifying a target RNA sequence of interest. Using RNA-targeting rather than DNA targeting offers several advantages relevant for therapeutic development. First, there are substantial safety benefits to targeting RNA: there will be fewer off-target events because the available sequence space in the transcriptome is significantly smaller than the genome, and if an off-target event does occur, it will be transient and less likely to induce negative side effects. Second, RNA-targeting therapeutics will be more efficient because they are cell-type independent and not have to enter the nucleus, making them easier to deliver.

[0008] At least a first aspect of the invention relates to a method of modifying an Adenine in a target RNA sequence of interest. In particular embodiments, the method comprises delivering to said target RNA: (a) a catalytically inactive (dead) Cas13 protein; (b) a guide molecule which comprises a guide sequence linked to a direct repeat sequence; and (c) an adenosine deaminase protein or catalytic domain thereof; wherein said adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to said dead Cas13

protein or said guide molecule or is adapted to link thereto after delivery; wherein guide molecule forms a complex with said dead Cas13 protein and directs said complex to bind said target RNA sequence of interest, wherein said guide sequence is capable of hybridizing with a target sequence comprising said Adenine to form an RNA duplex, wherein said guide sequence comprises a non-pairing Cytosine at a position corresponding to said Adenine resulting in an A-C mismatch in the RNA duplex formed; wherein said adenosine deaminase protein or catalytic domain thereof deaminates said Adenine in said RNA duplex.

[0009] In certain example embodiment the Cas13 protein is Cas13a, Cas13b or Cas 13c.

[0010] The adenosine deaminase protein or catalytic domain thereof is fused to N- or C-terminus of said dead Cas13 protein. In certain example embodiments, the adenosine deaminase protein or catalytic domain thereof is fused to said dead Cas13 protein by a linker. The linker may be $(GGGGS)_{3-11}$ (SEQ ID Nos. 1-9) GSG_5 (SEQ ID No. 10) or LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID No. 11).

[0011] In certain example embodiments, the adenosine deaminase protein or catalytic domain thereof is linked to an adaptor protein and said guide molecule or said dead Cas13 protein comprises an aptamer sequence capable of binding to said adaptor protein. The adaptor sequence may be selected from MS2, PP7, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1.

[0012] In certain example embodiments, the adenosine deaminase protein or catalytic domain thereof is inserted into an internal loop of said dead Cas13 protein. In certain example embodiments, the Cas13a protein comprises one or more mutations in the two HEPN domains, particularly at postion R474 and R1046 of Cas 13a protein originating from *Leptotrichia wadei* or amino acid positions corresponding thereto of a Cas13a ortholog.

[0013] In certain example embodiments, the Cas 13 protein is a Cas13b proteins, and the Cas13b comprises a mutation in one or more of positions R116, H121, R1177, H1182 of Cas13b protein originating from *Bergeyella zoohelcum* ATCC 43767 or amino acid positions corresponding thereto of a Cas13b ortholog. In certain other example embodiments, the mutation is one or more of R116A, H121A, R1177A, H1182A of Cas13b protein originating from *Bergeyella zoohelcum* ATCC 43767 or amino acid positions corresponding thereto of a Cas13b ortholog.

[0014] In certain example embodiments, the guide sequence has a length of about 29-53 nt capable of forming said RNA duplex with said target sequence. In certain other example embodiments, the guide sequence has a length of about 40-50 nt capable of forming said RNA

duplex with said target sequence. In certain example embodiments, the distance between said non-pairing C and the 5' end of said guide sequence is 20-30 nucleotides.

[0015] In certain example embodiments, the adenosine deaminase protein or catalytic domain thereof is a human, cephalopod, or Drosophila adenosine deaminase protein or catalytic domain thereof. In certain example embodiments, the adenosine deaminase protein or catalytic domain thereof has been modified to comprise a mutation at glutamic acid^{488} of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In certain example embodiments, the glutamic acid residue may be at position 488 or a corresponding position in a homologous ADAR protein is replaced by a glutamine residue (E488Q).

[0016] In certain other example embodiments, the adenosine deaminase protein or catalytic domain thereof is a mutated hADAR2d comprising mutation E488Q or a mutated hADAR1d comprising mutation E1008Q.

[0017] In certain example embodiments, the guide sequence comprises more than one mismatch corresponding to different adenosine sites in the target RNA sequence or wherein two guide molecules are used, each comprising a mismatch corresponding to a different adenosine sites in the target RNA sequence.

[0018] In certain example embodiments, the Cas13 protein and optionally said adenosine deaminase protein or catalytic domain thereof comprise one or more heterologous nuclear localization signal(s) (NLS(s)).

[0019] In certain example embodiments, the method further comprises, determining the target sequence of interest and selecting an adenosine deaminase protein or catalytic domain thereof which most efficiently deaminates said Adenine present in then target sequence.

[0020] The target RNA sequence of interest may be within a cell. The cell may be a eukaryotic cell, a non-human animal cell, a human cell, a plant cell. The target locus of interest may be within an animal or plant.

[0021] The target RNA sequence of interest may comprise in an RNA polynucleotide *in vitro*.

[0022] The components of the systems described herein may be delivered to said cell as a ribonucleoprotein complex or as one or more polynucleotide molecules. The one or more polynucleotide molecules may comprise one or more mRNA molecules encoding the components. The one or more polynucleotide molecules may be comprised within one or more vectors. The one or more polynucleotide molecules may further comprise one or more regulatory elements operably configured to express said Cas13 protein, said guide molecule, and said adenosine deaminase protein or catalytic domain thereof, optionally wherein said one

or more regulatory elements comprise inducible promoters. The one or more polynucleotide molecules or said ribonucleoprotein complex may be delivered via particles, vesicles, or one or more viral vectors. The particles may comprise a lipid, a sugar, a metal or a protein. The particles may comprise lipid nanoparticles. The vesicles may comprise exosomes or liposomes. The one or more viral vectors may comprise one or more of adenovirus, one or more lentivirus or one or more adeno-associated virus.

[0023] The methods disclosed herein may be used to modify a cell, a cell line or an organism by manipulation of one or more target RNA sequences.

[0024] In certain example embodiments, the deamination of said Adenine in said target RNA of interest remedies a disease caused by transcripts containing a pathogenic G→A or C→T point mutation.

[0025] The methods maybe be used to treat a disase. In certain example embodiments, the disease is selected from Meier-Gorlin syndrome, Seckel syndrome 4, Joubert syndrome 5, Leber congenital amaurosis 10; Charcot-Marie-Tooth disease, type 2; Charcot-Marie-Tooth disease, type 2; Usher syndrome, type 2C; Spinocerebellar ataxia 28; Spinocerebellar ataxia 28; Spinocerebellar ataxia 28; Long QT syndrome 2; Sjögren-Larsson syndrome; Hereditary fructosuria; Hereditary fructosuria; Neuroblastoma; Neuroblastoma; Kallmann syndrome 1; Kallmann syndrome 1; Kallmann syndrome 1; Metachromatic leukodystrophy, Rett syndrome, Amyotrophic lateral sclerosis type 10, Li-Fraumeni syndrome, or a disease listed in Table 5. The disease may be a premature termination disease.

[0026] The methods disclosed herein, may be used to make a modification that affects the fertility of an organism. The modification may affects splicing of said target RNA sequence. The modification mayintroduces a mutation in a transcript introducing an amino acid change and causing expression of a new antigen in a cancer cell.

[0027] In certain example embodiments, the target RNA may be a microRNA or comprised within a microRNA. In certain example embodiments, the deamination of said Adenine in said target RNA of interest causes a gain of function or a loss of function of a gene.In certain example embodiments, the gene is a gene expressed by a cancer cell.

[0028] In another aspect, the invention comprises a modified cell or progeny thereof that is obtained using the methods disclosed herein, wherein said cell comprises a hypoxanthine or a guanine in replace of said Adenine in said target RNA of interest compared to a corresponding cell not subjected to the method. The modified cell or progeny thereof may be a eukaryotic cell an animal cell, a human cell, a therapeutic T cell, an antibody-producing B cell, a plant cell.

[0029] In another aspect, the invention comprises a non-human animal comprising said modified cell or progeny therof. The modified may be a plant cell.

[0030] In another aspect, the invention comprises a method for cell therapy, comprising administering to a patient in need thereof the modified cells disclosed herein, wherein the presence of said modified cell remedies a disease in the patient.

[0031] In another aspect, the invention is directed to an engineered, non-naturally occurring system suitable for modifying an Adenine in a target locus of interest, comprising A) a guide molecule which comprises a guide sequence linked to a direct repeat sequence, or a nucleotide sequence encoding said guide molecule; B) a catalytically inactive Cas13 protein, or a nucleotide sequence encoding said catalytically inactive Cas13 protein; C) an adenosine deaminase protein or catalytic domain thereof, or a nucleotide sequence encoding said adenosine deaminase protein or catalytic domain thereof; wherein said adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to said Cas13 protein or said guide molecule or is adapted to link thereto after delivery; wherein said guide sequence is capable of hybridizing with a target RNA sequence comprising an Adenine to form an RNA duplex, wherein said guide sequence comprises a non-pairing Cytosine at a position corresponding to said Adenine resulting in an A-C mismatch in the RNA duplex formed.

[0032] In another aspect, the invention is directed to an engineered, non-naturally occurring vector system suitable for modifying an Adenine in a target locus of interest, comprising the nucleotide sequences of a), b) and c)

[0033] In another aspect, the invention is directed to an engineered, non-naturally occurring vector system, comprising one or more vectors comprising: a first regulatory element operably linked to a nucleotide sequence encoding said guide molecule which comprises said guide sequence, a second regulatory element operably linked to a nucleotide sequence encoding said catalytically inactive Cas13 protein; and a nucleotide sequence encoding an adenosine deaminase protein or catalytic domain thereof which is under control of said first or second regulatory element or operably linked to a third regulatory element; wherein, if said nucleotide sequence encoding an adenosine deaminase protein or catalytic domain thereof is operably linked to a third regulatory element, said adenosine deaminase protein or catalytic domain thereof is adapted to link to said guide molecule or said Cas13 protein after expression; wherein components A), B) and C) are located on the same or different vectors of the system.

[0034] As the methods disclosed herein demonstate the ability of Cas13 proteins to function in mammalian cells for binding and specificity of cleaving RNA, additional extended

applications include editing splice variants, and measuring how RNA-binding proteins interact with RNA.

[0035] In another aspect, the invention is directed to *in vitro* or *ex vivo* host cell or progeny thereof or cell line or progeny thereof comprising the systems disclosed herein. The host cell or progeny thereof may be a a eukaryotice cell, an animal cell, a human cell, or a plant cell.

[0036] In another aspect, the invention relates to an adenosine deaminase protein or catalytic domain thereof and comprising one or more mutations as described herein elsewhere.

[0037] In certain embodiments, such adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to a nucleic acid binding molecule or targeting domain as described herein elsewhere. Accordingly, the invention further relates to compositions comprising said adenosine deaminase protein or catalytic domain and a nucleic acid binding molecule and to fusion proteins of said adenosine deaminase protein or catalytic domain and said nucleic acid binding molecule.

[0038] In another aspect the invention relates to an engineered composition for site directed base editing comprising a targeting domain and an adenosine deaminase, or catalytic domain thereof. In particular embodiments, the targeting domain is an oligonucleotide targeting domain. In particular embodiments, the adenosine deaminase, or catalytic domain thereof, comprises one or more mutations that increase activity or specificity of the adenosine deaminase relative to wild type. In particular embodiments, the adenosine deaminase comprises one or more mutations that changes the functionality of the adenosine deaminase relative to wild type, preferably an ability of the adenosine deaminase to deaminate cytodine as described elsewhere herein. In particular embodiments, the targeting domain is a CRISPR system comprising a CRISPR effector protein, or functional domain thereof, and a guide molecule, more particularly the CRISPR system is catalytically inactive. In particular embodiments, the CRISPR system comprises an RNA-binding protein, preferably Cas13, preferably the Cas13 protein is Cas13a, Cas13b or Cas13c, preferably wherein said Cas13 a Cas13 listed in any of Tables 1, 2, 3, 4, or 6 or is from a bacterial species listed in any of Tables 1, 2, 3, 4, or 6, preferably wherein said Cas13 protein is Prevotella sp.P5-125 Cas13b, Porphyromas gulae Cas13b, or Riemerella anatipestifer Cas13b; preferably Prevotella sp.P5-125 Cas13b. In particular embodiments, the Cas13 protein is a Cas13a protein and said Cas13a comprises one or more mutations the two HEPN domains, particularly at position R474 and R1046 of Cas13a protein originating from *Leptotrichia wadei* or amino acid positions corresponding thereto of a Cas13a ortholog, or wherein said Cas13 protein is a Cas13b protein and said Cas13b comprises a mutation in one or more of positions R116, H121, R1177, H1182, preferably

R116A, H121A, R1177A, H1182A of Cas13b protein originating from *Bergeyella zoohelcum* ATCC 43767 or amino acid positions corresponding thereto of a Cas13b ortholog, or wherein said Cas13 protein is a Cas13b protein and said Cas13b comprises a mutation in one or more of positions R128, H133, R1053, H1058, preferably H133 and H1058, preferably H133A and H1058A, of a Cas13b protein originating from Prevotella sp. P5-125 or amino acid positions corresponding thereto of a Cas13b ortholog as described elsewhere herein or the Cas 13 is truncated, preferably C-terminally truncated, preferably wherein said Cas13 is a truncated functional variant of the corresponding wild type Cas13, optionally wherein said truncated Cas13b is encoded by nt 1-984 of Prevotella sp.P5-125 Cas13b or the corresponding nt of a Cas13b orthologue or homologue.

[0039] In particular embodiments, the guide molecule of the targeting domain comprises a guide sequence is capable of hybridizing with a target RNA sequence comprising an Adenine to form an RNA duplex, wherein said guide sequence comprises a non-pairing Cytosine at a position corresponding to said Adenine resulting in an A-C mismatch in the RNA duplex formed. In particular embodiments, the guide sequence has a length of about 20-53 nt, preferably 25-53 nt, more preferably 29-53 nt or 40-50 nt capable of forming said RNA duplex with said target sequence, and/or wherein the distance between said non-pairing C and the 5' end of said guide sequence is 20-30 nucleotides. In particular embodiments, the guide sequence comprises more than one mismatch corresponding to different adenosine sites in the target RNA sequence or wherein two guide molecules are used, each comprising a mismatch corresponding to a different adenosine sites in the target RNA sequence.

[0040] In particular embodiments, of the composition the adenosine deaminase protein or catalytic domain thereof is fused to a N- or C-terminus of said oligonucleotide targeting protein, optionally by a linker as described elsewhere herein. Alternatively, said adenosine deaminase protein or catalytic domain thereof is inserted into an internal loop of said dead Cas13 protein. In a further alternative embodiment, the adenosine deaminase protein or catalytic domain thereof is linked to an adaptor protein and said guide molecule or said dead Cas13 protein comprises an aptamer sequence capable of binding to said adaptor protein as described elsewhere herein.

[0041] In particular embodiments of the composition the adenosine deaminase protein or catalytic domain thereof capable of deaminating adenosine or cytodine in RNA or is an RNA specific adenosine deaminase and/or is a bacterial, human, cephalopod, or Drosophila adenosine deaminase protein or catalytic domain thereof, preferably TadA, more preferably

ADAR, optionally huADAR, optionally (hu)ADAR1 or (hu)ADAR2, preferably huADAR2 or catalytic domain thereof.

[0042] In particular embodiments of the composition, the targeting domain and optionally the adenosine protein or catalytic domain thereof comprise one or more heterologous nuclear export signal(s) (NES(s)) or nuclear localization signal(s) (NLS(s)), preferably an HIV Rev NES or MAPK NES, preferably C-terminal.

[0043] A further aspect of the invention relates to the composition as envisaged herein for use in prophylactic or therapeutic treatment, preferably wherein said target locus of interest is within a human or animal and to methods of modifying an Adenine or Cytidine in a target RNA sequence of interest, comprising delivering to said target RNA, the composition as described hereinabove. In particular embodiments, the CRISPR system and the adenonsine deaminase, or catalytic domain thereof, are delivered as one or more polynucleotide molecules, as a ribonucleoprotein complex, optionally via particles, vesicles, or one or more viral vectors. In particular embodiments, the composition is for use in the treatment or prevention of a disease caused by transcripts containing a pathogenic G→A or C→T point mutation. In particular embodiments, the invention thus comprises compositions for use in therapy. This implies that the methods can be performed in vivo, ex vivo or in vitro. In particular embodiments, the methods are not methods of treatment of the animal or human body or a method for modifying the germ line genetic identity of a human cell. In particular embodiments; when carrying out the method, the target RNa is not comprised within a human or animal cell. In particular embodiments, when the target is a human or animal target, the method is carried out ex vivo or in vitro

[0044] A further aspects relates to an isolated cell obtained or obtainable from the methods described above and/or comprising the composition described above or progeny of said modified cell, preferably wherein said cell comprises a hypoxanthine or a guanine in replace of said Adenine in said target RNA of interest compared to a corresponding cell not subjected to the method. In particular embodiments, the cell is a eukaryotic cell, preferably a human or non-human animal cell, optionally a therapeutic T cell or an antibody-producing B-cell or wherein said cell is a plant cell. A further aspect provides a non-human animal or a plant comprising said modified cell or progeny thereof . Yet a further aspect provides the modified cell as described hereinabove for use in therapy, preferably cell therapy.

BRIEF DESCRIPTION OF THE DRAWINGS

[0045] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0046] **FIG. 1** illustrates an example embodiment of the invention for targeted deamination of adenine at a target RNA sequence of interest, exemplified herein with a Cas13b protein.

[0047] **FIG. 2** illustrates the Development of RNA editing as a therapeutic strategy to treat human disease at the transcript level such as when using Cas13b. Schematic of RNA base editing by Cas13-ADAR2 fusion targeting an engineered pre-termination stop codon in the luciferase transcript.

[0048] **FIG. 3** Guide position and length optimization to restore luciferase expression.

[0049] **FIG. 4** Exemplary sequences of adenine deaminase proteins. (SEQ ID Nos. 650 - 656)

[0050] **FIG. 5** Guides used in an exemplary emodiment (SEQ ID Nos. 657 – 660 and 703)

[0051] **FIG. 6** : Editing efficiency correlates to edited base being further away from the DR and having a long RNA duplex, which is accomplished by extending the guide length

[0052] **FIG. 7** Greater editing efficiency the further the editing site is away from the DR/protein binding area.

[0053] **FIG. 8** Distance of edited site from DR

[0054] **FIG. 9A and B**: Fused ADAR1 or ADAR2 to Cas13b12 (double R HEPN mutant) on the N or C-terminus. Guides are perfect matches to the stop codon in luciferase. Signal appears correlated with distance between edited base and 5' end of the guide, with shorter distances providing better editing.

[0055] **FIG. 10**: Cluc/Gluc tiling for Cas13a/Cas13b interference

[0056] **FIG. 11**: ADAR editing quantification by NGS (luciferase reporter).

[0057] **FIG. 12**: ADAR editing quantification by NGS (KRAS and PPIB).

[0058] **FIG. 13**: Cas13a/b + shRNA specificity from RNA Seq

[0059] **FIG. 14**: Mismatch specificity to reduce off targets (A:A or A:G) (SEQ ID Nos. 661 - 668)

[0060] **FIG. 15**: Mismatch for on-target activity

[0061] **FIG. 16**: ADAR Motif preference

[0062] **FIG. 17**: Larger bubbles to enhance RNA editing efficiency

[0063] **FIG. 18**: Editing of multiple A's in a transcript (SEQ ID Nos. 669-672)

[0064] **FIG. 19**: Guide length titration for RNA editing

[0065] **FIG. 20**: Mammalian codon-optimized Cas13b orthologs mediate highly efficient RNA knockdown. (A) Schematic of representative Cas13a, Cas13b, and Cas13c loci and associated crRNAs. (B) Schematic of luciferase assay to measure Cas13a cleavage activity in HEK293FT cells. (C) RNA knockdown efficiency using two different guides targeting Cluc with 19 Cas13a, 15 Cas13b, and 5 Cas13c orthologs. Luciferase expression is normalized to the expression in non-targeting guide control conditions. (D) The top 7 orthologs performing in part C are assayed for activity with three different NLS and NES tags with two different guide RNAs targeting Cluc. (E) Cas13b12 and Cas13a2 (LwCas13a) are compared for knockdown activity against Gluc and Cluc. Guides are tiled along the transcripts and guides between Cas13b12 and Cas13a2 are position matched. (F) Guide knockdown for Cas13a2, Cas13b6, Cas13b11, and Cas13b12 against the endogenous KRAS transcript and are compared against corresponding shRNAs.

[0066] **FIG. 21**: Cas13 enzymes mediate specific RNA knockdown in mammalian cells. (A) Schematic of semi-degenerate target sequences for Cas13a/b mismatch specificity testing. (SEQ ID Nos. 673-694) (B) Heatmap of single mismatch knockdown data for Cas13 a/b. Knockdown is normalized to non-targeting (NT) guides for each enzyme. (C) Double mismatch knockdown data for Cas13a. The position of each mismatch is indicated on the X and Y axes. Knockdown data is the sum of all double mismatches for a given set of positions. Data is normalized to NT guides for each enzyme. (D) Double mismatch knockdown data for Cas13b. See C for description. (E) RNA-seq data comparing transcriptome-wide specificity for Cas13 a/b and shRNA for position-matched guides. The Y axis represents read counts for the targeting condition and the X axis represents counts for the non-targeting condition. (F) RNA expression as calculated from RNA-seq data for Cas13 a/b and shRNA. (G) Significant off-targets for Cas13 a/b and shRNA from RNA-seq data. Significant off-targets were calculated using FDR <0.05.

[0067] **FIG. 22**: Catalytically inactive Cas13b-ADAR fusions enable targeted RNA editing in mammalian cells. (A) Schematic of RNA editing with Cas13b-ADAR fusion proteins to remove stop codons on the Cypridina luciferase transcript. (B) RNA editing comparison between Cas13b fused with wild-type ADAR2 and Cas13b fused with the hyperactive ADAR2 E488Q mutant for multiple guide positions. Luciferase expression is normalized to Gaussia luciferase control values. (C) RNA editing comparisons between 30, 50, 70, and 84 nt guides designed to target various positions surrounding the editing site. (D) Effect of surrounding

motif sequence on ADAR editing efficiency on the Cypridina luciferase transcript. (SEQ ID No. 695) (E) Schematic showing the position and length of guides used for sequencing quantification relative to the stop codon on the Cypridina luciferase transcript. (F) On- and off-target editing efficiencies for each guide design at the corresponding adenine bases on the Cypridina luciferase transcript as quantified by sequencing. (G) Luciferase readout of guides with varied bases opposite to the targeted adenine.

[0068] **FIG. 23**: Endogenous RNA editing with Cas13b-ADAR fusions. (A) Next generation sequencing of endogenous Cas13b12-ADAR editing of endogenous KRAS and PPIB loci. Two different regions per transcript were targeted and A->G editing was quantified at all adenines in the vicinity of the targeted adenine.

[0069] **FIG. 24**: Strategy for determining optimal guide position.

[0070] **FIG. 25**: (A) Cas13b-huADAR2 promotes repair of mutated luciferase transcripts. (B) Cas13b-huADAR1 promotes repair of mutated luciferase transcripts. (C) Comparison of human ADAR1 and human ADAR2.

[0071] **FIG. 26**: Comparison of E488Q vs. wt dADAR2 editing. E488Q is a hyperactive mutant of dADAR2.

[0072] **FIG. 27**: Transcripts targeted by Cas13b-huADAR2-E488Q contain the expected A-G edit. (A) heatmap. (B) Positions in template. Only A sites are shown with the editing rate to G as in heatmap.

[0073] **FIG. 28**: Endogenous tiling of guides. (A) KRAS: heatmap. Only A sites are shown with the editing rate to G as in heatmap. (B) Positions in template (bottom). (C) PPIB: heatmap. Only A sites are shown with the editing rate to G as in heatmap. Positions in template (D).

[0074] **FIG. 29**: Non-targeting editing.

[0075] **FIG. 30**: Linker optimization.

[0076] **FIG. 31**: Cas13b ADAR can be used to correct pathogenic A>G mutations from patients in expressed cDNAs.

[0077] **FIG. 32**: Cas13b-ADAR has a slight restriction on 5' G motifs.

[0078] **FIG. 33**: Screening degenerate PFS locations for effect on editing efficiency. All PFS (4-N) identities have higher editing than non-targeting. Fig A. (SEQ ID Nos. 696 - 699)

[0079] **FIG. 34**: Reducing off-target editing in the target transcript.

[0080] **FIG. 35**: Reducing off-target editing in the target transcript.

[0081] **FIG. 36**: Cas13b-ADAR transcriptome specificity. On-target editing is 71%. (A) targeting guide; 482 significant sites. (B) non-targeting guide; 949 significant sites. Note that

chromosome 0 is Gluc and chromosome 1 is Cluc; human chromosomes are then in order after that.

[0082] **FIG. 37**: Cas13b-ADAR transcriptome specificity. (A) targeting guide. (B) non-targeting guide.

[0083] **FIG. 38**: Cas13b has the highest efficiency compared to competing ADAR editing strategies.

[0084] **FIG. 39**: Competing RNA editing systems. (A-B) BoxB; on-target editing is 63%; (A) targeting guide – 2020 significant sites; (B) non-targeting guide – 1805 significant sites. (C-D) Stafforst; on-target editing is 36%; (C) targeting guide – 176 significant sites; (D) non-targeting guide – 186 significant sites.

[0085] **FIG. 40**: Dose titration of ADAR. crRNA amount is constant.

[0086] **FIG. 41**: Dose response effect on specificity. (A-B) 150 ng Cas13-ADAR; on-target editing is 83%; (A) targeting guide – 1231 significant sites; (B) non-targeting guide – 520 significant sites. (C-D) 10 ng Cas13-ADAR; on-target editing is 80%; (C) targeting guide – 347 significant sites; (D) non-targeting guide – 223 significant sites.

[0087] **FIG. 42**: ADAR1 seems more specific than ADAR2. On-target editing is 29%. (A) targeting guide; 11 significant sites. (B) non-targeting guide; 6 significant sites. Note that chromosome 0 is Gluc and chromosome 1 is Cluc; human chromosomes are then in order after that.

[0088] **FIG. 43**: ADAR specificity mutants have enhanced specificity. (A) Targeting guide. (B) Non-targeting guide. (C) Targeting to non-targeting ratio. (D) Targeting and non-targeting guide.

[0089] **FIG. 44**: ADAR mutant luciferase results plotted along the contact points of each residue with the RNA target.

[0090] **FIG. 45**: ADAR specificity mutants have enhanced specificity. Purple points are mutants selected for whole transcriptome off-target NGS analysis. Red point is the starting point (i.e. E488Q mutant). Note that all additional mutants also have the E488Q mutation.

[0091] **FIG. 46**: ADAR mutants are more specific according to NGS. (A) on target. (B) Off-target.

[0092] **FIG. 47**: Luciferase data on ADAR specificity mutants matches the NGS. (A) Targeting guide selected for NGS. (B) Non-targeting guide selected for NGS. Luciferase data matches the NGS data in FIG.46. The orthologs that have fewer activity with non-targeting guide have fewer off-targets across the transcriptome and their on-target editing efficiency can be predicted by the targeting guide luciferase condition.

[0093] **FIG. 48**: C-terminal truncations of Cas13b 12 are still highly active in ADAR editing.

[0094] **FIG. 49**: Characterization of a highly active Cas13b ortholog for RNA knockdown A) Schematic of stereotypical Cas13 loci and corresponding crRNA structure. B) Evaluation of 19 Cas13a, 15 Cas13b, and 7 Cas13c orthologs for luciferase knockdown using two different guides. Orthologs with efficient knockdown using both guides are labeled with their host organism name. Values are normalized to a non-targeting guide with designed against the *E. coli LacZ* transcript, with no homology to the human transcriptome.C) PspCas13b and LwaCas13a knockdown activity are compared by tiling guides against Gluc and measuring luciferase expression. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 49B. D) PspCas13b and LwaCas13a knockdown activity are compared by tiling guides against Cluc and measuring luciferase expression. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 49B. E) Expression levels in log2(transcripts per million (TPM)) values of all genes detected in RNA-seq libraries of non-targeting control (x-axis) compared to Gluc-targeting condition (y-axis) for LwaCas13a (red) and shRNA (black). Shown is the mean of three biological replicates. The Gluc transcript data point is labeled. Non-targeting guide is the same as in Fig. 49B. F) Expression levels in log2(transcripts per million (TPM)) values of all genes detected in RNA-seq libraries of non-targeting control (x-axis) compared to Gluc-targeting condition (y-axis) for PspCas13b (blue) and shRNA (black). Shown is the mean of three biological replicates. The Gluc transcript data point is labeled. Non-targeting guide is the same as in Fig. 49B. G) Number of significant off-targets from Gluc knockdown for LwaCas13a, PspCas13b, and shRNA from the transcriptome wide analysis in E and F.

[0095] **FIG. 50**: Engineering dCas13b-ADAR fusions for RNA editing A) Schematic of RNA editing by dCas13b-ADAR fusion proteins. Catalytically dead Cas13b (dCas13b) is fused to the deaminase domain of human ADAR ($ADAR_{DD}$), which naturally deaminates adenosines to insosines in dsRNA. The crRNA specifies the target site by hybridizing to the bases surrounding the target adenosine, creating a dsRNA structure for editing, and recruiting the dCas13b-$ADAR_{DD}$ fusion. A mismatched cytidine in the crRNA opposite the target adenosine enhances the editing reaction, promoting target adenosine deamination to inosine, a base that functionally mimics guanosine in many cellular reactions. B) Schematic of *Cypridina* luciferase W85X target and targeting guide design. (SEQ ID Nos. 700 and 701) Deamination of the target adenosine restores the stop codon to the wildtype tryptophan. Spacer length is the region of the guide that contains homology to the target sequence. Mismatch distance is the

number of bases between the 3' end of the spacer and the mismatched cytidine. The cytidine mismatched base is included as part of the mismatch distance calculation. C) Quantification of luciferase activity restoration for Cas13b-dADAR1 (left) and Cas13b-ADAR2-cd (right) with tiling guides of length 30, 50, 70, or 84 nt. All guides with even mismatch distances are tested for each guide length. Values are background subtracted relative to a 30nt non-targeting guide that is randomized with no sequence homology to the human transcriptome. D) Schematic of target site for targeting *Cypridinia* luciferase W85X. (SEQ ID No. 702) E) Sequencing quantification of A->I editing for 50 nt guides targeting *Cypridinia* luciferase W85X. Blue triangle indicates the targeted adenosine. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C.

[0096] **FIG. 51**: Measuring sequence flexibility for RNA editing by REPAIRv1 Schematic of screen for determining Protospacer Flanking Site (PFS) preferences of RNA editing by REPAIRv1. A randomized PFS sequence is cloned 5' to a target site for REPAIR editing. Following exposure to REPAIR, deep sequencing of reverse transcribed RNA from the target site and PFS is used to associate edited reads with PFS sequences. B) Distributions of RNA editing efficiencies for all 4-N PFS combinations at two different editing sites. C) Quantification of the percent editing of REPAIRv1 at Cluc W85 across all possible 3 base motifs. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C. D) Heatmap of 5' and 3' base preferences of RNA editing at Cluc W85 for all possible 3 base motifs

[0097] **FIG. 52**: Correction of disease-relevant mutations with REPAIRv1 A) Schematic of target and guide design for targeting *AVPR2* 878G>A. (SEQ ID Nos. 705-708) B) The 878G>A mutation in *AVPR2* is corrected to varying percentages using REPAIRv1 with three different guide designs. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C. C) Schematic of target and guide design for targeting *FANCC* 1517G>A. (SEQ ID Nos. 709-712) D) The 1517G>A mutation in *FANCC* is corrected to varying percentages using REPAIRv1 with three different guide designs. For each guide, the region of duplex RNA is outlined in red. The heatmap scale bar is the same as in panel B. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C. E) Quantification of the percent editing of 34 different disease-relevant G>A mutations using REPAIRv1. Non-targeting guide is the same as in Fig. 50C. F) Analysis of all the possible G>A mutations that could be corrected as annotated by the ClinVar database. The distribution of editing motifs for all G>A mutations in ClinVar is shown versus the editing efficiency by REPAIRv1 per motif as quantified on the Gluc transcript. G) The

distribution of editing motifs for all G>A mutations in ClinVar is shown versus the editing efficiency by REPAIRv1 per motif as quantified on the *Gluc* transcript. Values represent mean +/– S.E.M.

[0098] **FIG. 53**: Characterizing specificity of REPAIRv1 A) Schematic of *KRAS* target site and guide design. (SEQ ID Nos. 713-720) B) Quantification of percent editing for tiled *KRAS*-targeting guides. Editing percentages are shown at the on-target and neighboring adenosine sites. For each guide, the region of duplex RNA is indicated by a red rectangle. Values represent mean +/– S.E.M. C) Transcriptome-wide sites of significant RNA editing by REPAIRv1 with Cluc targeting guide. The on-target site Cluc site (254 A>G) is highlighted in orange. D) Transcriptome-wide sites of significant RNA editing by REPAIRv1 (150ng REPAIR vector transfected) with non-targeting guide. Non-targeting guide is the same as in Fig. 50C.

[0099] **FIG. 54**: Rational mutagenesis of ADAR2 to improve the specificity of REPAIRv1 A) Quantification of luciferase signal restoration by various dCas13-ADAR2 mutants as well as their specificity score plotted along a schematic for the contacts between key ADAR2 deaminase residues and the dsRNA target. All deaminase mutations were made on the dCas13-$ADAR2_{DD}$(E488Q) background. The specificity score is defined as the ratio of the luciferase signal between targeting guide and non-targeting guide conditions. Schematic of ADAR2 deaminase domain contacts with dsRNA is adapted from ref (*20*) B) Quantification of luciferase signal restoration by various dCas13-ADAR2 mutants versus their specificity score. Non-targeting guide is the same as in Fig. 50C. C) Measurement of the on-target editing fraction as well as the number of significant off-targets for each dCas13-ADAR2 mutant by transcriptome wide sequencing of mRNAs. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C. D) Transcriptome-wide sites of significant RNA editing by REPAIRv1 and REPAIRv2 with a guide targeting a pretermination site in Cluc. The on-target Cluc site (254 A>G) is highlighted in orange. 10 ng of REPAIR vector was transfected for each condition. E) RNA sequencing reads surrounding the on-target Cluc editing site (SEQ ID No. 721) (254 A>G) highlighting the differences in off-target editing between REPAIRv1 and REPAIRv2. All A>G edits are highlighted in red while sequencing errors are highlighted in blue. Gaps reflect spaces between aligned reads. Non-targeting guide is the same as in Fig. 50C. F) RNA editing by REPAIRv1 and REPAIRv2 with guides targeting an out-of-frame UAG site in the endogenous *KRAS* and *PPIB* transcripts. The on-target editing fraction is shown as a sideways bar chart on the right for each condition row. The duplex region formed by the guide RNA is shown by a red outline box. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C.

[00100] **FIG. 55**: Bacterial screening of Cas13b orthologs for *in vivo* efficiency and PFS determination. A) Schematic of bacterial assay for determining the PFS of Cas13b orthologs. Cas13b orthologs with beta-lactamase targeting spacers (SEQ ID No. 722) are co-transformed with beta-lactamase expression plasmids containing randomized PFS sequences and subjected to double selection. PFS sequences that are depleted during co-transformation with Cas13b suggest targeting activity and are used to infer PFS preferences. B) Quantitation of interference activity of Cas13b orthologs targeting beta-lactamase as measured by colony forming units (cfu). Values represent mean +/– S.D. C) PFS logos for Cas13b orthologs as determined by depleted sequences from the bacterial assay. PFS preferences are derived from sequences depleted in the Cas13b condition relative to empty vector controls. Depletion values used to calculate PFS weblogos are listed in table 7.

[0100] **FIG. 56**: Optimization of Cas13b knockdown and further characterization of mismatch specificity. A) Gluc knockdown with two different guides is measured using the top 2 Cas13a and top 4 Cas13b orthologs fused to a variety of nuclear localization and nuclear export tags. B) Knockdown of *KRAS* is measured for LwaCas13a, RanCas13b, PguCas13b, and PspCas13b with four different guides and compared to four position-matched shRNA controls. Non-targeting guide is the same as in Figure 49B. shRNA non-targeting guide sequence is listed in table 11. C) Schematic of the single and double mismatch plasmid libraries used for evaluating the specificity of LwaCas13a and PspCas13b knockdown. Every possible single and double mismatch is present in the target sequence as well as in 3 positions directly flanking the 5' and 3' ends of the target site. (SEQ ID Nos. 723-734) D) The depletion level of transcripts with the indicated single mismatches are plotted as a heatmap for both the LwaCas13a and PspCas13b conditions. (SEQ ID Nos. 723 and 736) The wildtype base is outlined by a green box. E) The depletion level of transcripts with the indicated double mismatches are plotted as a heatmap for both the LwaCas13a and PspCas13b conditions (SEQ ID Nos. 723 and 736). Each box represents the average of all possible double mismatches for the indicated position.

[0101] **FIG. 57**: Characterization of design parameters for dCas13-ADAR2 RNA editing A) Knockdown efficiency of Gluc targeting for wildtype Cas13b and catalytically inactive H133A/H1058A Cas13b (dCas13b). B) Quantification of luciferase activity restoration by dCas13b fused to either the wildtype ADAR2 catalytic domain or the hyperactive E488Q mutant ADAR2 catalytic catalytic domain, tested with tiling Cluc targeting guides. C) Guide design and sequencing quantification of A->I editing for 30 nt guides targeting *Cypridinia* luciferase W85X (SEQ ID Nos. 737-745). D) Guide design and sequencing quantification of

AMENDED SHEET - IPEA/US

A->I editing for 50 nt

AMENDED SHEET - IPEA/US

guides targeting *PPIB* (SEQ ID Nos. 746-753). E) Influence of linker choice on luciferase activity restoration by REPAIRv1. F) Influence of base identify opposite the targeted adenosine on luciferase activity restoration by REPAIRv1 (SEQ ID Nos. 754 and 755). Values represent mean +/– S.E.M.

[0102] **FIG. 58**: ClinVar motif distribution for G>A mutations. The number of each possible triplet motif observed in the ClinVar database for all G>A mutations.

[0103] **FIG. 59**: Truncations of dCas13b still have functional RNA editing. Various N-terminal and C-terminal truncations of dCas13b allow for RNA editing as measured by restoration of luciferase signal for the *Cluc* W85X reporter. Values represent mean +/– S.E.M. The construct length refers to the coding sequence of the REPAIR constructs..

[0104] **FIG. 60**: Comparison of other programmable ADAR systems with the dCas13-ADAR2 editor. A) Schematic of two programmable ADAR schemes: BoxB-based targeting and full length ADAR2 targeting. In the BoxB scheme (top), the ADAR2 deaminase domain ($ADAR2_{DD}$(E488Q)) is fused to a small bacterial virus protein called lambda N (ƛN), which binds specifically a small RNA sequence called BoxB-ƛ, and the fusion protein is recruited to target adenosines by a guide RNA containing homology to the target site and hairpins that BoxB-ƛ binds to. Full length ADAR2 targeting utilizes a guide RNA with homology to the target site and a motif recognized by the double strand RNA binding domains of ADAR2.. A guide RNA containing two BoxB-ƛ hairpins can then guide the ADAR2 $_{DD}$(E488Q), -ƛN for site specific editing. In the full length ADAR2 scheme (bottom), the dsRNA binding domains of ADAR2 bind a hairpin in the guide RNA, allowing for programmable ADAR2 editing (SEQ ID Nos. 756-760). B) Transcriptome-wide sites of significant RNA editing by BoxB-ADAR2 $_{DD}$(E488Q) with a guide targeting Cluc and a non-targeting guide. The on-target Cluc site (254 A>G) is highlighted in orange. C) Transcriptome-wide sites of significant RNA editing by ADAR2 with a guide targeting Cluc and a non-targeting guide. The on-target Cluc site (254 A>G) is highlighted in orange. D) Transcriptome-wide sites of significant RNA editing by REPAIRv1 with a guide targeting Cluc and a non-targeting guide. The on-target Cluc site (254 A>G) is highlighted in orange. The non-targeting guide is the same as in Fig50C. E) Quantitation of on-target editing rate percentage for BoxB-ADAR2 $_{DD}$(E488Q), ADAR2, and REPAIRv1 for targeting guides against Cluc. F) Overlap of off-target sites between different targeting and non-targeting conditions for programmable ADAR systems. The values plotted are the percent of the maximum possible intersection of the two off-target data sets.

[0105] **FIG. 61**: Efficiency and specificity of dCas13b-ADAR2 mutants A) Quantitation of luciferase activity restoration by dCas13b-ADAR2 $_{DD}$(E488Q) mutants for Cluc-targeting

and non-targeting guides. Non-targeting guide is the same as in Fig50C. B) Relationship

AMENDED SHEET - IPEA/US

between the ratio of targeting and non-targeting guides and the number of RNA-editing off-targets as quantified by transcriptome-wide sequencing C) Quantification of number of transcriptome-wide off-target RNA editing sites versus on-target Cluc editing efficiency for dCas13b-ADAR2 $_{DD}$(E488Q) mutants.

[0106] **FIG. 62**: Transcriptome-wide specificity of RNA editing by dCas13b-ADAR2 $_{DD}$(E488Q) mutants A) Transcriptome-wide sites of significant RNA editing by dCas13b-ADAR2 $_{DD}$(E488Q) mutants with a guide targeting Cluc. The on-target Cluc site (254 A>G) is highlighted in orange. B) Transcriptome-wide sites of significant RNA editing by dCas13b-ADAR2 $_{DD}$(E488Q) mutants with a non-targeting guide.

[0107] **FIG. 63**: Characterization of motif biases in the off-targets of dCas13b-ADAR2 $_{DD}$(E488Q) editing. A) For each dCas13b-ADAR2 $_{DD}$(E488Q) mutant, the motif present across all A>G off-target edits in the transcriptome is shown. B) The distribution of off-target A>I edits per motif identity is shown for REPAIRv1 with targeting and non-targeting guide. C) The distribution of off-target A>I edits per motif identity is shown for REPAIRv2 with targeting and non-targeting guide.

[0108] **FIG. 64**: Further characterization of REPAIRv1 and REPAIRv2 off-targets. A) Histogram of the number of off-targets per transcript for REPAIRv1. B) Histogram of the number of off-targets per transcript for REPAIRv2.C) Variant effect prediction of REPAIRv1 off targets. D) Distribution of REPAIRv1 off targets in cancer-related genes. TSG, tumor suppressor gene.. E) Variant effect prediction of REPAIRv2 off targets. F) Distribution of REPAIRv2 off targets in cancer-related genes.

[0109] **FIG. 65**: RNA editing efficiency and specificity of REPAIRv1 and REPAIRv2. A) Quantification of percent editing of *KRAS* with *KRAS*-targeting guide 1 at the targeted adenosine and neighboring sites for REPAIRv1 and REPAIRv2. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C. B) Quantification of percent editing of *KRAS* with *KRAS*-targeting guide 3 at the targeted adenosine and neighboring sites for REPAIRv1 and REPAIRv2. Non-targeting guide is the same as in Fig. 50C. C) Quantification of percent editing of *PPIB* with *PPIB*-targeting guide 2 at the targeted adenosine and neighboring sites for REPAIRv1 and REPAIRv2. Non-targeting guide is the same as in Fig. 50C.

[0110] **FIG. 66:** Demonstration of all potential codon changes with a A>I RNA editor. A) Table of all potential codon transitions enabled by A>I editing. B) A codon table demonstrating all the potential codon transitions enabled by A>I editing. Adapted and modified based on J. D. Watson, Molecular biology of the gene. (Pearson, Boston, ed. Seventh edition,

2014), pp. xxxiv, 872 pages.(38). C) Model of REPAIR A to I editing of a precisely encoded nucleotide via a mismatch in the guide sequence. The A to I transition is mediated by the catalytic activity of the ADAR2 deaminase domain and will be read as a guanosine by translational machinery. The base change does not rely on endogenous repair machinery and is permanent for as long as the RNA molecule exists in the cell. D) REPAIR can be used for correction of Mendelian disease mutations. E) REPAIR can be used for multiplexed A to I editing of multiple variants for engineering pathways or modifying disease. Multiplexed guide delivery can be achieved by delivering a single CRISPR array expression cassette since the Cas13b enzyme processes its own array. F) REPAIR can be used for modifying protein function through amino acid changes that affect enzyme domains, such as kinases. G) REPAIR can modulate splicing of transcripts by modifying the splice acceptor site.

[0111] **FIG. 67**: Additional truncations of Psp dCas13b.

[0112] **FIG. 68**: Potential effect of dosage on off target activity.

[0113] **FIG. 69**: Relative expression of Cas13 orthologs in mammalian cells and correlation of expression with interference activity. A) Expression of Cas13 orthologs as measured by msfGFP fluoresence. Cas13 orthologs C-terminally tagged with msfGFP were transfected into HEK293FT cells and their fluorescence measured 48 hours post transfection. B) Correlation of Cas13 expression to interference activity. The average RLU of two Gluc targeting guides for Cas13 orthologs, separated by subfamily, is plotted versus expression as determined by msfGFP fluoresence. The RLU for targeting guides are normalized to RLU for a non-targeting guide, whose value is set to 1. The non-targeting guide is the same as in Figure 49B for Cas13b.

[0114] **FIG. 70**: Comparison of RNA editing activity of dCas13b and REPAIRv1. A) Schematic of guides used to target the W85X mutation in the Cluc reporter (SEQ ID Nos. 911-917) B) Sequencing quantification of A to I editing for indicated guides transfected with dCas13b. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig50C. C) Sequencing quantification of A to I editing for indicated guides transfected with REPAIRv1. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig50C. D) Comparison of on-target A to I editing rates for dCas13b and dCas13b-ADAR2DD(E488Q) for guides tested in panel B and C. E) Influence of base identify opposite the targeted adenosine on luciferase activity restoration by REPAIRv1. Values represent mean +/– S.E.M. (SEQ ID Nos. 754 and 755)

[0115] **FIG. 71**: REPAIRv1 editing activity evaluated without a guide and in comparison to ADAR2 deaminase domain alone. A) Quantification of A to I editing of the Cluc W85X mutation by REPAIRv1 with and without guide as well as the ADAR2 deaminase domain only without guide. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig50C. B) Number of differentially expressed genes in the REPAIRv1 and ADAR2DD conditions from panel A. C) The number of significant off-targets from the REPAIRv1 and ADAR2DD conditions from panel A. D) Overlap of off-target A to I editing events between the REPAIRv1 and ADAR2DD conditions from panel A. The values plotted are the percent of the maximum possible intersection of the two off-target data sets.

[0116] **FIG. 72**: Evaluation of off-target sequence similarity to the guide sequence. A) Distribution of the number of mismatches (hamming distance) between the targeting guide sequence and the off-target editing sites for REPAIRv1 with a Cluc targeting guide. B) Distribution of the number of mismatches (hamming distance) between the targeting guide sequence and the off-target editing sites for REPAIRv2 with a Cluc targeting guide.

[0117] **FIG. 73**: Comparison of REPAIRv1, REPAIRv2, ADAR2 RNA targeting, and BoxB RNA targeting at two different doses of vector (150ng and 10ng effector). A) Quantification of RNA editing activity at the Cluc W85X (254 A>I) on-target editing site by REPAIRv1, REPAIRv2, ADAR2 RNA targeting, and BoxB RNA targeting approaches. Each of the four methods were tested with a targeting or non-targeting guide. Values shown are the mean of the three replicates. B) Quantification of RNA editing off-targets by REPAIRv1, REPAIRv2, ADAR2 RNA targeting, and BoxB RNA targeting approaches. Each of the four methods were tested with a targeting guide for the Cluc W85X (254 A>I) site or non-targeting guide. For REPAIR constructs, non-targeting guide is the same as in Fig. 50C.

[0118] **FIG. 74**: RNA editing efficiency and genome-wide specificity of REPAIRv1 and REPAIRv2. A) Quantification of RNA editing activity at the PPIB guide 1 on-target editing site by REPAIRv1, REPAIRv2 with targeting and non-targeting guides. Values represent mean +/– S.E.M. B) Quantification of RNA editing activity at the PPIB guide 2 on-target editing site by REPAIRv1, REPAIRv2 with targeting and non-targeting guides. Values represent mean +/– S.E.M. C) Quantification of RNA editing off-targets by REPAIRv1 or REPAIRv2 with PPIB guide 1, PPIB guide 2, or non-targeting guide. D) Overlap of off-targets between REPAIRv1 for PPIB targeting, Cluc targeting, and non-targeting guides. The values plotted are the percent of the maximum possible intersection of the two off-target data sets.

[0119] **FIG. 75**: High coverage sequencing of REPAIRv1 and REPAIRv2 off-targets. A) Quantitation of off-target edits for REPAIRv1 and REPAIRv2 as a function of read depth with

a total of 5 million reads (12.5x coverage), 15 million reads (37.5x coverage) and 50 million reads (125x coverage) per condition. B) Overlap of off-target sites at different read depths of the following conditions: REPAIRv1 versus REPAIRv1 (left), REPAIRv2 versus REPAIRv2 (middle), and REPAIRv1 versus REPAIRv2 (right). The values plotted are the percent of the maximum possible intersection of the two off-target data sets. C) Editing rate of off-target sites compared to the coverage (log2(number of reads)) of the off-target for REPAIRv1 and REPAIRv2 targeting conditions at different read depths. D) Editing rate of off-target sites compared to the log2(TPM+1) of the off-target gene expression for REPAIRv1 and REPAIRv2 targeting conditions at different read depths.

[0120] **FIG. 76**: Quantification of REPAIRv2 activity and off-targets in the U2OS cell line. A) Transcriptome-wide sites of significant RNA editing by REPAIRv2 with a guide targeting Cluc in the U2OS cell line. The on-target Cluc site (254 A>I) is highlighted in orange. B) Transcriptome-wide sites of significant RNA editing by REPAIRv2 with a non-targeting guide in the U2OS cell line. C) The on-target editing rate at the Cluc W85X (254 A>I) by REPAIRv2 with a targeting guide or non-targeting guide in the U2OS cell line. D) Quantification of off-targets by REPAIRv2 with a guide targeting Cluc or non-targeting guide in the U2OS cell line.

[0121] **FIG. 77:** Identifying additional ADAR mutants with increased efficiency and specificity. Cas13b-ADAR fusions with mutations in the ADAR deaminase domain, assayed on the luciferase target. Lower non-targeting RLU is indicative of more specificity.

[0122] **FIG. 78:** Identifying additional ADAR mutants with increased efficiency and specificity. Mutants were chosen from flow cytometry data for low, medium, and high-disrupting mutantions.

[0123] **FIG. 79:** Identifying additional ADAR mutants with increased efficiency and specificity.

[0124] **FIG. 80:** Identifying additional ADAR mutants with increased efficiency and specificity.

[0125] **FIG. 81:** Identifying additional ADAR mutants with increased efficiency and specificity through saturating mutagenesis on V351.

[0126] **FIG. 82:** Identifying additional ADAR mutants with increased efficiency and specificity through saturating mutagenesis on T375.

[0127] **FIG. 83:** Identifying additional ADAR mutants with increased efficiency and specificity through saturating mutagenesis on R455.

[0128] **FIG. 84:** Identifying additional ADAR mutants with increased efficiency and specificity through saturating mutagenesis.

[0129] **FIG. 85:** 3' binding loop residue saturation mutagenesis.

[0130] **FIG. 86:** Select ADAR mutants with increased efficiency and specificity. Screening has identified multiple mutants with increased specificity compared to REPAIRv1 and increased activity compared to REPAIRv1 and REPAIRv2.

[0131] **FIG. 87:** Second round saturating mutagenesis performed on promising residues with additional E488 mutations.

[0132] **FIG. 88:** Second round saturating mutagenesis performed on promising residues with additional E488 mutations.

[0133] **FIG. 89:** Combinations of ADAR mutants identified through screening.

[0134] **FIG. 90:** Combinations of ADAR mutants identified through screening.

[0135] **FIG. 91:** Testing most promising mutants by NGS.

[0136] **FIG. 92:** Testing most promising mutants by NGS.

[0137] **FIG. 93:** Testing most promising mutants by NGS.

[0138] **FIG. 94:** Testing most promising mutants by NGS.

[0139] **FIG. 95:** Finding most promising base flip for C-U activity on existing constructs.

[0140] **FIG. 96:** Testing ADAR mutants with best guide for C->U activity.

[0141] **FIG. 97:** Validation of V351 mutants for C>U activity.

[0142] **FIG. 98:** Testing Cas13b-cytidine deaminase fusions with testing panning guides across construct:

[0143] **FIG. 99:** Testing Cas13b-cytidine deaminase fusions with testing panning guides across construct.

[0144] **FIG. 100** is a graph depicting that Cas13b orthologs fused to ADAR exhibit variable protein recovery and off-target effects. 15 dCas13b orthologs were fused to ADAR and targeted to edit a Cypridina luciferase reporter with an introduced pretermination site that, when corrected, restores luciferase function. A nontargeting guide was additionally used to evaluate off target effects. REPAIRv1 and REPAIRv2 are as published in Cox et al. (2017). Different orthologs fused to ADAR exhibit different ability to recover functional luciferase, as well as different off-target effects. In particular, Cas12b6 (Riemerella anatipestifer (RanCas13b)) appears to have a better ability to recover functional luciferase as well as fewer off-target events than REPAIRv1. Points marked in red were selected for further engineering and analysis as these were the two orthologs that exhibited the highest functional protein recovery other than Cas13b12 (REPAIRv1).

[0145] **FIG. 101** is a graph showing targeted sequencing of editing locus for all orthologs. Targeted next generation sequencing of the editing locus shows that most Cas13b orthologs fused to ADAR mediate bona fide editing events at the target adenosine. Orthologs are ordered from lowest to highest editing percentage from top to bottom. In particular, although Cas13b6 is observed to exhibit higher functional luciferase recovery (FIG. 100), REPAIRv1 still shows a higher percentage of editing events at the target adenosine. Additionally, different orthologs show different percentages of off target edits at other adenosines within the sequencing window, and, in particular, Cas13b6 shows much lower editing at A33 both in the targeting and non-targeting condition than REPAIRv1, which is consistent with the lower off-target signal observed in the luciferase assay (FIG. 100). The ratio between on target and off-target editing is not consistent between orthologs, and in particular, Cas13b6 seems to maximize the amount of on-target edits per off-target edit.

[0146] **FIG. 102** is a schematic illustrating design constraints for delivery with Adeno-associated virus (AAV). AAV, a clinically relevant viral delivery vector, has a packaging limit of about 4.7 kilobases for efficient packaging and titering of the virus. However, REPAIR is much larger than this when the promoter is included. Additionally, it would be ideal to deliver the entire system (REPAIR fusion protein + guide RNA) in a single vector for ease of production and delivery. Therefore Cas13b orthologs are chosen to be truncated down

[0147] **FIG. 103A** is a graph showing results of truncating N-terminus of Cas13b6. Each ortholog was truncated down in 20 amino acid (60 base pair) intervals up to 300 amino acids (900 base pairs) from each of the N and C termini of the protein. RNA editing activity was then evaluated via the luciferase correction assay previously described. Luciferase recovery in the targeting guideRNA condition is shown on the y-axis, versus the size in amino acids of the truncated Cas13b ortholog on the x-axis. Truncating at different points changes the ability of the REPAIR fusion to recover luciferase function - some are better and some are worse than the full length Cas13b protein, and different patterns are observed with different orthologs. **FIG. 103B** is a graph showing results of truncating C-terminus of Cas13b6. For Cas13b6, the CΔ300 truncation was chosen as having the best activity with a sufficiently small size.

[0148] **FIG. 104A** is a graph showing results of truncating N-terminus of Cas13b11. **FIG. 104B** is a graph showing results of truncating C-terminus of Cas13b11. For Cas13b11, the NΔ280 truncation was chosen as having the best activity with a sufficiently small size.

[0149] **FIG. 105A** is a graph showing results of truncating N-terminus of Cas13b12. **FIG. 104B** is a graph showing results of truncating C-terminus of Cas13b12. For Cas13b12, the CΔ300 truncation was chosen as having the best activity with a sufficiently small size.

[0150] **FIG. 106** is a graph showing tiling guide RNAs across a single editing site. Editing is targeted to an adenosine in an introduced premature stop codon in a luciferase reporter, which, if corrected, will restore the amino acid at this position to a tryptophan and thus restore function of the luciferase. Guide RNAs with both 50 and 30 nucleotide spacers are tiled across this editing site such that the target adenosine is at a different position within the guide RNA. Each of these guides were evaluated with both the full length and best truncations previously noted on the preceding three slides. (SEQ ID Nos. 700 and 701)

[0151] **FIG. 107** is a graph showing Cas13b6 results with different guide RNAs. The results show that target adenosine position within the spacer sequence does have an effect on editing. Interestingly, both the full length and truncated Cas13b exhibit very similar patterns of which position within the guide is optimal, but different orthologs exhibit slightly different patterns, though still relatively similar (FIGs. 108 and 109). In general, 50 bp guides seem to be slightly better for A to I editing. shown here, B11 and B12 (REPAIRv1) on the following two slides.

[0152] **FIG. 108** is a graph showing Cas13b11 results with different guide RNAs.

[0153] **FIG. 109** is a graph showing Cas13b12 (REPAIRv1) with different guide RNAs.

[0154] **FIG. 110** is a graph showing results of Cas13b6-REPAIR targeting KRAS. In this figure, instead of moving the guide across a single editing site, the sequence of the guide is fixed and each guide RNA targets a different adenosine within the fixed sequence. Two sites were evaluated for both Cas13b6 and the Cas13b6CΔ300 truncation, with both 30 and 50 nucleotide guides as indicated in the schematic at the top (SEQ ID No. 918). Editing is evaluated by targeted next generation sequencing across the editing loci. Again, different target positions within the guide show different editing rates and patterns for both the full length and truncated Cas13b6s.

[0155] **FIG. 111** is a graph depicting that localization tags may affect on-target editing. Different localization tags (both nuclear localization and nuclear export tags) with Cas13b6 seem to affect the ability of Cas13b6-REPAIR to recover luciferase activity, but does not appear to affect off-target activity appreciably. Red points are REPAIRv1 and REPAIRv2, which are with the Cas13b12 ortholog and using the HIV NES, blue points with Cas13b6 ortholog.

[0156] **FIG. 112** is a graph showing results of RfxCas13d. Cas13d is a recently discovered class of Cas13 proteins that are on average smaller than Cas13b proteins. A characterized Cas13d ortholog known as RfxCas13d is tested in this figure for REPAIR activity using the same tiling guide scheme shown in Fig. 106. crRNA refers to mature CRISPR RNA and pre-

crRNA refers to unprocessed version. Although most guide RNAs with RfxCas13d-REPAIR show no RNA editing activity, there are a few that seem to mediate relatively good editing when compared to existing systems shown in black.

[0157] **FIG. 113** is a graph showing results of guide RNA-mediated editing with RfxCas13d. The data show that even without the RfxCas13d-REPAIR or even ADAR, the guide RNA (mismatch position 33) by itself is somehow able to mediate editing events (left-most condition), which is not the case with a Cas13b12 guide. Furthermore, it appears that the introduction of ADAR or RfxCas13d-REPAIR does not seem to have much effect on the editing mediated by this guide RNA.

[0158] **FIG. 114** is a schematic illustrating the dual vector system design for evaluating RNA editing in cultures of primary rat cortical neurons.

[0159] **FIG. 115** is a graph showing that up to 35% editing is achieved in neurons with dual vector system. Using two guides as indicated in the schematic at the top (SEQ ID No. 761, guide 1 has one base flip/targeted adenosine at the indicated position, while guide 2 has two targeted adenosine), REPAIR with B6/B11/B12 was packaged into AAV using the dual vector system in FIG. 114. Guide 2 was found to mediate up to 35% editing at A57 with B6-REPAIR (~30% for B11-REPAIR) with targeted next generation sequencing 14 days after transduction with AAV, showing that AAV-delivered REPAIR can mediate RNA base editing in post-mitotic cell types.

[0160] **FIG. 116** is a graph depicting that single vector AAV B6-REPAIR system is able to edit RNA in neuron cultures. Using the single vector system in FIG. 102 with the Cas13b6CΔ300 truncation, the guide that has two target adenosines in FIG. 115 was used, as well as a guide across the same sequence but only targeting A48 as indicated. 5 days after transduction with AAV, targeted next-generation sequencing shows approximately 6% editing with guide 2 at A24 (Same as A57 in FIG. 115), demonstrating the viability of the single vector approach.

[0161] **FIG. 117** is a graph is a graph depicting that different Cas13b orthologs fused to ADAR.

[0162] **FIG. 118** is a graph showing that V351G editing greatly increases REPAIR editing. The V351G mutation (pAB316) was introduced into the E488Q PspCas13b (Cas13b12) REPAIR construct (REPAIR v1, pAB0048) and tested for C-U activity on a gauss luciferase construct with a TCG motif (TCG). Editing was read out by next generation sequencing, revealing increased C-U activity.

[0163] **FIG. 119** is a graph showing endogenous KRAS and PPIB targeting. The V351G mutation (pAB316) was introduced into the E488Q PspCas13b REPAIR construct (REPAIR v1, pAB0048) and tested for C-U activity on a gauss four sites, two in each gene, with different motifs. Editing was read out by next generation sequencing, revealing increased C-U activity.

[0164] **FIG. 120** is a graph showing optimal V351G combination mutants. Selected sites (S486, G489) were mutagenized to all 20 possible residues and tested on a background of REPAIR[E488Q, V351G]. Constructs were tested on two luciferase motifs, TCG and GCG, and selected on the basis of luciferase activity.

[0165] **FIG. 121** is a graph showing S486A and V351G combination C-to-U activity. S486A was tested against the [V351G, E488Q] background and the E488Q background on all four motifs, with luciferase activity as a readout. S486A performs better on all motifs, especially ACG and TCG.

[0166] **FIG. 122** is a graph showing that S486A improves C-to-U editing across all motifs. S486A improves targeting over the [V351G, E488Q] background on all motifs, when measured by luciferase activity.

[0167] **FIG. 123A** is a graph showing S486 mutants C-to-U activity with both TCG and CCG targeting. **FIG. 123B** is a graph showing S486 mutants C-to-U activity with CCG targeting only. S486A was tested against the [V351G, E488Q] background and the E488Q background on all four motifs, with NGS as a readout. S486A performs better on all motifs, especially ACG and TCG.

[0168] **FIG. 124** is a graph showing S486A A-to-I activity. The data shows that S486A mutations maintain A-to-I activity of the previous constructs when measured on a luciferase reporter.

[0169] **FIG. 125** is a graph showing S486A A-to-I off-target activity. The data shows that S486A has comparable A-to-I off-target activity when measured on a luciferase reporter.

[0170] **FIG. 126A** is a graph showing that targeting by S486A/V351G/E488Q (pAB493), V351G/E488Q (pAB316), and E488Q (REPAIRv1) is comparable when read out by luciferase activity (Gluc/Cluc RLU). **FIG. 126B** is a graph showing that targeting by S486A/V351G/E488Q (pAB493), V351G/E488Q (pAB316), and E488Q (REPAIRv1) is comparable when assayed by NGS (fraction editing).

[0171] **FIG. 127A** is a graph showing S486A C-to-U activity by NGS on Cluc reporter constructs. **FIG. 127B** is a graph showing S486A C-to-U activity by NGS on endogenous gene PPIB.

[0172] **FIG. 128** is a graph depicting identification of new T375 and K376 mutants. Selected sites (T375, K376) were mutagenized to all 20 possible residues and tested on a background of REPAIR[E488Q, V351G]. Constructs were tested on the TCG luciferase motif and selected on the basis of luciferase activity.

[0173] **FIG. 129** is a graph showing that T375S has relaxed motif. T375S was tested against the [S486A,V351G, E488Q] background (pAB493), [V351G, E488Q] background (pAB316), and the E488Q background (pAB48) on all TCG and GCG motifs, with luciferase activity as a readout. T375S improves GCG motif.

[0174] **FIG. 130** is a graph showing that T375S has relaxed motif. T375S was tested against the [S486A,V351G, E488Q] background (pAB493), [V351G, E488Q] background (pAB316), and the E488Q background (pAB48) on GCG motifs, with luciferase activity as a readout. T375S improves GCG motif.

[0175] **FIG. 131** is a graph depicting that B6 and B11 orthologs show improved RESCUE activity. Cas13b orthologs Cas13b6 (RanCas13b) and Cas13b11 (PguCas13b) were tested with T375S mutation, and show improved activity as measured by luciferase assay. Mutations shows are on corresponding backgrounds (T375S = T375S/S486A/V351G/E448Q).

[0176] **FIG. 132** is a graph showing that DNA2.0 vectors has comparable luciferase to transient transfection vectors. RESCUE vectors based off of either DNA2.0 (now Atum) constructs compared to a non-lenti vector, with Cas13b11 (PguCas13b) show improved luciferase activity. The Atum vector map (https://benchling.com/s/seq-DENgx9izDhsRTFFgy71K) has additional EES elements for expression. Mutations shows are on corresponding backgrounds (V351G = V351G/E448Q, S486A = S486A/V351G/E448Q).

[0177] **FIG. 133A** is a graph showing luciferase results of testing truncations validated by REPAIR (B6 Cdelta300) with RESCUE using 30bp guides. **FIG. 133B** is a graph showing luciferase results of testing truncations validated by REPAIR (B6 Cdelta300) with RESCUE using 50bp guides. The 26 mismatch distance (as measured by the 5' end) shows the optimal activity with both full length and truncated versions).

[0178] **FIG. 134A** is a graph showing luciferase results of testing truncations validated by REPAIR (B11 Ndelta280) with RESCUE using 30bp guides. **FIG. 134B** is a graph showing luciferase results of testing truncations validated by REPAIR (B11 Ndelta280) with RESCUE using 50bp guides. The 26 mismatch distance (as measured by the 5' end) shows the optimal activity with both full length and truncated versions).

[0179] **FIG. 135** is a graph showing results of testing all B6 truncations. Iterative truncations were generated from the N and C termini on RanCas13b (B6), with the

T375S/S486A/V351G/E448Q mutation, with optimal activity up to C-delta 200, and activity at C-delta 320. Truncations are tested on luciferase, and editing is read out as luciferase activity. Missing bars indicate no data. The pAB0642 is an untruncated N-term control, T375S/S486A/V351G/E448Q. The pAB0440 is an untruncated C-term control, E448Q. All N-term constructs, and pAB0642, have an mark NES linker. All C-term constrcuts, and pAB0440, have a HIV-NES linker.

[0180] **FIG. 136** is a graph showing results of testing all B11 truncations. Iterative truncations were generated from the N and C termini on PguCas13b (B11), with the T375S/S486A/V351G/E448Q mutation. Truncations are tested on luciferase, and editing is read out as luciferase activity.

[0181] **FIG. 137A** is a graph showing Beta catenin modulation with REPAIR/RESCUE as measured by Beta-catenin activity via the TCF-LEF RE Wnt pathway reporter (Promega). **FIG. 137B** is a graph showing Beta catenin modulation with REPAIR/RESCUE as measured by the M50 Super 8x TOPFlash reporter (Addgene). Beta-catenin/Wnt pathway induction is tested by using RNA editing to remove phosphorylation sites on Beta catenin. Guides targeting beta-catenin for either REPAIR (RanCas13b ortholog, E488Q mutation) or RESCUE (RanCas13b ortholog, T375S/S486A/V351G/E448Q mutation) were tested for phenotypic activity. The T41A guide shows activity on both reporters.

[0182] **FIG. 138** is a graph showing NGS results of Beta catenin modulation. NGS readouts of either A-I (A) or C-U (C) activity at targeted sites by either REPAIR (RanCas13b ortholog, E488Q mutation) or RESCUE (RanCas13b ortholog, T375S/S486A/V351G/E448Q mutation. REPAIR was used on A targets, and RESCUE was used on C targets.

[0183] **FIG. 139** is a graph depicting that tiling different guides shows improved motif activity at the 30_5 mutation (mismatch is 26 nt away from the 5' of the guide). All four motifs were tested with various tiling guides for luciferase activity. Nomenclature corresponds to distance from the 3' end of the spacer (i.e., 26 nt mismatch is 30_5). The 26 mismatch distance (as measured by the 5' end) shows the optimal activity with most motifs. Guides were tested with RESCUE (RanCas13b ortholog, T375S/S486A/V351G/E448Q mutation.

[0184] **FIG. 140A** is a graph showing that REPAIR allows for editing residues associated with PTMs. **FIG. 140B** is a graph showing that RESCUE allows for editing residues associated with PTMs.

[0185] The appended claims are herein explicitly incorporated by reference.

[0186] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION

General Definitions

[0187] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.): Antibodies, A Laboraotry Manual (1988) (Harlow and Lane, eds.): Antibodies A Laboraotry Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011).

[0188] Reference is made to US Provisional 62/351,662 and 62/351,803, filed on June 17, 2016, US Provisional 62/376,377, filed on August 17, 2016, US Provisional 62/410,366, filed October 19, 2016, US Provisional 62/432,240, filed December 9, 2016, US provisional 62/471,792 filed March 15, 2017, and US Provisional 62/484,786 filed April 12, 2017. Reference is made to International PCT application PCT/US2017/038154, filed June 19, 2017. Reference is made to US Provisional 62/471,710, filed March 15, 2017 (entitled, "Novel Cas13B Orthologues CRISPR Enzymes and Systems," Attorney Ref: BI-10157 VP 47627.04.2149). Reference is further made to US Provisional 62/432,553, filed December 9, 2016, US Provisional 62/456,645, filed February 8, 2017, and US Provisional 62/471,930, filed March 15, 2017 (entitled "CRISPR Effector System Based Diagnostics," Attorney Ref. BI-10121 BROD 0842P) and US Provisional To Be Assigned, filed April 12, 2017 (entitled "CRISPR Effector System Based Diagnostics," Attorney Ref. BI-10121 BROD 0843P)

[0189] As used herein, the singular forms “a”, “an”, and “the” include both singular and plural referents unless the context clearly dictates otherwise.

[0190] The term “optional” or “optionally” means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[0191] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0192] The terms “about” or “approximately” as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier “about” or “approximately” refers is itself also specifically, and preferably, disclosed.

[0193] Reference throughout this specification to “one embodiment”, “an embodiment,” “an example embodiment,” means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases “in one embodiment,” “in an embodiment,” or “an example embodiment” in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

[0194] C2c2 is now known as Cas13a. It will be understood that the term “C2c2” herein is used interchangeably with “Cas13a”.

[0195] All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

[0196] Various embodiments are described hereinafter. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader

aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s).

OVERVIEW

[0197] The embodiments disclosed herein provide systems, constructs, and methods for targeted base editing. In general the systems disclosed herein comprise a targeting component and a base editing component. The targeting component functions to specifically target the base editing component to a target nucleotide sequence in which one or more nucleotides are to be edited. The base editing component may then catalyze a chemical reaction to convert a first nucleotide in the target sequence to a second nucleotide. For example, the base editor may catalyze conversion of an adenine such that it is read as guanine by a cell's transcription or translation macchinery, or vice versa. Likewise, the base editing component may catalyze conversion of cytidine to a uracil, or vice versa. In certain example embodiments, the base editor may be derived by starting with a known base editor, such as an adenine deaminase or cytodine deaminase, and modified using methods such as directed evolution to derive new functionalities. Directed evolution techniques are known in the art and may include those described in WO 2015/184016 "High-Throughput Assembly of Genetic Permuatations." In will be understood that the present invention in certain aspects equally relates to deaminases per se as described herein and having undergone directed evolution, such as the mutated deaminases described herein elsewhere, as well as polynucleotides encoding such deaminases (including vectors and expression and/or delivery systems), as well as fusions between such mutated deaminases and targeting component, such as polynucleotide binding molecules or systems, as described herein elsewhere.

[0198] In one aspect the present invention provides methods for targeted deamination of adenine or cytodine in RNA or DNA by an adenosine deaminase or modified variant thereof. According to the methods of the invention, the adenosine deaminase (AD) protein is recruited specifically to the nucleic acid to be modified. The term "AD functionalized compositions" refers to the engineered compositions for site directed base editing disclosed herein, comprising a targeting domain complexed to an adenosine deaminase, or catalytic domain thereof.

[0199] In particular embodiments of the methods of the present invention, recruitment of the adenosine deaminase to the target locus is ensured by fusing the adenosine deaminase or catalytic domain thereof to the targeting domain. Methods of generating a fusion protein from two separate proteins are known in the art and typically involve the use of spacers or linkers. The target domain can be fused to the adenosine deaminase protein or catalytic domain thereof on either the N- or C-terminal end thereof.

[0200] The term "linker" as used in reference to a fusion protein refers to a molecule which joins the proteins to form a fusion protein. Generally, such molecules have no specific biological activity other than to join or to preserve some minimum distance or other spatial relationship between the proteins. However, in certain embodiments, the linker may be selected to influence some property of the linker and/or the fusion protein such as the folding, net charge, or hydrophobicity of the linker.

[0201] Suitable linkers for use in the methods of the present invention are well known to those of skill in the art and include, but are not limited to, straight or branched-chain carbon linkers, heterocyclic carbon linkers, or peptide linkers. However, as used herein the linker may also be a covalent bond (carbon-carbon bond or carbon-heteroatom bond). In particular embodiments, the linker is used to separate the targeting domain and the adenosine deaminase by a distance sufficient to ensure that each protein retains its required functional property. Preferred peptide linker sequences adopt a flexible extended conformation and do not exhibit a propensity for developing an ordered secondary structure. In certain embodiments, the linker can be a chemical moiety which can be monomeric, dimeric, multimeric or polymeric. Preferably, the linker comprises amino acids. Typical amino acids in flexible linkers include Gly, Asn and Ser. Accordingly, in particular embodiments, the linker comprises a combination of one or more of Gly, Asn and Ser amino acids. Other near neutral amino acids, such as Thr and Ala, also may be used in the linker sequence. Exemplary linkers are disclosed in Maratea et al. (1985), Gene 40: 39-46; Murphy et al. (1986) Proc. Nat'l. Acad. Sci. USA 83: 8258-62; U.S. Pat. No. 4,935,233; and U.S. Pat. No. 4,751,180. For example, GlySer linkers GGS, GGGS or GSG can be used. GGS, GSG, GGGS or GGGGS linkers can be used in repeats of 3 (such as (GGS)3 (SEQ ID No. 12), (GGGGS)3) or 5, 6, 7, 9 or even 12 (SEQ ID No. 13) or more, to provide suitable lengths. In particular embodiments, linkers such as (GGGGS)3 are preferably used herein. (GGGGS)6 (GGGGS)9 or (GGGGS)12 may preferably be used as alternatives. Other preferred alternatives are (GGGGS)1 (SEQ ID No 14), (GGGGS)2 (SEQ ID No. 15), (GGGGS)4, (GGGGS)5, (GGGGS)7, (GGGGS)8, (GGGGS)10, or (GGGGS)11. In yet a further embodiment, LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID No:11) is used as a linker. In yet an additional embodiment, the linker is XTEN linker (SEQ ID No. 919). The invention also relates to a method for treating or preventing a disease by the targeted deamination or a disease causing variant using the AD-functionalized compositions. For example, the deamination of an A, may remedy a disease caused by transcripts containing a pathogenic G→A or C→T point mutation. Examples of disease that can be treated or prevented with the present invention include cancer, Meier-Gorlin syndrome, Seckel syndrome

4, Joubert syndrome 5, Leber congenital amaurosis 10; Charcot-Marie-Tooth disease, type 2; Charcot-Marie-Tooth disease, type 2; Usher syndrome, type 2C; Spinocerebellar ataxia 28; Spinocerebellar ataxia 28; Spinocerebellar ataxia 28; Long QT syndrome 2; Sjögren-Larsson syndrome; Hereditary fructosuria; Hereditary fructosuria; Neuroblastoma; Neuroblastoma; Kallmann syndrome 1; Kallmann syndrome 1; Kallmann syndrome 1; Metachromatic leukodystrophy.

[0202] In particular embodiments, the invention thus comprises compositions for use in therapy. This implies that the methods can be performed *in vivo, ex vivo* or *in vitro*. In particular embodiments, the methods are not methods of treatment of the animal or human body or a method for modifying the germ line genetic identity of a human cell. In particular embodiments; when carrying out the method, the target RNA is not comprised within a human or animal cell. In particular embodiments, when the target is a human or animal target, the method is carried out *ex vivo* or *in vitro*.

[0203] The invention also relates to a method for knocking-out or knocking-down an undesirable activity of a gene, wherein the deamination of an A or C at the transcript of the gene results in a loss of function. For example, in one embodiment, the targeted deamination by the AD-functionalized CRISPR system can cause a nonsense mutation resulting in a premature stop codon in an endogenous gene. This may alter the expression of the endogenous gene and can lead to a desirable trait in the edited cell. In another embodiment, the targeted deamination by the AD-functionalized compositions can cause a nonconservative missense mutation resulting in a code for a different amino acid residue in an endogenous gene. This may alter the function of the endogenous gene expressed and can also lead to a desirable trait in the edited cell.

[0204] The invention also relates to a modified cell obtained by the targeted deamination using the AD-functionalized composition, or progeny thereof, wherein the modified cell comprises an I or G in replace of the A, or a T in replace of the C in the target RNA sequence of interest compared to a corresponding cell before the targeted deamination. The modified cell can be a eukaryotic cell, such as an animal cell, a plant cell, an mammalian cell, or a human cell.

[0205] In some embodiments, the modified cell is a therapeutic T cell, such as a T cell sutiable for CAR-T therapies. The modification may result in one or more desirable traits in the therapeutic T cell, including but not limited to, reduced expression of an immune checkpoint receptor (e.g., PDA, CTLA4), reduced expression of HLA proteins (e.g., B2M, HLA-A), and reduced expression of an endogenous TCR.

[0206] In some embodiments, the modified cell is an antibody-producing B cell. The modification may results in one or more desirable traits in the B cell, including but not limited to, enhanced antibody production.

[0207] The invention also relates to a modified non-human animal or a modified plant. The modified non-human animal can be a farm animal. The modified plant can be an agricultural crop.

[0208] The invention further relates to a method for cell therapy, comprising administering to a patient in need thereof the modified cell described herein, wherein the presence of the modified cell remedies a disease in the patient. In one embodiment, the modified cell for cell therapy is a CAR-T cell capable of recognizing and/or attacking a tumor cell. In another embodiment, the modified cell for cell therapy is a stem cell, such as a neural stem cell, a mesenchymal stem cell, a hematopoietic stem cell, or an iPSC cell.

[0209] The invention additionally relates to an engineered, non-naturally occurring system suitable for modifying an Adenine or Cytodine in a target locus of interest, comprising: a targeteting domain; an adenosine deaminase protein or catalytic domain thereof, or one or more nucleotide sequences encoding; wherein the adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to the targeting domain or is adapted to link thereto after delivery; wherein the targeting domain is capable of hybridizing with a target sequence comprising an Adenine or Cytidine within an RNA or DNA polynucleotide of interest.

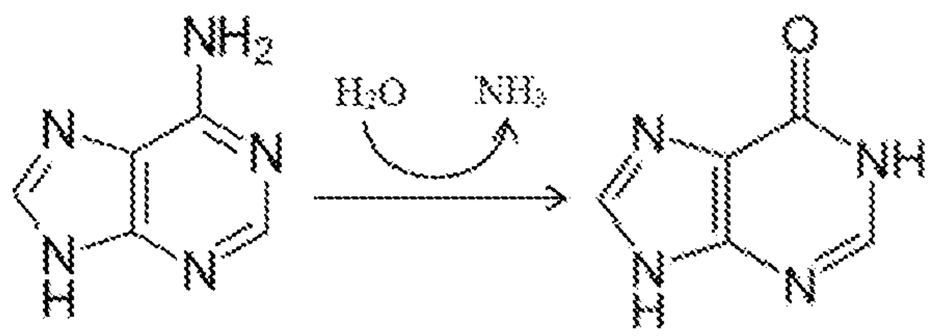
[0210] The invention additionally relates to an engineered, non-naturally occurring vector system suitable for modifying an Adenine or Cytodine in a target locus of interest, comprising one or more vectors comprising: (a) a first regulatory element operably linked to one or more nucleotide sequences encoding encoding a targeting domain; and (b) optionally a nucleotide sequence encoding an adenosine deaminase protein or catalytic domain thereof which is under control of the first or operably linked to a second regulatory element; wherein, if the nucleotide sequence encoding an adenosine deaminase protein or catalytic domain thereof is operably linked to a second regulatory element, the adenosine deaminase protein or catalytic domain thereof is adapted to link to the targeting domain after expression; wherein the targeting domain is capable of hybridizing with a target sequence comprising an Adenine or Cytodine within the target locus; wherein components (a) and (b) are located on the same or different vectors of the system.

[0211] The invention additionally relates to *in vitro*, *ex vivo* or *in vivo* host cell or cell line or progeny thereof comprising the engineered, non-naturally occurring system or vector system

described herein. The host cell can be a eukaryotic cell, such as an animal cell, a plant cell, an mammalian cell, or a human cell.

Adenosine Deaminase

[0212] The term "adenosine deaminase" or "adenosine deaminase protein" as used herein refers to a protein, a polypeptide, or one or more functional domain(s) of a protein or a polypeptide that is capable of catalyzing a hydrolytic deamination reaction that converts an adenine (or an adenine moiety of a molecule) to a hypoxanthine (or a hypoxanthine moiety of a molecule), as shown below. In some embodiments, the adenine-containing molecule is an adenosine (A), and the hypoxanthine-containing molecule is an inosine (I). The adenine-containing molecule can be deoxyribonucleic acid (DNA) or ribonucleic acid (RNA).



Adenine Hypoxanthine

[0213] According to the present disclosure, adenosine deaminases that can be used in connection with the present disclosure include, but are not limited to, members of the enzyme family known as adenosine deaminases that act on RNA (ADARs), members of the enzyme family known as adenosine deaminases that act on tRNA (ADATs), and other adenosine deaminase domain-containing (ADAD) family members. According to the present disclosure, the adenosine deaminase is capable of targeting adenine in a RNA/DNA and RNA duplexes. Indeed, Zheng et al. (Nucleic Acids Res. 2017, 45(6): 3369-3377) demonstrate that ADARs can cary out adenosine to inosine editing reactions on RNA/DNA and RNA/RNA duplexes. In particular embodiments, the adenosine deaminase has been modified to increase its ability to edit DNA in a RNA/DNAn RNA duplex as detailed herein below.

[0001] In some embodiments, the adenosine deaminase is derived from one or more metazoa species, including but not limited to, mammals, birds, frogs, squids, fish, flies and worms. In some embodiments, the adenosine deaminase is a human, squid or Drosophila adenosine deaminase.

[0214] In some embodiments, the adenosine deaminase is a human ADAR, including hADAR1, hADAR2, hADAR3. In some embodiments, the adenosine deaminase is a Caenorhabditis elegans ADAR protein, including ADR-1 and ADR-2. In some embodiments, the adenosine deaminase is a Drosophila ADAR protein, including dAdar. In some

embodiments, the adenosine deaminase is a squid *Loligo pealeii* ADAR protein, including sqADAR2a and sqADAR2b. In some embodiments, the adenosine deaminase is a human ADAT protein. In some embodiments, the adenosine deaminase is a Drosophila ADAT protein. In some embodiments, the adenosine deaminase is a human ADAD protein, including TENR (hADAD1) and TENRL (hADAD2).

[0215] In some embodiments, the adenosine deaminase is a TadA protein such as E. coli TadA. See Kim et al., Biochemistry 45:6407-6416 (2006); Wolf et al., EMBO J. 21:3841-3851 (2002). In some embodiments, the adenosine deaminase is mouse ADA. See Grunebaum et al., Curr. Opin. Allergy Clin. Immunol. 13:630-638 (2013). In some embodiments, the adenosine deaminase is human ADAT2. See Fukui et al., J. Nucleic Acids 2010:260512 (2010).

[0216] In some embodiments, the adenosine deaminase protein recognizes and converts one or more target adenosine residue(s) in a double-stranded nucleic acid substrate into inosine residues (s). In some embodiments, the double-stranded nucleic acid substrate is a RNA-DNA hybrid duplex. In some embodiments, the adenosine deaminase protein recognizes a binding window on the double-stranded substrate. In some embodiments, the binding window contains at least one target adenosine residue(s). In some embodiments, the binding window is in the range of about 3 bp to about 100 bp. In some embodiments, the binding window is in the range of about 5 bp to about 50 bp. In some embodiments, the binding window is in the range of about 10 bp to about 30 bp. In some embodiments, the binding window is about 1 bp, 2 bp, 3 bp, 5 bp, 7 bp, 10 bp, 15 bp, 20 bp, 25 bp, 30 bp, 40 bp, 45 bp, 50 bp, 55 bp, 60 bp, 65 bp, 70 bp, 75 bp, 80 bp, 85 bp, 90 bp, 95 bp, or 100 bp.

[0217] In some embodiments, the adenosine deaminase protein comprises one or more deaminase domains. Not intended to be bound by a particular theory, it is contemplated that the deaminase domain functions to recognize and convert one or more target adenosine (A) residue(s) contained in a double-stranded nucleic acid substrate into inosine (I) residue(s). In some embodiments, the deaminase domain comprises an active center. In some embodiments, the active center comprises a zinc ion. In some embodiments, during the A-to-I editing process, base pairing at the target adenosine residue is disrupted, and the target adenosine residue is "flipped" out of the double helix to become accessible by the adenosine deaminase. In some embodiments, amino acid residues in or near the active center interact with one or more nucleotide(s) 5' to a target adenosine residue. In some embodiments, amino acid residues in or near the active center interact with one or more nucleotide(s) 3' to a target adenosine residue. In some embodiments, amino acid residues in or near the active center further interact with the nucleotide complementary to the target adenosine residue on the opposite strand. In some

embodiments, the amino acid residues form hydrogen bonds with the 2' hydroxyl group of the nucleotides.

[0218] In some embodiments, the adenosine deaminase comprises human ADAR2 full protein (hADAR2) or the deaminase domain thereof (hADAR2-D). In some embodiments, the adenosine deaminase is an ADAR family member that is homologous to hADAR2 or hADAR2-D.

[0219] Particularly, in some embodiments, the homologous ADAR protein is human ADAR1 (hADAR1) or the deaminase domain thereof (hADAR1-D). In some embodiments, glycine 1007 of hADAR1-D corresponds to glycine 487 hADAR2-D, and glutamic Acid 1008 of hADAR1-D corresponds to glutamic acid 488 of hADAR2-D.

[0220] In some embodiments, the adenosine deaminase comprises the wild-type amino acid sequence of hADAR2-D. In some embodiments, the adenosine deaminase comprises one or more mutations in the hADAR2-D sequence, such that the editing efficiency, and/or substrate editing preference of hADAR2-D is changed according to specific needs.

[0221] Certain mutations of hADAR1 and hADAR2 proteins have been described in Kuttan et al., Proc Natl Acad Sci U S A. (2012) 109(48):E3295-304; Want et al. ACS Chem Biol. (2015) 10(11):2512-9; and Zheng et al. Nucleic Acids Res. (2017) 45(6):3369-337, each of which is incorporated herein by reference in its entirety.

[0222] In some embodiments, the adenosine deaminase comprises a mutation at glycine336 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 336 is replaced by an aspartic acid residue (G336D).

[0223] In some embodiments, the adenosine deaminase comprises a mutation at Glycine487 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 487 is replaced by a non-polar amino acid residue with relatively small side chains. For example, in some embodiments, the glycine residue at position 487 is replaced by an alanine residue (G487A). In some embodiments, the glycine residue at position 487 is replaced by a valine residue (G487V). In some embodiments, the glycine residue at position 487 is replaced by an amino acid residue with relatively large side chains. In some embodiments, the glycine residue at position 487 is replaced by a arginine residue (G487R). In some embodiments, the glycine residue at position 487 is replaced by a lysine residue (G487K). In some embodiments, the glycine residue at position 487 is replaced by a tryptophan residue (G487W). In some embodiments, the glycine residue at position 487 is replaced by a tyrosine residue (G487Y).

[0224] In some embodiments, the adenosine deaminase comprises a mutation at glutamic acid488 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glutamic acid residue at position 488 is replaced by a glutamine residue (E488Q). In some embodiments, the glutamic acid residue at position 488 is replaced by a histidine residue (E488H). In some embodiments, the glutamic acid residue at position 488 is replace by an arginine residue (E488R). In some embodiments, the glutamic acid residue at position 488 is replace by a lysine residue (E488K). In some embodiments, the glutamic acid residue at position 488 is replace by an asparagine residue (E488N). In some embodiments, the glutamic acid residue at position 488 is replace by an alanine residue (E488A). In some embodiments, the glutamic acid residue at position 488 is replace by a Methionine residue (E488M). In some embodiments, the glutamic acid residue at position 488 is replace by a serine residue (E488S). In some embodiments, the glutamic acid residue at position 488 is replace by a phenylalanine residue (E488F). In some embodiments, the glutamic acid residue at position 488 is replace by a lysine residue (E488L). In some embodiments, the glutamic acid residue at position 488 is replace by a tryptophan residue (E488W).

[0225] In some embodiments, the adenosine deaminase comprises a mutation at threonine490 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the threonine residue at position 490 is replaced by a cysteine residue (T490C). In some embodiments, the threonine residue at position 490 is replaced by a serine residue (T490S). In some embodiments, the threonine residue at position 490 is replaced by an alanine residue (T490A). In some embodiments, the threonine residue at position 490 is replaced by a phenylalanine residue (T490F). In some embodiments, the threonine residue at position 490 is replaced by a tyrosine residue (T490Y). In some embodiments, the threonine residue at position 490 is replaced by a serine residue (T490R). In some embodiments, the threonine residue at position 490 is replaced by an alanine residue (T490K). In some embodiments, the threonine residue at position 490 is replaced by a phenylalanine residue (T490P). In some embodiments, the threonine residue at position 490 is replaced by a tyrosine residue (T490E).

[0226] In some embodiments, the adenosine deaminase comprises a mutation at valine493 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the valine residue at position 493 is replaced by an alanine residue (V493A). In some embodiments, the valine residue at position 493 is replaced by a serine residue (V493S). In some embodiments, the valine residue at position 493 is replaced by a threonine residue (V493T). In some embodiments, the valine residue at position 493 is

replaced by an arginine residue (V493R). In some embodiments, the valine residue at position 493 is replaced by an aspartic acid residue (V493D). In some embodiments, the valine residue at position 493 is replaced by a proline residue (V493P). In some embodiments, the valine residue at position 493 is replaced by a glycine residue (V493G).

[0227] In some embodiments, the adenosine deaminase comprises a mutation at alanine589 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the alanine residue at position 589 is replaced by a valine residue (A589V).

[0228] In some embodiments, the adenosine deaminase comprises a mutation at asparagine597 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the asparagine residue at position 597 is replaced by a lysine residue (N597K). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by an arginine residue (N597R). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by an alanine residue (N597A). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by a glutamic acid residue (N597E). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by a histidine residue (N597H). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by a glycine residue (N597G). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by a tyrosine residue (N597Y). In some embodiments, the asparagine residue at position 597 is replaced by a phenylalanine residue (N597F). In some embodiments, the adenosine deaminase comprises mutation N597I. In some embodiments, the adenosine deaminase comprises mutation N597L. In some embodiments, the adenosine deaminase comprises mutation N597V. In some embodiments, the adenosine deaminase comprises

mutation N597M. In some embodiments, the adenosine deaminase comprises mutation N597C. In some embodiments, the adenosine deaminase comprises mutation N597P. In some embodiments, the adenosine deaminase comprises mutation N597T. In some embodiments, the adenosine deaminase comprises mutation N597S. In some embodiments, the adenosine deaminase comprises mutation N597W. In some embodiments, the adenosine deaminase comprises mutation N597Q. In some embodiments, the adenosine deaminase comprises mutation N597D. In certain example embodiments, the mutations at N597 described above are further made in the context of an E488Q background

[0229] In some embodiments, the adenosine deaminase comprises a mutation at serine599 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the serine residue at position 599 is replaced by a threonine residue (S599T).

[0230] In some embodiments, the adenosine deaminase comprises a mutation at asparagine613 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the asparagine residue at position 613 is replaced by a lysine residue (N613K). In some embodiments, the adenosine deaminase comprises a mutation at position 613 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 613 is replaced by an arginine residue (N613R). In some embodiments, the adenosine deaminase comprises a mutation at position 613 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 613 is replaced by an alanine residue (N613A) In some embodiments, the adenosine deaminase comprises a mutation at position 613 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 613 is replaced by a glutamic acid residue (N613E). In some embodiments, the adenosine deaminase comprises mutation N613I. In some embodiments, the adenosine deaminase comprises mutation N613L. In some embodiments, the adenosine deaminase comprises mutation N613V. In some embodiments, the adenosine deaminase comprises mutation N613F. In some embodiments, the adenosine deaminase comprises mutation N613M. In some embodiments, the adenosine deaminase comprises mutation N613C. In some embodiments, the adenosine deaminase comprises mutation N613G. In some embodiments, the adenosine deaminase comprises mutation N613P. In some embodiments, the adenosine deaminase comprises mutation N613T. In some embodiments, the adenosine deaminase comprises mutation N613S. In some embodiments, the adenosine deaminase comprises mutation N613Y.

In some embodiments, the adenosine deaminase comprises mutation N613W. In some embodiments, the adenosine deaminase comprises mutation N613Q. In some embodiments, the adenosine deaminase comprises mutation N613H. In some embodiments, the adenosine deaminase comprises mutation N613D. In some embodiments, the mutations at N613 described above are further made in combination with a E488Q mutation.

[0231] In some embodiments, to improve editing efficiency, the adenosine deaminase may comprise one or more of the mutations: G336D, G487A, G487V, E488Q, E488H, E488R, E488N, E488A, E488S, E488M, T490C, T490S, V493T, V493S, V493A, V493R, V493D, V493P, V493G, N597K, N597R, N597A, N597E, N597H, N597G, N597Y, A589V, S599T, N613K, N613R, N613A, N613E, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above.

[0232] In some embodiments, to reduce editing efficiency, the adenosine deaminase may comprise one or more of the mutations: E488F, E488L, E488W, T490A, T490F, T490Y, T490R, T490K, T490P, T490E, N597F, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In particular embodiments, it can be of interest to use an adenosine deaminase enzyme with reduced efficicay to reduce off-target effects.

[0233] In some embodiments, to reduce off-target effects, the adenosine deaminase comprises one or more of mutations at R348, V351, T375, K376, E396, C451, R455, N473, R474, K475, R477, R481, S486, E488, T490, S495, R510, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase comprises mutation at E488 and one or more additional positions selected from R348, V351, T375, K376, E396, C451, R455, N473, R474, K475, R477, R481, S486, T490, S495, R510. In some embodiments, the adenosine deaminase comprises mutation at T375, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at N473, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at V351, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at E488 and T375, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at E488 and N473, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation E488 and V351, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at E488 and one or more of T375, N473, and V351.

[0234] In some embodiments, to reduce off-target effects, the adenosine deaminase comprises one or more of mutations selected from R348E, V351L, T375G, T375S, R455G, R455S, R455E, N473D, R474E, K475Q, R477E, R481E, S486T, E488Q, T490A, T490S, S495T, and R510E, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase comprises mutation E488Q and one or more additional mutations selected from R348E, V351L, T375G, T375S, R455G, R455S, R455E, N473D, R474E, K475Q, R477E, R481E, S486T, T490A, T490S, S495T, and R510E. In some embodiments, the adenosine deaminase comprises mutation T375G or T375S, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation N473D, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation V351L, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation E488Q, and T375G or T375G, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation E488Q and N473D, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation E488Q and V351L, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation E488Q and one or more of T375G/S, N473D and V351L.

[0235] Crystal structures of the human ADAR2 deaminase domain bound to duplex RNA reveal a protein loop that binds the RNA on the 5' side of the modification site. This 5' binding loop is one contributor to substrate specificity differences between ADAR family members. See Wang et al., Nucleic Acids Res., 44(20):9872-9880 (2016), the content of which is incorporated herein by reference in its entirety. In addition, an ADAR2-specific RNA-binding loop was identified near the enzyme active site. See Mathews et al., Nat. Struct. Mol. Biol., 23(5):426-33 (2016), the content of which is incorporated herein by reference in its entirety. In some embodiments, the adenosine deaminase comprises one or more mutations in the RNA binding loop to improve editing specificity and/or efficiency.

[0236] In some embodiments, the adenosine deaminase comprises a mutation at alanine454 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the alanine residue at position 454 is replaced by a serine residue (A454S). In some embodiments, the alanine residue at position 454 is replaced by a cysteine residue (A454C). In some embodiments, the alanine residue at position 454 is replaced by an aspartic acid residue (A454D).

[0237] In some embodiments, the adenosine deaminase comprises a mutation at arginine455 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 455 is replaced by an alanine residue (R455A). In some embodiments, the arginine residue at position 455 is replaced by a valine residue (R455V). In some embodiments, the arginine residue at position 455 is replaced by a histidine residue (R455H). In some embodiments, the arginine residue at position 455 is replaced by a glycine residue (R455G). In some embodiments, the arginine residue at position 455 is replaced by a serine residue (R455S). In some embodiments, the arginine residue at position 455 is replaced by a glutamic acid residue (R455E). In some embodiments, the adenosine deaminase comprises mutation R455C. In some embodiments, the adenosine deaminase comprises mutation R455I. In some embodiments, the adenosine deaminase comprises mutation R455K. In some embodiments, the adenosine deaminase comprises mutation R455L. In some embodiments, the adenosine deaminase comprises mutation R455M. In some embodiments, the adenosine deaminase comprises mutation R455N. In some embodiments, the adenosine deaminase comprises mutation R455Q. In some embodiments, the adenosine deaminase comprises mutation R455F. In some embodiments, the adenosine deaminase comprises mutation R455W. In some embodiments, the adenosine deaminase comprises mutation R455P. In some embodiments, the adenosine deaminase comprises mutation R455Y. In some embodiments, the adenosine deaminase comprises mutation R455E. In some embodiments, the adenosine deaminase comprises mutation R455D. In some embodiments, the mutations at at R455 described above are further made in combination with a E488Q mutation.

[0238] In some embodiments, the adenosine deaminase comprises a mutation at isoleucine456 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the isoleucine residue at position 456 is replaced by a valine residue (I456V). In some embodiments, the isoleucine residue at position 456 is replaced by a leucine residue (I456L). In some embodiments, the isoleucine residue at position 456 is replaced by an aspartic acid residue (I456D).

[0239] In some embodiments, the adenosine deaminase comprises a mutation at phenylalanine457 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the phenylalanine residue at position 457 is replaced by a tyrosine residue (F457Y). In some embodiments, the phenylalanine residue at position 457 is replaced by an arginine residue (F457R). In some embodiments, the phenylalanine residue at position 457 is replaced by a glutamic acid residue (F457E).

[0240] In some embodiments, the adenosine deaminase comprises a mutation at serine458 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the serine residue at position 458 is replaced by a valine residue (S458V). In some embodiments, the serine residue at position 458 is replaced by a phenylalanine residue (S458F). In some embodiments, the serine residue at position 458 is replaced by a proline residue (S458P). In some embodiments, the adenosine deaminase comprises mutation S458I. In some embodiments, the adenosine deaminase comprises mutation S458L. In some embodiments, the adenosine deaminase comprises mutation S458M. In some embodiments, the adenosine deaminase comprises mutation S458C. In some embodiments, the adenosine deaminase comprises mutation S458A. In some embodiments, the adenosine deaminase comprises mutation S458G. In some embodiments, the adenosine deaminase comprises mutation S458T. In some embodiments, the adenosine deaminase comprises mutation S458Y. In some embodiments, the adenosine deaminase comprises mutation S458W. In some embodiments, the adenosine deaminase comprises mutation S458Q. In some embodiments, the adenosine deaminase comprises mutation S458N. In some embodiments, the adenosine deaminase comprises mutation S458H. In some embodiments, the adenosine deaminase comprises mutation S458E. In some embodiments, the adenosine deaminase comprises mutation S458D. In some embodiments, the adenosine deaminase comprises mutation S458K. In some embodiments, the adenosine deaminase comprises mutation S458R. In some embodiments, the mutations at S458 described above are further made in combination with a E488Q mutation.

[0241] In some embodiments, the adenosine deaminase comprises a mutation at proline459 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the proline residue at position 459 is replaced by a cysteine residue (P459C). In some embodiments, the proline residue at position 459 is replaced by a histidine residue (P459H). In some embodiments, the proline residue at position 459 is replaced by a tryptophan residue (P459W).

[0242] In some embodiments, the adenosine deaminase comprises a mutation at histidine460 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the histidine residue at position 460 is replaced by an arginine residue (H460R). In some embodiments, the histidine residue at position 460 is replaced by an isoleucine residue (H460I). In some embodiments, the histidine residue at position 460 is replaced by a proline residue (H460P). In some embodiments, the adenosine deaminase comprises mutation H460L. In some embodiments, the adenosine

deaminase comprises mutation H460V. In some embodiments, the adenosine deaminase comprises mutation H460F. In some embodiments, the adenosine deaminase comprises mutation H460M. In some embodiments, the adenosine deaminase comprises mutation H460C. In some embodiments, the adenosine deaminase comprises mutation H460A. In some embodiments, the adenosine deaminase comprises mutation H460G. In some embodiments, the adenosine deaminase comprises mutation H460T. In some embodiments, the adenosine deaminase comprises mutation H460S. In some embodiments, the adenosine deaminase comprises mutation H460Y. In some embodiments, the adenosine deaminase comprises mutation H460W. In some embodiments, the adenosine deaminase comprises mutation H460Q. In some embodiments, the adenosine deaminase comprises mutation H460N. In some embodiments, the adenosine deaminase comprises mutation H460E. In some embodiments, the adenosine deaminase comprises mutation H460D. In some embodiments, the adenosine deaminase comprises mutation H460K. In some embodiments, the mutations at H460 described above are further made in combination with a E488Q mutation.

[0243] In some embodiments, the adenosine deaminase comprises a mutation at proline462 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the proline residue at position 462 is replaced by a serine residue (P462S). In some embodiments, the proline residue at position 462 is replaced by a tryptophan residue (P462W). In some embodiments, the proline residue at position 462 is replaced by a glutamic acid residue (P462E).

[0244] In some embodiments, the adenosine deaminase comprises a mutation at aspartic acid469 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the aspartic acid residue at position 469 is replaced by a glutamine residue (D469Q). In some embodiments, the aspartic acid residue at position 469 is replaced by a serine residue (D469S). In some embodiments, the aspartic acid residue at position 469 is replaced by a tyrosine residue (D469Y).

[0245] In some embodiments, the adenosine deaminase comprises a mutation at arginine470 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 470 is replaced by an alanine residue (R470A). In some embodiments, the arginine residue at position 470 is replaced by an isoleucine residue (R470I). In some embodiments, the arginine residue at position 470 is replaced by an aspartic acid residue (R470D).

[0246] In some embodiments, the adenosine deaminase comprises a mutation at histidine471 of the hADAR2-D amino acid sequence, or a corresponding position in a

homologous ADAR protein. In some embodiments, the histidine residue at position 471 is replaced by a lysine residue (H471K). In some embodiments, the histidine residue at position 471 is replaced by a threonine residue (H471T). In some embodiments, the histidine residue at position 471 is replaced by a valine residue (H471V).

[0247] In some embodiments, the adenosine deaminase comprises a mutation at proline472 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the proline residue at position 472 is replaced by a lysine residue (P472K). In some embodiments, the proline residue at position 472 is replaced by a threonine residue (P472T). In some embodiments, the proline residue at position 472 is replaced by an aspartic acid residue (P472D).

[0248] In some embodiments, the adenosine deaminase comprises a mutation at asparagine473 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the asparagine residue at position 473 is replaced by an arginine residue (N473R). In some embodiments, the asparagine residue at position 473 is replaced by a tryptophan residue (N473W). In some embodiments, the asparagine residue at position 473 is replaced by a proline residue (N473P). In some embodiments, the asparagine residue at position 473 is replaced by an aspartic acid residue (N473D).

[0249] In some embodiments, the adenosine deaminase comprises a mutation at arginine474 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 474 is replaced by a lysine residue (R474K). In some embodiments, the arginine residue at position 474 is replaced by a glycine residue (R474G). In some embodiments, the arginine residue at position 474 is replaced by an aspartic acid residue (R474D). In some embodiments, the arginine residue at position 474 is replaced by a glutamic acid residue (R474E).

[0250] In some embodiments, the adenosine deaminase comprises a mutation at lysine475 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the lysine residue at position 475 is replaced by a glutamine residue (K475Q). In some embodiments, the lysine residue at position 475 is replaced by an asparagine residue (K475N). In some embodiments, the lysine residue at position 475 is replaced by an aspartic acid residue (K475D).

[0251] In some embodiments, the adenosine deaminase comprises a mutation at alanine476 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the alanine residue at position 476 is

replaced by a serine residue (A476S). In some embodiments, the alanine residue at position 476 is replaced by an arginine residue (A476R). In some embodiments, the alanine residue at position 476 is replaced by a glutamic acid residue (A476E).

[0252] In some embodiments, the adenosine deaminase comprises a mutation at arginine477 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 477 is replaced by a lysine residue (R477K). In some embodiments, the arginine residue at position 477 is replaced by a threonine residue (R477T). In some embodiments, the arginine residue at position 477 is replaced by a phenylalanine residue (R477F). In some embodiments, the arginine residue at position 474 is replaced by a glutamic acid residue (R477E).

[0253] In some embodiments, the adenosine deaminase comprises a mutation at glycine478 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 478 is replaced by an alanine residue (G478A). In some embodiments, the glycine residue at position 478 is replaced by an arginine residue (G478R). In some embodiments, the glycine residue at position 478 is replaced by a tyrosine residue (G478Y). In some embodiments, the adenosine deaminase comprises mutation G478I. In some embodiments, the adenosine deaminase comprises mutation G478L. In some embodiments, the adenosine deaminase comprises mutation G478V. In some embodiments, the adenosine deaminase comprises mutation G478F. In some embodiments, the adenosine deaminase comprises mutation G478M. In some embodiments, the adenosine deaminase comprises mutation G478C. In some embodiments, the adenosine deaminase comprises mutation G478P. In some embodiments, the adenosine deaminase comprises mutation G478T. In some embodiments, the adenosine deaminase comprises mutation G478S. In some embodiments, the adenosine deaminase comprises mutation G478W. In some embodiments, the adenosine deaminase comprises mutation G478Q. In some embodiments, the adenosine deaminase comprises mutation G478N. In some embodiments, the adenosine deaminase comprises mutation G478H. In some embodiments, the adenosine deaminase comprises mutation G478E. In some embodiments, the adenosine deaminase comprises mutation G478D. In some embodiments, the adenosine deaminase comprises mutation G478K. In some embodiments, the mutations at G478 described above are further made in combination with a E488Q mutation.

[0254] In some embodiments, the adenosine deaminase comprises a mutation at glutamine479 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glutamine residue at position 479 is

replaced by an asparagine residue (Q479N). In some embodiments, the glutamine residue at position 479 is replaced by a serine residue (Q479S). In some embodiments, the glutamine residue at position 479 is replaced by a proline residue (Q479P).

[0255] In some embodiments, the adenosine deaminase comprises a mutation at arginine348 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 348 is replaced by an alanine residue (R348A). In some embodiments, the arginine residue at position 348 is replaced by a glutamic acid residue (R348E).

[0256] In some embodiments, the adenosine deaminase comprises a mutation at valine351 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the valine residue at position 351 is replaced by a leucine residue (V351L). In some embodiments, the adenosine deaminase comprises mutation V351Y. In some embodiments, the adenosine deaminase comprises mutation V351M. In some embodiments, the adenosine deaminase comprises mutation V351T. In some embodiments, the adenosine deaminase comprises mutation V351G. In some embodiments, the adenosine deaminase comprises mutation V351A. In some embodiments, the adenosine deaminase comprises mutation V351F. In some embodiments, the adenosine deaminase comprises mutation V351E. In some embodiments, the adenosine deaminase comprises mutation V351I. In some embodiments, the adenosine deaminase comprises mutation V351C. In some embodiments, the adenosine deaminase comprises mutation V351H. In some embodiments, the adenosine deaminase comprises mutation V351P. In some embodiments, the adenosine deaminase comprises mutation V351S. In some embodiments, the adenosine deaminase comprises mutation V351K. In some embodiments, the adenosine deaminase comprises mutation V351N. In some embodiments, the adenosine deaminase comprises mutation V351W. In some embodiments, the adenosine deaminase comprises mutation V351Q. In some embodiments, the adenosine deaminase comprises mutation V351D. In some embodiments, the adenosine deaminase comprises mutation V351R. In some embodiments, the mutations at V351 described above are further made in combination with a E488Q mutation.

[0257] In some embodiments, the adenosine deaminase comprises a mutation at threonine375 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the threonine residue at position 375 is replaced by a glycine residue (T375G). In some embodiments, the threonine residue at position 375 is replaced by a serine residue (T375S). In some embodiments, the adenosine deaminase

comprises mutation T375H. In some embodiments, the adenosine deaminase comprises mutation T375Q. In some embodiments, the adenosine deaminase comprises mutation T375C. In some embodiments, the adenosine deaminase comprises mutation T375N. In some embodiments, the adenosine deaminase comprises mutation T375M. In some embodiments, the adenosine deaminase comprises mutation T375A. In some embodiments, the adenosine deaminase comprises mutation T375W. In some embodiments, the adenosine deaminase comprises mutation T375V. In some embodiments, the adenosine deaminase comprises mutation T375R. In some embodiments, the adenosine deaminase comprises mutation T375E. In some embodiments, the adenosine deaminase comprises mutation T375K. In some embodiments, the adenosine deaminase comprises mutation T375F. In some embodiments, the adenosine deaminase comprises mutation T375I. In some embodiments, the adenosine deaminase comprises mutation T375D. In some embodiments, the adenosine deaminase comprises mutation T375P. In some embodiments, the adenosine deaminase comprises mutation T375L. In some embodiments, the adenosine deaminase comprises mutation T375Y. In some embodiments, the mutations at T375Y described above are further made in combination with an E488Q mutation.

[0258] In some embodiments, the adenosine deaminase comprises a mutation at arginine481 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 481 is replaced by a glutamic acid residue (R481E).

[0259] In some embodiments, the adenosine deaminase comprises a mutation at serine486 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the serine residue at position 486 is replaced by a threonine residue (S486T).

[0260] In some embodiments, the adenosine deaminase comprises a mutation at threonine490 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the threonine residue at position 490 is replaced by an alanine residue (T490A). In some embodiments, the threonine residue at position 490 is replaced by a serine residue (T490S).

[0261] In some embodiments, the adenosine deaminase comprises a mutation at serine495 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the serine residue at position 495 is replaced by a threonine residue (S495T).

[0262] In some embodiments, the adenosine deaminase comprises a mutation at arginine510 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 510 is replaced by a glutamine residue (R510Q). In some embodiments, the arginine residue at position 510 is replaced by an alanine residue (R510A). In some embodiments, the arginine residue at position 510 is replaced by a glutamic acid residue (R510E).

[0263] In some embodiments, the adenosine deaminase comprises a mutation at glycine593 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 593 is replaced by an alanine residue (G593A). In some embodiments, the glycine residue at position 593 is replaced by a glutamic acid residue (G593E).

[0264] In some embodiments, the adenosine deaminase comprises a mutation at lysine594 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the lysine residue at position 594 is replaced by an alanine residue (K594A).

[0265] In some embodiments, the adenosine deaminase comprises a mutation at any one or more of positions A454, R455, I456, F457, S458, P459, H460, P462, D469, R470, H471, P472, N473, R474, K475, A476, R477, G478, Q479, R348, R510, G593, K594 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein.

[0266] In some embodiments, the adenosine deaminase comprises any one or more of mutations A454S, A454C, A454D, R455A, R455V, R455H, I456V, I456L, I456D, F457Y, F457R, F457E, S458V, S458F, S458P, P459C, P459H, P459W, H460R, H460I, H460P, P462S, P462W, P462E, D469Q, D469S, D469Y, R470A, R470I, R470D, H471K, H471T, H471V, P472K, P472T, P472D, N473R, N473W, N473P, R474K, R474G, R474D, K475Q, K475N, K475D, A476S, A476R, A476E, R477K, R477T, R477F, G478A, G478R, G478Y, Q479N, Q479S, Q479P, R348A, R510Q, R510A, G593A, G593E, K594A of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein.

[0267] In some embodiments, the adenosine deaminase comprises a mutation at any one or more of positions T375, V351, G478, S458, H460 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein, optionally in combination a mutation at E488. In some embodiments, the adenosine deaminase comprises one or more of mutations selected from T375G, T375C, T375H, T375Q, V351M, V351T, V351Y, G478R, S458F, H460I, optionally in combination with E488Q.

[0268] In some embodiments, the adenosine deaminase comprises one or more of mutations selected from T375H, T375Q, V351M, V351Y, H460P, optionally in combination with E488Q.

[0269] In some embodiments, the adenosine deaminase comprises mutations T375S and S458F, optionally in combination with E488Q.

[0270] In some embodiments, the adenosine deaminase comprises a mutation at two or more of positions T375, N473, R474, G478, S458, P459, V351, R455, R455, T490, R348, Q479 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein, optionally in combination a mutation at E488. In some embodiments, the adenosine deaminase comprises two or more of mutations selected from T375G, T375S, N473D, R474E, G478R, S458F, P459W, V351L, R455G, R455S, T490A, R348E, Q479P, optionally in combination with E488Q.

[0271] In some embodiments, the adenosine deaminase comprises mutations T375G and V351L. In some embodiments, the adenosine deaminase comprises mutations T375G and R455G. In some embodiments, the adenosine deaminase comprises mutations T375G and R455S. In some embodiments, the adenosine deaminase comprises mutations T375G and T490A. In some embodiments, the adenosine deaminase comprises mutations T375G and R348E. In some embodiments, the adenosine deaminase comprises mutations T375S and V351L. In some embodiments, the adenosine deaminase comprises mutations T375S and R455G. In some embodiments, the adenosine deaminase comprises mutations T375S and R455S. In some embodiments, the adenosine deaminase comprises mutations T375S and T490A. In some embodiments, the adenosine deaminase comprises mutations T375S and R348E. In some embodiments, the adenosine deaminase comprises mutations N473D and V351L. In some embodiments, the adenosine deaminase comprises mutations N473D and R455G. In some embodiments, the adenosine deaminase comprises mutations N473D and R455S. In some embodiments, the adenosine deaminase comprises mutations N473D and T490A. In some embodiments, the adenosine deaminase comprises mutations N473D and R348E. In some embodiments, the adenosine deaminase comprises mutations R474E and V351L. In some embodiments, the adenosine deaminase comprises mutations R474E and R455G. In some embodiments, the adenosine deaminase comprises mutations R474E and R455S. In some embodiments, the adenosine deaminase comprises mutations R474E and T490A. In some embodiments, the adenosine deaminase comprises mutations R474E and R348E. In some embodiments, the adenosine deaminase comprises mutations S458F and T375G. In some embodiments, the adenosine deaminase comprises mutations S458F and

T375S. In some embodiments, the adenosine deaminase comprises mutations S458F and N473D. In some embodiments, the adenosine deaminase comprises mutations S458F and R474E. In some embodiments, the adenosine deaminase comprises mutations S458F and G478R. In some embodiments, the adenosine deaminase comprises mutations G478R and T375G. In some embodiments, the adenosine deaminase comprises mutations G478R and T375S. In some embodiments, the adenosine deaminase comprises mutations G478R and N473D. In some embodiments, the adenosine deaminase comprises mutations G478R and R474E. In some embodiments, the adenosine deaminase comprises mutations P459W and T375G. In some embodiments, the adenosine deaminase comprises mutations P459W and T375S. In some embodiments, the adenosine deaminase comprises mutations P459W and N473D. In some embodiments, the adenosine deaminase comprises mutations P459W and R474E. In some embodiments, the adenosine deaminase comprises mutations P459W and G478R. In some embodiments, the adenosine deaminase comprises mutations Q479P and T375G. In some embodiments, the adenosine deaminase comprises mutations Q479P and T375S. In some embodiments, the adenosine deaminase comprises mutations Q479P and N473D. In some embodiments, the adenosine deaminase comprises mutations Q479P and R474E. In some embodiments, the adenosine deaminase comprises mutations Q479P and G478R. In some embodiments, the adenosine deaminase comprises mutations Q479P and S458F. In some embodiments, the adenosine deaminase comprises mutations Q479P and P459W. All mutations described in this paragraph may also further be made in cominbation with a E488Q mutations.

[0272] In some embodiments, the adenosine deaminase comprises a mutation at any one or more of positions K475, Q479, P459, G478, S458of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein, optionally in combination a mutation at E488. In some embodiments, the adenosine deaminase comprises one or more of mutations selected from K475N, Q479N, P459W, G478R, S458P, S458F, optionally in combination with E488Q.

[0273] In some embodiments, the adenosine deaminase comprises a mutation at any one or more of positions T375, V351, R455, H460, A476 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein, optionally in combination a mutation at E488. In some embodiments, the adenosine deaminase comprises one or more of mutations selected from T375G, T375C, T375H, T375Q, V351M, V351T, V351Y, R455H, H460P, H460I, A476E, optionally in combination with E488Q.

[0274] In certain embodiments, improvement of editing and reduction of off-target modification is achieved by chemical modification of gRNAs. gRNAs which are chemically modified as exemplified in Vogel et al. (2014), Angew Chem Int Ed, 53:6267-6271, doi:10.1002/anie.201402634 (incorporated herein by reference in its entirety) reduce off-target activity and improve on-target efficiency. 2′-O-methyl and phosphothioate modified guide RNAs in general improve editing efficiency in cells.

[0275] ADAR has been known to demonstrate a preference for neighboring nucleotides on either side of the edited A (www.nature.com/nsmb/journal/v23/n5/full/nsmb.3203.html, Matthews et al. (2017), Nature Structural Mol Biol, 23(5): 426-433, incorporated herein by reference in its entirety). Accordingly, in certain embodiments, the gRNA, target, and/or ADAR is selected optimized for motif preference.

[0276] Intentional mismatches have been demonstrated in vitro to allow for editing of non-preferred motifs (https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku272; Schneider et al (2014), Nucleic Acid Res, 42(10):e87); Fukuda et al. (2017), Scienticic Reports, 7, doi:10.1038/srep41478, incorporated herein by reference in its entirety). Accordingly, in certain embodiments, to enhance RNA editing efficiency on non-preferred 5’ or 3’ neighboring bases, intentional mismatches in neighboring bases are introduced.

[0277] Results suggest that A’s opposite C’s in the targeting window of the ADAR deaminase domain are preferentially edited over other bases. Additionally, A’s base-paired with U’s within a few bases of the targeted base show low levels of editing by Cas13b-ADAR fusions, suggesting that there is flexibility for the enzyme to edit multiple A’s. See e.g. FIG. 18. These two observations suggest that multiple A’s in the activity window of Cas13b-ADAR fusions could be specified for editing by mismatching all A’s to be edited with C’s. Accordingly, in certain embodiments, multiple A:C mismatches in the activity window are designed to create multiple A:I edits. In certain embodiments, to suppress potential off-target editing in the activity window, non-target A’s are paired with A’s or G’s.

[0278] The terms “editing specificity” and “editing preference” are used interchangeably herein to refer to the extent of A-to-I editing at a particular adenosine site in a double-stranded substrate. In some embodiment, the substrate editing preference is determined by the 5’ nearest neighbor and/or the 3’ nearest neighbor of the target adenosine residue. In some embodiments, the adenosine deaminase has preference for the 5’ nearest neighbor of the substrate ranked as U>A>C>G (“>” indicates greater preference). In some embodiments, the adenosine deaminase has preference for the 3’ nearest neighbor of the substrate ranked as G>C~A>U (“>” indicates greater preference; “~” indicates similar preference). In some embodiments, the adenosine

deaminase has preference for the 3' nearest neighbor of the substrate ranked as G>C>U~A (">" indicates greater preference; "~" indicates similar preference). In some embodiments, the adenosine deaminase has preference for the 3' nearest neighbor of the substrate ranked as G>C>A>U (">" indicates greater preference). In some embodiments, the adenosine deaminase has preference for the 3' nearest neighbor of the substrate ranked as C~G~A>U (">" indicates greater preference; "~" indicates similar preference). In some embodiments, the adenosine deaminase has preference for a triplet sequence containing the target adenosine residue ranked as TAG>AAG>CAC>AAT>GAA>GAC (">" indicates greater preference), the center A being the target adenosine residue.

[0279] In some embodiments, the substrate editing preference of an adenosine deaminase is affected by the presence or absence of a nucleic acid binding domain in the adenosine deaminase protein. In some embodiments, to modify substrate editing preference, the deaminase domain is connected with a double-strand RNA binding domain (dsRBD) or a double-strand RNA binding motif (dsRBM). In some embodiments, the dsRBD or dsRBM may be derived from an ADAR protein, such as hADAR1 or hADAR2. In some embodiments, a full length ADAR protein that comprises at least one dsRBD and a deaminase domain is used. In some embodiments, the one or more dsRBM or dsRBD is at the N-terminus of the deaminase domain. In other embodiments, the one or more dsRBM or dsRBD is at the C-terminus of the deaminase domain.

[0280] In some embodiments, the substrate editing preference of an adenosine deaminase is affected by amino acid residues near or in the active center of the enzyme. In some embodiments, to modify substrate editing preference, the adenosine deaminase may comprise one or more of the mutations: G336D, G487R, G487K, G487W, G487Y, E488Q, E488N, T490A, V493A, V493T, V493S, N597K, N597R, A589V, S599T, N613K, N613R, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above.

[0281] Particularly, in some embodiments, to reduce editing specificity, the adenosine deaminase can comprise one or more of mutations E488Q, V493A, N597K, N613K, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, to increase editing specificity, the adenosine deaminase can comprise mutation T490A.

[0282] In some embodiments, to increase editing preference for target adenosine (A) with an immediate 5' G, such as substrates comprising the triplet sequence GAC, the center A being the target adenosine residue, the adenosine deaminase can comprise one or more of mutations

G336D, E488Q, E488N, V493T, V493S, V493A, A589V, N597K, N597R, S599T, N613K, N613R, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above.

[0283] Particularly, in some embodiments, the adenosine deaminase comprises mutation E488Q or a corresponding mutation in a homologous ADAR protein for editing substrates comprising the following triplet sequences: GAC, GAA, GAU, GAG, CAU, AAU, UAC, the center A being the target adenosine residue.

[0284] In some embodiments, the adenosine deaminase comprises the wild-type amino acid sequence of hADAR1-D as defined in SEQ ID No. 704. In some embodiments, the adenosine deaminase comprises one or more mutations in the hADAR1-D sequence, such that the editing efficiency, and/or substrate editing preference of hADAR1-D is changed according to specific needs.

[0285] In some embodiments, the adenosine deaminase comprises a mutation at Glycine1007 of the hADAR1-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 1007 is replaced by a non-polar amino acid residue with relatively small side chains. For example, in some embodiments, the glycine residue at position 1007 is replaced by an alanine residue (G1007A). In some embodiments, the glycine residue at position 1007 is replaced by a valine residue (G1007V). In some embodiments, the glycine residue at position 1007 is replaced by an amino acid residue with relatively large side chains. In some embodiments, the glycine residue at position 1007 is replaced by an arginine residue (G1007R). In some embodiments, the glycine residue at position 1007 is replaced by a lysine residue (G1007K). In some embodiments, the glycine residue at position 1007 is replaced by a tryptophan residue (G1007W). In some embodiments, the glycine residue at position 1007 is replaced by a tyrosine residue (G1007Y). Additionally, in other embodiments, the glycine residue at position 1007 is replaced by a leucine residue (G1007L). In other embodiments, the glycine residue at position 1007 is replaced by a threonine residue (G1007T). In other embodiments, the glycine residue at position 1007 is replaced by a serine residue (G1007S).

[0286] In some embodiments, the adenosine deaminase comprises a mutation at glutamic acid1008 of the hADAR1-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glutamic acid residue at position 1008 is replaced by a polar amino acid residue having a relatively large side chain. In some embodiments, the glutamic acid residue at position 1008 is replaced by a glutamine residue (E1008Q). In some embodiments, the glutamic acid residue at position 1008 is replaced by a histidine residue

(E1008H). In some embodiments, the glutamic acid residue at position 1008 is replaced by an arginine residue (E1008R). In some embodiments, the glutamic acid residue at position 1008 is replaced by a lysine residue (E1008K). In some embodiments, the glutamic acid residue at position 1008 is replaced by a nonpolar or small polar amino acid residue. In some embodiments, the glutamic acid residue at position 1008 is replaced by a phenylalanine residue (E1008F). In some embodiments, the glutamic acid residue at position 1008 is replaced by a tryptophan residue (E1008W). In some embodiments, the glutamic acid residue at position 1008 is replaced by a glycine residue (E1008G). In some embodiments, the glutamic acid residue at position 1008 is replaced by an isoleucine residue (E1008I). In some embodiments, the glutamic acid residue at position 1008 is replaced by a valine residue (E1008V). In some embodiments, the glutamic acid residue at position 1008 is replaced by a proline residue (E1008P). In some embodiments, the glutamic acid residue at position 1008 is replaced by a serine residue (E1008S). In other embodiments, the glutamic acid residue at position 1008 is replaced by an asparagine residue (E1008N). In other embodiments, the glutamic acid residue at position 1008 is replaced by an alanine residue (E1008A). In other embodiments, the glutamic acid residue at position 1008 is replaced by a Methionine residue (E1008M). In some embodiments, the glutamic acid residue at position 1008 is replaced by a leucine residue (E1008L).

[0287] In some embodiments, to improve editing efficiency, the adenosine deaminase may comprise one or more of the mutations: E1007S, E1007A, E1007V, E1008Q, E1008R, E1008H, E1008M, E1008N, E1008K, based on amino acid sequence positions of hADAR1-D, and mutations in a homologous ADAR protein corresponding to the above.

[0288] In some embodiments, to reduce editing efficiency, the adenosine deaminase may comprise one or more of the mutations: E1007R, E1007K, E1007Y, E1007L, E1007T, E1008G, E1008I, E1008P, E1008V, E1008F, E1008W, E1008S, E1008N, E1008K, based on amino acid sequence positions of hADAR1-D, and mutations in a homologous ADAR protein corresponding to the above.

[0289] In some embodiments, the substrate editing preference, efficiency and/or selectivity of an adenosine deaminase is affected by amino acid residues near or in the active center of the enzyme. In some embodiments, the adenosine deaminase comprises a mutation at the glutamic acid 1008 position in hADAR1-D sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the mutation is E1008R, or a corresponding mutation in a homologous ADAR protein. In some embodiments, the E1008R mutant has an increased

editing efficiency for target adenosine residue that has a mismatched G residue on the opposite strand.

[0290] In some embodiments, the adenosine deaminase protein further comprises or is connected to one or more double-stranded RNA (dsRNA) binding motifs (dsRBMs) or domains (dsRBDs) for recognizing and binding to double-stranded nucleic acid substrates. In some embodiments, the interaction between the adenosine deaminase and the double–stranded substrate is mediated by one or more additional protein factor(s), including a CRISPR/CAS protein factor. In some embodiments, the interaction between the adenosine deaminase and the double-stranded substrate is further mediated by one or more nucleic acid component(s), including a guide RNA.

Modified Adenosine Deaminase Having C-to U Deamination Activity

[0291] In certain example embodiments, directed evolution may be used to design modified ADAR proteins capable of catalyzing additional reactions besides deamination of an adenine to a hypoxanthine. For example, the modified ADAR protein may be capable of catalyzing deamination of a cytidine to a uracil. While not bound by a particular theory, mutations that improve C to U activity may alter the shape of the binding pocket to be more amenable to the smaller cytidine base.

[0292] In some embodiments, the modified adenosine deaminase having C-to-U deamination activity comprises a mutation at any one or more of positions V351, T375, R455, and E488 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the adenosine deaminase comprises mutation E488Q. In some embodiments, the adenosine deaminase comprises one or more of mutations selected from V351I, V351L, V351F, V351M, V351C, V351A, V351G, V351P, V351T, V351S, V351Y, V351W, V351Q, V351N, V351H, V351E, V351D, V351K, V351R, T375I, T375L, T375V, T375F, T375M, T375C, T375A, T375G, T375P, T375S, T375Y, T375W, T375Q, T375N, T375H, T375E, T375D, T375K, T375R, R455I, R455L, R455V, R455F, R455M, R455C, R455A, R455G, R455P, R455T, R455S, R455Y, R455W, R455Q, R455N, R455H, R455E, R455D, R455K. In some embodiments, the adenosine deaminase comprises mutation E488Q, and further comprises one or more of mutations selected from V351I, V351L, V351F, V351M, V351C, V351A, V351G, V351P, V351T, V351S, V351Y, V351W, V351Q, V351N, V351H, V351E, V351D, V351K, V351R, T375I, T375L, T375V, T375F, T375M, T375C, T375A, T375G, T375P, T375S, T375Y, T375W, T375Q, T375N, T375H, T375E, T375D, T375K, T375R, R455I, R455L, R455V, R455F, R455M, R455C, R455A, R455G, R455P, R455T, R455S, R455Y, R455W, R455Q, R455N, R455H, R455E, R455D, R455K.

[0293] In connection with the aforementioned modified ADAR protein having C-to-U deamination activity, the invention described herein also relates to a method for deaminating a C in a target RNA sequence of interest, comprising delivering to a target RNA or DNA an AD-functoinalized composition disclosed herein.

[0294] In certain example embodiments, the method for deaminating a C in a target RNA sequencecomprising delivering to said target RNA: (a) a catalytically inactive (dead) Cas; (b) a guide molecule which comprises a guide sequence linked to a direct repeat sequence; and (c) a modified ADAR protein having C-to-U deamination activity or catalytic domain thereof; wherein said modified ADAR protein or catalytic domain thereof is covalently or non-covalently linked to said dead Cas protein or said guide molecule or is adapted to link thereto after delivery; wherein guide molecule forms a complex with said dead Cas protein and directs said complex to bind said target RNA sequence of interest; wherein said guide sequence is capable of hybridizing with a target sequence comprising said C to form an RNA duplex; wherein, optionally, said guide sequence comprises a non-pairing A or U at a position corresponding to said C resulting in a mismatch in the RNA duplex formed; and wherein said modified ADAR protein or catalytic domain thereof deaminates said C in said RNA duplex.

[0295] In connection with the aforementioned modified ADAR protein having C-to-U deamination activity, the invention described herein further relates to an engineered, non-naturally occurring system suitable for deaminating a C in a target locus of interest, comprising: (a) a guide molecule which comprises a guide sequence linked to a direct repeat sequence, or a nucleotide sequence encoding said guide molecule; (b) a catalytically inactive Cas13 protein, or a nucleotide sequence encoding said catalytically inactive Cas13 protein; (c) a modified ADAR protein having C-to-U deamination activity or catalytic domain thereof, or a nucleotide sequence encoding said modified ADAR protein or catalytic domain thereof; wherein said modified ADAR protein or catalytic domain thereof is covalently or non-covalently linked to said Cas13 protein or said guide molecule or is adapted to link thereto after delivery; wherein said guide sequence is capable of hybridizing with a target RNA sequence comprising a C to form an RNA duplex; wherein, optionally, said guide sequence comprises a non-pairing A or U at a position corresponding to said C resulting in a mismatch in the RNA duplex formed; wherein, optionally, the system is a vector system comprising one or more vectors comprising: (a) a first regulatory element operably linked to a nucleotide sequence encoding said guide molecule which comprises said guide sequence, (b) a second regulatory element operably linked to a nucleotide sequence encoding said catalytically inactive Cas13 protein; and (c) a nucleotide sequence encoding a modified ADAR protein having C-to-U deamination activity

or catalytic domain thereof which is under control of said first or second regulatory element or operably linked to a third regulatory element; wherein, if said nucleotide sequence encoding a modified ADAR protein or catalytic domain thereof is operably linked to a third regulatory element, said modified ADAR protein or catalytic domain thereof is adapted to link to said guide molecule or said Cas13 protein after expression; wherein components (a), (b) and (c) are located on the same or different vectors of the system, optionally wherein said first, second, and/or third regulatory element is an inducible promoter.

[0296] According to the present invention, the substrate of the adenosine deaminase is an RNA/DNAn RNA duplex formed upon binding of the guide molecule to its DNA target which then forms the CRISPR-Cas complex with the CRISPR-Cas enzyme. The substrate of the adenosine deaminase can also be an RNA/RNA duplex formed upon binding of the guide molecule to its RNA target which then forms the CRISPR-Cas complex with the CRISPR-Cas enzyme. The RNA/DNA or DNA/RNAn RNA duplex is also referred to herein as the "RNA/DNA hybrid", "DNA/RNA hybrid" or "double-stranded substrate". The particular features of the guide molecule and CRISPR-Cas enzyme are detailed below.

[0297] The term "editing selectivity" as used herein refers to the fraction of all sites on a double-stranded substrate that is edited by an adenosine deaminase. Without being bound by theory, it is contemplated that editing selectivity of an adenosine deaminase is affected by the double-stranded substrate's length and secondary structures, such as the presence of mismatched bases, bulges and/or internal loops.

[0298] In some embodiments, when the substrate is a perfectly base-paired duplex longer than 50 bp, the adenosine deaminase may be able to deaminate multiple adenosine residues within the duplex (e.g., 50% of all adenosine residues). In some embodiments, when the substrate is shorter than 50 bp, the editing selectivity of an adenosine deaminase is affected by the presence of a mismatch at the target adenosine site. Particularly, in some embodiments, adenosine (A) residue having a mismatched cytidine (C) residue on the opposite strand is deaminated with high efficiency. In some embodiments, adenosine (A) residue having a mismatched guanosine (G) residue on the opposite strand is skipped without editing.

TARGETING DOMAIN

[0299] The methods, tools, and compositions of the invention comprise or make use of a targeting component which can be referred to as a targeting domain. The targeting domain is preferably a DNA or RNA targeting domain, more particularly an oligonucleotide targeting domain, or a variant or fragment theofe which retains DNA and/or RNA binding activity. The oligonucleotide targeting domain may bind a sequence, motif, or structural feature of the RNA

or DNA of interest at or adajacent to the target locus. A structural feature may include hairpins, tetraloops, or other secondary structural features of a nucleic acid. As used herein "adjacent" means within a distance and/or orientation of the target locus in which the adenosine deaminase can complete its base editing function. In certain example embodiments, the oligonucleotide binding protein may be a RNA-binding protein or functional domain thereof, or a DNA-binding protein or functional domain thereof.

[0300] In particular embodiments, the targeting domain further comprises a guide RNA (as will be detailed below). The nucleic acid binding protein can be an (endo)nuclease or any other (oligo)nucleotide binding protein. In particular embodiments, the nucleotide binding protein is modified to inactivate any other function not required for said DNA or RNA binding. In particular embodiments, where the nucleotide binding protein is an (endo)nuclease, preferably the (endo)nuclease has altered or modified activity (i.e. a modified nuclease, as described herein elsewhere) compared to the wildtype DNA or RNA binding protein. In certain embodiments, said nuclease is a targeted or site-specific or homing nuclease or a variant thereof having altered or modified activity. In certain embodiments, said (oligo)nucleotide binding protein is the (oligo)nucleotide binding domain of said (oligo)nucleotide binding protein and does not comprise one or more domains of said protein not required for DNA and/or RNA binding (more particular does not comprise one or more other functional domains).

RNA-binding proteins

[0301] In certain example embodiments, the oligonucleotide binding domain may comprise or consist of a RNA-binding protein, or functional domain thereof, that comprises a RNA recognition motif. Example RNA-binding proteins comprising a RNA recognition motif include, but are not limited to, A2BP1; ACF; BOLL; BRUNOL4; BRUNOL5; BRUNOL6; CCBL2; CGI96; CIRBP; CNOT 4; CPEB2; CPEB3; CPEB4; CPSF7; CSTF2; CSTF2T; CUGBP1; CUGBP2; D10S102; DAZ 1; DAZ2; DAZ3; DAZ4; DAZAP1; DAZL; DNAJC17; DND1; EIF3S4; EIF3S9; EIF4B; EI F4H; ELAVL1; ELAVL2; ELAVL3; ELAVL4; ENOX1; ENOX2; EWSR1; FUS; FUSIP1; G3BP; G3BP1; G3BP2; GRSF1; HNRNPL; HNRPA0; HNRPA1; HNRPA2B1; HNRPA3; H NRPAB; HNRPC; HNRPCL1; HNRPD; HNRPDL; HNRPF; HNRPH1; HNRPH2; HNRPH 3; HNRPL; HNRPLL; HNRPM; HNRPR; HRNBP1; HSU53209; HTATSF1; IGF2BP1; IGF 2BP2; IGF2BP3; LARP7; MKI67IP; MSI1; MSI2; MSSP2; MTHFSD; MYEF2; NCBP2; N CL; NOL8; NONO; P14; PABPC1; PABPC1L; PABPC3; PABPC4; PABPC5; PABPN1; PO LDIP3; PPARGC1; PPARGC1A; PPARGC1B; PPIE; PPIL4; PPRC1; PSPC1; PTBP1; PTB P2; PUF60; RALY; RALYL; RAVER1; RAVER2; RBM10; RBM11; RBM12; RBM12B; R

BM14; RBM15; RBM15B; RBM16; RBM17; RBM18; RBM19; RBM22; RBM23; RBM24; RBM25; RBM26; RBM27; RBM28; RBM3; RBM32B; RBM33; RBM34; RBM35A; RBM35B; RBM38; RBM39; RBM4; RBM41; RBM42; RBM44; RBM45; RBM46; RBM47; RBM4B; RBM5; RBM7; RBM8A; RBM9; RBMS1; RBMS2; RBMS3; RBMX; RBMX2; RBMXL2; RBMY1A1; RBMY1B; RBMY1E; RBMY1F; RBMY2FP; RBPMS; RBPMS2; RDBP; RNPC3; RNPC4; RNPS1; ROD1; SAFB; SAFB2; SART3; SETD1A; SF3B14; SF3B4; SFPQ; SFRS1; SFRS10; SFRS11; SFRS12; SFRS15; SFRS2; SFRS2B; SFRS3; SFRS4; SFRS5; SFRS6; SFRS7; SFRS9; SLIRP; SLTM; SNRP70; SNRPA; SNRPB2; SPEN; SR140; SRRP35; SSB; SYNCRIP; TAF15; TARDBP; THOC4; TIA1; TIAL1; TNRC4; TNRC6C; TRA2A; TRSPAP1; TUT1; U1SNRNPBP; U2AF1; U2AF2; UHMK1; ZCRB1; ZNF638; ZRSR1; and ZRSR2.

[0302] In certain example embodiments, the RNA-binding protein or function domain thereof may comprise a K homology domain. Example RNA-binding proteins comprising a K homology domain include, but are not limited to, AKAP1; ANKHD1; ANKRD17; ASCC1; BICC1; DDX43; DDX53; DPPA5; FMR1; FUBP1; FUBP3; FXR1; FXR2; GLD1; HDLBP; HNRPK; IGF2BP1; IGF2BP2; IGF2BP3; KHDRBS1; KHDRBS2; KHDRBS3; KHSRP; KRR1; MEX3A; MEX3B; MEX3C; MEX3D; NOVA1; NOVA2; PCBP1; PCBP2; PCBP3; PCBP4; PNO1; PNPT1; QKI; SF1; and TDRKH

[0303] In certain example embodiments, the RNA-binding protein comprises a zinc finger motif. RNA-binding proteins or functional domains thereof may comprise a Cys2-His2, Gag-knuckle, Treble-clet, Zinc ribbon, Zn2/Cys6 class motif.

[0304] In certain example embodiments, the RNA-binding protein may comprise a Pumilio homology domain.

TALENS

[0305] In certain embodiments, the nucleic acid binding protein is a (modified) transcription activator-like effector nuclease (TALEN) system. Transcription activator-like effectors (TALEs) can be engineered to bind practically any desired DNA sequence. Exemplary methods of genome editing using the TALEN system can be found for example in Cermak T. Doyle EL. Christian M. Wang L. Zhang Y. Schmidt C, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 2011;39:e82; Zhang F. Cong L. Lodato S. Kosuri S. Church GM. Arlotta P Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol. 2011;29:149–153 and US Patent Nos. 8,450,471, 8,440,431 and 8,440,432, all of which are specifically incorporated by reference. By means of further

guidance, and without limitation, naturally occurring TALEs or "wild type TALEs" are nucleic acid binding proteins secreted by numerous species of proteobacteria. TALE polypeptides contain a nucleic acid binding domain composed of tandem repeats of highly conserved monomer polypeptides that are predominantly 33, 34 or 35 amino acids in length and that differ from each other mainly in amino acid positions 12 and 13. In advantageous embodiments the nucleic acid is DNA. As used herein, the term "polypeptide monomers", or "TALE monomers" will be used to refer to the highly conserved repetitive polypeptide sequences within the TALE nucleic acid binding domain and the term "repeat variable di-residues" or "RVD" will be used to refer to the highly variable amino acids at positions 12 and 13 of the polypeptide monomers. As provided throughout the disclosure, the amino acid residues of the RVD are depicted using the IUPAC single letter code for amino acids. A general representation of a TALE monomer which is comprised within the DNA binding domain is X1-11-(X12X13)-X14-33 or 34 or 35, where the subscript indicates the amino acid position and X represents any amino acid. X12X13 indicate the RVDs. In some polypeptide monomers, the variable amino acid at position 13 is missing or absent and in such polypeptide monomers, the RVD consists of a single amino acid. In such cases the RVD may be alternatively represented as X*, where X represents X12 and (*) indicates that X13 is absent. The DNA binding domain comprises several repeats of TALE monomers and this may be represented as (X1-11-(X12X13)-X14-33 or 34 or 35)z, where in an advantageous embodiment, z is at least 5 to 40. In a further advantageous embodiment, z is at least 10 to 26. The TALE monomers have a nucleotide binding affinity that is determined by the identity of the amino acids in its RVD. For example, polypeptide monomers with an RVD of NI preferentially bind to adenine (A), polypeptide monomers with an RVD of NG preferentially bind to thymine (T), polypeptide monomers with an RVD of HD preferentially bind to cytosine (C) and polypeptide monomers with an RVD of NN preferentially bind to both adenine (A) and guanine (G). In yet another embodiment of the invention, polypeptide monomers with an RVD of IG preferentially bind to T. Thus, the number and order of the polypeptide monomer repeats in the nucleic acid binding domain of a TALE determines its nucleic acid target specificity. In still further embodiments of the invention, polypeptide monomers with an RVD of NS recognize all four base pairs and may bind to A, T, G or C. The structure and function of TALEs is further described in, for example, Moscou et al., Science 326:1501 (2009); Boch et al., Science 326:1509-1512 (2009); and Zhang et al., Nature Biotechnology 29:149-153 (2011), each of which is incorporated by reference in its entirety. In certain embodiments, targeting is effected by a polynucleic acid binding TALEN fragment.

In certain embodiments, the targeting domain comprises or consists of a catalytically inactive TALEN or nucleic acid binding fragment thereof.

Zn-Finger Nucleases

[0306] In certain embodiments, the targeting domain comprises or consists of a (modified) zinc-finger nuclease (ZFN) system. The ZFN system uses artificial restriction enzymes generated by fusing a zinc finger DNA-binding domain to a DNA-cleavage domain that can be engineered to target desired DNA sequences. Exemplary methods of genome editing using ZFNs can be found for example in U.S. Patent Nos. 6,534,261, 6,607,882, 6,746,838, 6,794,136, 6,824,978, 6,866,997, 6,933,113, 6,979,539, 7,013,219, 7,030,215, 7,220,719, 7,241,573, 7,241,574, 7,585,849, 7,595,376, 6,903,185, and 6,479,626, all of which are specifically incorporated by reference. By means of further guidance, and without limitation, artificial zinc-finger (ZF) technology involves arrays of ZF modules to target new DNA-binding sites in the genome. Each finger module in a ZF array targets three DNA bases. A customized array of individual zinc finger domains is assembled into a ZF protein (ZFP). ZFPs can comprise a functional domain. The first synthetic zinc finger nucleases (ZFNs) were developed by fusing a ZF protein to the catalytic domain of the Type IIS restriction enzyme FokI. (Kim, Y. G. et al., 1994, Chimeric restriction endonuclease, Proc. Natl. Acad. Sci. U.S.A. 91, 883−887; Kim, Y. G. et al., 1996, Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc. Natl. Acad. Sci. U.S.A. 93, 1156−1160). Increased cleavage specificity can be attained with decreased off target activity by use of paired ZFN heterodimers, each targeting different nucleotide sequences separated by a short spacer. (Doyon, Y. et al., 2011, Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat. Methods 8, 74−79). ZFPs can also be designed as transcription activators and repressors and have been used to target many genes in a wide variety of organisms. In certain embodiments, the targeting domain comprises or consists of a nucleic acid binding zinc finger nuclease or a nucleic acid binding fragment thereof. In certain embodiments, the nucleic acid binding (fragment of) a zinc finger nuclease is catalytically inactive.

Meganuclease

[0307] In certain embodiments, the targeting domain comprises a (modified) meganuclease, which are endodeoxyribonucleases characterized by a large recognition site (double-stranded DNA sequences of 12 to 40 base pairs). Exemplary method for using meganucleases can be found in US Patent Nos: 8,163,514; 8,133,697; 8,021,867; 8,119,361; 8,119,381; 8,124,369; and 8,129,134, which are specifically incorporated by reference. In certain embodiments, targeting is effected by a polynucleic acid binding meganuclease

fragment. In certain embodiments, targeting is effected by a polynucleic acid binding catalytically inactive meganuclease (fragment). Accordingly in particular embodiments, the targeting domain comprises or consists of a nucleic acid binding meganuclease or a nucleic acid binding fragment thereof.

CRISPR-Cas Systems

[0308] In certain embodiments, the targeting domain comprises a (modified) CRISPR/Cas complex or system. General information on CRISPR/Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, and making and using thereof, including as to amounts and formulations, as well as CRISPR/Cas-expressing eukaryotic cells, CRISPR/Cas expressing eukaryotes, such as a mouse, is described herein elsewhere. In certain embodiments, targeting is effected by an oligonucleic acid binding CRISPR protein fragment and/or a gRNA. In certain embodiments, targeting is effected by a nucleic acid binding catalytically inactive CRISPR protein (fragment). Accordingly in particular embodiments, the targeting domain comprises oligonucleic acid binding CRISPR protein or an oligonucleic acid binding fragment of a CRISPR protein and/or a gRNA.

[0309] As used herein, the term "Cas" generally refers to a (modified) effector protein of the CRISPR/Cas system or complex, and can be without limitation a (modified) Cas9, or other enzymes such as Cpf1, C2c1, C2c2, C2c3, group 29, or group 30 protein The term "Cas" may be used herein interchangeably with the terms "CRISPR" protein, "CRISPR/Cas protein", "CRISPR effector", "CRISPR/Cas effector", "CRISPR enzyme", "CRISPR/Cas enzyme" and the like, unless otherwise apparent, such as by specific and exclusive reference to Cas9. It is to be understood that the term "CRISPR protein" may be used interchangeably with "CRISPR enzyme", irrespective of whether the CRISPR protein has altered, such as increased or decreased (or no) enzymatic activity, compared to the wild type CRISPR protein. Likewise, as used herein, in certain embodiments, where appropriate and which will be apparent to the skilled person, the term "nuclease" may refer to a modified nuclease wherein catalytic activity has been altered, such as having increased or decreased nuclease activity, or no nuclease activity at all, as well as nickase activity, as well as otherwise modified nuclease as defined herein elsewhere, unless otherwise apparent, such as by specific and exclusive reference to unmodified nuclease.

[0310] In some embodiments, the CRISPR effector protein is Cas9, Cpf1, C2c1, C2c2, or Cas13a, Cas13b, Cas13c, or Cas13d. In some embodiments, the CRISPR effector protein is a DNA-targeting CRISPR effector protein. In some embodiments, the CRISPR effector protein

is a Type-II CRISPR effector protein such as Cas9. In some embodiments, the CRISPR effector protein is a Type-V CRISPR effector protein such as Cpf1 or C2c1. In some embodiments, the CRISPR effector protein is a RNA-targeting CRISPR effector protein. In some embodiments, the CRISPR effector protein is a Type-VI CRISPR effector protein such as Cas13a, Cas13b, Cas13c, or Cas13d.

[0311] In some embodiments, the CRISPR effector protein is a Cas9, for instance SaCas9, SpCas9, StCas9, CjCas9 and so forth – any ortholog is envisaged. In some embodiments, the CRISPR effector protein is a Cpf1, for instance AsCpf1, LbCpf1, FnCpf1 and so forth – any ortholog is envisaged.In certain embodiments, the targeting component as described herein according to the invention is a (endo)nuclease or a variant thereof having altered or modified activity (i.e. a modified nuclease, as described herein elsewhere). In certain embodiments, said nuclease is a targeted or site-specific or homing nuclease or a variant thereof having altered or modified activity. In certain embodiments, said nuclease or targeted/site-specific/homing nuclease is, comprises, consists essentially of, or consists of a (modified) CRISPR/Cas system or complex, a (modified) Cas protein, a (modified) zinc finger, a (modified) zinc finger nuclease (ZFN), a (modified) transcription factor-like effector (TALE), a (modified) transcription factor-like effector nuclease (TALEN), or a (modified) meganuclease. In certain embodiments, said (modified) nuclease or targeted/site-specific/homing nuclease is, comprises, consists essentially of, or consists of a (modified) RNA-guided nuclease.

[0312] In particular embodiments, more particularly where the nuclease is a CRISPR protein, the targeting domain further comprises a guide molecule which targets a selected nucleic acid. For instance, in the context of the CRISPR/Cas system, the guide RNA is capable of hybridizing with a selected nucleic acid sequence. As uses herein, "hybridization" or “hybridizing” refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of PGR, or the cleavage of a polynucleotide by an enzyme. A sequence capable of hybridizing with a given sequence is referred to as the "complement" of the given sequence

[0313] In the methods and systems of the present invention use is made of a CRISPR-Cas protein and corresponding guide molecule. More particularly, the CRISPR-Cas protein is a

class 2 CRISPR-Cas protein. In certain embodiments, said CRISPR-Cas protein is a Cas13. The CRISPR-Cas system does not require the generation of customized proteins to target specific sequences but rather a single Cas protein can be programmed by guide molecule to recognize a specific nucleic acid target, in other words the Cas enzyme protein can be recruited to a specific nucleic acid target locus of interest using said guide molecule.

[0314] The term "AD-functionalized CRISPR system" as used here refers to a nucleic acid targeting and editing system comprising (a) a CRISPR-Cas protein, more particularly a Cas13 protein which is catalytically inactive; (b) a guide molecule which comprises a guide sequence; and (c) an adenosine deaminase protein or catalytic domain thereof; wherein the adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to the CRISPR-Cas protein or the guide molecule or is adapted to link thereto after delivery; wherein the guide sequence is substantially complementary to the target sequence but comprises a non-pairing C corresponding to the A being targeted for deamination, resulting in an A-C mismatch in an RNA duplex formed by the guide sequence and the target sequence. For application in eukaryotic cells, the CRISPR-Cas protein and/or the adenosine deaminase are preferably NLS-tagged.

[0315] In particular embodiments, the targeting domain is a CRISPR-cas protein. In certain example embodiments, the CRISPR-cas protein is linked to the deaminase protein or its catalytic domain by means of an LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID No. 11) linker. In further particular embodiments, the CRISPR-Cas protein is linked C-terminally to the N-terminus of a deaminase protein or its catalytic domain by means of an LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID No. 11) linker. In addition, N- and C-terminal NLSs can also function as linker (e.g., PKKKRKVEASSPKKRKVEAS (SEQ ID No. 16)). In particular embodiments of the methods of the present invention, the adenosine deaminase protein or catalytic domain thereof is delivered to the cell or expressed within the cell as a separate protein, but is modified so as to be able to link to the targeting domain or the guide molecule. In those embodiments in which the targeting domain is a CRISPR-Cas system, the adenosine deaminase may link to either the Cas protein or the guide moledule. In particular embodiments, this is ensured by the use of orthogonal RNA-binding protein or adaptor protein / aptamer combinations that exist within the diversity of bacteriophage coat proteins. Examples of such coat proteins include but are not limited to: MS2, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1. Aptamers can be naturally occurring or synthetic

oligonucleotides that have been engineered through repeated rounds of in vitro selection or SELEX (systematic evolution of ligands by exponential enrichment) to bind to a specific target.

[0316] In particular embodiments of the methods and systems of the present invention, the guide molecule is provided with one or more distinct RNA loop(s) or disctinct sequence(s) that can recruit an adaptor protein. For example, a guide molecule may be extended without colliding with the Cas protein by the insertion of distinct RNA loop(s) or distinct sequence(s) that may recruit adaptor proteins that can bind to the distinct RNA loop(s) or distinct sequence(s). Examples of modified guides and their use in recruiting effector domains to the CRISPR-Cas complex are provided in Konermann (Nature 2015, 517(7536): 583–588). In particular embodiments, the aptamer is a minimal hairpin aptamer which selectively binds dimerized MS2 bacteriophage coat proteins in mammalian cells and is introduced into the guide molecule, such as in the stemloop and/or in a tetraloop. In these embodiments, the adenosine deaminase protein is fused to MS2. The adenosine deaminase protein is then co-delivered together with the CRISPR-Cas protein and corresponding guide RNA.

[0317] In some embodiments, the components (a), (b) and (c) are delivered to the cell as a ribonucleoprotein complex. The ribonucleoprotein complex can be delivered via one or more lipid nanoparticles.

[0318] In some embodiments, the components (a), (b) and (c) are delivered to the cell as one or more RNA molecules, such as one or more guide RNAs and one or more mRNA molecules encoding the CRISPR-Cas protein, the adenosine deaminase protein, and optionally the adaptor protein. The RNA molecules can be delivered via one or more lipid nanoparticles.

[0319] In some embodiments, the components (a), (b) and (c) are delivered to the cell as one or more DNA molecules. In some embodiments, the one or more DNA molecules are comprised within one or more vectors such as viral vectors (e.g., AAV). In some embodiments, the one or more DNA molecules comprise one or more regulatory elements operably configured to express the CRISPR-Cas protein, the guide molecule, and the adenosine deaminase protein or catalytic domain thereof, optionally wherein the one or more regulatory elements comprise inducible promoters.

[0320] In some embodiments, the CRISPR-Cas protein is a dead Cas13. In some embodiments, the dead Cas13 is a dead Cas13a protein which comprises one or more mutations in the HEPN domain. In some embodiments, the dead Cas13a comprises a mutation corresponding to R474A and R1046A in Leptotrichia wadei (LwaCas13a). In some embodiments, the dead Cas13 is a dead Cas13b protein which comprises one or more of

R116A, H121A, R1177A, H1182A of a Cas13b protein originating from Bergeyella zoohelcum ATCC 43767 or amino acid positions corresponding thereto of a Cas13b ortholog.

[0321] In some embodiments of the guide molecule is capable of hybridizing with a target sequence comprising the Adenine to be deaminated within an RNA sequence to form an RNA duplex which comprises a non-pairing Cytosine opposite to said Adenine. Upon RNA duplex formation, the guide molecule forms a complex with the Cas13 protein and directs the complex to bind the RNA polynucleotide at the target RNA sequence of interest. Details on the aspect of the guide of the AD-functionalized CRISPR-Cas system are provided herein below.

[0322] In some embodiments, a Cas13 guide RNA having a canonical length of, e.g. LawCas13 is used to form an RNA duplex with the target DNA. In some embodiments, a Cas13 guide molecule longer than the canonical length for, e.g. LawCas13a is used to form an RNA duplex with the target DNA including outside of the Cas13-guide RNA-target DNA complex.

[0323] In at least a first design, the AD-functionalized CRISPR system comprises (a) an adenosine deaminase fused or linked to a CRISPR-Cas protein, wherein the CRISPR-Cas protein is catalytically inactive, and (b) a guide molecule comprising a guide sequence designed to introduce an A-C mismatch in an RNA duplex formed between the guide sequence and the target sequence. In some embodiments, the CRISPR-Cas protein and/or the adenosine deaminase are NLS-tagged, on either the N- or C-terminus or both.

[0324] In at least a second design, the AD-functionalized CRISPR system comprises (a) a CRISPR-Cas protein that is catalytically inactive, (b) a guide molecule comprising a guide sequence designed to introduce an A-C mismatch in an RNA duplex formed between the guide sequence and the target sequence, and an aptamer sequence (e.g., MS2 RNA motif or PP7 RNA motif) capable of binding to an adaptor protein (e.g., MS2 coating protein or PP7 coat protein), and (c) an adenosine deaminase fused or linked to an adaptor protein, wherein the binding of the aptamer and the adaptor protein recruits the adenosine deaminase to the RNA duplex formed between the guide sequence and the target sequence for targeted deamination at the A of the A-C mismatch. In some embodiments, the adaptor protein and/or the adenosine deaminase are NLS-tagged, on either the N- or C-terminus or both. The CRISPR-Cas protein can also be NLS-tagged.

[0325] The use of different aptamers and corresponding adaptor proteins also allows orthogonal gene editing to be implemented. In one example in which adenosine deaminase are used in combination with cytidine deaminase for orthogonal gene editing/deamination, sgRNA targeting different loci are modified with distinct RNA loops in order to recruit MS2-adenosine deaminase and PP7-cytidine deaminase (or PP7-adenosine deaminase and MS2-cytidine

deaminase), respectively, resulting in orthogonal deamination of A or C at the target loci of interested, respectively. PP7 is the RNA-binding coat protein of the bacteriophage Pseudomonas. Like MS2, it binds a specific RNA sequence and secondary structure. The PP7 RNA-recognition motif is distinct from that of MS2. Consequently, PP7 and MS2 can be multiplexed to mediate distinct effects at different genomic loci simultaneously. For example, an sgRNA targeting locus A can be modified with MS2 loops, recruiting MS2-adenosine deaminase, while another sgRNA targeting locus B can be modified with PP7 loops, recruiting PP7-cytidine deaminase. In the same cell, orthogonal, locus-specific modifications are thus realized. This principle can be extended to incorporate other orthogonal RNA-binding proteins.

[0326] In at least a third design, the AD-functionalized CRISPR system comprises (a) an adenosine deaminase inserted into an internal loop or unstructured region of a CRISPR-Cas protein, wherein the CRISPR-Cas protein is catalytically inactive or a nickase, and (b) a guide molecule comprising a guide sequence designed to introduce an A-C mismatch in an RNA duplex formed between the guide sequence and the target sequence.

[0327] CRISPR-Cas protein split sites that are suitable for inseration of adenosine deaminase can be identified with the help of a crystal structure. One can use the crystal structure of an ortholog if a relatively high degree of homology exists between the ortholog and the intended CRISPR-Cas protein.

[0328] The split position may be located within a region or loop. Preferably, the split position occurs where an interruption of the amino acid sequence does not result in the partial or full destruction of a structural feature (e.g. alpha-helixes or β-sheets). Unstructured regions (regions that did not show up in the crystal structure because these regions are not structured enough to be "frozen" in a crystal) are often preferred options. The positions within the unstructured regions or outside loops may not need to be exactly the numbers provided above, but may vary by, for example 1, 2, 3, 4, 5, 6, 7, 8, 9, or even 10 amino acids either side of the position given above, depending on the size of the loop, so long as the split position still falls within an unstructured region of outside loop.

[0329] The AD-functionalized CRISPR system described herein can be used to target a specific Adenine or Cytidine within an RNA polynucleotide sequence for deamination. For example, the guide molecule can form a complex with the CRISPR-Cas protein and directs the complex to bind a target RNA sequence in the RNA polynucleotide of interest. In certain example embodiments, because the guide sequence is designed to have a non-pairing C, the RNA duplex formed between the guide sequence and the target sequence comprises an A-C mismatch, which directs the adenosine deaminase to contact and deaminate the A opposite to

the non-pairing C, converting it to a Inosine (I). Since Inosine (I) base pairs with C and functions like G in cellular process, the targeted deamination of A described herein are useful for correction of undesirable G-A and C-T mutations, as well as for obtaining desirable A-G and T-C mutations.

[0330] In some embodiments, the AD-functionalized CRISPR system is used for targeted deamination in an RNA polynucleotide molecule in vitro. In some embodiments, the AD-functionalized CRISPR system is used for targeted deamination in a DNA molecule within a cell. The cell can be a eukaryotic cell, such as a animal cell, a mammalian cell, a human, or a plant cell.

Guide molecule

[0331] The guide molecule or guide RNA of a Class 2 type V CRISPR-Cas protein comprises a tracr-mate sequence (encompassing a "direct repeat" in the context of an endogenous CRISPR system) and a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system). Indeed, in contrast to the type II CRISPR-Cas proteins, the Cas13 protein does not rely on the presence of a tracr sequence. In some embodiments, the CRISPR-Cas system or complex as described herein does not comprise and/or does not rely on the presence of a tracr sequence (e.g. if the Cas protein is Cas13). In certain embodiments, the guide molecule may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence.

[0332] In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence. In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target DNA sequence and a guide sequence promotes the formation of a CRISPR complex.

[0333] The terms "guide molecule" and "guide RNA" are used interchangeably herein to refer to RNA-based molecules that are capable of forming a complex with a CRISPR-Cas protein and comprises a guide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of the complex to the target nucleic acid sequence. The guide molecule or guide RNA specifically encompasses RNA-based molecules having one or more chemically modifications (e.g., by chemical linking two ribonucleotides or by replacement of one or more ribonucleotides with one or more deoxyribonucleotides), as described herein.

[0334] As used herein, the term "crRNA" or "guide RNA" or "single guide RNA" or "sgRNA" or "one or more nucleic acid components" of a Type V or Type VI CRISPR-Cas

locus effector protein comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. In some embodiments, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid-targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting guide may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmatic RNA (scRNA). In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a

sequence within a RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[0335] In some embodiments, the guide molecule comprises a guide sequence that is designed to have at least one mismatch with the taret sequence, such that an RNA duplex formed between the guide sequence and the target sequence comprises a non-pairing C in the guide sequence opposite to the target A for deamination on the target sequence. In some embodiments, aside from this A-C mismatch, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more.

[0336] As used herein, the term "crRNA" or "guide RNA" or "single guide RNA" or "sgRNA" or "one or more nucleic acid components" of a Type V or Type VI CRISPR-Cas locus effector protein comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. In some embodiments, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid-targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from

the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting guide may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmatic RNA (scRNA). In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[0337] In some embodiments, a nucleic acid-targeting guide is selected to reduce the degree secondary structure within the nucleic acid-targeting guide. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the nucleic acid-targeting guide 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).

[0338] In certain embodiments, a guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat (DR) sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence. In certain embodiments, the direct repeat sequence may be located upstream (i.e., 5') from the guide sequence or spacer sequence. In other embodiments, the direct repeat sequence may be located downstream (i.e., 3') from the guide sequence or spacer sequence.

[0339] In certain embodiments, the crRNA comprises a stem loop, preferably a single stem loop. In certain embodiments, the direct repeat sequence forms a stem loop, preferably a single stem loop.

[0340] In certain embodiments, the spacer length of the guide RNA is from 15 to 35 nt. In certain embodiments, the spacer length of the guide RNA is at least 15 nucleotides. In certain embodiments, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27-30 nt, e.g., 27, 28, 29, or 30 nt, from 30-35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer.

[0341] The "tracrRNA" sequence or analogous terms includes any polynucleotide sequence that has sufficient complementarity with a crRNA sequence to hybridize. In some embodiments, the degree of complementarity between the tracrRNA sequence and crRNA sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and crRNA sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. In an embodiment of the invention, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In preferred embodiments, the transcript has two, three, four or five hairpins. In a further embodiment of the invention, the transcript has at most five hairpins. In a hairpin structure the portion of the sequence 5' of the final "N" and upstream of the loop corresponds to the tracr mate sequence, and the portion of the sequence 3' of the loop corresponds to the tracr sequence.

[0342] In general, degree of complementarity is with reference to the optimal alignment of the sca sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the sca sequence or tracr sequence. In some embodiments, the degree of complementarity between the tracr sequence and sca sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher.

[0343] In general, the CRISPR-Cas or CRISPR system may be as used in the foregoing documents, such as WO 2014/093622 (PCT/US2013/074667) and refers collectively to

transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, in particular a Cas13 gene in the case of CRISPR-Cas13, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas13, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. The section of the guide sequence through which complementarity to the target sequence is important for cleavage activity is referred to herein as the seed sequence. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell, and may include nucleic acids in or from mitochondrial, organelles, vesicles, liposomes or particles present within the cell. In some embodiments, especially for non-nuclear uses, NLSs are not preferred. In some embodiments, a CRISPR system comprises one or more nuclear exports signals (NESs). In some embodiments, a CRISPR system comprises one or more NLSs and one or more NESs. In some embodiments, direct repeats may be identified in silico by searching for repetitive motifs that fulfill any or all of the following criteria: 1. found in a 2Kb window of genomic sequence flanking the type II CRISPR locus; 2. span from 20 to 50 bp; and 3. interspaced by 20 to 50 bp. In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[0344] In embodiments of the invention the terms guide sequence and guide RNA, i.e. RNA capable of guiding Cas to a target genomic locus, are used interchangeably as in foregoing cited documents such as WO 2014/093622 (PCT/US2013/074667). In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when

optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. Preferably the guide sequence is 10 30 nucleotides long. The ability of a guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[0345] In some embodiments of CRISPR-Cas systems, the degree of complementarity between a guide sequence and its corresponding target sequence can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or 100%; a guide or RNA or sgRNA can be about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length; or guide or RNA or sgRNA can be less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length; and advantageously tracr RNA is 30 or 50 nucleotides in length. However, an aspect of the invention is to reduce off-target interactions, e.g., reduce the guide interacting with a target sequence having low complementarity. Indeed, in the examples, it is shown that the invention involves mutations that result in the CRISPR-Cas system being able to distinguish between target and off-target sequences that have greater than 80% to about 95%

complementarity, e.g., 83%-84% or 88-89% or 94-95% complementarity (for instance, distinguishing between a target having 18 nucleotides from an off-target of 18 nucleotides having 1, 2 or 3 mismatches). Accordingly, in the context of the present invention the degree of complementarity between a guide sequence and its corresponding target sequence is greater than 94.5% or 95% or 95.5% or 96% or 96.5% or 97% or 97.5% or 98% or 98.5% or 99% or 99.5% or 99.9%, or 100%. Off target is less than 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% or 94% or 93% or 92% or 91% or 90% or 89% or 88% or 87% or 86% or 85% or 84% or 83% or 82% or 81% or 80% complementarity between the sequence and the guide, with it advantageous that off target is 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% complementarity between the sequence and the guide.

[0346] In particularly preferred embodiments according to the invention, the guide RNA (capable of guiding Cas to a target locus) may comprise (1) a guide sequence capable of hybridizing to a genomic target locus in the eukaryotic cell; (2) a tracr sequence; and (3) a tracr mate sequence. All (1) to (3) may reside in a single RNA, i.e. an sgRNA (arranged in a 5' to 3' orientation), or the tracr RNA may be a different RNA than the RNA containing the guide and tracr sequence. The tracr hybridizes to the tracr mate sequence and directs the CRISPR/Cas complex to the target sequence. Where the tracr RNA is on a different RNA than the RNA containing the guide and tracr sequence, the length of each RNA may be optimized to be shortened from their respective native lengths, and each may be independently chemically modified to protect from degradation by cellular RNase or otherwise increase stability.

[0347] The methods according to the invention as described herein comprehend inducing one or more mutations in a eukaryotic cell (in vitro, i.e. in an isolated eukaryotic cell) as herein discussed comprising delivering to cell a vector as herein discussed. The mutation(s) can include the introduction, deletion, or substitution of one or more nucleotides at each target sequence of cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1-75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations include the introduction, deletion, or substitution of 10, 11, 12, 13,

14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 40, 45, 50, 75, 100, 200, 300, 400 or 500 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s).

[0348] For minimization of toxicity and off-target effect, it may be important to control the concentration of Cas mRNA and guide RNA delivered. Optimal concentrations of Cas mRNA and guide RNA can be determined by testing different concentrations in a cellular or non-human eukaryote animal model and using deep sequencing the analyze the extent of modification at potential off-target genomic loci. Alternatively, to minimize the level of toxicity and off-target effect, Cas nickase mRNA (for example S. pyogenes Cas9 with the D10A mutation) can be delivered with a pair of guide RNAs targeting a site of interest. Guide sequences and strategies to minimize toxicity and off-target effects can be as in WO 2014/093622 (PCT/US2013/074667); or, via mutation as herein.

[0349] Typically, in the context of an endogenous CRISPR system, formation of a CRISPR complex (comprising a guide sequence hybridized to a target sequence and complexed with one or more Cas proteins) results in cleavage of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. Without wishing to be bound by theory, the tracr sequence, which may comprise or consist of all or a portion of a wild-type tracr sequence (e.g. about or more than about 20, 26, 32, 45, 48, 54, 63, 67, 85, or more nucleotides of a wild-type tracr sequence), may also form part of a CRISPR complex, such as by hybridization along at least a portion of the tracr sequence to all or a portion of a tracr mate sequence that is operably linked to the guide sequence.

Guide Modifications

[0350] In certain embodiments, guides of the invention comprise non-naturally occurring nucleic acids and/or non-naturally occurring nucleotides and/or nucleotide analogs, and/or chemically modifications. Non-naturally occurring nucleic acids can include, for example, mixtures of naturally and non-naturally occurring nucleotides. Non-naturally occurring nucleotides and/or nucleotide analogs may be modified at the ribose, phosphate, and/or base moiety. In an embodiment of the invention, a guide nucleic acid comprises ribonucleotides and non-ribonucleotides. In one such embodiment, a guide comprises one or more ribonucleotides and one or more deoxyribonucleotides. In an embodiment of the invention, the guide

comprises one or more non-naturally occurring nucleotide or nucleotide analog such as a nucleotide with phosphorothioate linkage, boranophosphate linkage, a locked nucleic acid (LNA) nucleotides comprising a methylene bridge between the 2,Ä≤ and 4,Ä≤ carbons of the ribose ring, peptide nucleic acids (PNA), or bridged nucleic acids (BNA). Other examples of modified nucleotides include 2'-O-methyl analogs, 2'-deoxy analogs, 2-thiouridine analogs, N6-methyladenosine analogs, or 2'-fluoro analogs. Further examples of modified nucleotides include linkage of chemical moieties at the 2' position, including but not limited to peptides, nuclear localization sequence (NLS), peptide nucleic acid (PNA), polyethylene glycol (PEG), triethylene glycol, or tetraethyleneglycol (TEG). Further examples of modified bases include, but are not limited to, 2-aminopurine, 5-bromo-uridine, pseudouridine (Œ®), N1-methylpseudouridine (me1Œ®), 5-methoxyuridine(5moU), inosine, 7-methylguanosine. Examples of guide RNA chemical modifications include, without limitation, incorporation of 2'-O-methyl (M), 2'-O-methyl-3'-phosphorothioate (MS), phosphorothioate (PS), S-constrained ethyl(cEt), 2'-O-methyl-3'-thioPACE (MSP), or 2'-O-methyl-3'-phosphonoacetate (MP) at one or more terminal nucleotides. Such chemically modified guides can comprise increased stability and increased activity as compared to unmodified guides, though on-target vs. off-target specificity is not predictable. (See, Hendel, 2015, Nat Biotechnol. 33(9):985-9, doi: 10.1038/nbt.3290, published online 29 June 2015; Ragdarm et al., 0215, PNAS, E7110-E7111; Allerson et al., J. Med. Chem. 2005, 48:901-904; Bramsen et al., Front. Genet., 2012, 3:154; Deng et al., PNAS, 2015, 112:11870-11875; Sharma et al., MedChemComm., 2014, 5:1454-1471; Hendel et al., Nat. Biotechnol. (2015) 33(9): 985-989; Li et al., Nature Biomedical Engineering, 2017, 1, 0066 DOI:10.1038/s41551-017-0066; Ryan et al., Nucleic Acids Res. (2018) 46(2): 792-803). In some embodiments, the 5' and/or 3' end of a guide RNA is modified by a variety of functional moieties including fluorescent dyes, polyethylene glycol, cholesterol, proteins, or detection tags. (See Kelly et al., 2016, J. Biotech. 233:74-83). In certain embodients, a guide comprises ribonucleotides in a region that binds to a target DNA and one or more deoxyribonucletides and/or nucleotide analogs in a region that binds to Cas9, Cpf1, C2c1, or Cas13. In an embodiment of the invention, deoxyribonucleotides and/or nucleotide analogs are incorporated in engineered guide structures, such as, without limitation, 5' and/or 3' end, stem-loop regions, and the seed region. In certain embodiments, the modification is not in the 5'-handle of the stem-loop regions. Chemical modification in the 5'-handle of the stem-loop region of a guide may abolish its function (see Li, et al., Nature Biomedical Engineering, 2017, 1:0066). In certain embodiments, 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, 35, 40, 45, 50, or 75 nucleotides of a

guide is chemically modified. In some embodiments, 3-5 nucleotides at either the 3' or the 5' end of a guide is chemically modified. In some embodiments, only minor modifications are introduced in the seed region, such as 2'-F modifications. In some embodiments, 2'-F modification is introduced at the 3' end of a guide. In certain embodiments, three to five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-O-methyl (M), 2'-O-methyl-3'-phosphorothioate (MS), S-constrained ethyl(cEt), 2'-O-methyl-3'-thioPACE (MSP), or 2'-O-methyl-3'-phosphonoacetate (MP). Such modification can enhance genome editing efficiency (see Hendel et al., Nat. Biotechnol. (2015) 33(9): 985-989; Ryan et al., Nucleic Acids Res. (2018) 46(2): 792-803). In certain embodiments, all of the phosphodiester bonds of a guide are substituted with phosphorothioates (PS) for enhancing levels of gene disruption. In certain embodiments, more than five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-O-Me, 2'-F or S-constrained ethyl(cEt). Such chemically modified guide can mediate enhanced levels of gene disruption (see Ragdarm et al., 0215, PNAS, E7110-E7111). In an embodiment of the invention, a guide is modified to comprise a chemical moiety at its 3' and/or 5' end. Such moieties include, but are not limited to amine, azide, alkyne, thio, dibenzocyclooctyne (DBCO), Rhodamine, peptides, nuclear localization sequence (NLS), peptide nucleic acid (PNA), polyethylene glycol (PEG), triethylene glycol, or tetraethyleneglycol (TEG). In certain embodiment, the chemical moiety is conjugated to the guide by a linker, such as an alkyl chain. In certain embodiments, the chemical moiety of the modified guide can be used to attach the guide to another molecule, such as DNA, RNA, protein, or nanoparticles. Such chemically modified guide can be used to identify or enrich cells generically edited by a CRISPR system (see Lee et al., eLife, 2017, 6:e25312, DOI:10.7554). In some embodiments, 3 nucleotides at each of the 3' and 5' ends are chemically modified. In a specific embodiment, the modifications comprise 2'-O-methyl or phosphorothioate analogs. In a specific embodiment, 12 nucleotides in the tetraloop and 16 nucleotides in the stem-loop region are replaced with 2'-O-methyl analogs. Such chemical modifications improve in vivo editing and stability (see Finn et al., Cell Reports (2018), 22: 2227-2235). In some embodiments, more than 60 or 70 nucleotides of the guide are chemically modified. In some embodiments, this modification comprises replacement of nucleotides with 2'-O-methyl or 2'-fluoro nucleotide analogs or phosphorothioate (PS) modification of phosphodiester bonds. In some embodiments, the chemical modification comprises 2'-O-methyl or 2'-fluoro modification of guide nucleotides extending outside of the nuclease protein when the CRISPR complex is formed or PS modification of 20 to 30 or more nucleotides of the 3'-terminus of the guide. In a particular embodiment, the chemical modification further

comprises 2'-O-methyl analogs at the 5' end of the guide or 2'-fluoro analogs in the seed and tail regions. Such chemical modifications improve stability to nuclease degradation and maintain or enhance genome-editing activity or efficiency, but modification of all nucleotides may abolish the function of the guide (see Yin et al., Nat. Biotech. (2018), 35(12): 1179-1187). Such chemical modifications may be guided by knowledge of the structure of the CRISPR complex, including knowledge of the limited number of nuclease and RNA 2'-OH interactions (see Yin et al., Nat. Biotech. (2018), 35(12): 1179-1187). In some embodiments, one or more guide RNA nucleotides may be replaced with DNA nucleotides. In some embodiments, up to 2, 4, 6, 8, 10, or 12 RNA nucleotides of the 5'-end tail/seed guide region are replaced with DNA nucleotides. In certain embodiments, the majority of guide RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, 16 guide RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, 8 guide RNA nucleotides of the 5'-end tail/seed region and 16 RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, guide RNA nucleotides that extend outside of the nuclease protein when the CRISPR complex is formed are replaced with DNA nucleotides. Such replacement of multiple RNA nucleotides with DNA nucleotides leads to decreased off-target activity but similar on-target activity compared to an unmodified guide; however, replacement of all RNA nucleotides at the 3' end may abolish the function of the guide (see Yin et al., Nat. Chem. Biol. (2018) 14, 311-316). Such modifications may be guided by knowledge of the structure of the CRISPR complex, including knowledge of the limited number of nuclease and RNA 2'-OH interactions (see Yin et al., Nat. Chem. Biol. (2018) 14, 311-316).

[0351] In one aspect of the invention, the guide comprises a modified crRNA for Cpf1, having a 5'-handle and a guide segment further comprising a seed region and a 3'-terminus. In some embodiments, the modified guide can be used with a Cpf1 of any one of Acidaminococcus sp. BV3L6 Cpf1 (AsCpf1); Francisella tularensis subsp. Novicida U112 Cpf1 (FnCpf1); L. bacterium MC2017 Cpf1 (Lb3Cpf1); Butyrivibrio proteoclasticus Cpf1 (BpCpf1); Parcubacteria bacterium GWC2011_GWC2_44_17 Cpf1 (PbCpf1); Peregrinibacteria bacterium GW2011_GWA_33_10 Cpf1 (PeCpf1); Leptospira inadai Cpf1 (LiCpf1); Smithella sp. SC_K08D17 Cpf1 (SsCpf1); L. bacterium MA2020 Cpf1 (Lb2Cpf1); Porphyromonas crevioricanis Cpf1 (PcCpf1); Porphyromonas macacae Cpf1 (PmCpf1); Candidatus Methanoplasma termitum Cpf1 (CMtCpf1); Eubacterium eligens Cpf1 (EeCpf1); Moraxella bovoculi 237 Cpf1 (MbCpf1); Prevotella disiens Cpf1 (PdCpf1); or L. bacterium ND2006 Cpf1 (LbCpf1).

[0352] In some embodiments, the modification to the guide is a chemical modification, an insertion, a deletion or a split. In some embodiments, the chemical modification includes, but is not limited to, incorporation of 2'-O-methyl (M) analogs, 2'-deoxy analogs, 2-thiouridine analogs, N6-methyladenosine analogs, 2'-fluoro analogs, 2-aminopurine, 5-bromo-uridine, pseudouridine (Œ®), N1-methylpseudouridine (me1Œ®), 5-methoxyuridine(5moU), inosine, 7-methylguanosine, 2'-O-methyl-3'-phosphorothioate (MS), S-constrained ethyl(cEt), phosphorothioate (PS), 2'-O-methyl-3'-thioPACE (MSP), or 2'-O-methyl-3'-phosphonoacetate (MP). In some embodiments, the guide comprises one or more of phosphorothioate modifications. In certain embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 25 nucleotides of the guide are chemically modified. In some embodiments, all nucleotides are chemically modified. In certain embodiments, one or more nucleotides in the seed region are chemically modified. In certain embodiments, one or more nucleotides in the 3'-terminus are chemically modified. In certain embodiments, none of the nucleotides in the 5'-handle is chemically modified. In some embodiments, the chemical modification in the seed region is a minor modification, such as incorporation of a 2'-fluoro analog. In a specific embodiment, one nucleotide of the seed region is replaced with a 2'-fluoro analog. In some embodiments, 5 or 10 nucleotides in the 3'-terminus are chemically modified. Such chemical modifications at the 3'-terminus of the Cpf1 CrRNA improve gene cutting efficiency (see Li, et al., Nature Biomedical Engineering, 2017, 1:0066). In a specific embodiment, 5 nucleotides in the 3'-terminus are replaced with 2'-fluoro analogues. In a specific embodiment, 10 nucleotides in the 3'-terminus are replaced with 2'-fluoro analogues. In a specific embodiment, 5 nucleotides in the 3'-terminus are replaced with 2'- O-methyl (M) analogs. In some embodiments, 3 nucleotides at each of the 3' and 5' ends are chemically modified. In a specific embodiment, the modifications comprise 2'-O-methyl or phosphorothioate analogs. In a specific embodiment, 12 nucleotides in the tetraloop and 16 nucleotides in the stem-loop region are replaced with 2'-O-methyl analogs. Such chemical modifications improve in vivo editing and stability (see Finn et al., Cell Reports (2018), 22: 2227-2235).

[0353] In some embodiments, the loop of the 5'-handle of the guide is modified. In some embodiments, the loop of the 5'-handle of the guide is modified to have a deletion, an insertion, a split, or chemical modifications. In certain embodiments, the loop comprises 3, 4, or 5 nucleotides. In certain embodiments, the loop comprises the sequence of UCUU, UUUU, UAUU, or UGUU. In some embodiments, the guide molecule forms a stemloop with a separate non-covalently linked sequence, which can be DNA or RNA.

Synthetically linked guide

[0354] In one aspect, the guide comprises a tracr sequence and a tracr mate sequence that are chemically linked or conjugated via a non-phosphodiester bond. In one aspect, the guide comprises a tracr sequence and a tracr mate sequence that are chemically linked or conjugated via a non-nucleotide loop. In some embodiments, the tracr and tracr mate sequences are joined via a non-phosphodiester covalent linker. Examples of the covalent linker include but are not limited to a chemical moiety selected from the group consisting of carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, fulfones, sulfoxides, ureas, thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[0355] In some embodiments, the tracr and tracr mate sequences are first synthesized using the standard phosphoramidite synthetic protocol (Herdewijn, P., ed., Methods in Molecular Biology Col 288, Oligonucleotide Synthesis: Methods and Applications, Humana Press, New Jersey (2012)). In some embodiments, the tracr or tracr mate sequences can be functionalized to contain an appropriate functional group for ligation using the standard protocol known in the art (Hermanson, G. T., Bioconjugate Techniques, Academic Press (2013)). Examples of functional groups include, but are not limited to, hydroxyl, amine, carboxylic acid, carboxylic acid halide, carboxylic acid active ester, aldehyde, carbonyl, chlorocarbonyl, imidazolylcarbonyl, hydrozide, semicarbazide, thio semicarbazide, thiol, maleimide, haloalkyl, sufonyl, ally, propargyl, diene, alkyne, and azide. Once the tracr and the tracr mate sequences are functionalized, a covalent chemical bond or linkage can be formed between the two oligonucleotides. Examples of chemical bonds include, but are not limited to, those based on carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, fulfones, sulfoxides, ureas, thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[0356] In some embodiments, the tracr and tracr mate sequences can be chemically synthesized. In some embodiments, the chemical synthesis uses automated, solid-phase oligonucleotide synthesis machines with 2'-acetoxyethyl orthoester (2'-ACE) (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989).

[0357] In some embodiments, the tracr and tracr mate sequences can be covalently linked using various bioconjugation reactions, loops, bridges, and non-nucleotide links via modifications of sugar, internucleotide phosphodiester bonds, purine and pyrimidine residues. Sletten et al., Angew. Chem. Int. Ed. (2009) 48:6974-6998; Manoharan, M. Curr. Opin. Chem. Biol. (2004) 8: 570-9; Behlke et al., Oligonucleotides (2008) 18: 305-19; Watts, et al., Drug. Discov. Today (2008) 13: 842-55; Shukla, et al., ChemMedChem (2010) 5: 328-49.

[0358] In some embodiments, the tracr and tracr mate sequences can be covalently linked using click chemistry. In some embodiments, the tracr and tracr mate sequences can be covalently linked using a triazole linker. In some embodiments, the tracr and tracr mate sequences can be covalently linked using Huisgen 1,3-dipolar cycloaddition reaction involving an alkyne and azide to yield a highly stable triazole linker (He et al., ChemBioChem (2015) 17: 1809-1812; WO 2016/186745). In some embodiments, the tracr and tracr mate sequences are covalently linked by ligating a 5'-hexyne tracrRNA and a 3'-azide crRNA. In some embodiments, either or both of the 5'-hexyne tracrRNA and a 3'-azide crRNA can be protected with 2'-acetoxyethl orthoester (2'-ACE) group, which can be subsequently removed using Dharmacon protocol (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18).

[0359] In some embodiments, the tracr and tracr mate sequences can be covalently linked via a linker (e.g., a non-nucleotide loop) that comprises a moiety such as spacers, attachments, bioconjugates, chromophores, reporter groups, dye labeled RNAs, and non-naturally occurring nucleotide analogues. More specifically, suitable spacers for purposes of this invention include, but are not limited to, polyethers (e.g., polyethylene glycols, polyalcohols, polypropylene glycol or mixtures of efhylene and propylene glycols), polyamines group (e.g., spennine, spermidine and polymeric derivatives thereof), polyesters (e.g., poly(ethyl acrylate)), polyphosphodiesters, alkylenes, and combinations thereof. Suitable attachments include any moiety that can be added to the linker to add additional properties to the linker, such as but not limited to, fluorescent labels. Suitable bioconjugates include, but are not limited to, peptides, glycosides, lipids, cholesterol, phospholipids, diacyl glycerols and dialkyl glycerols, fatty acids, hydrocarbons, enzyme substrates, steroids, biotin, digoxigenin, carbohydrates, polysaccharides. Suitable chromophores, reporter groups, and dye-labeled RNAs include, but are not limited to, fluorescent dyes such as fluorescein and rhodamine, chemiluminescent, electrochemiluminescent, and bioluminescent marker compounds. The design of example linkers conjugating two RNA components are also described in WO 2004/015075.

[0360] The linker (e.g., a non-nucleotide loop) can be of any length. In some embodiments, the linker has a length equivalent to about 0-16 nucleotides. In some embodiments, the linker has a length equivalent to about 0-8 nucleotides. In some embodiments, the linker has a length equivalent to about 0-4 nucleotides. In some embodiments, the linker has a length equivalent to about 2 nucleotides. Example linker design is also described in WO2011/008730.

[0361] A typical Type II Cas sgRNA comprises (in 5' to 3' direction): a guide sequence, a poly U tract, a first complimentary stretch (the "repeat"), a loop (tetraloop), a second complimentary stretch (the "anti-repeat" being complimentary to the repeat), a stem, and further stem loops and stems and a poly A (often poly U in RNA) tail (terminator). In preferred embodiments, certain aspects of guide architecture are retained, certain aspect of guide architecture cam be modified, for example by addition, subtraction, or substitution of features, whereas certain other aspects of guide architecture are maintained. Preferred locations for engineered sgRNA modifications, including but not limited to insertions, deletions, and substitutions include guide termini and regions of the sgRNA that are exposed when complexed with CRISPR protein and/or target, for example the tetraloop and/or loop2.

[0362] In certain embodiments, guides of the invention comprise specific binding sites (e.g. aptamers) for adapter proteins, which may comprise one or more functional domains (e.g. via fusion protein). When such a guide forms a CRISPR complex (i.e. CRISPR enzyme binding to guide and target) the adapter proteins bind and, the functional domain associated with the adapter protein is positioned in a spatial orientation which is advantageous for the attributed function to be effective. For example, if the functional domain is a transcription activator (e.g. VP64 or p65), the transcription activator is placed in a spatial orientation which allows it to affect the transcription of the target. Likewise, a transcription repressor will be advantageously positioned to affect the transcription of the target and a nuclease (e.g. Fok1) will be advantageously positioned to cleave or partially cleave the target.

[0363] The skilled person will understand that modifications to the guide which allow for binding of the adapter + functional domain but not proper positioning of the adapter + functional domain (e.g. due to steric hindrance within the three dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified guide may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein, preferably at either the tetra loop or stem loop 2, and most preferably at both the tetra loop and stem loop 2.

[0364] The repeat:anti repeat duplex will be apparent from the secondary structure of the sgRNA. It may be typically a first complimentary stretch after (in 5' to 3' direction) the poly U tract and before the tetraloop; and a second complimentary stretch after (in 5' to 3' direction) the tetraloop and before the poly A tract. The first complimentary stretch (the "repeat") is complimentary to the second complimentary stretch (the "anti-repeat"). As such, they Watson-Crick base pair to form a duplex of dsRNA when folded back on one another. As such, the anti-repeat sequence is the complimentary sequence of the repeat and in terms to A-U or C-G base pairing, but also in terms of the fact that the anti-repeat is in the reverse orientation due to the tetraloop.

[0365] In an embodiment of the invention, modification of guide architecture comprises replacing bases in stemloop 2. For example, in some embodiments, "actt" ("acuu" in RNA) and "aagt" ("aagu" in RNA) bases in stemloop2 are replaced with "cgcc" and "gcgg". In some embodiments, "actt" and "aagt" bases in stemloop2 are replaced with complimentary GC-rich regions of 4 nucleotides. In some embodiments, the complimentary GC-rich regions of 4 nucleotides are "cgcc" and "gcgg" (both in 5' to 3' direction). In some embodiments, the complimentary GC-rich regions of 4 nucleotides are "gcgg" and "cgcc" (both in 5' to 3' direction). Other combination of C and G in the complimentary GC-rich regions of 4 nucleotides will be apparent including CCCC and GGGG.

[0366] In one aspect, the stemloop 2, e.g., "ACTTgtttAAGT" can be replaced by any "XXXXgtttYYYY", e.g., where XXXX and YYYY represent any complementary sets of nucleotides that together will base pair to each other to create a stem.

[0367] In one aspect, the stem comprises at least about 4bp comprising complementary X and Y sequences, although stems of more, e.g., 5, 6, 7, 8, 9, 10, 11 or 12 or fewer, e.g., 3, 2, base pairs are also contemplated. Thus, for example X2-12 and Y2-12 (wherein X and Y represent any complementary set of nucleotides) may be contemplated. In one aspect, the stem made of the X and Y nucleotides, together with the "gttt," will form a complete hairpin in the overall secondary structure; and, this may be advantageous and the amount of base pairs can be any amount that forms a complete hairpin. In one aspect, any complementary X:Y basepairing sequence (e.g., as to length) is tolerated, so long as the secondary structure of the entire sgRNA is preserved. In one aspect, the stem can be a form of X:Y basepairing that does not disrupt the secondary structure of the whole sgRNA in that it has a DR:tracr duplex, and 3 stemloops. In one aspect, the "gttt" tetraloop that connects ACTT and AAGT (or any alternative stem made of X:Y basepairs) can be any sequence of the same length (e.g., 4 basepair) or longer that does not interrupt the overall secondary structure of the sgRNA. In one aspect, the

stemloop can be something that further lengthens stemloop2, e.g. can be MS2 aptamer. In one aspect, the stemloop3 "GGCACCGagtCGGTGC" can likewise take on a "XXXXXXXagtYYYYYYY" form, e.g., wherein X7 and Y7 represent any complementary sets of nucleotides that together will base pair to each other to create a stem. In one aspect, the stem comprises about 7bp comprising complementary X and Y sequences, although stems of more or fewer basepairs are also contemplated. In one aspect, the stem made of the X and Y nucleotides, together with the "agt", will form a complete hairpin in the overall secondary structure. In one aspect, any complementary X:Y basepairing sequence is tolerated, so long as the secondary structure of the entire sgRNA is preserved. In one aspect, the stem can be a form of X:Y basepairing that doesn't disrupt the secondary structure of the whole sgRNA in that it has a DR:tracr duplex, and 3 stemloops. In one aspect, the "agt" sequence of the stemloop 3 can be extended or be replaced by an aptamer, e.g., a MS2 aptamer or sequence that otherwise generally preserves the architecture of stemloop3. In one aspect for alternative Stemloops 2 and/or 3, each X and Y pair can refer to any basepair. In one aspect, non-Watson Crick basepairing is contemplated, where such pairing otherwise generally preserves the architecture of the stemloop at that position.

[0368] In one aspect, the DR:tracrRNA duplex can be replaced with the form: gYYYYag(N)NNNNxxxxNNNN(AAN)uuRRRRu (using standard IUPAC nomenclature for nucleotides), wherein (N) and (AAN) represent part of the bulge in the duplex, and "xxxx" represents a linker sequence. NNNN on the direct repeat can be anything so long as it basepairs with the corresponding NNNN portion of the tracrRNA. In one aspect, the DR:tracrRNA duplex can be connected by a linker of any length (xxxx...), any base composition, as long as it doesn't alter the overall structure.

[0369] In one aspect, the sgRNA structural requirement is to have a duplex and 3 stemloops. In most aspects, the actual sequence requirement for many of the particular base requirements are lax, in that the architecture of the DR:tracrRNA duplex should be preserved, but the sequence that creates the architecture, i.e., the stems, loops, bulges, etc., may be alterred.

Aptamers

[0370] One guide with a first aptamer/RNA-binding protein pair can be linked or fused to an activator, whilst a second guide with a second aptamer/RNA-binding protein pair can be linked or fused to a repressor. The guides are for different targets (loci), so this allows one gene to be activated and one repressed. For example, the following schematic shows such an approach:

[0371] Guide 1- MS2 aptamer-------MS2 RNA-binding protein-------VP64 activator; and

[0372] Guide 2 - PP7 aptamer-------PP7 RNA-binding protein-------SID4x repressor.

[0373] The present invention also relates to orthogonal PP7/MS2 gene targeting. In this example, sgRNA targeting different loci are modified with distinct RNA loops in order to recruit MS2-VP64 or PP7-SID4X, which activate and repress their target loci, respectively. PP7 is the RNA-binding coat protein of the bacteriophage Pseudomonas. Like MS2, it binds a specific RNA sequence and secondary structure. The PP7 RNA-recognition motif is distinct from that of MS2. Consequently, PP7 and MS2 can be multiplexed to mediate distinct effects at different genomic loci simultaneously. For example, an sgRNA targeting locus A can be modified with MS2 loops, recruiting MS2-VP64 activators, while another sgRNA targeting locus B can be modified with PP7 loops, recruiting PP7-SID4X repressor domains. In the same cell, dCas13 can thus mediate orthogonal, locus-specific modifications. This principle can be extended to incorporate other orthogonal RNA-binding proteins such as Q-beta.

[0374] An alternative option for orthogonal repression includes incorporating non-coding RNA loops with transactive repressive function into the guide (either at similar positions to the MS2/PP7 loops integrated into the guide or at the 3' terminus of the guide). For instance, guides were designed with non-coding (but known to be repressive) RNA loops (e.g. using the Alu repressor (in RNA) that interferes with RNA polymerase II in mammalian cells). The Alu RNA sequence was located: in place of the MS2 RNA sequences as used herein (e.g. at tetraloop and/or stem loop 2); and/or at 3' terminus of the guide. This gives possible combinations of MS2, PP7 or Alu at the tetraloop and/or stemloop 2 positions, as well as, optionally, addition of Alu at the 3' end of the guide (with or without a linker).

[0375] The use of two different aptamers (distinct RNA) allows an activator-adaptor protein fusion and a repressor-adaptor protein fusion to be used, with different guides, to activate expression of one gene, whilst repressing another. They, along with their different guides can be administered together, or substantially together, in a multiplexed approach. A large number of such modified guides can be used all at the same time, for example 10 or 20 or 30 and so forth, whilst only one (or at least a minimal number) of Cas13s to be delivered, as a comparatively small number of Cas13s can be used with a large number modified guides. The adaptor protein may be associated (preferably linked or fused to) one or more activators or one or more repressors. For example, the adaptor protein may be associated with a first activator and a second activator. The first and second activators may be the same, but they are preferably different activators. For example, one might be VP64, whilst the other might be p65, although these are just examples and other transcriptional activators are envisaged. Three or more or even four or more activators (or repressors) may be used, but package size may limit

the number being higher than 5 different functional domains. Linkers are preferably used, over a direct fusion to the adaptor protein, where two or more functional domains are associated with the adaptor protein. Suitable linkers might include the GlySer linker.

[0376] It is also envisaged that the enzyme-guide complex as a whole may be associated with two or more functional domains. For example, there may be two or more functional domains associated with the enzyme, or there may be two or more functional domains associated with the guide (via one or more adaptor proteins), or there may be one or more functional domains associated with the enzyme and one or more functional domains associated with the guide (via one or more adaptor proteins).

[0377] The fusion between the adaptor protein and the activator or repressor may include a linker. For example, GlySer linkers GGGS can be used. They can be used in repeats of 3 ((GGGGS)3) or 6, 9 or even 12 or more, to provide suitable lengths, as required. Linkers can be used between the RNA-binding protein and the functional domain (activator or repressor), or between the CRISPR Enzyme (Cas13) and the functional domain (activator or repressor). The linkers the user to engineer appropriate amounts of "mechanical flexibility".

[0378] *Dead guides: Guide RNAs comprising a dead guide sequence may be used in the present invention*

[0379] In one aspect, the invention provides guide sequences which are modified in a manner which allows for formation of the CRISPR complex and successful binding to the target, while at the same time, not allowing for successful nuclease activity (i.e. without nuclease activity / without indel activity). For matters of explanation such modified guide sequences are referred to as "dead guides" or "dead guide sequences". These dead guides or dead guide sequences can be thought of as catalytically inactive or conformationally inactive with regard to nuclease activity. Nuclease activity may be measured using surveyor analysis or deep sequencing as commonly used in the art, preferably surveyor analysis. Similarly, dead guide sequences may not sufficiently engage in productive base pairing with respect to the ability to promote catalytic activity or to distinguish on-target and off-target binding activity. Briefly, the surveyor assay involves purifying and amplifying a CRISPR target site for a gene and forming heteroduplexes with primers amplifying the CRISPR target site. After re-anneal, the products are treated with SURVEYOR nuclease and SURVEYOR enhancer S (Transgenomics) following the manufacturer's recommended protocols, analyzed on gels, and quantified based upon relative band intensities.

[0380] Hence, in a related aspect, the invention provides a non-naturally occurring or engineered composition Cas13 CRISPR-Cas system comprising a functional Cas13 as

described herein, and guide RNA (gRNA) wherein the gRNA comprises a dead guide sequence whereby the gRNA is capable of hybridizing to a target sequence such that the Cas13 CRISPR-Cas system is directed to a genomic locus of interest in a cell without detectable indel activity resultant from nuclease activity of a non-mutant Cas13 enzyme of the system as detected by a SURVEYOR assay. For shorthand purposes, a gRNA comprising a dead guide sequence whereby the gRNA is capable of hybridizing to a target sequence such that the Cas13 CRISPR-Cas system is directed to a genomic locus of interest in a cell without detectable indel activity resultant from nuclease activity of a non-mutant Cas13 enzyme of the system as detected by a SURVEYOR assay is herein termed a "dead gRNA". It is to be understood that any of the gRNAs according to the invention as described herein elsewhere may be used as dead gRNAs / gRNAs comprising a dead guide sequence as described herein below. Any of the methods, products, compositions and uses as described herein elsewhere is equally applicable with the dead gRNAs / gRNAs comprising a dead guide sequence as further detailed below. By means of further guidance, the following particular aspects and embodiments are provided.

[0381] The ability of a dead guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the dead guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the dead guide sequence to be tested and a control guide sequence different from the test dead guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A dead guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a genome of a cell.

[0382] As explained further herein, several structural parameters allow for a proper framework to arrive at such dead guides. Dead guide sequences are shorter than respective guide sequences which result in active Cas13-specific indel formation. Dead guides are 5%, 10%, 20%, 30%, 40%, 50%, shorter than respective guides directed to the same Cas13 leading to active Cas13-specific indel formation.

[0383] As explained below and known in the art, one aspect of gRNA - Cas specificity is the direct repeat sequence, which is to be appropriately linked to such guides. In particular, this implies that the direct repeat sequences are designed dependent on the origin of the Cas. Thus, structural data available for validated dead guide sequences may be used for designing Cas specific equivalents. Structural similarity between, e.g., the orthologous nuclease domains RuvC of two or more Cas effector proteins may be used to transfer design equivalent dead guides. Thus, the dead guide herein may be appropriately modified in length and sequence to reflect such Cas specific equivalents, allowing for formation of the CRISPR complex and successful binding to the target, while at the same time, not allowing for successful nuclease activity.

[0384] The use of dead guides in the context herein as well as the state of the art provides a surprising and unexpected platform for network biology and/or systems biology in both in vitro, ex vivo, and in vivo applications, allowing for multiplex gene targeting, and in particular bidirectional multiplex gene targeting. Prior to the use of dead guides, addressing multiple targets, for example for activation, repression and/or silencing of gene activity, has been challenging and in some cases not possible. With the use of dead guides, multiple targets, and thus multiple activities, may be addressed, for example, in the same cell, in the same animal, or in the same patient. Such multiplexing may occur at the same time or staggered for a desired timeframe.

[0385] For example, the dead guides now allow for the first time to use gRNA as a means for gene targeting, without the consequence of nuclease activity, while at the same time providing directed means for activation or repression. Guide RNA comprising a dead guide may be modified to further include elements in a manner which allow for activation or repression of gene activity, in particular protein adaptors (e.g. aptamers) as described herein elsewhere allowing for functional placement of gene effectors (e.g. activators or repressors of gene activity). One example is the incorporation of aptamers, as explained herein and in the state of the art. By engineering the gRNA comprising a dead guide to incorporate protein-interacting aptamers (Konermann et al., "Genome-scale transcription activation by an engineered CRISPR-Cas9 complex," doi:10.1038/nature14136, incorporated herein by reference), one may assemble a synthetic transcription activation complex consisting of multiple distinct effector domains. Such may be modeled after natural transcription activation processes. For example, an aptamer, which selectively binds an effector (e.g. an activator or repressor; dimerized MS2 bacteriophage coat proteins as fusion proteins with an activator or repressor), or a protein which itself binds an effector (e.g. activator or repressor) may be

appended to a dead gRNA tetraloop and/or a stem-loop 2. In the case of MS2, the fusion protein MS2-VP64 binds to the tetraloop and/or stem-loop 2 and in turn mediates transcriptional up-regulation, for example for Neurog2. Other transcriptional activators are, for example, VP64. P65, HSF1, and MyoD1. By mere example of this concept, replacement of the MS2 stem-loops with PP7-interacting stem-loops may be used to recruit repressive elements.

[0386] Thus, one aspect is a gRNA of the invention which comprises a dead guide, wherein the gRNA further comprises modifications which provide for gene activation or repression, as described herein. The dead gRNA may comprise one or more aptamers. The aptamers may be specific to gene effectors, gene activators or gene repressors. Alternatively, the aptamers may be specific to a protein which in turn is specific to and recruits / binds a specific gene effector, gene activator or gene repressor. If there are multiple sites for activator or repressor recruitment, it is preferred that the sites are specific to either activators or repressors. If there are multiple sites for activator or repressor binding, the sites may be specific to the same activators or same repressors. The sites may also be specific to different activators or different repressors. The gene effectors, gene activators, gene repressors may be present in the form of fusion proteins.

[0387] In an embodiment, the dead gRNA as described herein or the Cas13 CRISPR-Cas complex as described herein includes a non-naturally occurring or engineered composition comprising two or more adaptor proteins, wherein each protein is associated with one or more functional domains and wherein the adaptor protein binds to the distinct RNA sequence(s) inserted into the at least one loop of the dead gRNA.

[0388] Hence, an aspect provides a non-naturally occurring or engineered composition comprising a guide RNA (gRNA) comprising a dead guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell, wherein the dead guide sequence is as defined herein, a Cas13 comprising at least one or more nuclear localization sequences, wherein the Cas13 optionally comprises at least one mutation wherein at least one loop of the dead gRNA is modified by the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins, and wherein the adaptor protein is associated with one or more functional domains; or, wherein the dead gRNA is modified to have at least one non-coding functional loop, and wherein the composition comprises two or more adaptor proteins, wherein the each protein is associated with one or more functional domains.

[0389] In certain embodiments, the adaptor protein is a fusion protein comprising the functional domain, the fusion protein optionally comprising a linker between the adaptor protein and the functional domain, the linker optionally including a GlySer linker.

[0390] In certain embodiments, the at least one loop of the dead gRNA is not modified by the insertion of distinct RNA sequence(s) that bind to the two or more adaptor proteins.

[0391] In certain embodiments, the one or more functional domains associated with the adaptor protein is a transcriptional activation domain.

[0392] In certain embodiments, the one or more functional domains associated with the adaptor protein is a transcriptional activation domain comprising VP64, p65, MyoD1, HSF1, RTA or SET7/9.

[0393] In certain embodiments, the one or more functional domains associated with the adaptor protein is a transcriptional repressor domain.

[0394] In certain embodiments, the transcriptional repressor domain is a KRAB domain.

[0395] In certain embodiments, the transcriptional repressor domain is a NuE domain, NcoR domain, SID domain or a SID4X domain.

[0396] In certain embodiments, at least one of the one or more functional domains associated with the adaptor protein have one or more activities comprising methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, DNA integration activity RNA cleavage activity, DNA cleavage activity or nucleic acid binding activity.

[0397] In certain embodiments, the DNA cleavage activity is due to a Fok1 nuclease.

[0398] In certain embodiments, the dead gRNA is modified so that, after dead gRNA binds the adaptor protein and further binds to the Cas13 and target, the functional domain is in a spatial orientation allowing for the functional domain to function in its attributed function.

[0399] In certain embodiments, the at least one loop of the dead gRNA is tetra loop and/or loop2. In certain embodiments, the tetra loop and loop 2 of the dead gRNA are modified by the insertion of the distinct RNA sequence(s).

[0400] In certain embodiments, the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins is an aptamer sequence. In certain embodiments, the aptamer sequence is two or more aptamer sequences specific to the same adaptor protein. In certain embodiments, the aptamer sequence is two or more aptamer sequences specific to different adaptor protein.

[0001] In certain embodiments, the adaptor protein comprises MS2, PP7, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s, PRR1.

[0401] In certain embodiments, the cell is a eukaryotic cell. In certain embodiments, the eukaryotic cell is a mammalian cell, optionally a mouse cell. In certain embodiments, the mammalian cell is a human cell.

[0402] In certain embodiments, a first adaptor protein is associated with a p65 domain and a second adaptor protein is associated with a HSF1 domain.

[0403] In certain embodiments, the composition comprises a Cas13 CRISPR-Cas complex having at least three functional domains, at least one of which is associated with the Cas13 and at least two of which are associated with dead gRNA.

[0404] In certain embodiments, the composition further comprises a second gRNA, wherein the second gRNA is a live gRNA capable of hybridizing to a second target sequence such that a second Cas13 CRISPR-Cas system is directed to a second genomic locus of interest in a cell with detectable indel activity at the second genomic locus resultant from nuclease activity of the Cas13 enzyme of the system.

[0405] In certain embodiments, the composition further comprises a plurality of dead gRNAs and/or a plurality of live gRNAs.

[0406] One aspect of the invention is to take advantage of the modularity and customizability of the gRNA scaffold to establish a series of gRNA scaffolds with different binding sites (in particular aptamers) for recruiting distinct types of effectors in an orthogonal manner. Again, for matters of example and illustration of the broader concept, replacement of the MS2 stem-loops with PP7-interacting stem-loops may be used to bind / recruit repressive elements, enabling multiplexed bidirectional transcriptional control. Thus, in general, gRNA comprising a dead guide may be employed to provide for multiplex transcriptional control and preferred bidirectional transcriptional control. This transcriptional control is most preferred of genes. For example, one or more gRNA comprising dead guide(s) may be employed in targeting the activation of one or more target genes. At the same time, one or more gRNA comprising dead guide(s) may be employed in targeting the repression of one or more target genes. Such a sequence may be applied in a variety of different combinations, for example the target genes are first repressed and then at an appropriate period other targets are activated, or select genes are repressed at the same time as select genes are activated, followed by further activation and/or repression. As a result, multiple components of one or more biological systems may advantageously be addressed together.

[0407] In an aspect, the invention provides nucleic acid molecule(s) encoding dead gRNA or the Cas13 CRISPR-Cas complex or the composition as described herein.

[0408] In an aspect, the invention provides a vector system comprising: a nucleic acid molecule encoding dead guide RNA as defined herein. In certain embodiments, the vector system further comprises a nucleic acid molecule(s) encoding Cas13. In certain embodiments, the vector system further comprises a nucleic acid molecule(s) encoding (live) gRNA. In

certain embodiments, the nucleic acid molecule or the vector further comprises regulatory element(s) operable in a eukaryotic cell operably linked to the nucleic acid molecule encoding the guide sequence (gRNA) and/or the nucleic acid molecule encoding Cas13 and/or the optional nuclear localization sequence(s).

[0409] In another aspect, structural analysis may also be used to study interactions between the dead guide and the active Cas nuclease that enable DNA binding, but no DNA cutting. In this way amino acids important for nuclease activity of Cas are determined. Modification of such amino acids allows for improved Cas enzymes used for gene editing.

[0410] A further aspect is combining the use of dead guides as explained herein with other applications of CRISPR, as explained herein as well as known in the art. For example, gRNA comprising dead guide(s) for targeted multiplex gene activation or repression or targeted multiplex bidirectional gene activation / repression may be combined with gRNA comprising guides which maintain nuclease activity, as explained herein. Such gRNA comprising guides which maintain nuclease activity may or may not further include modifications which allow for repression of gene activity (e.g. aptamers). Such gRNA comprising guides which maintain nuclease activity may or may not further include modifications which allow for activation of gene activity (e.g. aptamers). In such a manner, a further means for multiplex gene control is introduced (e.g. multiplex gene targeted activation without nuclease activity / without indel activity may be provided at the same time or in combination with gene targeted repression with nuclease activity).

[0411] For example, 1) using one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) comprising dead guide(s) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene activators; 2) may be combined with one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) comprising dead guide(s) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene repressors. 1) and/or 2) may then be combined with 3) one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) targeted to one or more genes. This combination can then be carried out in turn with 1) + 2) + 3) with 4) one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene activators. This combination can then be carried in turn with 1) + 2) + 3) + 4) with 5) one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene repressors. As a result various uses and combinations are included in the

invention. For example, combination 1) + 2); combination 1) + 3); combination 2) + 3); combination 1) + 2) + 3); combination 1) + 2) +3) +4); combination 1) + 3) + 4); combination 2) + 3) +4); combination 1) + 2) + 4); combination 1) + 2) +3) +4) + 5); combination 1) + 3) + 4) +5); combination 2) + 3) +4) +5); combination 1) + 2) + 4) +5); combination 1) + 2) +3) + 5); combination 1) + 3) +5); combination 2) + 3) +5); combination 1) + 2) +5).

[0412] In an aspect, the invention provides an algorithm for designing, evaluating, or selecting a dead guide RNA targeting sequence (dead guide sequence) for guiding a Cas13 CRISPR-Cas system to a target gene locus. In particular, it has been determined that dead guide RNA specificity relates to and can be optimized by varying i) GC content and ii) targeting sequence length. In an aspect, the invention provides an algorithm for designing or evaluating a dead guide RNA targeting sequence that minimizes off-target binding or interaction of the dead guide RNA. In an embodiment of the invention, the algorithm for selecting a dead guide RNA targeting sequence for directing a CRISPR system to a gene locus in an organism comprises a) locating one or more CRISPR motifs in the gene locus, analyzing the 20 nt sequence downstream of each CRISPR motif by i) determining the GC content of the sequence; and ii) determining whether there are off-target matches of the 15 downstream nucleotides nearest to the CRISPR motif in the genome of the organism, and c) selecting the 15 nucleotide sequence for use in a dead guide RNA if the GC content of the sequence is 70% or less and no off-target matches are identified. In an embodiment, the sequence is selected for a targeting sequence if the GC content is 60% or less. In certain embodiments, the sequence is selected for a targeting sequence if the GC content is 55% or less, 50% or less, 45% or less, 40% or less, 35% or less or 30% or less. In an embodiment, two or more sequences of the gene locus are analyzed and the sequence having the lowest GC content, or the next lowest GC content, or the next lowest GC content is selected. In an embodiment, the sequence is selected for a targeting sequence if no off-target matches are identified in the genome of the organism. In an embodiment, the targeting sequence is selected if no off-target matches are identified in regulatory sequences of the genome.

[0413] In an aspect, the invention provides a method of selecting a dead guide RNA targeting sequence for directing a functionalized CRISPR system to a gene locus in an organism, which comprises: a) locating one or more CRISPR motifs in the gene locus; b) analyzing the 20 nt sequence downstream of each CRISPR motif by: i) determining the GC content of the sequence; and ii) determining whether there are off-target matches of the first 15 nt of the sequence in the genome of the organism; c) selecting the sequence for use in a guide RNA if the GC content of the sequence is 70% or less and no off-target matches are identified.

In an embodiment, the sequence is selected if the GC content is 50% or less. In an embodiment, the sequence is selected if the GC content is 40% or less. In an embodiment, the sequence is selected if the GC content is 30% or less. In an embodiment, two or more sequences are analyzed and the sequence having the lowest GC content is selected. In an embodiment, off-target matches are determined in regulatory sequences of the organism. In an embodiment, the gene locus is a regulatory region. An aspect provides a dead guide RNA comprising the targeting sequence selected according to the aforementioned methods.

[0414] In an aspect, the invention provides a dead guide RNA for targeting a functionalized CRISPR system to a gene locus in an organism. In an embodiment of the invention, the dead guide RNA comprises a targeting sequence wherein the CG content of the target sequence is 70% or less, and the first 15 nt of the targeting sequence does not match an off-target sequence downstream from a CRISPR motif in the regulatory sequence of another gene locus in the organism. In certain embodiments, the GC content of the targeting sequence 60% or less, 55% or less, 50% or less, 45% or less, 40% or less, 35% or less or 30% or less. In certain embodiments, the GC content of the targeting sequence is from 70% to 60% or from 60% to 50% or from 50% to 40% or from 40% to 30%. In an embodiment, the targeting sequence has the lowest CG content among potential targeting sequences of the locus.

[0415] In an embodiment of the invention, the first 15 nt of the dead guide match the target sequence. In another embodiment, first 14 nt of the dead guide match the target sequence. In another embodiment, the first 13 nt of the dead guide match the target sequence. In another embodiment first 12 nt of the dead guide match the target sequence. In another embodiment, first 11 nt of the dead guide match the target sequence. In another embodiment, the first 10 nt of the dead guide match the target sequence. In an embodiment of the invention the first 15 nt of the dead guide does not match an off-target sequence downstream from a CRISPR motif in the regulatory region of another gene locus. In other embodiments, the first 14 nt, or the first 13 nt of the dead guide, or the first 12 nt of the guide, or the first 11 nt of the dead guide, or the first 10 nt of the dead guide, does not match an off-target sequence downstream from a CRISPR motif in the regulatory region of another gene locus. In other embodiments, the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt of the dead guide do not match an off-target sequence downstream from a CRISPR motif in the genome.

[0416] In certain embodiments, the dead guide RNA includes additional nucleotides at the 3'-end that do not match the target sequence. Thus, a dead guide RNA that includes the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt downstream of a CRISPR motif can be extended in length at the 3' end to 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, or longer.

[0417] The invention provides a method for directing a Cas13 CRISPR-Cas system, including but not limited to a dead Cas13 (dCas13) or functionalized Cas13 system (which may comprise a functionalized Cas13 or functionalized guide) to a gene locus. In an aspect, the invention provides a method for selecting a dead guide RNA targeting sequence and directing a functionalized CRISPR system to a gene locus in an organism. In an aspect, the invention provides a method for selecting a dead guide RNA targeting sequence and effecting gene regulation of a target gene locus by a functionalized Cas13 CRISPR-Cas system. In certain embodiments, the method is used to effect target gene regulation while minimizing off-target effects. In an aspect, the invention provides a method for selecting two or more dead guide RNA targeting sequences and effecting gene regulation of two or more target gene loci by a functionalized Cas13 CRISPR-Cas system. In certain embodiments, the method is used to effect regulation of two or more target gene loci while minimizing off-target effects.

[0418] In an aspect, the invention provides a method of selecting a dead guide RNA targeting sequence for directing a functionalized Cas13 to a gene locus in an organism, which comprises: a) locating one or more CRISPR motifs in the gene locus; b) analyzing the sequence downstream of each CRISPR motif by: i) selecting 10 to 15 nt adjacent to the CRISPR motif, ii) determining the GC content of the sequence; and c) selecting the 10 to 15 nt sequence as a targeting sequence for use in a guide RNA if the GC content of the sequence is 40% or more. In an embodiment, the sequence is selected if the GC content is 50% or more. In an embodiment, the sequence is selected if the GC content is 60% or more. In an embodiment, the sequence is selected if the GC content is 70% or more. In an embodiment, two or more sequences are analyzed and the sequence having the highest GC content is selected. In an embodiment, the method further comprises adding nucleotides to the 3' end of the selected sequence which do not match the sequence downstream of the CRISPR motif. An aspect provides a dead guide RNA comprising the targeting sequence selected according to the aforementioned methods.

[0419] In an aspect, the invention provides a dead guide RNA for directing a functionalized CRISPR system to a gene locus in an organism wherein the targeting sequence of the dead guide RNA consists of 10 to 15 nucleotides adjacent to the CRISPR motif of the gene locus, wherein the CG content of the target sequence is 50% or more. In certain embodiments, the dead guide RNA further comprises nucleotides added to the 3' end of the targeting sequence which do not match the sequence downstream of the CRISPR motif of the gene locus.

[0420] In an aspect, the invention provides for a single effector to be directed to one or more, or two or more gene loci. In certain embodiments, the effector is associated with a Cas13, and one or more, or two or more selected dead guide RNAs are used to direct the Cas13-associated effector to one or more, or two or more selected target gene loci. In certain embodiments, the effector is associated with one or more, or two or more selected dead guide RNAs, each selected dead guide RNA, when complexed with a Cas13 enzyme, causing its associated effector to localize to the dead guide RNA target. One non-limiting example of such CRISPR systems modulates activity of one or more, or two or more gene loci subject to regulation by the same transcription factor.

[0421] In an aspect, the invention provides for two or more effectors to be directed to one or more gene loci. In certain embodiments, two or more dead guide RNAs are employed, each of the two or more effectors being associated with a selected dead guide RNA, with each of the two or more effectors being localized to the selected target of its dead guide RNA. One non-limiting example of such CRISPR systems modulates activity of one or more, or two or more gene loci subject to regulation by different transcription factors. Thus, in one non-limiting embodiment, two or more transcription factors are localized to different regulatory sequences of a single gene. In another non-limiting embodiment, two or more transcription factors are localized to different regulatory sequences of different genes. In certain embodiments, one transcription factor is an activator. In certain embodiments, one transcription factor is an inhibitor. In certain embodiments, one transcription factor is an activator and another transcription factor is an inhibitor. In certain embodiments, gene loci expressing different components of the same regulatory pathway are regulated. In certain embodiments, gene loci expressing components of different regulatory pathways are regulated.

[0422] In an aspect, the invention also provides a method and algorithm for designing and selecting dead guide RNAs that are specific for target DNA cleavage or target binding and gene regulation mediated by an active Cas13 CRISPR-Cas system. In certain embodiments, the Cas13 CRISPR-Cas system provides orthogonal gene control using an active Cas13 which cleaves target DNA at one gene locus while at the same time binds to and promotes regulation of another gene locus.

[0423] In an aspect, the invention provides an method of selecting a dead guide RNA targeting sequence for directing a functionalized Cas13 to a gene locus in an organism, without cleavage, which comprises a) locating one or more CRISPR motifs in the gene locus; b) analyzing the sequence downstream of each CRISPR motif by i) selecting 10 to 15 nt adjacent to the CRISPR motif, ii) determining the GC content of the sequence, and c) selecting the 10

to 15 nt sequence as a targeting sequence for use in a dead guide RNA if the GC content of the sequence is 30% more, 40% or more. In certain embodiments, the GC content of the targeting sequence is 35% or more, 40% or more, 45% or more, 50% or more, 55% or more, 60% or more, 65% or more, or 70% or more. In certain embodiments, the GC content of the targeting sequence is from 30% to 40% or from 40% to 50% or from 50% to 60% or from 60% to 70%. In an embodiment of the invention, two or more sequences in a gene locus are analyzed and the sequence having the highest GC content is selected.

[0424] In an embodiment of the invention, the portion of the targeting sequence in which GC content is evaluated is 10 to 15 contiguous nucleotides of the 15 target nucleotides nearest to the PAM. In an embodiment of the invention, the portion of the guide in which GC content is considered is the 10 to 11 nucleotides or 11 to 12 nucleotides or 12 to 13 nucleotides or 13, or 14, or 15 contiguous nucleotides of the 15 nucleotides nearest to the PAM.

[0425] In an aspect, the invention further provides an algorithm for identifying dead guide RNAs which promote CRISPR system gene locus cleavage while avoiding functional activation or inhibition. It is observed that increased GC content in dead guide RNAs of 16 to 20 nucleotides coincides with increased DNA cleavage and reduced functional activation.

[0426] It is also demonstrated herein that efficiency of functionalized Cas13 can be increased by addition of nucleotides to the 3' end of a guide RNA which do not match a target sequence downstream of the CRISPR motif. For example, of dead guide RNA 11 to 15 nt in length, shorter guides may be less likely to promote target cleavage, but are also less efficient at promoting CRISPR system binding and functional control. In certain embodiments, addition of nucleotides that don't match the target sequence to the 3' end of the dead guide RNA increase activation efficiency while not increasing undesired target cleavage. In an aspect, the invention also provides a method and algorithm for identifying improved dead guide RNAs that effectively promote CRISPRP system function in DNA binding and gene regulation while not promoting DNA cleavage. Thus, in certain embodiments, the invention provides a dead guide RNA that includes the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt downstream of a CRISPR motif and is extended in length at the 3' end by nucleotides that mismatch the target to 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, or longer.

[0427] In an aspect, the invention provides a method for effecting selective orthogonal gene control. As will be appreciated from the disclosure herein, dead guide selection according to the invention, taking into account guide length and GC content, provides effective and selective transcription control by a functional Cas13 CRISPR-Cas system, for example to regulate transcription of a gene locus by activation or inhibition and minimize off-target effects.

Accordingly, by providing effective regulation of individual target loci, the invention also provides effective orthogonal regulation of two or more target loci.

[0428] In certain embodiments, orthogonal gene control is by activation or inhibition of two or more target loci. In certain embodiments, orthogonal gene control is by activation or inhibition of one or more target locus and cleavage of one or more target locus.

[0429] In one aspect, the invention provides a cell comprising a non-naturally occurring Cas13 CRISPR-Cas system comprising one or more dead guide RNAs disclosed or made according to a method or algorithm described herein wherein the expression of one or more gene products has been altered. In an embodiment of the invention, the expression in the cell of two or more gene products has been altered. The invention also provides a cell line from such a cell.

[0430] In one aspect, the invention provides a multicellular organism comprising one or more cells comprising a non-naturally occurring Cas13 CRISPR-Cas system comprising one or more dead guide RNAs disclosed or made according to a method or algorithm described herein. In one aspect, the invention provides a product from a cell, cell line, or multicellular organism comprising a non-naturally occurring Cas13 CRISPR-Cas system comprising one or more dead guide RNAs disclosed or made according to a method or algorithm described herein.

[0431] A further aspect of this invention is the use of gRNA comprising dead guide(s) as described herein, optionally in combination with gRNA comprising guide(s) as described herein or in the state of the art, in combination with systems e.g. cells, transgenic animals, transgenic mice, inducible transgenic animals, inducible transgenic mice) which are engineered for either overexpression of Cas13 or preferably knock in Cas13. As a result a single system (e.g. transgenic animal, cell) can serve as a basis for multiplex gene modifications in systems / network biology. On account of the dead guides, this is now possible in both in vitro, ex vivo, and in vivo.

[0432] For example, once the Cas13 is provided for, one or more dead gRNAs may be provided to direct multiplex gene regulation, and preferably multiplex bidirectional gene regulation. The one or more dead gRNAs may be provided in a spatially and temporally appropriate manner if necessary or desired (for example tissue specific induction of Cas13 expression). On account that the transgenic / inducible Cas13 is provided for (e.g. expressed) in the cell, tissue, animal of interest, both gRNAs comprising dead guides or gRNAs comprising guides are equally effective. In the same manner, a further aspect of this invention is the use of gRNA comprising dead guide(s) as described herein, optionally in combination with gRNA comprising guide(s) as described herein or in the state of the art, in combination

with systems (e.g. cells, transgenic animals, transgenic mice, inducible transgenic animals, inducible transgenic mice) which are engineered for knockout Cas13 CRISPR-Cas.

[0433] As a result, the combination of dead guides as described herein with CRISPR applications described herein and CRISPR applications known in the art results in a highly efficient and accurate means for multiplex screening of systems (e.g. network biology). Such screening allows, for example, identification of specific combinations of gene activities for identifying genes responsible for diseases (e.g. on/off combinations), in particular gene related diseases. A preferred application of such screening is cancer. In the same manner, screening for treatment for such diseases is included in the invention. Cells or animals may be exposed to aberrant conditions resulting in disease or disease like effects. Candidate compositions may be provided and screened for an effect in the desired multiplex environment. For example a patient's cancer cells may be screened for which gene combinations will cause them to die, and then use this information to establish appropriate therapies.

[0434] In one aspect, the invention provides a kit comprising one or more of the components described herein. The kit may include dead guides as described herein with or without guides as described herein.

[0435] The structural information provided herein allows for interrogation of dead gRNA interaction with the target DNA and the Cas13 permitting engineering or alteration of dead gRNA structure to optimize functionality of the entire Cas13 CRISPR-Cas system. For example, loops of the dead gRNA may be extended, without colliding with the Cas13 protein by the insertion of adaptor proteins that can bind to RNA. These adaptor proteins can further recruit effector proteins or fusions which comprise one or more functional domains.

[0436] In some preferred embodiments, the functional domain is a transcriptional activation domain, preferably VP64. In some embodiments, the functional domain is a transcription repression domain, preferably KRAB. In some embodiments, the transcription repression domain is SID, or concatemers of SID (e.g. SID4X). In some embodiments, the functional domain is an epigenetic modifying domain, such that an epigenetic modifying enzyme is provided. In some embodiments, the functional domain is an activation domain, which may be the P65 activation domain.

[0437] An aspect of the invention is that the above elements are comprised in a single composition or comprised in individual compositions. These compositions may advantageously be applied to a host to elicit a functional effect on the genomic level.

[0438] In general, the dead gRNA is modified in a manner that provides specific binding sites (e.g. aptamers) for adapter proteins comprising one or more functional domains (e.g. via

fusion protein) to bind to. The modified dead gRNA is modified such that once the dead gRNA forms a CRISPR complex (i.e. Cas13 binding to dead gRNA and target) the adapter proteins bind and, the functional domain on the adapter protein is positioned in a spatial orientation which is advantageous for the attributed function to be effective. For example, if the functional domain is a transcription activator (e.g. VP64 or p65), the transcription activator is placed in a spatial orientation which allows it to affect the transcription of the target. Likewise, a transcription repressor will be advantageously positioned to affect the transcription of the target and a nuclease (e.g. Fok1) will be advantageously positioned to cleave or partially cleave the target.

[0439] The skilled person will understand that modifications to the dead gRNA which allow for binding of the adapter + functional domain but not proper positioning of the adapter + functional domain (e.g. due to steric hindrance within the three dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified dead gRNA may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein, preferably at either the tetra loop or stem loop 2, and most preferably at both the tetra loop and stem loop 2.

[0440] As explained herein the functional domains may be, for example, one or more domains from the group consisting of methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, DNA cleavage activity, nucleic acid binding activity, and molecular switches (e.g. light inducible). In some cases it is advantageous that additionally at least one NLS is provided. In some instances, it is advantageous to position the NLS at the N terminus. When more than one functional domain is included, the functional domains may be the same or different.

[0441] The dead gRNA may be designed to include multiple binding recognition sites (e.g. aptamers) specific to the same or different adapter protein. The dead gRNA may be designed to bind to the promoter region -1000 - +1 nucleic acids upstream of the transcription start site (i.e. TSS), preferably -200 nucleic acids. This positioning improves functional domains which affect gene activation (e.g. transcription activators) or gene inhibition (e.g. transcription repressors). The modified dead gRNA may be one or more modified dead gRNAs targeted to one or more target loci (e.g. at least 1 gRNA, at least 2 gRNA, at least 5 gRNA, at least 10 gRNA, at least 20 gRNA, at least 30 gRNA, at least 50 gRNA) comprised in a composition.

[0442] The adaptor protein may be any number of proteins that binds to an aptamer or recognition site introduced into the modified dead gRNA and which allows proper positioning of one or more functional domains, once the dead gRNA has been incorporated into the CRISPR complex, to affect the target with the attributed function. As explained in detail in this application such may be coat proteins, preferably bacteriophage coat proteins. The functional domains associated with such adaptor proteins (e.g. in the form of fusion protein) may include, for example, one or more domains from the group consisting of methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, DNA cleavage activity, nucleic acid binding activity, and molecular switches (e.g. light inducible). Preferred domains are Fok1, VP64, P65, HSF1, MyoD1. In the event that the functional domain is a transcription activator or transcription repressor it is advantageous that additionally at least an NLS is provided and preferably at the N terminus. When more than one functional domain is included, the functional domains may be the same or different. The adaptor protein may utilize known linkers to attach such functional domains.

[0443] Thus, the modified dead gRNA, the (inactivated) Cas13 (with or without functional domains), and the binding protein with one or more functional domains, may each individually be comprised in a composition and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or described herein for delivery to a host (e.g. lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g. for lentiviral gRNA selection) and concentration of gRNA (e.g. dependent on whether multiple gRNAs are used) may be advantageous for eliciting an improved effect.

[0444] On the basis of this concept, several variations are appropriate to elicit a genomic locus event, including DNA cleavage, gene activation, or gene deactivation. Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic locus events. The compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling in vivo (e.g. gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling; the use the compositions of the invention to establish cell lines and transgenic animals for optimization and screening purposes).

[0445] The current invention comprehends the use of the compositions of the current invention to establish and utilize conditional or inducible CRISPR transgenic cell /animals, which are not believed prior to the present invention or application. For example, the target cell comprises Cas13 conditionally or inducibly (e.g. in the form of Cre dependent constructs) and/or the adapter protein conditionally or inducibly and, on expression of a vector introduced into the target cell, the vector expresses that which induces or gives rise to the condition of Cas13 expression and/or adaptor expression in the target cell. By applying the teaching and compositions of the current invention with the known method of creating a CRISPR complex, inducible genomic events affected by functional domains are also an aspect of the current invention. One example of this is the creation of a CRISPR knock-in / conditional transgenic animal (e.g. mouse comprising e.g. a Lox-Stop-polyA-Lox(LSL) cassette) and subsequent delivery of one or more compositions providing one or more modified dead gRNA (e.g. -200 nucleotides to TSS of a target gene of interest for gene activation purposes) as described herein (e.g. modified dead gRNA with one or more aptamers recognized by coat proteins, e.g. MS2), one or more adapter proteins as described herein (MS2 binding protein linked to one or more VP64) and means for inducing the conditional animal (e.g. Cre recombinase for rendering Cas13 expression inducible). Alternatively, the adaptor protein may be provided as a conditional or inducible element with a conditional or inducible Cas13 to provide an effective model for screening purposes, which advantageously only requires minimal design and administration of specific dead gRNAs for a broad number of applications.

[0446] In another aspect the dead guides are further modified to improve specificity. Protected dead guides may be synthesized, whereby secondary structure is introduced into the 3' end of the dead guide to improve its specificity. A protected guide RNA (pgRNA) comprises a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell and a protector strand, wherein the protector strand is optionally complementary to the guide sequence and wherein the guide sequence may in part be hybridizable to the protector strand. The pgRNA optionally includes an extension sequence. The thermodynamics of the pgRNA-target DNA hybridization is determined by the number of bases complementary between the guide RNA and target DNA. By employing 'thermodynamic protection', specificity of dead gRNA can be improved by adding a protector sequence. For example, one method adds a complementary protector strand of varying lengths to the 3' end of the guide sequence within the dead gRNA. As a result, the protector strand is bound to at least a portion of the dead gRNA and provides for a protected gRNA (pgRNA). In turn, the dead gRNA references herein may be easily protected using the described embodiments, resulting in

pgRNA. The protector strand can be either a separate RNA transcript or strand or a chimeric version joined to the 3' end of the dead gRNA guide sequence.

Tandem guides and uses in a multiplex (tandem) targeting approach

[0447] The inventors have shown that CRISPR enzymes as defined herein can employ more than one RNA guide without losing activity. This enables the use of the CRISPR enzymes, systems or complexes as defined herein for targeting multiple DNA targets, genes or gene loci, with a single enzyme, system or complex as defined herein. The guide RNAs may be tandemly arranged, optionally separated by a nucleotide sequence such as a direct repeat as defined herein. The position of the different guide RNAs is the tandem does not influence the activity. It is noted that the terms "CRISPR-Cas system", "CRISP-Cas complex" "CRISPR complex" and "CRISPR system" are used interchangeably. Also the terms "CRISPR enzyme", "Cas enzyme", or "CRISPR-Cas enzyme", can be used interchangeably. In preferred embodiments, said CRISPR enzyme, CRISP-Cas enzyme or Cas enzyme is Cas13, or any one of the modified or mutated variants thereof described herein elsewhere.

[0448] In one aspect, the invention provides a non-naturally occurring or engineered CRISPR enzyme, preferably a class 2 CRISPR enzyme, preferably a Type V or VI CRISPR enzyme as described herein, such as without limitation Cas13 as described herein elsewhere, used for tandem or multiplex targeting. It is to be understood that any of the CRISPR (or CRISPR-Cas or Cas) enzymes, complexes, or systems according to the invention as described herein elsewhere may be used in such an approach. Any of the methods, products, compositions and uses as described herein elsewhere are equally applicable with the multiplex or tandem targeting approach further detailed below. By means of further guidance, the following particular aspects and embodiments are provided.

[0449] In one aspect, the invention provides for the use of a Cas13 enzyme, complex or system as defined herein for targeting multiple gene loci. In one embodiment, this can be established by using multiple (tandem or multiplex) guide RNA (gRNA) sequences.

[0450] In one aspect, the invention provides methods for using one or more elements of a Cas13 enzyme, complex or system as defined herein for tandem or multiplex targeting, wherein said CRISP system comprises multiple guide RNA sequences. Preferably, said gRNA sequences are separated by a nucleotide sequence, such as a direct repeat as defined herein elsewhere.

[0451] The Cas13 enzyme, system or complex as defined herein provides an effective means for modifying multiple target polynucleotides. The Cas13 enzyme, system or complex as defined herein has a wide variety of utility including modifying (e.g., deleting, inserting,

translocating, inactivating, activating) one or more target polynucleotides in a multiplicity of cell types. As such the Cas13 enzyme, system or complex as defined herein of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis, including targeting multiple gene loci within a single CRISPR system.

[0452] In one aspect, the invention provides a Cas13 enzyme, system or complex as defined herein, i.e. a Cas13 CRISPR-Cas complex having a Cas13 protein having at least one destabilization domain associated therewith, and multiple guide RNAs that target multiple nucleic acid molecules such as DNA molecules, whereby each of said multiple guide RNAs specifically targets its corresponding nucleic acid molecule, e.g., DNA molecule. Each nucleic acid molecule target, e.g., DNA molecule can encode a gene product or encompass a gene locus. Using multiple guide RNAs hence enables the targeting of multiple gene loci or multiple genes. In some embodiments the Cas13 enzyme may cleave the RNA molecule encoding the gene product. In some embodiments expression of the gene product is altered. The Cas13 protein and the guide RNAs do not naturally occur together. The invention comprehends the guide RNAs comprising tandemly arranged guide sequences. The invention further comprehends coding sequences for the Cas13 protein being codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell and in a more preferred embodiment the mammalian cell is a human cell. Expression of the gene product may be decreased. The Cas13 enzyme may form part of a CRISPR system or complex, which further comprises tandemly arranged guide RNAs (gRNAs) comprising a series of 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 25, 25, 30, or more than 30 guide sequences, each capable of specifically hybridizing to a target sequence in a genomic locus of interest in a cell. In some embodiments, the functional Cas13 CRISPR system or complex binds to the multiple target sequences. In some embodiments, the functional CRISPR system or complex may edit the multiple target sequences, e.g., the target sequences may comprise a genomic locus, and in some embodiments there may be an alteration of gene expression. In some embodiments, the functional CRISPR system or complex may comprise further functional domains. In some embodiments, the invention provides a method for altering or modifying expression of multiple gene products. The method may comprise introducing into a cell containing said target nucleic acids, e.g., DNA molecules, or containing and expressing target nucleic acid, e.g., DNA molecules; for instance, the target nucleic acids may encode gene products or provide for expression of gene products (e.g., regulatory sequences).

[0453] In preferred embodiments the CRISPR enzyme used for multiplex targeting is Cas13, or the CRISPR system or complex comprises Cas13. In some embodiments, the

CRISPR enzyme used for multiplex targeting is AsCas13, or the CRISPR system or complex used for multiplex targeting comprises an AsCas13. In some embodiments, the CRISPR enzyme is an LbCas13, or the CRISPR system or complex comprises LbCas13. In some embodiments, the Cas enzyme used for multiplex targeting cleaves both strands of DNA to produce a double strand break (DSB). In some embodiments, the CRISPR enzyme used for multiplex targeting is a nickase. In some embodiments, the Cas13 enzyme used for multiplex targeting is a dual nickase. In some embodiments, the Cas13 enzyme used for multiplex targeting is a Cas13 enzyme such as a DD Cas13 enzyme as defined herein elsewhere.

[0454] In some general embodiments, the Cas13 enzyme used for multiplex targeting is associated with one or more functional domains. In some more specific embodiments, the CRISPR enzyme used for multiplex targeting is a deadCas13 as defined herein elsewhere.

[0455] In an aspect, the present invention provides a means for delivering the Cas13 enzyme, system or complex for use in multiple targeting as defined herein or the polynucleotides defined herein. Non-limiting examples of such delivery means are e.g. particle(s) delivering component(s) of the complex, vector(s) comprising the polynucleotide(s) discussed herein (e.g., encoding the CRISPR enzyme, providing the nucleotides encoding the CRISPR complex). In some embodiments, the vector may be a plasmid or a viral vector such as AAV, or lentivirus. Transient transfection with plasmids, e.g., into HEK cells may be advantageous, especially given the size limitations of AAV and that while Cas13 fits into AAV, one may reach an upper limit with additional guide RNAs.

[0456] Also provided is a model that constitutively expresses the Cas13 enzyme, complex or system as used herein for use in multiplex targeting. The organism may be transgenic and may have been transfected with the present vectors or may be the offspring of an organism so transfected. In a further aspect, the present invention provides compositions comprising the CRISPR enzyme, system and complex as defined herein or the polynucleotides or vectors described herein. Also provides are Cas13 CRISPR systems or complexes comprising multiple guide RNAs, preferably in a tandemly arranged format. Said different guide RNAs may be separated by nucleotide sequences such as direct repeats.

[0457] Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing gene editing by transforming the subject with the polynucleotide encoding the Cas13 CRISPR system or complex or any of polynucleotides or vectors described herein and administering them to the subject. A suitable repair template may also be provided, for example delivered by a vector comprising said repair template. Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing transcriptional activation

or repression of multiple target gene loci by transforming the subject with the polynucleotides or vectors described herein, wherein said polynucleotide or vector encodes or comprises the Cas13 enzyme, complex or system comprising multiple guide RNAs, preferably tandemly arranged. Where any treatment is occurring ex vivo, for example in a cell culture, then it will be appreciated that the term 'subject' may be replaced by the phrase "cell or cell culture."

[0458] Compositions comprising Cas13 enzyme, complex or system comprising multiple guide RNAs, preferably tandemly arranged, or the polynucleotide or vector encoding or comprising said Cas13 enzyme, complex or system comprising multiple guide RNAs, preferably tandemly arranged, for use in the methods of treatment as defined herein elsewhere are also provided. A kit of parts may be provided including such compositions. Use of said composition in the manufacture of a medicament for such methods of treatment are also provided. Use of a Cas13 CRISPR system in screening is also provided by the present invention, e.g., gain of function screens. Cells which are artificially forced to overexpress a gene are be able to down regulate the gene over time (re-establishing equilibrium) e.g. by negative feedback loops. By the time the screen starts the unregulated gene might be reduced again. Using an inducible Cas13 activator allows one to induce transcription right before the screen and therefore minimizes the chance of false negative hits. Accordingly, by use of the instant invention in screening, e.g., gain of function screens, the chance of false negative results may be minimized.

[0459] In one aspect, the invention provides an engineered, non-naturally occurring CRISPR system comprising a Cas13 protein and multiple guide RNAs that each specifically target a DNA molecule encoding a gene product in a cell, whereby the multiple guide RNAs each target their specific DNA molecule encoding the gene product and the Cas13 protein cleaves the target DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the CRISPR protein and the guide RNAs do not naturally occur together. The invention comprehends the multiple guide RNAs comprising multiple guide sequences, preferably separated by a nucleotide sequence such as a direct repeat and optionally fused to a tracr sequence. In an embodiment of the invention the CRISPR protein is a type V or VI CRISPR-Cas protein and in a more preferred embodiment the CRISPR protein is a Cas13 protein. The invention further comprehends a Cas13 protein being codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of the gene product is decreased.

[0460] In another aspect, the invention provides an engineered, non-naturally occurring vector system comprising one or more vectors comprising a first regulatory element operably linked to the multiple Cas13 CRISPR system guide RNAs that each specifically target a DNA molecule encoding a gene product and a second regulatory element operably linked coding for a CRISPR protein. Both regulatory elements may be located on the same vector or on different vectors of the system. The multiple guide RNAs target the multiple DNA molecules encoding the multiple gene products in a cell and the CRISPR protein may cleave the multiple DNA molecules encoding the gene products (it may cleave one or both strands or have substantially no nuclease activity), whereby expression of the multiple gene products is altered; and, wherein the CRISPR protein and the multiple guide RNAs do not naturally occur together. In a preferred embodiment the CRISPR protein is Cas13 protein, optionally codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of each of the multiple gene products is altered, preferably decreased.

[0461] In one aspect, the invention provides a vector system comprising one or more vectors. In some embodiments, the system comprises: (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the one or more guide sequence(s) direct(s) sequence-specific binding of the CRISPR complex to the one or more target sequence(s) in a eukaryotic cell, wherein the CRISPR complex comprises a Cas13 enzyme complexed with the one or more guide sequence(s) that is hybridized to the one or more target sequence(s); and (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas13 enzyme, preferably comprising at least one nuclear localization sequence and/or at least one NES; wherein components (a) and (b) are located on the same or different vectors of the system. Where applicable, a tracr sequence may also be provided. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a Cas13 CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the CRISPR complex comprises one or more nuclear localization sequences and/or one or more NES of sufficient strength to drive accumulation of said Cas13 CRISPR complex in a detectable amount in or out of the nucleus of a eukaryotic cell. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments,

the second regulatory element is a polymerase II promoter. In some embodiments, each of the guide sequences is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length.

[0462] Recombinant expression vectors can comprise the polynucleotides encoding the Cas13 enzyme, system or complex for use in multiple targeting as defined herein in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) 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).

[0463] In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors comprising the polynucleotides encoding the Cas13 enzyme, system or complex for use in multiple targeting as defined herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art and exemplified herein elsewhere. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In some embodiments, a cell transfected with one or more vectors comprising the polynucleotides encoding the Cas13 enzyme, system or complex for use in multiple targeting as defined herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the components of a Cas13 CRISPR system or complex for use in multiple targeting as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a Cas13 CRISPR system or complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors comprising the polynucleotides encoding the Cas13 enzyme, system or complex for use in multiple targeting as defined herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[0464] The term "regulatory element" is as defined herein elsewhere.

[0465] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[0466] In one aspect, the invention provides a eukaryotic host cell comprising (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide RNA sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the guide sequence(s) direct(s) sequence-specific binding of the Cas13 CRISPR complex to the respective target sequence(s) in a eukaryotic cell, wherein the Cas13 CRISPR complex comprises a Cas13 enzyme complexed with the one or more guide sequence(s) that is hybridized to the respective target sequence(s); and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas13 enzyme comprising preferably at least one nuclear localization sequence and/or NES. In some embodiments, the host cell comprises components (a) and (b). Where applicable, a tracr sequence may also be provided. In some embodiments, component (a), component (b), or components (a) and (b) are stably integrated into a genome of the host eukaryotic cell. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, and optionally separated by a direct repeat, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a Cas13 CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas13 enzyme comprises one or more nuclear localization sequences and/or nuclear export sequences or NES of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in and/or out of the nucleus of a eukaryotic cell.

[0467] In some embodiments, the Cas13 enzyme is a type V or VI CRISPR system enzyme. In some embodiments, the Cas enzyme is a Cas13 enzyme. In some embodiments, the Cas13 enzyme is derived from 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, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, or Porphyromonas macacae Cas13, and may include further alterations or mutations of the Cas13 as defined herein elsewhere, and can be a chimeric Cas13. In some embodiments, the Cas13 enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage

of one or two strands at the location of the target sequence. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the one or more guide sequence(s) is (are each) at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length. When multiple guide RNAs are used, they are preferably separated by a direct repeat sequence. In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example a mammal. Also, the organism may be an arthropod such as an insect. The organism also may be a plant. Further, the organism may be a fungus.

[0468] In one aspect, the invention provides a kit comprising one or more of the components described herein. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a Cas13 CRISPR complex to a target sequence in a eukaryotic cell, wherein the Cas13 CRISPR complex comprises a Cas13 enzyme complexed with the guide sequence that is hybridized to the target sequence; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas13 enzyme comprising a nuclear localization sequence. Where applicable, a tracr sequence may also be provided. In some embodiments, the kit comprises components (a) and (b) located on the same or different vectors of the system. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas13 enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the CRISPR enzyme is a type V or VI CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas13 enzyme. In some embodiments, the Cas13 enzyme is derived from 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, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, or Porphyromonas macacae Cas13 (e.g., modified to have or be associated with at least one DD), and may include further alteration or mutation of the Cas13, and can be a chimeric Cas13. In some embodiments, the DD-CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the DD-CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the DD-CRISPR enzyme lacks or substantially DNA strand cleavage activity (e.g., no more than 5% nuclease activity as compared with a wild type enzyme or enzyme not having the mutation or alteration that decreases nuclease activity). In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length.

[0469] In one aspect, the invention provides a method of modifying multiple target polynucleotides in a host cell such as a eukaryotic cell. In some embodiments, the method comprises allowing a Cas13 CRISPR complex to bind to multiple target polynucleotides, e.g., to effect cleavage of said multiple target polynucleotides, thereby modifying multiple target polynucleotides, wherein the Cas13 CRISPR complex comprises a Cas13 enzyme complexed with multiple guide sequences each of the being hybridized to a specific target sequence within said target polynucleotide, wherein said multiple guide sequences are linked to a direct repeat sequence. Where applicable, a tracr sequence may also be provided (e.g. to provide a single guide RNA, sgRNA). In some embodiments, said cleavage comprises cleaving one or two strands at the location of each of the target sequence by said Cas13 enzyme. In some embodiments, said cleavage results in decreased transcription of the multiple target genes. In some embodiments, the method further comprises repairing one or more of said cleaved target polynucleotide by homologous recombination with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of one or more of said target polynucleotides. In some embodiments, said mutation results in one or more amino acid changes in a protein expressed from a gene comprising one or more of the target sequence(s). In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cell, wherein the one or more

vectors drive expression of one or more of: the Cas13 enzyme and the multiple guide RNA sequence linked to a direct repeat sequence. Where applicable, a tracr sequence may also be provided. In some embodiments, said vectors are delivered to the eukaryotic cell in a subject. In some embodiments, said modifying takes place in said eukaryotic cell in a cell culture. In some embodiments, the method further comprises isolating said eukaryotic cell from a subject prior to said modifying. In some embodiments, the method further comprises returning said eukaryotic cell and/or cells derived therefrom to said subject.

[0470] In one aspect, the invention provides a method of modifying expression of multiple polynucleotides in a eukaryotic cell. In some embodiments, the method comprises allowing a Cas13 CRISPR complex to bind to multiple polynucleotides such that said binding results in increased or decreased expression of said polynucleotides; wherein the Cas13 CRISPR complex comprises a Cas13 enzyme complexed with multiple guide sequences each specifically hybridized to its own target sequence within said polynucleotide, wherein said guide sequences are linked to a direct repeat sequence. Where applicable, a tracr sequence may also be provided. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cells, wherein the one or more vectors drive expression of one or more of: the Cas13 enzyme and the multiple guide sequences linked to the direct repeat sequences. Where applicable, a tracr sequence may also be provided.

[0471] In one aspect, the invention provides a recombinant polynucleotide comprising multiple guide RNA sequences up- or downstream (whichever applicable) of a direct repeat sequence, wherein each of the guide sequences when expressed directs sequence-specific binding of a Cas13 CRISPR complex to its corresponding target sequence present in a eukaryotic cell. In some embodiments, the target sequence is a viral sequence present in a eukaryotic cell. Where applicable, a tracr sequence may also be provided. In some embodiments, the target sequence is a proto-oncogene or an oncogene.

[0472] Aspects of the invention encompass a non-naturally occurring or engineered composition that may comprise a guide RNA (gRNA) comprising a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell and a Cas13 enzyme as defined herein that may comprise at least one or more nuclear localization sequences.

[0473] An aspect of the invention encompasses methods of modifying a genomic locus of interest to change gene expression in a cell by introducing into the cell any of the compositions described herein.

[0474] An aspect of the invention is that the above elements are comprised in a single composition or comprised in individual compositions. These compositions may advantageously be applied to a host to elicit a functional effect on the genomic level.

[0475] As used herein, the term "guide RNA" or "gRNA" has the leaning as used herein elsewhere and comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. Each gRNA may be designed to include multiple binding recognition sites (e.g., aptamers) specific to the same or different adapter protein. Each gRNA may be designed to bind to the promoter region -1000 - +1 nucleic acids upstream of the transcription start site (i.e. TSS), preferably -200 nucleic acids. This positioning improves functional domains which affect gene activation (e.g., transcription activators) or gene inhibition (e.g., transcription repressors). The modified gRNA may be one or more modified gRNAs targeted to one or more target loci (e.g., at least 1 gRNA, at least 2 gRNA, at least 5 gRNA, at least 10 gRNA, at least 20 gRNA, at least 30 g RNA, at least 50 gRNA) comprised in a composition. Said multiple gRNA sequences can be tandemly arranged and are preferably separated by a direct repeat.

[0476] Thus, gRNA, the CRISPR enzyme as defined herein may each individually be comprised in a composition and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or described herein for delivery to a host (e.g., lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g., for lentiviral sgRNA selection) and concentration of gRNA (e.g., dependent on whether multiple gRNAs are used) may be advantageous for eliciting an improved effect. On the basis of this concept, several variations are appropriate to elicit a genomic locus event, including DNA cleavage, gene activation, or gene deactivation. Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic locus events. The compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling in vivo (e.g., gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling; the use the compositions of the invention to establish cell lines and transgenic animals for optimization and screening purposes).

[0477] The current invention comprehends the use of the compositions of the current invention to establish and utilize conditional or inducible CRISPR transgenic cell /animals;

see, e.g., Platt et al., Cell (2014), 159(2): 440-455, or PCT patent publications cited herein, such as WO 2014/093622 (PCT/US2013/074667). For example, cells or animals such as non-human animals, e.g., vertebrates or mammals, such as rodents, e.g., mice, rats, or other laboratory or field animals, e.g., cats, dogs, sheep, etc., may be 'knock-in' whereby the animal conditionally or inducibly expresses Cas13 akin to Platt et al. The target cell or animal thus comprises the CRISPR enzyme (e.g., Cas13) conditionally or inducibly (e.g., in the form of Cre dependent constructs), on expression of a vector introduced into the target cell, the vector expresses that which induces or gives rise to the condition of the CRISPR enzyme (e.g., Cas13) expression in the target cell. By applying the teaching and compositions as defined herein with the known method of creating a CRISPR complex, inducible genomic events are also an aspect of the current invention. Examples of such inducible events have been described herein elsewhere.

[0478] In some embodiments, phenotypic alteration is preferably the result of genome modification when a genetic disease is targeted, especially in methods of therapy and preferably where a repair template is provided to correct or alter the phenotype.

[0479] In some embodiments diseases that may be targeted include those concerned with disease-causing splice defects.

[0480] In some embodiments, cellular targets include Hemopoietic Stem/Progenitor Cells (CD34+); Human T cells; and Eye (retinal cells) - for example photoreceptor precursor cells.

[0481] In some embodiments Gene targets include: Human Beta Globin - HBB (for treating Sickle Cell Anemia, including by stimulating gene-conversion (using closely related HBD gene as an endogenous template)); CD3 (T-Cells); and CEP920 - retina (eye).

[0482] In some embodiments disease targets also include: cancer; Sickle Cell Anemia (based on a point mutation); HBV, HIV; Beta-Thalassemia; and ophthalmic or ocular disease - for example Leber Congenital Amaurosis (LCA)-causing Splice Defect.

[0483] In some embodiments delivery methods include: Cationic Lipid Mediated "direct" delivery of Enzyme-Guide complex (RiboNucleoProtein) and electroporation of plasmid DNA.

[0484] Methods, products and uses described herein may be used for non-therapeutic purposes. Furthermore, any of the methods described herein may be applied in vitro and ex vivo.

[0485] In an aspect, provided is a non-naturally occurring or engineered composition comprising:

I. two or more CRISPR-Cas system polynucleotide sequences comprising

(a) a first guide sequence capable of hybridizing to a first target sequence in a polynucleotide locus,

(b) a second guide sequence capable of hybridizing to a second target sequence in a polynucleotide locus,

(c) a direct repeat sequence,

and

II. a Cas13 enzyme or a second polynucleotide sequence encoding it,

wherein when transcribed, the first and the second guide sequences direct sequence-specific binding of a first and a second Cas13 CRISPR complex to the first and second target sequences respectively,

wherein the first CRISPR complex comprises the Cas13 enzyme complexed with the first guide sequence that is hybridizable to the first target sequence,

wherein the second CRISPR complex comprises the Cas13 enzyme complexed with the second guide sequence that is hybridizable to the second target sequence, and

wherein the first guide sequence directs cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directs cleavage of the other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human or non-animal organism. Similarly, compositions comprising more than two guide RNAs can be envisaged e.g. each specific for one target, and arranged tandemly in the composition or CRISPR system or complex as described herein.

[0486] In another embodiment, the Cas13 is delivered into the cell as a protein. In another and particularly preferred embodiment, the Cas13 is delivered into the cell as a protein or as a nucleotide sequence encoding it. Delivery to the cell as a protein may include delivery of a Ribonucleoprotein (RNP) complex, where the protein is complexed with the multiple guides.

[0487] In an aspect, host cells and cell lines modified by or comprising the compositions, systems or modified enzymes of present invention are provided, including stem cells, and progeny thereof.

[0488] In an aspect, methods of cellular therapy are provided, where, for example, a single cell or a population of cells is sampled or cultured, wherein that cell or cells is or has been modified ex vivo as described herein, and is then re-introduced (sampled cells) or introduced (cultured cells) into the organism. Stem cells, whether embryonic or induce pluripotent or totipotent stem cells, are also particularly preferred in this regard. But, of course, in vivo embodiments are also envisaged.

[0489] Inventive methods can further comprise delivery of templates, such as repair templates, which may be dsODN or ssODN, see below. Delivery of templates may be via the cotemporaneous or separate from delivery of any or all the CRISPR enzyme or guide RNAs and via the same delivery mechanism or different. In some embodiments, it is preferred that the template is delivered together with the guide RNAs and, preferably, also the CRISPR enzyme. An example may be an AAV vector where the CRISPR enzyme is AsCas or LbCas.

[0490] Inventive methods can further comprise: (a) delivering to the cell a double-stranded oligodeoxynucleotide (dsODN) comprising overhangs complimentary to the overhangs created by said double strand break, wherein said dsODN is integrated into the locus of interest; or -(b) delivering to the cell a single-stranded oligodeoxynucleotide (ssODN), wherein said ssODN acts as a template for homology directed repair of said double strand break. Inventive methods can be for the prevention or treatment of disease in an individual, optionally wherein said disease is caused by a defect in said locus of interest. Inventive methods can be conducted in vivo in the individual or ex vivo on a cell taken from the individual, optionally wherein said cell is returned to the individual.

[0491] The invention also comprehends products obtained from using CRISPR enzyme or Cas enzyme or Cas13 enzyme or CRISPR-CRISPR enzyme or CRISPR-Cas system or CRISPR-Cas13 system for use in tandem or multiple targeting as defined herein.

Escorted guides for the Cas13 CRISPR-Cas system according to the invention

[0492] In one aspect the invention provides escorted Cas13 CRISPR-Cas systems or complexes, especially such a system involving an escorted Cas13 CRISPR-Cas system guide. By "escorted" is meant that the Cas13 CRISPR-Cas system or complex or guide is delivered to a selected time or place within a cell, so that activity of the Cas13 CRISPR-Cas system or complex or guide is spatially or temporally controlled. For example, the activity and destination of the Cas13 CRISPR-Cas system or complex or guide may be controlled by an escort RNA aptamer sequence that has binding affinity for an aptamer ligand, such as a cell surface protein or other localized cellular component. Alternatively, the escort aptamer may for example be responsive to an aptamer effector on or in the cell, such as a transient effector, such as an external energy source that is applied to the cell at a particular time.

[0493] The escorted Cas13 CRISPR-Cas systems or complexes have a gRNA with a functional structure designed to improve gRNA structure, architecture, stability, genetic expression, or any combination thereof. Such a structure can include an aptamer.

[0494] Aptamers are biomolecules that can be designed or selected to bind tightly to other ligands, for example using a technique called systematic evolution of ligands by exponential

enrichment (SELEX; Tuerk C, Gold L: "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science 1990, 249:505-510). Nucleic acid aptamers can for example be selected from pools of random-sequence oligonucleotides, with high binding affinities and specificities for a wide range of biomedically relevant targets, suggesting a wide range of therapeutic utilities for aptamers (Keefe, Anthony D., Supriya Pai, and Andrew Ellington. "Aptamers as therapeutics." Nature Reviews Drug Discovery 9.7 (2010): 537-550). These characteristics also suggest a wide range of uses for aptamers as drug delivery vehicles (Levy-Nissenbaum, Etgar, et al. "Nanotechnology and aptamers: applications in drug delivery." Trends in biotechnology 26.8 (2008): 442-449; and, Hicke BJ, Stephens AW. "Escort aptamers: a delivery service for diagnosis and therapy." J Clin Invest 2000, 106:923-928.). Aptamers may also be constructed that function as molecular switches, responding to a que by changing properties, such as RNA aptamers that bind fluorophores to mimic the activity of green fluorescent protein (Paige, Jeremy S., Karen Y. Wu, and Samie R. Jaffrey. "RNA mimics of green fluorescent protein." Science 333.6042 (2011): 642-646). It has also been suggested that aptamers may be used as components of targeted siRNA therapeutic delivery systems, for example targeting cell surface proteins (Zhou, Jiehua, and John J. Rossi. "Aptamer-targeted cell-specific RNA interference." Silence 1.1 (2010): 4).

[0495] Accordingly, provided herein is a gRNA modified, e.g., by one or more aptamer(s) designed to improve gRNA delivery, including delivery across the cellular membrane, to intracellular compartments, or into the nucleus. Such a structure can include, either in addition to the one or more aptamer(s) or without such one or more aptamer(s), moiety(ies) so as to render the guide deliverable, inducible or responsive to a selected effector. The invention accordingly comprehends an gRNA that responds to normal or pathological physiological conditions, including without limitation pH, hypoxia, O2 concentration, temperature, protein concentration, enzymatic concentration, lipid structure, light exposure, mechanical disruption (e.g. ultrasound waves), magnetic fields, electric fields, or electromagnetic radiation.

[0496] An aspect of the invention provides non-naturally occurring or engineered composition comprising an escorted guide RNA (egRNA) comprising:

an RNA guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell; and,

an escort RNA aptamer sequence, wherein the escort aptamer has binding affinity for an aptamer ligand on or in the cell, or the escort aptamer is responsive to a localized aptamer

effector on or in the cell, wherein the presence of the aptamer ligand or effector on or in the cell is spatially or temporally restricted.

[0497] The escort aptamer may for example change conformation in response to an interaction with the aptamer ligand or effector in the cell.

[0498] The escort aptamer may have specific binding affinity for the aptamer ligand.

[0499] The aptamer ligand may be localized in a location or compartment of the cell, for example on or in a membrane of the cell. Binding of the escort aptamer to the aptamer ligand may accordingly direct the egRNA to a location of interest in the cell, such as the interior of the cell by way of binding to an aptamer ligand that is a cell surface ligand. In this way, a variety of spatially restricted locations within the cell may be targeted, such as the cell nucleus or mitochondria.

[0500] Once intended alterations have been introduced, such as by editing intended copies of a gene in the genome of a cell, continued CRISPR/Cas13 expression in that cell is no longer necessary. Indeed, sustained expression would be undesirable in certain casein case of off-target effects at unintended genomic sites, etc. Thus time-limited expression would be useful. Inducible expression offers one approach, but in addition Applicants have engineered a Self-Inactivating Cas13 CRISPR-Cas system that relies on the use of a non-coding guide target sequence within the CRISPR vector itself. Thus, after expression begins, the CRISPR system will lead to its own destruction, but before destruction is complete it will have time to edit the genomic copies of the target gene (which, with a normal point mutation in a diploid cell, requires at most two edits). Simply, the self inactivating Cas13 CRISPR-Cas system includes additional RNA (i.e., guide RNA) that targets the coding sequence for the CRISPR enzyme itself or that targets one or more non-coding guide target sequences complementary to unique sequences present in one or more of the following: (a) within the promoter driving expression of the non-coding RNA elements, (b) within the promoter driving expression of the Cas13 gene, (c) within 100bp of the ATG translational start codon in the Cas13 coding sequence, (d) within the inverted terminal repeat (iTR) of a viral delivery vector, e.g., in an AAV genome.

[0501] The egRNA may include an RNA aptamer linking sequence, operably linking the escort RNA sequence to the RNA guide sequence.

[0502] In embodiments, the egRNA may include one or more photolabile bonds or non-naturally occurring residues.

[0503] In one aspect, the escort RNA aptamer sequence may be complementary to a target miRNA, which may or may not be present within a cell, so that only when the target miRNA is present is there binding of the escort RNA aptamer sequence to the target miRNA which

results in cleavage of the egRNA by an RNA-induced silencing complex (RISC) within the cell.

[0504] In embodiments, the escort RNA aptamer sequence may for example be from 10 to 200 nucleotides in length, and the egRNA may include more than one escort RNA aptamer sequence.

[0505] It is to be understood that any of the RNA guide sequences as described herein elsewhere can be used in the egRNA described herein. In certain embodiments of the invention, the guide RNA or mature crRNA comprises, consists essentially of, or consists of a direct repeat sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or mature crRNA comprises, consists essentially of, or consists of a direct repeat sequence linked to a guide sequence or spacer sequence. In certain embodiments the guide RNA or mature crRNA comprises 19 nts of partial direct repeat followed by 23-25 nt of guide sequence or spacer sequence. In certain embodiments, the effector protein is a FnCas13 effector protein and requires at least 16 nt of guide sequence to achieve detectable DNA cleavage and a minimum of 17 nt of guide sequence to achieve efficient DNA cleavage in vitro. In certain embodiments, the direct repeat sequence is located upstream (i.e., 5') from the guide sequence or spacer sequence. In a preferred embodiment the seed sequence (i.e. the sequence essential critical for recognition and/or hybridization to the sequence at the target locus) of the FnCas13 guide RNA is approximately within the first 5 nt on the 5' end of the guide sequence or spacer sequence.

[0506] The egRNA may be included in a non-naturally occurring or engineered Cas13 CRISPR-Cas complex composition, together with a Cas13 which may include at least one mutation, for example a mutation so that the Cas13 has no more than 5% of the nuclease activity of a Cas13 not having the at least one mutation, for example having a diminished nuclease activity of at least 97%, or 100% as compared with the Cas13 not having the at least one mutation. The Cas13 may also include one or more nuclear localization sequences. Mutated Cas13 enzymes having modulated activity such as diminished nuclease activity are described herein elsewhere.

[0507] The engineered Cas13 CRISPR-Cas composition may be provided in a cell, such as a eukaryotic cell, a mammalian cell, or a human cell.

[0508] In embodiments, the compositions described herein comprise a Cas13 CRISPR-Cas complex having at least three functional domains, at least one of which is associated with Cas13 and at least two of which are associated with egRNA.

[0509] The compositions described herein may be used to introduce a genomic locus event in a host cell, such as a eukaryotic cell, in particular a mammalian cell, or a non-human eukaryote, in particular a non-human mammal such as a mouse, in vivo. The genomic locus event may comprise affecting gene activation, gene inhibition, or cleavage in a locus. The compositions described herein may also be used to modify a genomic locus of interest to change gene expression in a cell. Methods of introducing a genomic locus event in a host cell using the Cas13 enzyme provided herein are described herein in detail elsewhere. Delivery of the composition may for example be by way of delivery of a nucleic acid molecule(s) coding for the composition, which nucleic acid molecule(s) is operatively linked to regulatory sequence(s), and expression of the nucleic acid molecule(s) in vivo, for example by way of a lentivirus, an adenovirus, or an AAV.

[0510] The present invention provides compositions and methods by which gRNA-mediated gene editing activity can be adapted. The invention provides gRNA secondary structures that improve cutting efficiency by increasing gRNA and/or increasing the amount of RNA delivered into the cell. The gRNA may include light labile or inducible nucleotides.

[0511] To increase the effectiveness of gRNA, for example gRNA delivered with viral or non-viral technologies, Applicants added secondary structures into the gRNA that enhance its stability and improve gene editing. Separately, to overcome the lack of effective delivery, Applicants modified gRNAs with cell penetrating RNA aptamers; the aptamers bind to cell surface receptors and promote the entry of gRNAs into cells. Notably, the cell-penetrating aptamers can be designed to target specific cell receptors, in order to mediate cell-specific delivery. Applicants also have created guides that are inducible.

[0512] Light responsiveness of an inducible system may be achieved via the activation and binding of cryptochrome-2 and CIB1. Blue light stimulation induces an activating conformational change in cryptochrome-2, resulting in recruitment of its binding partner CIB1. This binding is fast and reversible, achieving saturation in <15 sec following pulsed stimulation and returning to baseline <15 min after the end of stimulation. These rapid binding kinetics result in a system temporally bound only by the speed of transcription/translation and transcript/protein degradation, rather than uptake and clearance of inducing agents. Crytochrome-2 activation is also highly sensitive, allowing for the use of low light intensity stimulation and mitigating the risks of phototoxicity. Further, in a context such as the intact mammalian brain, variable light intensity may be used to control the size of a stimulated region, allowing for greater precision than vector delivery alone may offer.

[0513] The invention contemplates energy sources such as electromagnetic radiation, sound energy or thermal energy to induce the guide. Advantageously, the electromagnetic radiation is a component of visible light. In a preferred embodiment, the light is a blue light with a wavelength of about 450 to about 495 nm. In an especially preferred embodiment, the wavelength is about 488 nm. In another preferred embodiment, the light stimulation is via pulses. The light power may range from about 0-9 mW/cm2. In a preferred embodiment, a stimulation paradigm of as low as 0.25 sec every 15 sec should result in maximal activation.

[0514] Cells involved in the practice of the present invention may be a prokaryotic cell or a eukaryotic cell, advantageously an animal cell a plant cell or a yeast cell, more advantageously a mammalian cell.

[0515] The chemical or energy sensitive guide may undergo a conformational change upon induction by the binding of a chemical source or by the energy allowing it act as a guide and have the Cas13 CRISPR-Cas system or complex function. The invention can involve applying the chemical source or energy so as to have the guide function and the Cas13 CRISPR-Cas system or complex function; and optionally further determining that the expression of the genomic locus is altered.

[0516] There are several different designs of this chemical inducible system: 1. ABI-PYL based system inducible by Abscisic Acid (ABA) (see, e.g., http://stke.sciencemag.org/cgi/content/abstract/sigtrans;4/164/rs2), 2. FKBP-FRB based system inducible by rapamycin (or related chemicals based on rapamycin) (see, e.g., http://www.nature.com/nmeth/journal/v2/n6/full/nmeth763.html), 3. GID1-GAI based system inducible by Gibberellin (GA) (see, e.g., http://www.nature.com/nchembio/journal/v8/n5/full/nchembio.922.html).

[0517] Another system contemplated by the present invention is a chemical inducible system based on change in sub-cellular localization. Applicants also developed a system in which the polypeptide include a DNA binding domain comprising at least five or more Transcription activator-like effector (TALE) monomers and at least one or more half-monomers specifically ordered to target the genomic locus of interest linked to at least one or more effector domains are further linker to a chemical or energy sensitive protein. This protein will lead to a change in the sub-cellular localization of the entire polypeptide (i.e. transportation of the entire polypeptide from cytoplasm into the nucleus of the cells) upon the binding of a chemical or energy transfer to the chemical or energy sensitive protein. This transportation of the entire polypeptide from one sub-cellular compartments or organelles, in which its activity is sequestered due to lack of substrate for the effector domain, into another one in which the

substrate is present would allow the entire polypeptide to come in contact with its desired substrate (i.e. genomic DNA in the mammalian nucleus) and result in activation or repression of target gene expression.

[0518] This type of system could also be used to induce the cleavage of a genomic locus of interest in a cell when the effector domain is a nuclease.

[0519] A chemical inducible system can be an estrogen receptor (ER) based system inducible by 4-hydroxytamoxifen (4OHT) (see, e.g., http://www.pnas.org/content/104/3/1027.abstract). A mutated ligand-binding domain of the estrogen receptor called ERT2 translocates into the nucleus of cells upon binding of 4-hydroxytamoxifen. In further embodiments of the invention any naturally occurring or engineered derivative of any nuclear receptor, thyroid hormone receptor, retinoic acid receptor, estrogen receptor, estrogen-related receptor, glucocorticoid receptor, progesterone receptor, androgen receptor may be used in inducible systems analogous to the ER based inducible system.

[0520] Another inducible system is based on the design using Transient receptor potential (TRP) ion channel based system inducible by energy, heat or radio-wave (see, e.g., http://www.sciencemag.org/content/336/6081/604). These TRP family proteins respond to different stimuli, including light and heat. When this protein is activated by light or heat, the ion channel will open and allow the entering of ions such as calcium into the plasma membrane. This influx of ions will bind to intracellular ion interacting partners linked to a polypeptide including the guide and the other components of the Cas13 CRISPR-Cas complex or system, and the binding will induce the change of sub-cellular localization of the polypeptide, leading to the entire polypeptide entering the nucleus of cells. Once inside the nucleus, the guide protein and the other components of the Cas13 CRISPR-Cas complex will be active and modulating target gene expression in cells.

[0521] This type of system could also be used to induce the cleavage of a genomic locus of interest in a cell; and, in this regard, it is noted that the Cas13 enzyme is a nuclease. The light could be generated with a laser or other forms of energy sources. The heat could be generated by raise of temperature results from an energy source, or from nano-particles that release heat after absorbing energy from an energy source delivered in the form of radio-wave.

[0522] While light activation may be an advantageous embodiment, sometimes it may be disadvantageous especially for in vivo applications in which the light may not penetrate the skin or other organs. In this instance, other methods of energy activation are contemplated, in particular, electric field energy and/or ultrasound which have a similar effect.

[0523] Electric field energy is preferably administered substantially as described in the art, using one or more electric pulses of from about 1 Volt/cm to about 10 kVolts/cm under in vivo conditions. Instead of or in addition to the pulses, the electric field may be delivered in a continuous manner. The electric pulse may be applied for between 1 ¬μs and 500 milliseconds, preferably between 1 ¬μs and 100 milliseconds. The electric field may be applied continuously or in a pulsed manner for 5 about minutes.

[0524] As used herein, 'electric field energy' is the electrical energy to which a cell is exposed. Preferably the electric field has a strength of from about 1 Volt/cm to about 10 kVolts/cm or more under in vivo conditions (see WO97/49450).

[0525] As used herein, the term "electric field" includes one or more pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave and/or modulated square wave forms. References to electric fields and electricity should be taken to include reference the presence of an electric potential difference in the environment of a cell. Such an environment may be set up by way of static electricity, alternating current (AC), direct current (DC), etc, as known in the art. The electric field may be uniform, non-uniform or otherwise, and may vary in strength and/or direction in a time dependent manner.

[0526] Single or multiple applications of electric field, as well as single or multiple applications of ultrasound are also possible, in any order and in any combination. The ultrasound and/or the electric field may be delivered as single or multiple continuous applications, or as pulses (pulsatile delivery).

[0527] Electroporation has been used in both in vitro and in vivo procedures to introduce foreign material into living cells. With in vitro applications, a sample of live cells is first mixed with the agent of interest and placed between electrodes such as parallel plates. Then, the electrodes apply an electrical field to the cell/implant mixture. Examples of systems that perform in vitro electroporation include the Electro Cell Manipulator ECM600 product, and the Electro Square Porator T820, both made by the BTX Division of Genetronics, Inc (see U.S. Pat. No 5,869,326).

[0528] The known electroporation techniques (both in vitro and in vivo) function by applying a brief high voltage pulse to electrodes positioned around the treatment region. The electric field generated between the electrodes causes the cell membranes to temporarily become porous, whereupon molecules of the agent of interest enter the cells. In known electroporation applications, this electric field comprises a single square wave pulse on the order of 1000 V/cm, of about 100 .mu.s duration. Such a pulse may be generated, for example, in known applications of the Electro Square Porator T820.

[0529] Preferably, the electric field has a strength of from about 1 V/cm to about 10 kV/cm under in vitro conditions. Thus, the electric field may have a strength of 1 V/cm, 2 V/cm, 3 V/cm, 4 V/cm, 5 V/cm, 6 V/cm, 7 V/cm, 8 V/cm, 9 V/cm, 10 V/cm, 20 V/cm, 50 V/cm, 100 V/cm, 200 V/cm, 300 V/cm, 400 V/cm, 500 V/cm, 600 V/cm, 700 V/cm, 800 V/cm, 900 V/cm, 1 kV/cm, 2 kV/cm, 5 kV/cm, 10 kV/cm, 20 kV/cm, 50 kV/cm or more. More preferably from about 0.5 kV/cm to about 4.0 kV/cm under in vitro conditions. Preferably the electric field has a strength of from about 1 V/cm to about 10 kV/cm under in vivo conditions. However, the electric field strengths may be lowered where the number of pulses delivered to the target site are increased. Thus, pulsatile delivery of electric fields at lower field strengths is envisaged.

[0530] Preferably the application of the electric field is in the form of multiple pulses such as double pulses of the same strength and capacitance or sequential pulses of varying strength and/or capacitance. As used herein, the term "pulse" includes one or more electric pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave/square wave forms.

[0531] Preferably the electric pulse is delivered as a waveform selected from an exponential wave form, a square wave form, a modulated wave form and a modulated square wave form.

[0532] A preferred embodiment employs direct current at low voltage. Thus, Applicants disclose the use of an electric field which is applied to the cell, tissue or tissue mass at a field strength of between 1V/cm and 20V/cm, for a period of 100 milliseconds or more, preferably 15 minutes or more.

[0533] Ultrasound is advantageously administered at a power level of from about 0.05 W/cm2 to about 100 W/cm2. Diagnostic or therapeutic ultrasound may be used, or combinations thereof.

[0534] As used herein, the term "ultrasound" refers to a form of energy which consists of mechanical vibrations the frequencies of which are so high they are above the range of human hearing. Lower frequency limit of the ultrasonic spectrum may generally be taken as about 20 kHz. Most diagnostic applications of ultrasound employ frequencies in the range 1 and 15 MHz' (From Ultrasonics in Clinical Diagnosis, P. N. T. Wells, ed., 2nd. Edition, Publ. Churchill Livingstone [Edinburgh, London & NY, 1977]).

[0535] Ultrasound has been used in both diagnostic and therapeutic applications. When used as a diagnostic tool ("diagnostic ultrasound"), ultrasound is typically used in an energy density range of up to about 100 mW/cm2 (FDA recommendation), although energy densities of up to 750 mW/cm2 have been used. In physiotherapy, ultrasound is typically used as an

energy source in a range up to about 3 to 4 W/cm2 (WHO recommendation). In other therapeutic applications, higher intensities of ultrasound may be employed, for example, HIFU at 100 W/cm up to 1 kW/cm2 (or even higher) for short periods of time. The term "ultrasound" as used in this specification is intended to encompass diagnostic, therapeutic and focused ultrasound.

[0536] Focused ultrasound (FUS) allows thermal energy to be delivered without an invasive probe (see Morocz et al 1998 Journal of Magnetic Resonance Imaging Vol.8, No. 1, pp.136-142. Another form of focused ultrasound is high intensity focused ultrasound (HIFU) which is reviewed by Moussatov et al in Ultrasonics (1998) Vol.36, No.8, pp.893-900 and TranHuuHue et al in Acustica (1997) Vol.83, No.6, pp.1103-1106.

[0537] Preferably, a combination of diagnostic ultrasound and a therapeutic ultrasound is employed. This combination is not intended to be limiting, however, and the skilled reader will appreciate that any variety of combinations of ultrasound may be used. Additionally, the energy density, frequency of ultrasound, and period of exposure may be varied.

[0538] Preferably the exposure to an ultrasound energy source is at a power density of from about 0.05 to about 100 Wcm-2. Even more preferably, the exposure to an ultrasound energy source is at a power density of from about 1 to about 15 Wcm-2.

[0539] Preferably the exposure to an ultrasound energy source is at a frequency of from about 0.015 to about 10.0 MHz. More preferably the exposure to an ultrasound energy source is at a frequency of from about 0.02 to about 5.0 MHz or about 6.0 MHz. Most preferably, the ultrasound is applied at a frequency of 3 MHz.

[0540] Preferably the exposure is for periods of from about 10 milliseconds to about 60 minutes. Preferably the exposure is for periods of from about 1 second to about 5 minutes. More preferably, the ultrasound is applied for about 2 minutes. Depending on the particular target cell to be disrupted, however, the exposure may be for a longer duration, for example, for 15 minutes.

[0541] Advantageously, the target tissue is exposed to an ultrasound energy source at an acoustic power density of from about 0.05 Wcm-2 to about 10 Wcm-2 with a frequency ranging from about 0.015 to about 10 MHz (see WO 98/52609). However, alternatives are also possible, for example, exposure to an ultrasound energy source at an acoustic power density of above 100 Wcm-2, but for reduced periods of time, for example, 1000 Wcm-2 for periods in the millisecond range or less.

[0542] Preferably the application of the ultrasound is in the form of multiple pulses; thus, both continuous wave and pulsed wave (pulsatile delivery of ultrasound) may be employed in

any combination. For example, continuous wave ultrasound may be applied, followed by pulsed wave ultrasound, or vice versa. This may be repeated any number of times, in any order and combination. The pulsed wave ultrasound may be applied against a background of continuous wave ultrasound, and any number of pulses may be used in any number of groups.

[0543] Preferably, the ultrasound may comprise pulsed wave ultrasound. In a highly preferred embodiment, the ultrasound is applied at a power density of 0.7 Wcm-2 or 1.25 Wcm-2 as a continuous wave. Higher power densities may be employed if pulsed wave ultrasound is used.

[0544] Use of ultrasound is advantageous as, like light, it may be focused accurately on a target. Moreover, ultrasound is advantageous as it may be focused more deeply into tissues unlike light. It is therefore better suited to whole-tissue penetration (such as but not limited to a lobe of the liver) or whole organ (such as but not limited to the entire liver or an entire muscle, such as the heart) therapy. Another important advantage is that ultrasound is a non-invasive stimulus which is used in a wide variety of diagnostic and therapeutic applications. By way of example, ultrasound is well known in medical imaging techniques and, additionally, in orthopedic therapy. Furthermore, instruments suitable for the application of ultrasound to a subject vertebrate are widely available and their use is well known in the art.

[0545] The rapid transcriptional response and endogenous targeting of the instant invention make for an ideal system for the study of transcriptional dynamics. For example, the instant invention may be used to study the dynamics of variant production upon induced expression of a target gene. On the other end of the transcription cycle, mRNA degradation studies are often performed in response to a strong extracellular stimulus, causing expression level changes in a plethora of genes. The instant invention may be utilized to reversibly induce transcription of an endogenous target, after which point stimulation may be stopped and the degradation kinetics of the unique target may be tracked.

[0546] The temporal precision of the instant invention may provide the power to time genetic regulation in concert with experimental interventions. For example, targets with suspected involvement in long-term potentiation (LTP) may be modulated in organotypic or dissociated neuronal cultures, but only during stimulus to induce LTP, so as to avoid interfering with the normal development of the cells. Similarly, in cellular models exhibiting disease phenotypes, targets suspected to be involved in the effectiveness of a particular therapy may be modulated only during treatment. Conversely, genetic targets may be modulated only during a pathological stimulus. Any number of experiments in which timing of genetic cues to external

experimental stimuli is of relevance may potentially benefit from the utility of the instant invention.

[0547] The in vivo context offers equally rich opportunities for the instant invention to control gene expression. Photoinducibility provides the potential for spatial precision. Taking advantage of the development of optrode technology, a stimulating fiber optic lead may be placed in a precise brain region. Stimulation region size may then be tuned by light intensity. This may be done in conjunction with the delivery of the Cas13 CRISPR-Cas system or complex of the invention, or, in the case of transgenic Cas13 animals, guide RNA of the invention may be delivered and the optrode technology can allow for the modulation of gene expression in precise brain regions. A transparent Cas13 expressing organism, can have guide RNA of the invention administered to it and then there can be extremely precise laser induced local gene expression changes.

[0548] A culture medium for culturing host cells includes a medium commonly used for tissue culture, such as M199-earle base, Eagle MEM (E-MEM), Dulbecco MEM (DMEM), SC-UCM102, UP-SFM (GIBCO BRL), EX-CELL302 (Nichirei), EX-CELL293-S (Nichirei), TFBM-01 (Nichirei), ASF104, among others. Suitable culture media for specific cell types may be found at the American Type Culture Collection (ATCC) or the European Collection of Cell Cultures (ECACC). Culture media may be supplemented with amino acids such as L-glutamine, salts, anti-fungal or anti-bacterial agents such as Fungizone¬Æ, penicillin-streptomycin, animal serum, and the like. The cell culture medium may optionally be serum-free.

[0549] The invention may also offer valuable temporal precision in vivo. The invention may be used to alter gene expression during a particular stage of development. The invention may be used to time a genetic cue to a particular experimental window. For example, genes implicated in learning may be overexpressed or repressed only during the learning stimulus in a precise region of the intact rodent or primate brain. Further, the invention may be used to induce gene expression changes only during particular stages of disease development. For example, an oncogene may be overexpressed only once a tumor reaches a particular size or metastatic stage. Conversely, proteins suspected in the development of Alzheimer's may be knocked down only at defined time points in the animal's life and within a particular brain region. Although these examples do not exhaustively list the potential applications of the invention, they highlight some of the areas in which the invention may be a powerful technology.

Protected guides: Enzymes according to the invention can be used in combination with protected guide RNAs

[0550] In one aspect, an object of the current invention is to further enhance the specificity of Cas13 given individual guide RNAs through thermodynamic tuning of the binding specificity of the guide RNA to target DNA. This is a general approach of introducing mismatches, elongation or truncation of the guide sequence to increase / decrease the number of complimentary bases vs. mismatched bases shared between a genomic target and its potential off-target loci, in order to give thermodynamic advantage to targeted genomic loci over genomic off-targets.

[0551] In one aspect, the invention provides for the guide sequence being modified by secondary structure to increase the specificity of the Cas13 CRISPR-Cas system and whereby the secondary structure can protect against exonuclease activity and allow for 3' additions to the guide sequence.

[0552] In one aspect, the invention provides for hybridizing a "protector RNA" to a guide sequence, wherein the "protector RNA" is an RNA strand complementary to the 5' end of the guide RNA (gRNA), to thereby generate a partially double-stranded gRNA. In an embodiment of the invention, protecting the mismatched bases with a perfectly complementary protector sequence decreases the likelihood of target DNA binding to the mismatched base pairs at the 3' end. In embodiments of the invention, additional sequences comprising an extended length may also be present.

[0553] Guide RNA (gRNA) extensions matching the genomic target provide gRNA protection and enhance specificity. Extension of the gRNA with matching sequence distal to the end of the spacer seed for individual genomic targets is envisaged to provide enhanced specificity. Matching gRNA extensions that enhance specificity have been observed in cells without truncation. Prediction of gRNA structure accompanying these stable length extensions has shown that stable forms arise from protective states, where the extension forms a closed loop with the gRNA seed due to complimentary sequences in the spacer extension and the spacer seed. These results demonstrate that the protected guide concept also includes sequences matching the genomic target sequence distal of the 20mer spacer-binding region. Thermodynamic prediction can be used to predict completely matching or partially matching guide extensions that result in protected gRNA states. This extends the concept of protected gRNAs to interaction between X and Z, where X will generally be of length 17-20nt and Z is of length 1-30nt. Thermodynamic prediction can be used to determine the optimal extension state for Z, potentially introducing small numbers of mismatches in Z to promote the formation

of protected conformations between X and Z. Throughout the present application, the terms "X" and seed length (SL) are used interchangeably with the term exposed length (EpL) which denotes the number of nucleotides available for target DNA to bind; the terms "Y" and protector length (PL) are used interchangeably to represent the length of the protector; and the terms "Z", "E", "E'" and "EL" are used interchangeably to correspond to the term extended length (ExL) which represents the number of nucleotides by which the target sequence is extended.

[0554] An extension sequence which corresponds to the extended length (ExL) may optionally be attached directly to the guide sequence at the 3' end of the protected guide sequence. The extension sequence may be 2 to 12 nucleotides in length. Preferably ExL may be denoted as 0, 2, 4, 6, 8, 10 or 12 nucleotides in length.. In a preferred embodiment the ExL is denoted as 0 or 4 nucleotides in length. In a more preferred embodiment the ExL is 4 nucleotides in length. The extension sequence may or may not be complementary to the target sequence.

[0555] An extension sequence may further optionally be attached directly to the guide sequence at the 5' end of the protected guide sequence as well as to the 3' end of a protecting sequence. As a result, the extension sequence serves as a linking sequence between the protected sequence and the protecting sequence. Without wishing to be bound by theory, such a link may position the protecting sequence near the protected sequence for improved binding of the protecting sequence to the protected sequence. It will be understood that the above-described relationship of seed, protector, and extension applies where the distal end (i.e., the targeting end) of the guide is the 5' end, e.g. a guide that functions is a Cas13 system. In an embodiment wherein the distal end of the guide is the 3' end, the relationship will be the reverse. In such an embodiment, the invention provides for hybridizing a "protector RNA" to a guide sequence, wherein the "protector RNA" is an RNA strand complementary to the 3' end of the guide RNA (gRNA), to thereby generate a partially double-stranded gRNA.

[0556] Addition of gRNA mismatches to the distal end of the gRNA can demonstrate enhanced specificity. The introduction of unprotected distal mismatches in Y or extension of the gRNA with distal mismatches (Z) can demonstrate enhanced specificity. This concept as mentioned is tied to X, Y, and Z components used in protected gRNAs. The unprotected mismatch concept may be further generalized to the concepts of X, Y, and Z described for protected guide RNAs.

[0557] Cas13. In one aspect, the invention provides for enhanced Cas13 specificity wherein the double stranded 3' end of the protected guide RNA (pgRNA) allows for two

possible outcomes: (1) the guide RNA-protector RNA to guide RNA-target DNA strand exchange will occur and the guide will fully bind the target, or (2) the guide RNA will fail to fully bind the target and because Cas13 target cleavage is a multiple step kinetic reaction that requires guide RNA:target DNA binding to activate Cas13-catalyzed DSBs, wherein Cas13 cleavage does not occur if the guide RNA does not properly bind. According to particular embodiments, the protected guide RNA improves specificity of target binding as compared to a naturally occurring CRISPR-Cas system. According to particular embodiments the protected modified guide RNA improves stability as compared to a naturally occurring CRISPR-Cas. According to particular embodiments the protector sequence has a length between 3 and 120 nucleotides and comprises 3 or more contiguous nucleotides complementary to another sequence of guide or protector. According to particular embodiments, the protector sequence forms a hairpin. According to particular embodiments the guide RNA further comprises a protected sequence and an exposed sequence. According to particular embodiments the exposed sequence is 1 to 19 nucleotides. More particularly, the exposed sequence is at least 75%, at least 90% or about 100% complementary to the target sequence. According to particular embodiments the guide sequence is at least 90% or about 100% complementary to the protector strand. According to particular embodiments the guide sequence is at least 75%, at least 90% or about 100% complementary to the target sequence. According to particular embodiments, the guide RNA further comprises an extension sequence. More particularly, when the distal end of the guide is the 3' end, the extension sequence is operably linked to the 3' end of the protected guide sequence, and optionally directly linked to the 3' end of the protected guide sequence. According to particular embodiments the extension sequence is 1-12 nucleotides. According to particular embodiments the extension sequence is operably linked to the guide sequence at the 3' end of the protected guide sequence and the 5' end of the protector strand and optionally directly linked to the 3' end of the protected guide sequence and the 53' end of the protector strand, wherein the extension sequence is a linking sequence between the protected sequence and the protector strand. According to particular embodiments the extension sequence is 100% not complementary to the protector strand, optionally at least 95%, at least 90%, at least 80%, at least 70%, at least 60%, or at least 50% not complementary to the protector strand. According to particular embodiments the guide sequence further comprises mismatches appended to the end of the guide sequence, wherein the mismatches thermodynamically optimize specificity.

[0558] According to the invention, in certain embodiments, guide modifications that impede strand invasion will be desireable. For example, to minimize off-target actifity, in

certain embodiments, it will be desireable to design or modify a guide to impede strand invasiom at off-target sites. In certain such embodiments, it may be acceptable or useful to design or modify a guide at the expense of on-target binding efficiency. In certain embodiments, guide-target mismatches at the target site may be tolerated that substantially reduce off-target activity.

[0559] In certain embodiments of the invention, it is desirable to adjust the binding characteristics of the protected guide to minimize off-target CRISPR activity. Accordingly, thermodynamic prediction algoithms are used to predict strengths of binding on target and off target. Alternatively or in addition, selection methods are used to reduce or minimize off-target effects, by absolute measures or relative to on-target effects.

[0560] Design options include, without limitation, i) adjusting the length of protector strand that binds to the protected strand, ii) adjusting the length of the portion of the protected strand that is exposed, iii) extending the protected strand with a stem-loop located external (distal) to the protected strand (i.e. designed so that the stem loop is external to the protected strand at the distal end), iv) extending the protected strand by addition of a protector strand to form a stem-loop with all or part of the protected strand, v) adjusting binding of the protector strand to the protected strand by designing in one or more base mismatches and/or one or more non-canonical base pairings, vi) adjusting the location of the stem formed by hybridization of the protector strand to the protected strand, and vii) addition of a non-structured protector to the end of the protected strand.

[0561] In one aspect, the invention provides an engineered, non-naturally occurring CRISPR-Cas system comprising a Cas13 protein and a protected guide RNA that targets a DNA molecule encoding a gene product in a cell, whereby the protected guide RNA targets the DNA molecule encoding the gene product and the Cas13 protein cleaves the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas13 protein and the protected guide RNA do not naturally occur together. The invention comprehends the protected guide RNA comprising a guide sequence fused 3' to a direct repeat sequence. The invention further comprehends the Cas13 CRISPR protein being codon optimized for expression in a eEukaryotic cell. In a preferred embodiment the eEukaryotic cell is a mammalian cell, a plant cell or a yeast cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of the gene product is decreased. In some embodiments the CRISPR protein is Cas13. In some embodiments the CRISPR protein is Cas12a. In some embodiments,, the Cas13 or Cas12a enzyme protein is Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium or Francisella

Novicida Cas13 or Cas12a, and may include mutated Cas13 or Cas12a derived from these organisms. The enzyme protein may be a further Cas13 or Cas12a homolog or ortholog. In some embodiments, the nucleotide sequence encoding the Cfp1 Csa13 or Cas12a enzyme protein is codon-optimized for expression in a eukaryotic cell. In some embodiments, the Cas13 or Cas12a enzyme protein directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter.

In general, and throughout this specification, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. 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) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0562] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) 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).

[0563] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[0564] In one aspect, the invention provides a eukaryotic host cell comprising (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences downstream of the direct repeat sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a CRISPR enzyme complexed with the guide RNA comprising the guide sequence that is hybridized to the target sequence and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas13 enzyme comprising a nuclear localization sequence. In some embodiments, the host cell comprises components (a) and (b). In some embodiments, component (a), component (b), or components (a) and (b) are stably integrated into a genome of the host eukaryotic cell. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas13 enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the Cas13 enzyme lacks RNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter.

[0565] In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example a mammal. Also, the organism may be an arthropod such as an insect. The organism also may be a plant or a yeast. Further, the organism may be a fungus.

[0566] In one aspect, the invention provides a kit comprising one or more of the components described herein above. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences downstream of the direct repeat sequence,

wherein when expressed, the guide sequence directs sequence-specific binding of a Cas13 CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a Cas13 enzyme complexed with the protected guide RNA comprising the guide sequence that is hybridized to the target sequence and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas13 enzyme comprising a nuclear localization sequence. In some embodiments, the kit comprises components (a) and (b) located on the same or different vectors of the system. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas13 enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said Cas13 enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the Cas13 enzyme is Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020 or Francisella tularensis 1 Novicida Cas13, and may include mutated Cas13 derived from these organisms. The enzyme may be a Cas13 homolog or ortholog. In some embodiments, the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR enzyme lacks DNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter.

[0567] In one aspect, the invention provides a method of modifying a target polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a Cas13 enzyme complexed with protected guide RNA comprising a guide sequence hybridized to a target sequence within said target polynucleotide. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said Cas13 enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by non-homologous end joining (NHEJ)-based gene insertion mechanisms, more particularly with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid

changes in a protein expressed from a gene comprising the target sequence. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cell, wherein the one or more vectors drive expression of one or more of: the Cas13 enzyme, the protected guide RNA comprising the guide sequence linked to direct repeat sequence. In some embodiments, said vectors are delivered to the eukaryotic cell in a subject. In some embodiments, said modifying takes place in said eukaryotic cell in a cell culture. In some embodiments, the method further comprises isolating said eukaryotic cell from a subject prior to said modifying. In some embodiments, the method further comprises returning said eukaryotic cell and/or cells derived therefrom to said subject.

[0568] In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a Cas13 CRISPR complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein the CRISPR complex comprises a Cas13 enzyme complexed with a protected guide RNA comprising a guide sequence hybridized to a target sequence within said polynucleotide. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cells, wherein the one or more vectors drive expression of one or more of: the Cas13 enzyme and the protected guide RNA.

[0569] In one aspect, the invention provides a method of generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) introducing one or more vectors into a eukaryotic cell, wherein the one or more vectors drive expression of one or more of: a Cas13 enzyme and a protected guide RNA comprising a guide sequence linked to a direct repeat sequence; and (b) allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said disease gene, wherein the CRISPR complex comprises the Cas13 enzyme complexed with the guide RNA comprising the sequence that is hybridized to the target sequence within the target polynucleotide, thereby generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said Cas13 enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by non-homologous end joining (NHEJ)-based gene insertion mechanisms with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion,

deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid changes in a protein expression from a gene comprising the target sequence.

[0570] In one aspect, the invention provides a method for developing a biologically active agent that modulates a cell signaling event associated with a disease gene. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) contacting a test compound with a model cell of any one of the described embodiments; and (b) detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with said mutation in said disease gene, thereby developing said biologically active agent that modulates said cell signaling event associated with said disease gene.

[0571] In one aspect, the invention provides a recombinant polynucleotide comprising a protected guide sequence downstream of a direct repeat sequence, wherein the protected guide sequence when expressed directs sequence-specific binding of a CRISPR complex to a corresponding target sequence present in a eukaryotic cell. In some embodiments, the target sequence is a viral sequence present in a eukaryotic cell. In some embodiments, the target sequence is a proto-oncogene or an oncogene.

[0572] In one aspect the invention provides for a method of selecting one or more cell(s) by introducing one or more mutations in a gene in the one or more cell (s), the method comprising: introducing one or more vectors into the cell (s), wherein the one or more vectors drive expression of one or more of: a Cas13 enzyme, a protected guide RNA comprising a guide sequence, and an editing template; wherein the editing template comprises the one or more mutations that abolish Cas13 enzyme cleavage; allowing non-homologous end joining (NHEJ)-based gene insertion mechanisms of the editing template with the target polynucleotide in the cell(s) to be selected; allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said gene, wherein the CRISPR complex comprises the Cas13 enzyme complexed with the protected guide RNA comprising a guide sequence that is hybridized to the target sequence within the target polynucleotide, wherein binding of the CRISPR complex to the target polynucleotide induces cell death, thereby allowing one or more cell(s) in which one or more mutations have been introduced to be selected. In a preferred embodiment of the invention the cell to be selected may be a eukaryotic cell. Aspects of the invention allow for selection of specific cells without requiring a selection marker or a two-step process that may include a counter-selection system.

[0573] With respect to mutations of the Cas13 enzyme, when the enzyme is not FnCas13, mutations may be as described herein elsewhere; conservative substitution for any of the replacement amino acids is also envisaged. In an aspect the invention provides as to any or each or all embodiments herein-discussed wherein the CRISPR enzyme comprises at least one or more, or at least two or more mutations, wherein the at least one or more mutation or the at least two or more mutations are selected from those described herein elsewhere.

[0574] In a further aspect, the invention involves a computer-assisted method for identifying or designing potential compounds to fit within or bind to CRISPR-Cas13 system or a functional portion thereof or vice versa (a computer-assisted method for identifying or designing potential CRISPR-Cas13 systems or a functional portion thereof for binding to desired compounds) or a computer-assisted method for identifying or designing potential CRISPR-Cas13 systems (e.g., with regard to predicting areas of the CRISPR-Cas13 system to be able to be manipulated-for instance, based on crystal structure data or based on data of Cas13 orthologs, or with respect to where a functional group such as an activator or repressor can be attached to the CRISPR-Cas13 system, or as to Cas13 truncations or as to designing nickases), said method comprising:

using a computer system, e.g., a programmed computer comprising a processor, a data storage system, an input device, and an output device, the steps of:

(a) inputting into the programmed computer through said input device data comprising the three-dimensional co-ordinates of a subset of the atoms from or pertaining to the CRISPR-Cas13 crystal structure, e.g., in the CRISPR-Cas13 system binding domain or alternatively or additionally in domains that vary based on variance among Cas13 orthologs or as to Cas13s or as to nickases or as to functional groups, optionally with structural information from CRISPR-Cas13 system complex(es), thereby generating a data set;

(b) comparing, using said processor, said data set to a computer database of structures stored in said computer data storage system, e.g., structures of compounds that bind or putatively bind or that are desired to bind to a CRISPR-Cas13 system or as to Cas13 orthologs (e.g., as Cas13s or as to domains or regions that vary amongst Cas13 orthologs) or as to the CRISPR-Cas13 crystal structure or as to nickases or as to functional groups;

(c) selecting from said database, using computer methods, structure(s)-e.g., CRISPR-Cas13 structures that may bind to desired structures, desired structures that may bind to certain CRISPR-Cas13 structures, portions of the CRISPR-Cas13 system that may be manipulated, e.g., based on data from other portions of the CRISPR-Cas13 crystal structure and/or from

Cas13 orthologs, truncated Cas13s, novel nickases or particular functional groups, or positions for attaching functional groups or functional-group-CRISPR-Cas13 systems;

(d) constructing, using computer methods, a model of the selected structure(s); and

(e) outputting to said output device the selected structure(s);

and optionally synthesizing one or more of the selected structure(s);

and further optionally testing said synthesized selected structure(s) as or in a CRISPR-Cas13 system;

or, said method comprising: providing the co-ordinates of at least two atoms of the CRISPR-Cas13 crystal structure, e.g., at least two atoms of the herein Crystal Structure Table of the CRISPR-Cas13 crystal structure or co-ordinates of at least a sub-domain of the CRISPR-Cas13 crystal structure ("selected co-ordinates"), providing the structure of a candidate comprising a binding molecule or of portions of the CRISPR-Cas13 system that may be manipulated, e.g., based on data from other portions of the CRISPR-Cas13 crystal structure and/or from Cas13 orthologs, or the structure of functional groups, and fitting the structure of the candidate to the selected co-ordinates, to thereby obtain product data comprising CRISPR-Cas13 structures that may bind to desired structures, desired structures that may bind to certain CRISPR-Cas13 structures, portions of the CRISPR-Cas13 system that may be manipulated, truncated Cas13s, novel nickases, or particular functional groups, or positions for attaching functional groups or functional-group-CRISPR-Cas13 systems, with output thereof; and optionally synthesizing compound(s) from said product data and further optionally comprising testing said synthesized compound(s) as or in a CRISPR-Cas13 system.

[0575] The testing can comprise analyzing the CRISPR-Cas13 system resulting from said synthesized selected structure(s), e.g., with respect to binding, or performing a desired function.

[0576] The output in the foregoing methods can comprise data transmission, e.g., transmission of information via telecommunication, telephone, video conference, mass communication, e.g., presentation such as a computer presentation (e.g. POWERPOINT), internet, email, documentary communication such as a computer program (e.g. WORD) document and the like. Accordingly, the invention also comprehends computer readable media containing: atomic co-ordinate data according to the herein-referenced Crystal Structure, said data defining the three dimensional structure of CRISPR-Cas13 or at least one sub-domain thereof, or structure factor data for CRISPR-Cas13, said structure factor data being derivable from the atomic co-ordinate data of herein-referenced Crystal Structure. The computer readable media can also contain any data of the foregoing methods. The invention further comprehends methods a computer system for generating or performing rational design as in

the foregoing methods containing either: atomic co-ordinate data according to herein-referenced Crystal Structure, said data defining the three dimensional structure of CRISPR-Cas13 or at least one sub-domain thereof, or structure factor data for CRISPR-Cas13, said structure factor data being derivable from the atomic co-ordinate data of herein-referenced Crystal Structure. The invention further comprehends a method of doing business comprising providing to a user the computer system or the media or the three dimensional structure of CRISPR-Cas13 or at least one sub-domain thereof, or structure factor data for CRISPR-Cas13, said structure set forth in and said structure factor data being derivable from the atomic co-ordinate data of herein-referenced Crystal Structure, or the herein computer media or a herein data transmission.

[0577] A "binding site" or an "active site" comprises or consists essentially of or consists of a site (such as an atom, a functional group of an amino acid residue or a plurality of such atoms and/or groups) in a binding cavity or region, which may bind to a compound such as a nucleic acid molecule, which is/are involved in binding.

[0578] By "fitting", is meant determining by automatic, or semi-automatic means, interactions between one or more atoms of a candidate molecule and at least one atom of a structure of the invention, and calculating the extent to which such interactions are stable. Interactions include attraction and repulsion, brought about by charge, steric considerations and the like. Various computer-based methods for fitting are described further

[0579] By "root mean square (or rms) deviation", we mean the square root of the arithmetic mean of the squares of the deviations from the mean.

[0580] By a "computer system", is meant the hardware means, software means and data storage means used to analyze atomic coordinate data. The minimum hardware means of the computer-based systems of the present invention typically comprises a central processing unit (CPU), input means, output means and data storage means. Desirably a display or monitor is provided to visualize structure data. The data storage means may be RAM or means for accessing computer readable media of the invention. Examples of such systems are computer and tablet devices running Unix, Windows or Apple operating systems.

[0581] By "computer readable media", is meant any medium or media, which can be read and accessed directly or indirectly by a computer e.g., so that the media is suitable for use in the above-mentioned computer system. Such media include, but are not limited to: magnetic storage media such as floppy discs, hard disc storage medium and magnetic tape; optical storage media such as optical discs or CD-ROM; electrical storage media such as RAM and

ROM; thumb drive devices; cloud storage devices and hybrids of these categories such as magnetic/optical storage media.

[0582] The invention comprehends the use of the protected guides described herein above in the optimized functional CRISPR-Cas enzyme systems described herein.

[0583] In some embodiments, the guide RNA is a toehold based guide RNA. The toehold based guide RNAs allows for guide RNAs only becoming activated based on the RNA levels of other transcripts in a cell. In certain embodiments, the guide RNA has an extension that includes a loop and a complementary sequence that fold over onto the guide and block the guide. The loop can be complementary to transcripts or miRNA in the cell and bind these transcripts if present. This will unfold the guide RNA allowing it to bind a Cas13 molecule. This bound complex can then knockdown transcripts or edit transcripts depending on the application.

CRISPR-Cas Enzyme

[0584] In its unmodified form, a CRISPR-Cas protein is a catalytically active protein. This implies that upon formation of a nucleic acid-targeting complex (comprising a guide RNA hybridized to a target sequence one or both DNA strands in or near (e.g., within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence is modified (e.g. cleaved). As used herein the term "sequence(s) associated with a target locus of interest" refers to sequences near the vicinity of the target sequence (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from the target sequence, wherein the target sequence is comprised within a target locus of interest). The unmodified catalytically active Cas13 protein generates a staggered cut, whereby the cut sites are typically within the target sequence. More particularly, the staggered cut is typically 13–23 nucleotides distal to the PAM. In particular embodiments, the cut on the non-target strand is 17 nucleotides downstream of the PAM (i.e. between nucleotide 17 and 18 downstream of the PAM), while the cut on the target strand (i.e. strand hybridizing with the guide sequence) occurs a further 4 nucleotides further from the sequence complementary to the PAM (this is 21 nucleotides upstream of the complement of the PAM on the 3' strand or between nucleotide 21 and 22 upstream of the complement of the PAM).

[0585] In the methods according to the present invention, the CRISPR-Cas protein is preferably mutated with respect to a corresponding wild-type enzyme such that the mutated CRISPR-Cas protein lacks the ability to cleave one or both DNA strands of a target locus containing a target sequence. In particular embodiments, one or more catalytic domains of the Cas13 protein are mutated to produce a mutated Cas protein which cleaves only one DNA strand of a target sequence.

[0586] In particular embodiments, the CRISPR-Cas protein may be mutated with respect to a corresponding wild-type enzyme such that the mutated CRISPR-Cas protein lacks substantially all DNA cleavage activity. In some embodiments, a CRISPR-Cas protein may be considered to substantially lack all DNA and/or RNA cleavage activity when the cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the nucleic acid cleavage activity of the non-mutated form of the enzyme; an example can be when the nucleic acid cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form.

[0587] In certain embodiments of the methods provided herein the CRISPR-Cas protein is a mutated CRISPR-Cas protein which cleaves only one DNA strand, i.e. a nickase. More particularly, in the context of the present invention, the nickase ensures cleavage within the non-target sequence, i.e. the sequence which is on the opposite DNA strand of the target sequence and which is 3' of the PAM sequence. By means of further guidance, and without limitation, an arginine-to-alanine substitution (R1226A) in the Nuc domain of Cas13 from Acidaminococcus sp. converts Cas13 from a nuclease that cleaves both strands to a nickase (cleaves a single strand). It will be understood by the skilled person that where the enzyme is not AsCas13, a mutation may be made at a residue in a corresponding position. In particular embodiments, the Cas13 is FnCas13 and the mutation is at the arginine at position R1218. In particular embodiments, the Cas13 is LbCas13 and the mutation is at the arginine at position R1138. In particular embodiments, the Cas13 is MbCas13 and the mutation is at the arginine at position R1293.

[0588] In certain embodiments of the methods provided herein the CRISPR-Cas protein has reduced or no catalytic activity. Where the CRISPR-Cas protein is a Cas13 protein, the mutations may include but are not limited to one or more mutations in the catalytic RuvC-like domain, such as D908A or E993A with reference to the positions in AsCas13.

[0589] In some embodiments, a CRISPR-Cas protein is considered to substantially lack all DNA cleavage activity when the DNA cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the DNA cleavage activity of the non-mutated form of the enzyme; an example can be when the DNA cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form. In these embodiments, the CRISPR-Cas protein is used as a generic DNA binding protein. The mutations may be artificially introduced mutations or gain- or loss-of-function mutations.

[0590] In addition to the mutations described above, the CRISPR-Cas protein may be additionally modified. As used herein, the term "modified" with regard to a CRISPR-Cas

protein generally refers to a CRISPR-Cas protein having one or more modifications or mutations (including point mutations, truncations, insertions, deletions, chimeras, fusion proteins, etc.) compared to the wild type Cas protein from which it is derived. By derived is meant that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as known in the art or as described herein.

[0591] In some embodiments, to reduce the size of a fusion protein of the Cas13b effector and the one or more functional domains, the C-terminus of the Cas13b effector can be truncated while still maintaining its RNA binding function. For example, at least 20 amino acids, at least 50 amino acids, at least 80 amino acids, or at least 100 amino acids, or at least 150 amino acids, or at least 200 amino acids, or at least 250 amino acids, or at least 300 amino acids, or at least 350 amino acids, or up to 120 amino acids, or up to 140 amino acids, or up to 160 amino acids, or up to 180 amino acids, or up to 200 amino acids, or up to 250 amino acids, or up to 300 amino acids, or up to 350 amino acids, or up to 400 amino acids, may be truncated at the C-terminus of the Cas13b effector. Specific examples of Cas13b truncations include C-terminal Δ984-1090, C-terminal Δ1026-1090, and C-terminal Δ1053-1090, C-terminal Δ934-1090, C-terminal Δ884-1090, C-terminal Δ834-1090, C-terminal Δ784-1090, and C-terminal Δ734-1090, wherein amino acid positions correspond to amino acid positions of Prevotella sp. P5-125 Cas13b protein. See also FIG. 67.

[0592] The additional modifications of the CRISPR-Cas protein may or may not cause an altered functionality. By means of example, and in particular with reference to CRISPR-Cas protein, modifications which do not result in an altered functionality include for instance codon optimization for expression into a particular host, or providing the nuclease with a particular marker (e.g. for visualization). Modifications with may result in altered functionality may also include mutations, including point mutations, insertions, deletions, truncations (including split nucleases), etc.. Fusion proteins may without limitation include for instance fusions with heterologous domains or functional domains (e.g. localization signals, catalytic domains, etc.). In certain embodiments, various different modifications may be combined (e.g. a mutated nuclease which is catalytically inactive and which further is fused to a functional domain, such as for instance to induce DNA methylation or another nucleic acid modification, such as including without limitation a break (e.g. by a different nuclease (domain)), a mutation, a deletion, an insertion, a replacement, a ligation, a digestion, a break or a recombination). As used herein, “altered functionality” includes without limitation an altered specificity (e.g. altered target recognition, increased (e.g. “enhanced” Cas proteins) or decreased specificity, or

altered PAM recognition), altered activity (e.g. increased or decreased catalytic activity, including catalytically inactive nucleases or nickases), and/or altered stability (e.g. fusions with destalilization domains). Suitable heterologous domains include without limitation a nuclease, a ligase, a repair protein, a methyltransferase, (viral) integrase, a recombinase, a transposase, an argonaute, a cytidine deaminase, a retron, a group II intron, a phosphatase, a phosphorylase, a sulpfurylase, a kinase, a polymerase, an exonuclease, etc.. Examples of all these modifications are known in the art. It will be understood that a "modified" nuclease as referred to herein, and in particular a "modified" Cas or "modified" CRISPR-Cas system or complex preferably still has the capacity to interact with or bind to the polynucleic acid (e.g. in complex with theguide molecule). Such modified Cas protein can be combined with the deaminase protein or active domain thereof as described herein.

[0593] In certain embodiments, CRISPR-Cas protein may comprise one or more modifications resulting in enhanced activity and/or specificity, such as including mutating residues that stabilize the targeted or non-targeted strand (e.g. eCas9; "Rationally engineered Cas9 nucleases with improved specificity", Slaymaker et al. (2016), Science, 351(6268):84-88, incorporated herewith in its entirety by reference). In certain embodiments, the altered or modified activity of the engineered CRISPR protein comprises increased targeting efficiency or decreased off-target binding. In certain embodiments, the altered activity of the engineered CRISPR protein comprises modified cleavage activity. In certain embodiments, the altered activity comprises increased cleavage activity as to the target polynucleotide loci. In certain embodiments, the altered activity comprises decreased cleavage activity as to the target polynucleotide loci. In certain embodiments, the altered activity comprises decreased cleavage activity as to off-target polynucleotide loci. In certain embodiments, the altered or modified activity of the modified nuclease comprises altered helicase kinetics. In certain embodiments, the modified nuclease comprises a modification that alters association of the protein with the nucleic acid molecule comprising RNA (in the case of a Cas protein), or a strand of the target polynucleotide loci, or a strand of off-target polynucleotide loci. In an aspect of the invention, the engineered CRISPR protein comprises a modification that alters formation of the CRISPR complex. In certain embodiments, the altered activity comprises increased cleavage activity as to off-target polynucleotide loci. Accordingly, in certain embodiments, there is increased specificity for target polynucleotide loci as compared to off-target polynucleotide loci. In other embodiments, there is reduced specificity for target polynucleotide loci as compared to off-target polynucleotide loci. In certain embodiments, the mutations result in decreased off-target effects (e.g. cleavage or binding properties, activity, or kinetics), such as in case for Cas

proteins for instance resulting in a lower tolerance for mismatches between target and guide RNA. Other mutations may lead to increased off-target effects (e.g. cleavage or binding properties, activity, or kinetics). Other mutations may lead to increased or decreased on-target effects (e.g. cleavage or binding properties, activity, or kinetics). In certain embodiments, the mutations result in altered (e.g. increased or decreased) helicase activity, association or formation of the functional nuclease complex (e.g. CRISPR-Cas complex). In certain embodiments, as described above, the mutations result in an altered PAM recognition, i.e. a different PAM may be (in addition or in the alternative) be recognized, compared to the unmodified Cas protein. Particularly preferred mutations include positively charged residues and/or (evolutionary) conserved residues, such as conserved positively charged residues, in order to enhance specificity. In certain embodiments, such residues may be mutated to uncharged residues, such as alanine.

[0594] In certain embodiments, the methods, products, and uses as described herein can be expanded or adapted to implement any type of CRISPR effector.

[0595] In certain embodiments, the CRISPR effector is a class 2 CRISPR-Cas system effector. It is to be understood that the term "CRISPR effector" preferably refers to an RNA-guided endonuclease. The skilled person will understand that the CRISPR effector may be modified, as described herein elsewhere, and as known in the art. By means of example, and without limitation, CRISPR effector modifications include modifications affecting CRISPR effector functionality or nuclease activity (e.g. catalytically inactive variants (optionally fused or otherwise associated with heterologous functional domains), nickases, altered PAM specificity/recognition, split CRISPR effectors,…), specificity (e.g. enhanced specificity mutants), stability (e.g. destabilized variants), etc.

[0596] In certain embodiments, the CRISPR effector cleaves, binds to, or associates with RNA. In certain embodiments, the CRISPR effector cleaves, binds to, or associates with DNA. In certain embodiments, the CRISPR effector cleaves, binds to, or associates with single stranded RNA. In certain embodiments, the CRISPR effector cleaves, binds to, or associates with single stranded DNA. In certain embodiments, the CRISPR effector cleaves, binds to, or associates with double stranded RNA. In certain embodiments, the CRISPR effector cleaves, binds to, or associates with Double stranded DNA. In certain embodiments, the CRISPR effector cleaves, binds to, or associates with DNA/RNA hybrids.

[0597] In certain embodiments, the CRISPR effector is a class 2, type II CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type II-A CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type II-B CRISPR effector. In certain

embodiments, the CRISPR effector is a class 2, type II-C CRISPR effector. In certain embodiments, the CRISPR effector is Cas9.

[0598] In certain embodiments, the CRISPR effector is a class 2, type V CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type V-A CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type V-B CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type V-C CRISPR effector. In certain embodiments, the CRISPR effector is Cas12a (Cpf1). In certain embodiments, the CRISPR effector is Cas12b (C2c1). In certain embodiments, the CRISPR effector is Cas12c (C2c3). In certain embodiments, the CRISPR effector is a class 2, type V-U CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type V-U1 CRISPR effector (e.g. C2c4). In certain embodiments, the CRISPR effector is a class 2, type V-U2 CRISPR effector (e.g. C2c8). In certain embodiments, the CRISPR effector is a class 2, type V-U3 CRISPR effector (e.g. C2c10). In certain embodiments, the CRISPR effector is a class 2, type V-U4 CRISPR effector (e.g. C2c9). In certain embodiments, the CRISPR effector is a class 2, type V-U5 CRISPR effector (e.g. C2c5).

[0599] In certain embodiments, the CRISPR effector is a class 2, type VI CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type VI-A CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type VI-B CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type VI-B1 CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type VI-B2 CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type VI-C CRISPR effector. In certain embodiments, the CRISPR effector is Cas13a (C2c2). In certain embodiments, the CRISPR effector is Cas13b (C2c6). In certain embodiments, the CRISPR effector is Cas13c (C2c7).

[0600] In certain embodiments, the CRISPR effector comprises one or more RuvC domain. In certain embodiments, the CRISPR effector comprises a RuvC-I domain. In certain embodiments, the CRISPR effector comprises a RuvC-II domain. In certain embodiments, the CRISPR effector comprises a RuvC-III domain. In certain embodiments, the CRISPR effector comprises a RuvC-I, RuvC-II, and RuvC-III domain. In certain embodiments, one or more of RuvC-I, II, and/or III are contiguous motifs. In certain embodiments, one or more of RuvC-I, II, and/or III are non-contiguous or discrete motifs. In certain embodiments, the CRISPR effector comprises one or more HNH domain. In certain embodiments, the CRISPR effector comprises one or more RuvC domain and one or more HNH domain. In certain embodiments, the CRISPR effector comprises a RuvC-I domain and an HNH domain. In certain embodiments, the CRISPR effector comprises a RuvC-II domain and an HNH domain. In

certain embodiments, the CRISPR effector comprises a RuvC-III domain and an HNH domain. In certain embodiments, the CRISPR effector comprises a RuvC-I, RuvC-II, and RuvC-III domain and an HNH domain. In certain embodiments, the CRISPR effector comprises one or more Nuc (nuclease) domain. In certain embodiments, the CRISPR effector comprises one or more RuvC domain and one or more Nuc domain. In certain embodiments, the CRISPR effector comprises a RuvC-I domain and a Nuc domain. In certain embodiments, the CRISPR effector comprises a RuvC-II domain and a Nuc domain. In certain embodiments, the CRISPR effector comprises a RuvC-III domain and a Nuc domain.

[0601] In certain embodiments, the CRISPR effector comprises one or more HEPN domain. In certaim embodiments, the CRISPR effector comprises a HEPN I domain. In certain embodiments, the CRISPR effector comprises a HEPN II domain. In certain embodiments, the CRISPR effector comprises a HEPN I domain and a HEPN II domain. In certain embodiments, one or more of the HEPN domains are contiguous domains. In certain embodiments, one or more of the HEPN domains comprise non-contiguous or discrete motifs.

[0602] In certain embodiments, the CRISPR effector is a CRISPR effector as disclosed for instance in Shmakov et al. (2017), "Diversity and evolution of class 2 CRISPR-Cas systems", Nature Rev Microbiol, 15(3):169-182; Shmakov et al. (2015) "Discovery and functional characterization of diverse class 2 CRISPR-Cas systems", Mol Cell, 60(3):385-397; Makarova et al. (2015), "An updated evolutionary classification of CRISPR-Cas systems", Nat Rev Microbiol, 13(11):722-736; or Koonin et al. (2017), "Diversity, classification and evolution of CRISPR-Cas systems", Curr Opin Microbiol, 37:67-78. All are incorporated herein by reference in their entirety, as well as the references cited therein.

[0603] The skilled person will understand that the choice of CRISPR effector may depend on the application (e.g. knockout or suppression, activation,…) , as well as the target (e.g. RNA or DNA, single or double stranded, as well as target sequence, including associated PAM sequence and/or specificity,…). It will be understood, that the choice of CRISPR effector may determine the particulars of other CRISPR-Cas system components (e.g. spacer (or guide sequence) length, direct repeat (or tracr mate) sequence or length, the presence or absence of a tracr, as well as tracr sequence or length, etc.).

[0604] CRISPR-Cas systems have been identified in numerous archaeal and bacterial species. The skilled person will understand that CRISPR effector homologues or orthologues from any of the identified CRISPR-Cas systems may advantageously be used in certain embodiments. It will be understood that further homologues (e.g. additional class 2 types of CRISPR-Cas systems and CRISPR effectors) or orthologues (e.g. known or unknown

CRISPR-Cas systems or CRISPR effectors from additional archaeal or bacterial species) can be identified. Such may suitably be used in certain embodiments and aspects of the invention.

[0605] By means of example, CRISPR-Cas systems (and CRISPR effectors) may be identified for instance and without limitation as described in Shmakov et al. (2017), "Diversity and evolution of class 2 CRISPR-Cas systems", Nature Rev Microbiol, 15(3):169-182 or Shmakov et al. (2015) "Discovery and functional characterization of diverse class 2 CRISPR-Cas systems", Mol Cell, 60(3):385-397. The methodology for identifying CRISPR-Cas systems and effectors is explicitly incorporated herein by reference.

[0606] In certain embodiments, a method for the systematic detection of class 2 CRISPR-Cas systems may begin with the identification of a 'seed' that signifies the likely presence of a CRISPR-Cas locus in a given nucleotide sequence. For instance, Cas1 may be used as the seed, as it is the most common Cas protein in CRISPR-Cas systems and is most highly conserved at the sequence level. Sequence databases may be searched with this seed. To ensure the maximum sensitivity of detection, the search may be carried out by comparing a Cas1 sequence profile to translated genomic and metagenomic sequences. After the Cas1 genes are detected, their respective 'neighbourhoods' are examined for the presence of other Cas genes by searching with previously developed profiles for Cas proteins and applying the criteria for the classification of the CRISPR-Cas loci. In a complementary approach, to extend the search to non-autonomous CRISPR-Cas systems, the same procedure may be repeated using the CRISPR array as the seed. To ensure that the CRISPR array is detected at a high level of sensitivity, the predictions can be made for instance using the Piler-CR72 and CRISPRfinder methods, which predictions can be pooled and taken as the final CRISPR set. As illustrated in Shmakov et al. (2017), "Diversity and evolution of class 2 CRISPR-Cas systems", Nature Rev Microbiol, 15(3):169-182, this latter procedure (i.e. using the CRISPR array as seed) yielded 47,174 CRISPR arrays, which is more than twice the number of Cas1 genes that were detected, reflecting the fact that many CRISPR-Cas loci lack the adaptation module and that numerous 'orphan' arrays, some of which seem to be functional, also exist.

[0607] All loci can either subsequently be assigned to known CRISPR-Cas subtypes through the Cas protein profile search or alternatively can be assigned to new subtypes. In certain embodiments, among the Cas1 or CRISPR neighborhoods, those that encode large proteins (>500 amino acids) can be analyzed in detail, given that Cas9 and Cpf1 are large proteins (typically >1000 amino acids) and that their protein structures suggest that this large size is required to accommodate the CRISPR RNA (crRNA)-target DNA complex. The sequences of such large proteins can then be screened for known protein domains using

sensitive profile-based methods, such as HHpred, secondary structure prediction and manual examination of multiple alignments. Under the premise that class 2 effector proteins contain nuclease domains, even if they are distantly related or unrelated to known families of nucleases, the proteins that contain domains that are deemed irrelevant in the context of the CRISPR-Cas function (for example, membrane transporters or metabolic enzymes) can be discarded. The retained proteins either contain readily identifiable, or completely unknown, nuclease domains. The sequences of these proteins can then be analyzed using the most sensitive methods for domain detection, such as HHpred, with a curated multiple alignment of the respective protein sequences that can be used as the query. The use of sensitive methods is essential because proteins that are involved in antiviral defense, and the Cas proteins in particular, typically evolve extremely fast. The above procedure for the discovery of class 2 CRISPR-Cas systems, at least in principle, is expected to be exhaustive, because all loci that contain a gene that encodes a large protein (that is, a putative class 2 effector) in the vicinity of cas1 and/or CRISPR are analyzed in detail. The assumption of the structural requirements for a class 2 effector, which underlie the protein size cut-off that is used, and the precision of Cas1 and CRISPR detection, are the only limitations of this approach.

[0608] In certain embodiments, the CRISPR effector is a CRISPR effector as identified for instance according to the methodology presented above. It will be understood that functionality of the identified CRISPR effectors can be readily evaluated and validated by the skilled person.

Base Excision Repair Inhibitor

[0609] In some embodiments, the AD-functionalized CRISPR system further comprises a base excision repair (BER) inhibitor. Without wishing to be bound by any particular theory, cellular DNA-repair response to the presence of I:T pairing may be responsible for a decrease in nucleobase editing efficiency in cells. Alkyladenine DNA glycosylase (also known as DNA-3-methyladenine glycosylase, 3-alkyladenine DNA glycosylase, or N-methylpurine DNA glycosylase) catalyzes removal of hypoxanthine from DNA in cells, which may initiate base excision repair, with reversion of the I:T pair to a A:T pair as outcome.

[0610] In some embodiments, the BER inhibitor is an inhibitor of alkyladenine DNA glycosylase. In some embodiments, the BER inhibitor is an inhibitor of human alkyladenine DNA glycosylase. In some embodiments, the BER inhibitor is a polypeptide inhibitor. In some embodiments, the BER inhibitor is a protein that binds hypoxanthine. In some embodiments, the BER inhibitor is a protein that binds hypoxanthine in DNA. In some embodiments, the BER inhibitor is a catalytically inactive alkyladenine DNA glycosylase protein or binding domain

thereof. In some embodiments, the BER inhibitor is a catalytically inactive alkyladenine DNA glycosylase protein or binding domain thereof that does not excise hypoxanthine from the DNA. Other proteins that are capable of inhibiting (e.g., sterically blocking) an alkyladenine DNA glycosylase base-excision repair enzyme are within the scope of this disclosure. Additionally, any proteins that block or inhibit base-excision repair as also within the scope of this disclosure.

[0611] Without wishing to be bound by any particular theory, base excision repair may be inhibited by molecules that bind the edited strand, block the edited base, inhibit alkyladenine DNA glycosylase, inhibit base excision repair, protect the edited base, and/or promote fixing of the non-edited strand. It is believed that the use of the BER inhibitor described herein can increase the editing efficiency of an adenosine deaminase that is capable of catalyzing a A to I change.

[0612] Accordingly, in the first design of the AD-functionalized CRISPR system discussed above, the CRISPR-Cas protein or the adenosine deaminase can be fused to or linked to a BER inhibitor (e.g., an inhibitor of alkyladenine DNA glycosylase). In some embodiments, the BER inhibitor can be comprised in one of the following structures (nCas13=Cas13 nickase; dCas13=dead Cas13); [AD]-[optional linker]-[nCas13/dCas13]-[optional linker]-[BER inhibitor]; [AD]-[optional linker]-[BER inhibitor]-[optional linker]-[nCas13/dCas13]; [BER inhibitor]-[optional linker]-[AD]-[optional linker]-[nCas13/dCas13]; [BER inhibitor]-[optional linker]-[nCas13/dCas13]-[optional linker]-[AD]; [nCas13/dCas13]-[optional linker]-[AD]-[optional linker]-[BER inhibitor]; [nCas13/dCas13]-[optional linker]-[BER inhibitor]-[optional linker]-[AD].

[0613] Similarly, in the second design of the AD-functionalized CRISPR system discussed above, the CRISPR-Cas protein, the adenosine deaminase, or the adaptor protein can be fused to or linked to a BER inhibitor (e.g., an inhibitor of alkyladenine DNA glycosylase). In some embodiments, the BER inhibitor can be comprised in one of the following structures (nCas13=Cas13 nickase; dCas13=dead Cas13): [nCas13/dCas13]-[optional linker]-[BER inhibitor]; [BER inhibitor]-[optional linker]-[nCas13/dCas13]; [AD]-[optional linker]-[Adaptor]-[optional linker]-[BER inhibitor]; [AD]-[optional linker]-[BER inhibitor]-[optional linker]-[Adaptor]; [BER inhibitor]-[optional linker]-[AD]-[optional linker]-[Adaptor]; [BER inhibitor]-[optional linker]-[Adaptor]-[optional linker]-[AD];

[Adaptor]-[optional linker]-[AD]-[optional linker]-[BER inhibitor]; [Adaptor]-[optional linker]-[BER inhibitor]-[optional linker]-[AD].

[0614] In the third design of the AD-functionalized CRISPR system discussed above, the BER inhibitor can be inserted into an internal loop or unstructured region of a CRISPR-Cas protein.

Targeting to the Nucleus

[0615] In some embodiments, the methods of the present invention relate to modifying an Adenine in a target locus of interest, whereby the target locus is within a cell. In order to improve targeting of the CRISPR-Cas protein and/or the adenosine deaminase protein or catalytic domain thereof used in the methods of the present invention to the nucleus, it may be advantageous to provide one or both of these components with one or more nuclear localization sequences (NLSs).

[0616] In preferred embodiments, the NLSs used in the context of the present invention are heterologous to the proteins. Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV (SEQ ID No. 17) or PKKKRKVEAS (SEQ ID No. 18); the NLS from nucleoplasmin (e.g., the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK (SEQ ID No. 19)); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID No. 20) or RQRRNELKRSP (SEQ ID No. 21); the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID No. 22); the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID No. 23) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID No. 24) and PPKKARED (SEQ ID No. 25) of the myoma T protein; the sequence PQPKKKPL (SEQ ID No. 26) of human p53; the sequence SALIKKKKKMAP (SEQ ID No. 27) of mouse c-abl IV; the sequences DRLRR (SEQ ID No. 28) and PKQKKRK (SEQ ID No. 29) of the influenza virus NS1; the sequence RKLKKKIKKL (SEQ ID No. 30) of the Hepatitis virus delta antigen; the sequence REKKKFLKRR (SEQ ID No. 31) of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK (SEQ ID No. 32) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK (SEQ ID No. 33) of the steroid hormone receptors (human) glucocorticoid. In general, the one or more NLSs are of sufficient strength to drive accumulation of the DNA-targeting Cas protein in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the CRISPR-Cas protein, the particular NLS(s) used, or a combination of these factors. 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, such as in combination with a means for detecting the location of the nucleus (e.g., a stain specific for the nucleus such as DAPI). Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay for the effect of nucleic acid-targeting complex formation (e.g., assay for deaminase activity) at the target sequence, or assay for altered gene expression activity affected by DNA-targeting complex formation and/or DNA-targeting), as compared to a control not exposed to the CRISPR-Cas protein and deaminase protein, or exposed to a CRISPR-Cas and/or deaminase protein lacking the one or more NLSs.

[0617] The CRISPR-Cas and/or adenosine deaminase proteins may be provided with 1 or more, such as with, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more heterologous NLSs. In some embodiments, the proteins comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g., zero or at least one or more NLS at the amino-terminus and zero or at one or more NLS at the carboxy terminus). When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies. In some embodiments, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is 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- or C-terminus. In preferred embodiments of the CRISPR-Cas proteins, an NLS attached to the C-terminal of the protein.

[0618] In certain embodiments of the methods provided herein, the CRISPR-Cas protein and the deaminase protein are delivered to the cell or expressed within the cell as separate proteins. In these embodiments, each of the CRISPR-Cas and deaminase protein can be provided with one or more NLSs as described herein. In certain embodiments, the CRISPR-Cas and deaminase proteins are delivered to the cell or expressed with the cell as a fusion protein. In these embodiments one or both of the CRISPR-Cas and deaminase protein is provided with one or more NLSs. Where the adenosine deaminase is fused to an adaptor protein (such as MS2) as described above, the one or more NLS can be provided on the adaptor protein, provided that this does not interfere with aptamer binding. In particular embodiments, the one

or more NLS sequences may also function as linker sequences between the adenosine deaminase and the CRISPR-Cas protein.

[0619] In certain embodiments, guides of the invention comprise specific binding sites (e.g. aptamers) for adapter proteins, which may be linked to or fused to an adenosine deaminase or catalytic domain thereof. When such a guides forms a CRISPR complex (i.e. CRISPR-Cas protein binding to guide and target) the adapter proteins bind and, the adenosine deaminase or catalytic domain thereof associated with the adapter protein is positioned in a spatial orientation which is advantageous for the attributed function to be effective.

[0620] The skilled person will understand that modifications to the guide which allow for binding of the adapter + adenosine deaminase, but not proper positioning of the adapter + adenosine deaminase (e.g. due to steric hindrance within the three dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified guide may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein, preferably at either the tetra loop or stem loop 2, and most preferably at both the tetra loop and stem loop 2.

Use of orthogonal catalytically inactive CRISPR-Cas proteins

[0621] In particular embodiments, the Cas13 nickase is used in combination with an orthogonal catalytically inactive CRISPR-Cas protein to increase efficiency of said Cas13 nickase (as described in Chen et al. 2017, Nature Communications 8:14958; doi:10.1038/ncomms14958). More particularly, the orthogonal catalytically inactive CRISPR-Cas protein is characterized by a different PAM recognition site than the Cas13 nickase used in the AD-functionalized CRISPR system and the corresponding guide sequence is selected to bind to a target sequence proximal to that of the Cas13 nickase of the AD-functionalized CRISPR system. The orthogonal catalytically inactive CRISPR-Cas protein as used in the context of the present invention does not form part of the AD-functionalized CRISPR system but merely functions to increase the efficiency of said Cas13 nickase and is used in combination with a standard guide molecule as described in the art for said CRISPR-Cas protein. In particular embodiments, said orthogonal catalytically inactive CRISPR-Cas protein is a dead CRISPR-Cas protein, i.e. comprising one or more mutations which abolishes the nuclease activity of said CRISPR-Cas protein. In particular embodiments, the catalytically inactive orthogonal CRISPR-Cas protein is provided with two or more guide molecules which are capable of hybridizing to target sequences which are proximal to the target sequence of the Cas13 nickase. In particular embodiments, at least two guide molecules are used to target said catalytically inactive CRISPR-Cas protein, of which at least one guide molecule is capable of

hybridizing to a target sequence 5" of the target sequence of the Cas13 nickase and at least one guide molecule is capable of hybridizing to a target sequence 3' of the target sequence of the Cas13 nickase of the AD-functionalized CRISPR system, whereby said one or more target sequences may be on the same or the opposite DNA strand as the target sequence of the Cas13 nickase. In particular embodiments, the guide sequences for the one or more guide molecules of the orthogonal catalytically inactive CRISPR-Cas protein are selected such that the target sequences are proximal to that of the guide molecule for the targeting of the AD-functionalized CRISPR, i.e. for the targeting of the Cas13 nickase. In particular embodiments, the one or more target sequences of the orthogonal catalytically inactive CRISPR-Cas enzyme are each separated from the target sequence of the Cas13 nickase by more than 5 but less than 450 basepairs. Optimal distances between the target sequences of the guides for use with the orthogonal catalytically inactive CRISPR-Cas protein and the target sequence of the AD-functionalized CRISPR system can be determined by the skilled person. In particular embodiments, the orthogonal CRISPR-Cas protein is a Class II, type II CRISPR protein. In particular embodiments, the orthogonal CRISPR-Cas protein is a Class II, type V CRISPR protein. In particular embodiments, the catalytically inactive orthogonal CRISPR-Cas protein In particular embodiments, the catalytically inactive orthogonal CRISPR-Cas protein has been modified to alter its PAM specificity as described elsewhere herein. In particular embodiments, the Cas13 protein nickase is a nickase which, by itself has limited activity in human cells, but which, in combination with an inactive orthogonal CRISPR-Cas protein and one or more corresponding proximal guides ensures the required nickase activity.

CRISPR Development and Use

[0622] The present invention may be further illustrated and extended based on aspects of CRISPR-Cas development and use as set forth in the following articles and particularly as relates to delivery of a CRISPR protein complex and uses of an RNA guided endonuclease in cells and organisms:

- Multiplex genome engineering using CRISPR-Cas systems. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, F. Science Feb 15;339(6121):819-23 (2013);
- RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Jiang W., Bikard D., Cox D., Zhang F, Marraffini LA. Nat Biotechnol Mar;31(3):233-9 (2013);
- One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR-Cas-Mediated Genome Engineering. Wang H., Yang H., Shivalila CS., Dawlaty MM., Cheng AW., Zhang F., Jaenisch R. Cell May 9;153(4):910-8 (2013);

- Optical control of mammalian endogenous transcription and epigenetic states. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. Nature. Aug 22;500(7463):472-6. doi: 10.1038/Nature12466. Epub 2013 Aug 23 (2013);
- Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Ran, FA., Hsu, PD., Lin, CY., Gootenberg, JS., Konermann, S., Trevino, AE., Scott, DA., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. Cell Aug 28. pii: S0092-8674(13)01015-5 (2013-A);
- DNA targeting specificity of RNA-guided Cas9 nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. Nat Biotechnol doi:10.1038/nbt.2647 (2013);
- Genome engineering using the CRISPR-Cas9 system. Ran, FA., Hsu, PD., Wright, J., Agarwala, V., Scott, DA., Zhang, F. Nature Protocols Nov;8(11):2281-308 (2013-B);
- Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. Science Dec 12. (2013);
- Crystal structure of cas9 in complex with guide RNA and target DNA. Nishimasu, H., Ran, FA., Hsu, PD., Konermann, S., Shehata, SI., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. Cell Feb 27, 156(5):935-49 (2014);
- Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Wu X., Scott DA., Kriz AJ., Chiu AC., Hsu PD., Dadon DB., Cheng AW., Trevino AE., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp PA. Nat Biotechnol. Apr 20. doi: 10.1038/nbt.2889 (2014);
- CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, Graham DB, Jhunjhunwala S, Heidenreich M, Xavier RJ, Langer R, Anderson DG, Hacohen N, Regev A, Feng G, Sharp PA, Zhang F. Cell 159(2): 440-455 DOI: 10.1016/j.cell.2014.09.014(2014);
- Development and Applications of CRISPR-Cas9 for Genome Engineering, Hsu PD, Lander ES, Zhang F., Cell. Jun 5;157(6):1262-78 (2014).
- Genetic screens in human cells using the CRISPR-Cas9 system, Wang T, Wei JJ, Sabatini DM, Lander ES., Science. January 3; 343(6166): 80–84. doi:10.1126/science.1246981 (2014);

- Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation, Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE., (published online 3 September 2014) Nat Biotechnol. Dec;32(12):1262-7 (2014);
- *In vivo* interrogation of gene function in the mammalian brain using CRISPR-Cas9, Swiech L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, Sur M, Zhang F., (published online 19 October 2014) Nat Biotechnol. Jan;33(1):102-6 (2015);
- Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F., Nature. Jan 29;517(7536):583-8 (2015).
- A split-Cas9 architecture for inducible genome editing and transcription modulation, Zetsche B, Volz SE, Zhang F., (published online 02 February 2015) Nat Biotechnol. Feb;33(2):139-42 (2015);
- Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis, Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R, Lee H, Zhang F, Sharp PA. Cell 160, 1246–1260, March 12, 2015 (multiplex screen in mouse), and
- In *vivo* genome editing using *Staphylococcus aureus* Cas9, Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F., (published online 01 April 2015), Nature. Apr 9;520(7546):186-91 (2015).
- Shalem et al., "High-throughput functional genomics using CRISPR-Cas9," Nature Reviews Genetics 16, 299-311 (May 2015).
- Xu et al., "Sequence determinants of improved CRISPR sgRNA design," Genome Research 25, 1147-1157 (August 2015).
- Parnas et al., "A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks," Cell 162, 675-686 (July 30, 2015).
- Ramanan et al., CRISPR-Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus," Scientific Reports 5:10833. doi: 10.1038/srep10833 (June 2, 2015)
- Nishimasu et al., Crystal Structure of Staphylococcus aureus Cas9," Cell 162, 1113-1126 (Aug. 27, 2015)

- BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis, Canver et al., Nature 527(7577):192-7 (Nov. 12, 2015) doi: 10.1038/nature15521. Epub 2015 Sep 16.
- *Cas13 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System*, Zetsche et al., Cell 163, 759-71 (Sep 25, 2015).
- *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems*, Shmakov et al., Molecular Cell, 60(3), 385–397 doi: 10.1016/j.molcel.2015.10.008 Epub October 22, 2015.
- *Rationally engineered Cas9 nucleases with improved specificity*, Slaymaker et al., Science 2016 Jan 1 351(6268): 84-88 doi: 10.1126/science.aad5227. Epub 2015 Dec 1.
- Gao *et al*, "Engineered Cas13 Enzymes with Altered PAM Specificities," bioRxiv 091611; doi: http://dx.doi.org/10.1101/091611 (Dec. 4, 2016).

each of which is incorporated herein by reference, may be considered in the practice of the instant invention, and discussed briefly below:

- Cong *et al.* engineered type II CRISPR-Cas systems for use in eukaryotic cells based on both *Streptococcus thermophilus* Cas9 and also *Streptococcus pyogenes* Cas9 and demonstrated that Cas9 nucleases can be directed by short RNAs to induce precise cleavage of DNA in human and mouse cells. Their study further showed that Cas9 as converted into a nicking enzyme can be used to facilitate homology-directed repair in eukaryotic cells with minimal mutagenic activity. Additionally, their study demonstrated that multiple guide sequences can be encoded into a single CRISPR array to enable simultaneous editing of several at endogenous genomic loci sites within the mammalian genome, demonstrating easy programmability and wide applicability of the RNA-guided nuclease technology. This ability to use RNA to program sequence specific DNA cleavage in cells defined a new class of genome engineering tools. These studies further showed that other CRISPR loci are likely to be transplantable into mammalian cells and can also mediate mammalian genome cleavage. Importantly, it can be envisaged that several aspects of the CRISPR-Cas system can be further improved to increase its efficiency and versatility.
- Jiang *et al.* used the clustered, regularly interspaced, short palindromic repeats (CRISPR)–associated Cas9 endonuclease complexed with dual-RNAs to introduce precise mutations in the genomes of *Streptococcus pneumoniae* and *Escherichia coli.* The approach relied on dual-RNA:Cas9-directed cleavage at the targeted genomic site to kill unmutated cells and circumvents the need for selectable markers or counter-

selection systems. The study reported reprogramming dual-RNA:Cas9 specificity by changing the sequence of short CRISPR RNA (crRNA) to make single- and multinucleotide changes carried on editing templates. The study showed that simultaneous use of two crRNAs enabled multiplex mutagenesis. Furthermore, when the approach was used in combination with recombineering, in *S. pneumoniae*, nearly 100% of cells that were recovered using the described approach contained the desired mutation, and in *E. coli,* 65% that were recovered contained the mutation.

- Wang *et al.* (2013) used the CRISPR-Cas system for the one-step generation of mice carrying mutations in multiple genes which were traditionally generated in multiple steps by sequential recombination in embryonic stem cells and/or time-consuming intercrossing of mice with a single mutation. The CRISPR-Cas system will greatly accelerate the *in vivo* study of functionally redundant genes and of epistatic gene interactions.
- Konermann *et al.* (2013) addressed the need in the art for versatile and robust technologies that enable optical and chemical modulation of DNA-binding domains based CRISPR Cas9 enzyme and also Transcriptional Activator Like Effectors
- Ran *et al.* (2013-A) described an approach that combined a Cas9 nickase mutant with paired guide RNAs to introduce targeted double-strand breaks. This addresses the issue of the Cas9 nuclease from the microbial CRISPR-Cas system being targeted to specific genomic loci by a guide sequence, which can tolerate certain mismatches to the DNA target and thereby promote undesired off-target mutagenesis. Because individual nicks in the genome are repaired with high fidelity, simultaneous nicking *via* appropriately offset guide RNAs is required for double-stranded breaks and extends the number of specifically recognized bases for target cleavage. The authors demonstrated that using paired nicking can reduce off-target activity by 50- to 1,500-fold in cell lines and to facilitate gene knockout in mouse zygotes without sacrificing on-target cleavage efficiency. This versatile strategy enables a wide variety of genome editing applications that require high specificity.
- Hsu *et al.* (2013) characterized SpCas9 targeting specificity in human cells to inform the selection of target sites and avoid off-target effects. The study evaluated >700 guide RNA variants and SpCas9-induced indel mutation levels at >100 predicted genomic off-target loci in 293T and 293FT cells. The authors that SpCas9 tolerates mismatches between guide RNA and target DNA at different positions in a sequence-dependent manner, sensitive to the number, position and distribution of mismatches. The authors

further showed that SpCas9-mediated cleavage is unaffected by DNA methylation and that the dosage of SpCas9 and guide RNA can be titrated to minimize off-target modification. Additionally, to facilitate mammalian genome engineering applications, the authors reported providing a web-based software tool to guide the selection and validation of target sequences as well as off-target analyses.

- Ran *et al.* (2013-B) described a set of tools for Cas9-mediated genome editing *via* non-homologous end joining (NHEJ) or homology-directed repair (HDR) in mammalian cells, as well as generation of modified cell lines for downstream functional studies. To minimize off-target cleavage, the authors further described a double-nicking strategy using the Cas9 nickase mutant with paired guide RNAs. The protocol provided by the authors experimentally derived guidelines for the selection of target sites, evaluation of cleavage efficiency and analysis of off-target activity. The studies showed that beginning with target design, gene modifications can be achieved within as little as 1–2 weeks, and modified clonal cell lines can be derived within 2–3 weeks.
- Shalem *et al.* described a new way to interrogate gene function on a genome-wide scale. Their studies showed that delivery of a genome-scale CRISPR-Cas9 knockout (GeCKO) library targeted 18,080 genes with 64,751 unique guide sequences enabled both negative and positive selection screening in human cells. First, the authors showed use of the GeCKO library to identify genes essential for cell viability in cancer and pluripotent stem cells. Next, in a melanoma model, the authors screened for genes whose loss is involved in resistance to vemurafenib, a therapeutic that inhibits mutant protein kinase BRAF. Their studies showed that the highest-ranking candidates included previously validated genes NF1 and MED12 as well as novel hits NF2, CUL3, TADA2B, and TADA1. The authors observed a high level of consistency between independent guide RNAs targeting the same gene and a high rate of hit confirmation, and thus demonstrated the promise of genome-scale screening with Cas9.
- Nishimasu *et al.* reported the crystal structure of *Streptococcus pyogenes* Cas9 in complex with sgRNA and its target DNA at 2.5 A° resolution. The structure revealed a bilobed architecture composed of target recognition and nuclease lobes, accommodating the sgRNA:DNAn RNA duplex in a positively charged groove at their interface. Whereas the recognition lobe is essential for binding sgRNA and DNA, the nuclease lobe contains the HNH and RuvC nuclease domains, which are properly positioned for cleavage of the complementary and non-complementary strands of the target DNA, respectively. The nuclease lobe also contains a carboxyl-terminal domain

responsible for the interaction with the protospacer adjacent motif (PAM). This high-resolution structure and accompanying functional analyses have revealed the molecular mechanism of RNA-guided DNA targeting by Cas9, thus paving the way for the rational design of new, versatile genome-editing technologies.

- Wu *et al*. mapped genome-wide binding sites of a catalytically inactive Cas9 (dCas9) from *Streptococcus pyogenes* loaded with single guide RNAs (sgRNAs) in mouse embryonic stem cells (mESCs). The authors showed that each of the four sgRNAs tested targets dCas9 to between tens and thousands of genomic sites, frequently characterized by a 5-nucleotide seed region in the sgRNA and an NGG protospacer adjacent motif (PAM). Chromatin inaccessibility decreases dCas9 binding to other sites with matching seed sequences; thus 70% of off-target sites are associated with genes. The authors showed that targeted sequencing of 295 dCas9 binding sites in mESCs transfected with catalytically active Cas9 identified only one site mutated above background levels. The authors proposed a two-state model for Cas9 binding and cleavage, in which a seed match triggers binding but extensive pairing with target DNA is required for cleavage.
- Platt *et al*. established a Cre-dependent Cas9 knockin mouse. The authors demonstrated *in vivo* as well as *ex vivo* genome editing using adeno-associated virus (AAV)-, lentivirus-, or particle-mediated delivery of guide RNA in neurons, immune cells, and endothelial cells.
- Hsu *et al.* (2014) is a review article that discusses generally CRISPR-Cas9 history from yogurt to genome editing, including genetic screening of cells.
- Wang *et al.* (2014) relates to a pooled, loss-of-function genetic screening approach suitable for both positive and negative selection that uses a genome-scale lentiviral single guide RNA (sgRNA) library.
- Doench *et al*. created a pool of sgRNAs, tiling across all possible target sites of a panel of six endogenous mouse and three endogenous human genes and quantitatively assessed their ability to produce null alleles of their target gene by antibody staining and flow cytometry. The authors showed that optimization of the PAM improved activity and also provided an on-line tool for designing sgRNAs.
- Swiech *et al*. demonstrate that AAV-mediated SpCas9 genome editing can enable reverse genetic studies of gene function in the brain.

- Konermann *et al.* (2015) discusses the ability to attach multiple effector domains, e.g., transcriptional activator, functional and epigenomic regulators at appropriate positions on the guide such as stem or tetraloop with and without linkers.
- Zetsche *et al.* demonstrates that the Cas9 enzyme can be split into two and hence the assembly of Cas9 for activation can be controlled.
- Chen *et al.* relates to multiplex screening by demonstrating that a genome-wide *in vivo* CRISPR-Cas9 screen in mice reveals genes regulating lung metastasis.
- Ran *et al.* (2015) relates to SaCas9 and its ability to edit genomes and demonstrates that one cannot extrapolate from biochemical assays.
- Shalem *et al.* (2015) described ways in which catalytically inactive Cas9 (dCas9) fusions are used to synthetically repress (CRISPRi) or activate (CRISPRa) expression, showing. advances using Cas9 for genome-scale screens, including arrayed and pooled screens, knockout approaches that inactivate genomic loci and strategies that modulate transcriptional activity.
- Xu *et al.* (2015) assessed the DNA sequence features that contribute to single guide RNA (sgRNA) efficiency in CRISPR-based screens. The authors explored efficiency of CRISPR-Cas9 knockout and nucleotide preference at the cleavage site. The authors also found that the sequence preference for CRISPRi/a is substantially different from that for CRISPR-Cas9 knockout.
- Parnas *et al.* (2015) introduced genome-wide pooled CRISPR-Cas9 libraries into dendritic cells (DCs) to identify genes that control the induction of tumor necrosis factor (Tnf) by bacterial lipopolysaccharide (LPS). Known regulators of Tlr4 signaling and previously unknown candidates were identified and classified into three functional modules with distinct effects on the canonical responses to LPS.
- Ramanan *et al* (2015) demonstrated cleavage of viral episomal DNA (cccDNA) in infected cells. The HBV genome exists in the nuclei of infected hepatocytes as a 3.2kb double-stranded episomal DNA species called covalently closed circular DNA (cccDNA), which is a key component in the HBV life cycle whose replication is not inhibited by current therapies. The authors showed that sgRNAs specifically targeting highly conserved regions of HBV robustly suppresses viral replication and depleted cccDNA.
- Nishimasu *et al.* (2015) reported the crystal structures of SaCas9 in complex with a single guide RNA (sgRNA) and its double-stranded DNA targets, containing the 5'-TTGAAT-3' PAM and the 5'-TTGGGT-3' PAM. A structural comparison of SaCas9

with SpCas9 highlighted both structural conservation and divergence, explaining their distinct PAM specificities and orthologous sgRNA recognition.

- Canver *et al.* (2015) demonstrated a CRISPR-Cas9-based functional investigation of non-coding genomic elements. The authors we developed pooled CRISPR-Cas9 guide RNA libraries to perform *in situ* saturating mutagenesis of the human and mouse BCL11A enhancers which revealed critical features of the enhancers.
- Zetsche et al. (2015) reported characterization of Cas13, a class 2 CRISPR nuclease from Francisella novicida U112 having features distinct from Cas9. Cas13 is a single RNA-guided endonuclease lacking tracrRNA, utilizes a T-rich protospacer-adjacent motif, and cleaves DNA via a staggered DNA double-stranded break.
- Shmakov et al. (2015) reported three distinct Class 2 CRISPR-Cas systems. Two system CRISPR enzymes (C2c1 and C2c3) contain RuvC-like endonuclease domains distantly related to Cas13. Unlike Cas13, C2c1 depends on both crRNA and tracrRNA for DNA cleavage. The third enzyme (C2c2) contains two predicted HEPN RNase domains and is tracrRNA independent.
- Slaymaker et al (2016) reported the use of structure-guided protein engineering to improve the specificity of Streptococcus pyogenes Cas9 (SpCas9). The authors developed "enhanced specificity" SpCas9 (eSpCas9) variants which maintained robust on-target cleavage with reduced off-target effects.

[0623] The methods and tools provided herein are exemplified for Cas13, a type II nuclease that does not make use of tracrRNA. Orthologs of Cas13 have been identified in different bacterial species as described herein. Further type II nucleases with similar properties can be identified using methods described in the art (Shmakov et al. 2015, 60:385–397; Abudayeh et al. 2016, Science, 5;353(6299)) . In particular embodiments, such methods for identifying novel CRISPR effector proteins may comprise the steps of selecting sequences from the database encoding a seed which identifies the presence of a CRISPR Cas locus, identifying loci located within 10 kb of the seed comprising Open Reading Frames (ORFs) in the selected sequences, selecting therefrom loci comprising ORFs of which only a single ORF encodes a novel CRISPR effector having greater than 700 amino acids and no more than 90% homology to a known CRISPR effector. In particular embodiments, the seed is a protein that is common to the CRISPR-Cas system, such as Cas1. In further embodiments, the CRISPR array is used as a seed to identify new effector proteins.

[0624] The effectiveness of the present invention has been demonstrated. Preassembled recombinant CRISPR-Cas13 complexes comprising Cas13 and crRNA may be transfected, for

example by electroporation, resulting in high mutation rates and absence of detectable off-target mutations. Hur, J.K. et al, Targeted mutagenesis in mice by electroporation of Cas13 ribonucleoproteins, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3596. Genome-wide analyses shows that Cas13 is highly specific. By one measure, in vitro cleavage sites determined for Cas13 in human HEK293T cells were significantly fewer that for SpCas9. Kim, D. et al., Genome-wide analysis reveals specificities of Cas13 endonucleases in human cells, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3609. An efficient multiplexed system employing Cas13 has been demonstrated in Drosophila employing gRNAs processed from an array containing inventing tRNAs. Port, F. et al, Expansion of the CRISPR toolbox in an animal with tRNA-flanked Cas9 and Cas13 gRNAs. doi: http://dx.doi.org/10.1101/046417.

[0625] Also, "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing", Shengdar Q. Tsai, Nicolas Wyvekens, Cyd Khayter, Jennifer A. Foden, Vishal Thapar, Deepak Reyon, Mathew J. Goodwin, Martin J. Aryee, J. Keith Joung Nature Biotechnology 32(6): 569-77 (2014), relates to dimeric RNA-guided FokI Nucleases that recognize extended sequences and can edit endogenous genes with high efficiencies in human cells.

[0626] With respect to general information on CRISPR/Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, and making and using thereof, including as to amounts and formulations, as well as CRISPR-Cas-expressing eukaryotic cells, CRISPR-Cas expressing eukaryotes, such as a mouse, reference is made to: US Patents Nos. 8,999,641, 8,993,233, 8,697,359, 8,771,945, 8,795,965, 8,865,406, 8,871,445, 8,889,356, 8,889,418, 8,895,308, 8,906,616, 8,932,814, and 8,945,839; US Patent Publications US 2014-0310830 (US App. Ser. No. 14/105,031), US 2014-0287938 A1 (U.S. App. Ser. No. 14/213,991), US 2014-0273234 A1 (U.S. App. Ser. No. 14/293,674), US2014-0273232 A1 (U.S. App. Ser. No. 14/290,575), US 2014-0273231 (U.S. App. Ser. No. 14/259,420), US 2014-0256046 A1 (U.S. App. Ser. No. 14/226,274), US 2014-0248702 A1 (U.S. App. Ser. No. 14/258,458), US 2014-0242700 A1 (U.S. App. Ser. No. 14/222,930), US 2014-0242699 A1 (U.S. App. Ser. No. 14/183,512), US 2014-0242664 A1 (U.S. App. Ser. No. 14/104,990), US 2014-0234972 A1 (U.S. App. Ser. No. 14/183,471), US 2014-0227787 A1 (U.S. App. Ser. No. 14/256,912), US 2014-0189896 A1 (U.S. App. Ser. No. 14/105,035), US 2014-0186958 (U.S. App. Ser. No. 14/105,017), US 2014-0186919 A1 (U.S. App. Ser. No. 14/104,977), US 2014-0186843 A1 (U.S. App. Ser. No. 14/104,900), US 2014-0179770 A1 (U.S. App. Ser. No. 14/104,837) and US 2014-0179006 A1 (U.S. App. Ser. No. 14/183,486), US 2014-0170753 (US App Ser No 14/183,429); US 2015-0184139 (U.S. App.

Ser. No. 14/324,960); 14/054,414 European Patent Applications EP 2 771 468 (EP13818570.7), EP 2 764 103 (EP13824232.6), and EP 2 784 162 (EP14170383.5); and PCT Patent Publications WO2014/093661 (PCT/US2013/074743), WO2014/093694 (PCT/US2013/074790), WO2014/093595 (PCT/US2013/074611), WO2014/093718 (PCT/US2013/074825), WO2014/093709 (PCT/US2013/074812), WO2014/093622 (PCT/US2013/074667), WO2014/093635 (PCT/US2013/074691), WO2014/093655 (PCT/US2013/074736), WO2014/093712 (PCT/US2013/074819), WO2014/093701 (PCT/US2013/074800), WO2014/018423 (PCT/US2013/051418), WO2014/204723 (PCT/US2014/041790), WO2014/204724 (PCT/US2014/041800), WO2014/204725 (PCT/US2014/041803), WO2014/204726 (PCT/US2014/041804), WO2014/204727 (PCT/US2014/041806), WO2014/204728 (PCT/US2014/041808), WO2014/204729 (PCT/US2014/041809), WO2015/089351 (PCT/US2014/069897), WO2015/089354 (PCT/US2014/069902), WO2015/089364 (PCT/US2014/069925), WO2015/089427 (PCT/US2014/070068), WO2015/089462 (PCT/US2014/070127), WO2015/089419 (PCT/US2014/070057), WO2015/089465 (PCT/US2014/070135), WO2015/089486 (PCT/US2014/070175), WO2015/058052 (PCT/US2014/061077), WO2015/070083 (PCT/US2014/064663), WO2015/089354 (PCT/US2014/069902), WO2015/089351 (PCT/US2014/069897), WO2015/089364 (PCT/US2014/069925), WO2015/089427 (PCT/US2014/070068), WO2015/089473 (PCT/US2014/070152), WO2015/089486 (PCT/US2014/070175), WO2016/049258 (PCT/US2015/051830), WO2016/094867 (PCT/US2015/065385), WO2016/094872 (PCT/US2015/065393), WO2016/094874 (PCT/US2015/065396), WO2016/106244 (PCT/US2015/067177).

[0627] Mention is also made of US application 62/180,709, 17-Jun-15, PROTECTED GUIDE RNAS (PGRNAS); US application 62/091,455, filed, 12-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US application 62/096,708, 24-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US applications 62/091,462, 12-Dec-14, 62/096,324, 23-Dec-14, 62/180,681, 17-Jun-2015, and 62/237,496, 5-Oct-2015, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US application 62/091,456, 12-Dec-14 and 62/180,692, 17-Jun-2015, ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS; US application 62/091,461, 12-Dec-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs); US application 62/094,903, 19-Dec-14, UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE

SEQUENCING; US application 62/096,761, 24-Dec-14, ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION; US application 62/098,059, 30-Dec-14, 62/181,641, 18-Jun-2015, and 62/181,667, 18-Jun-2015, RNA-TARGETING SYSTEM; US application 62/096,656, 24-Dec-14 and 62/181,151, 17-Jun-2015, CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS; US application 62/096,697, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH AAV; US application 62/098,158, 30-Dec-14, ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS; US application 62/151,052, 22-Apr-15, CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING; US application 62/054,490, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS; US application 61/939,154, 12-F EB-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,484, 25-Sep-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,537, 4-Dec-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/054,651, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US application 62/067,886, 23-Oct-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US applications 62/054,675, 24-Sep-14 and 62/181,002, 17-Jun-2015, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES; US application 62/054,528, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS; US application 62/055,454, 25-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP); US application 62/055,460, 25-Sep-14, MULTIFUNCTIONAL-CRISPR

COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; US application 62/087,475, 4-Dec-14 and 62/181,690, 18-Jun-2015, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,487, 25-Sep-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,546, 4-Dec-14 and 62/181,687, 18-Jun-2015, MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; and US application 62/098,285, 30-Dec-14, CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS.

[0628] Mention is made of US applications 62/181,659, 18-Jun-2015 and 62/207,318, 19-Aug-2015, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS, ENZYME AND GUIDE SCAFFOLDS OF CAS9 ORTHOLOGS AND VARIANTS FOR SEQUENCE MANIPULATION. Mention is made of US applications 62/181,663, 18-Jun-2015 and 62/245,264, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS, US applications 62/181,675, 18-Jun-2015, 62/285,349, 22-Oct-2015, 62/296,522, 17-Feb-2016, and 62/320,231, 8-Apr-2016, NOVEL CRISPR ENZYMES AND SYSTEMS, US application 62/232,067, 24-Sep-2015, US Application 14/975,085, 18-Dec-2015, European application No. 16150428.7, US application 62/205,733, 16-Aug-2015, US application 62/201,542, 5-Aug-2015, US application 62/193,507, 16-Jul-2015, and US application 62/181,739, 18-Jun-2015, each entitled NOVEL CRISPR ENZYMES AND SYSTEMS and of US application 62/245,270, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS. Mention is also made of US application 61/939,256, 12-Feb-2014, and WO 2015/089473 (PCT/US2014/070152), 12-Dec-2014, each entitled ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED GUIDE COMPOSITIONS WITH NEW ARCHITECTURES FOR SEQUENCE MANIPULATION. Mention is also made of PCT/US2015/045504, 15-Aug-2015, US application 62/180,699, 17-Jun-2015, and US application 62/038,358, 17-Aug-2014, each entitled GENOME EDITING USING CAS9 NICKASES.

[0629] Each of these patents, patent publications, and applications, and all documents cited therein or during their prosecution ("appln cited documents") and all documents cited or referenced in the appln cited documents, together with any instructions, descriptions, product specifications, and product sheets for any products mentioned therein or in any document therein and incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention. All documents (e.g., these patents, patent publications and applications and the appln cited documents) are incorporated herein by

reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

Type-V CRISPR-Cas Protein

[0630] The application describes methods using Type-V CRISPR-Cas proteins. This is exemplified herein with Cas13, whereby a number of orthologs or homologs have been identified. It will be apparent to the skilled person that further orthologs or homologs can be identified and that any of the functionalities described herein may be engineered into other orthologs, incuding chimeric enzymes comprising fragments from multiple orthologs.

[0631] Computational methods of identifying novel CRISPR-Cas loci are described in EP3009511 or US2016208243 and may comprise the following steps: detecting all contigs encoding the Cas1 protein; identifying all predicted protein coding genes within 20kB of the cas1 gene; comparing the identified genes with Cas protein-specific profiles and predicting CRISPR arrays; selecting unclassified candidate CRISPR-Cas loci containing proteins larger than 500 amino acids (>500 aa); analyzing selected candidates using methods such as PSI-BLAST and HHPred to screen for known protein domains, thereby identifying novel Class 2 CRISPR-Cas loci (see also Schmakov et al. 2015, Mol Cell. 60(3):385-97). In addition to the above mentioned steps, additional analysis of the candidates may be conducted by searching metagenomics databases for additional homologs. Additionally or alternatively, to expand the search to non-autonomous CRISPR-Cas systems, the same procedure can be performed with the CRISPR array used as the seed.

[0632] In one aspect the detecting all contigs encoding the Cas1 protein is performed by GenemarkS which a gene prediction program as further described in "GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions." John Besemer, Alexandre Lomsadze and Mark Borodovsky, Nucleic Acids Research (2001) 29, pp 2607-2618, herein incorporated by reference.

[0633] In one aspect the identifying all predicted protein coding genes is carried out by comparing the identified genes with Cas protein-specific profiles and annotating them according to NCBI Conserved Domain Database (CDD) which is a protein annotation resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins. These are available as position-specific score matrices (PSSMs) for fast identification of conserved domains in protein sequences via RPS-BLAST. CDD content includes NCBI-curated domains, which use 3D-structure information to explicitly define domain boundaries and provide insights into sequence/structure/function

relationships, as well as domain models imported from a number of external source databases (Pfam, SMART, COG, PRK, TIGRFAM). In a further aspect, CRISPR arrays were predicted using a PILER-CR program which is a public domain software for finding CRISPR repeats as described in "PILER-CR: fast and accurate identification of CRISPR repeats", Edgar, R.C., BMC Bioinformatics, Jan 20;8:18(2007), herein incorporated by reference.

[0634] In a further aspect, the case by case analysis is performed using PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool). PSI-BLAST derives a position-specific scoring matrix (PSSM) or profile from the multiple sequence alignment of sequences detected above a given score threshold using protein–protein BLAST. This PSSM is used to further search the database for new matches, and is updated for subsequent iterations with these newly detected sequences. Thus, PSI-BLAST provides a means of detecting distant relationships between proteins.

[0635] In another aspect, the case by case analysis is performed using HHpred, a method for sequence database searching and structure prediction that is as easy to use as BLAST or PSI-BLAST and that is at the same time much more sensitive in finding remote homologs. In fact, HHpred's sensitivity is competitive with the most powerful servers for structure prediction currently available. HHpred is the first server that is based on the pairwise comparison of profile hidden Markov models (HMMs). Whereas most conventional sequence search methods search sequence databases such as UniProt or the NR, HHpred searches alignment databases, like Pfam or SMART. This greatly simplifies the list of hits to a number of sequence families instead of a clutter of single sequences. All major publicly available profile and alignment databases are available through HHpred. HHpred accepts a single query sequence or a multiple alignment as input. Within only a few minutes it returns the search results in an easy-to-read format similar to that of PSI-BLAST. Search options include local or global alignment and scoring secondary structure similarity. HHpred can produce pairwise query-template sequence alignments, merged query-template multiple alignments (e.g. for transitive searches), as well as 3D structural models calculated by the MODELLER software from HHpred alignments.

Orthologs of Cas13

[0636] The terms "orthologue" (also referred to as "ortholog" herein) and "homologue" (also referred to as "homolog" herein) are well known in the art. By means of further guidance, a "homologue" of a protein as used herein is a protein of the same species which performs the same or a similar function as the protein it is a homologue of. Homologous proteins may but need not be structurally related, or are only partially structurally related. An "orthologue" of a protein as used herein is a protein of a different species which performs the same or a similar

function as the protein it is an orthologue of. Orthologous proteins may but need not be structurally related, or are only partially structurally related. Homologs and orthologs may be identified by homology modelling (see, e.g., Greer, Science vol. 228 (1985) 1055, and Blundell et al. Eur J Biochem vol 172 (1988), 513) or "structural BLAST" (Dey F, Cliff Zhang Q, Petrey D, Honig B. Toward a "structural BLAST": using structural relationships to infer function. Protein Sci. 2013 Apr;22(4):359-66. doi: 10.1002/pro.2225.). See also Shmakov et al. (2015) for application in the field of CRISPR-Cas loci. Homologous proteins may but need not be structurally related, or are only partially structurally related.

[0637] The Cas13 gene is found in several diverse bacterial genomes, typically in the same locus with cas1, cas2, and cas4 genes and a CRISPR cassette (for example, FNFX1_1431-FNFX1_1428 of Francisella cf . novicida Fx1). Thus, the layout of this putative novel CRISPR-Cas system appears to be similar to that of type II-B. Furthermore, similar to Cas9, the Cas13 protein contains a readily identifiable C-terminal region that is homologous to the transposon ORF-B and includes an active RuvC-like nuclease, an arginine-rich region, and a Zn finger (absent in Cas9). However, unlike Cas9, Cas13 is also present in several genomes without a CRISPR-Cas context and its relatively high similarity with ORF-B suggests that it might be a transposon component. It was suggested that if this was a genuine CRISPR-Cas system and Cas13 is a functional analog of Cas9 it would be a novel CRISPR-Cas type, namely type V (See Annotation and Classification of CRISPR-Cas Systems. Makarova KS, Koonin EV. Methods Mol Biol. 2015;1311:47-75). However, as described herein, Cas13 is denoted to be in subtype V-A to distinguish it from C2c1p which does not have an identical domain structure and is hence denoted to be in subtype V-B.

[0638] The present invention encompasses the use of a Cas13 effector protein, derived from a Cas13 locus denoted as subtype V-A. Herein such effector proteins are also referred to as “Cas13p”, e.g., a Cas13 protein (and such effector protein or Cas13 protein or protein derived from a Cas13 locus is also called “CRISPR-Cas protein”).

[0639] In particular embodiments, the effector protein is a Cas13 effector protein from an organism from a genus comprising Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Methanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes, Helcococcus, Leptospira, Desulfovibrio, Desulfonatronum, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium, Butyvibrio, Perigrinibacterium,

Pareubacterium, Moraxella, Thiomicrospira or Acidaminococcus. In particular embodiments, the Cas13 effector protein is selected from an organism from a genus selected from Eubacterium, Lachnospiraceae, Leptotrichia, Francisella, Methanomethyophilus, Porphyromonas, Prevotella, Leptospira, Butyvibrio, Perigrinibacterium, Pareubacterium, Moraxella, Thiomicrospira or Acidaminococcus

[0640] In further particular embodiments, the Cas13 effector protein is from an organism selected from S. mutans, S. agalactiae, S. equisimilis, S. sanguinis, S. pneumonia; C. jejuni, C. coli; N. salsuginis, N. tergarcus; S. auricularis, S. carnosus; N. meningitides, N. gonorrhoeae; L. monocytogenes, L. ivanovii; C. botulinum, C. difficile, C. tetani, C. sordellii, L inadai, F. tularensis 1, P. albensis, L. bacterium, B. proteoclasticus, P. bacterium, P. crevioricanis, P. disiens and P. macacae .

[0641] The effector protein may comprise a chimeric effector protein comprising a first fragment from a first effector protein (e.g., a Cas13) ortholog and a second fragment from a second effector (e.g., a Cas13) protein ortholog, and wherein the first and second effector protein orthologs are different. At least one of the first and second effector protein (e.g., a Cas13) orthologs may comprise an effector protein (e.g., a Cas13) from an organism comprising Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Methanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes, Helcococcus, Letospira, Desulfovibrio, Desulfonatronum, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium, Butyvibrio, Perigrinibacterium, Pareubacterium, Moraxella, Thiomicrospira or Acidaminococcus; e.g., a chimeric effector protein comprising a first fragment and a second fragment wherein each of the first and second fragments is selected from a Cas13 of an organism comprising Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Methanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes, Helcococcus, Letospira, Desulfovibrio, Desulfonatronum, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium, Butyvibrio, Perigrinibacterium, Pareubacterium, Moraxella, Thiomicrospira or Acidaminococcus wherein the first and second fragments are not from the same bacteria; for instance a chimeric effector protein comprising

a first fragment and a second fragment wherein each of the first and second fragments is selected from a Cas13 of S. mutans, S. agalactiae, S. equisimilis, S. sanguinis, S. pneumonia; C. jejuni, C. coli; N. salsuginis, N. tergarcus; S. auricularis, S. carnosus; N. meningitides, N. gonorrhoeae; L. monocytogenes, L. ivanovii; C. botulinum, C. difficile, C. tetani, C. sordellii; Francisella tularensis 1, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens and Porphyromonas macacae, wherein the first and second fragments are not from the same bacteria.

[0642] In a more preferred embodiment, the Cas13p is derived from a bacterial species selected from Francisella tularensis 1, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Moraxella bovoculi AAX08_00205, Moraxella bovoculi AAX11_00205, Butyrivibrio sp. NC3005, Thiomicrospira sp. XS5, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens and Porphyromonas macacae. In certain embodiments, the Cas13p is derived from a bacterial species selected from Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020.In certain embodiments, the effector protein is derived from a subspecies of Francisella tularensis 1, including but not limited to Francisella tularensis subsp. Novicida. In certain preferred embodiments, the Cas13p is derived from a bacterial species selected from Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium ND2006, Lachnospiraceae bacterium MA2020, Moraxella bovoculi AAX08_00205, Moraxella bovoculi AAX11_00205, Butyrivibrio sp. NC3005, or Thiomicrospira sp. XS5.

[0643] In particular embodiments, the homologue or orthologue of Cas13 as referred to herein has a sequence homology or identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the example Cas13 proteins disclosed herein. In further embodiments, the homologue or orthologue of Cas13 as referred to herein has a sequence identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the wild type Cas13. Where the Cas13 has one or more mutations (mutated), the homologue or orthologue of said Cas13 as

referred to herein has a sequence identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the mutated Cas13.

[0644] In an ambodiment, the Cas13 protein may be an ortholog of an organism of a genus which includes, but is not limited to Acidaminococcus sp, Lachnospiraceae bacterium or Moraxella bovoculi; in particular embodiments, the type V Cas protein may be an ortholog of an organism of a species which includes, but is not limited to Acidaminococcus sp. BV3L6; Lachnospiraceae bacterium ND2006 (LbCas13) or Moraxella bovoculi 237.In particular embodiments, the homologue or orthologue of Cas13 as referred to herein has a sequence homology or identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with one or more of the Cas13 sequences disclosed herein. In further embodiments, the homologue or orthologue of Cas13 as referred to herein has a sequence identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the wild type FnCas13, AsCas13 or LbCas13.

[0645] In particular embodiments, the Cas13 protein of the invention has a sequence homology or identity of at least 60%, more particularly at least 70, such as at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with FnCas13, AsCas13 or LbCas13. In further embodiments, the Cas13 protein as referred to herein has a sequence identity of at least 60%, such as at least 70%, more particularly at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the wild type AsCas13 or LbCas13. In particular embodiments, the Cas13 protein of the present invention has less than 60% sequence identity with FnCas13. The skilled person will understand that this includes truncated forms of the Cas13 protein whereby the sequence identity is determined over the length of the truncated form. In particular embodiments, the Cas13 enzyme is not FnCas13.

Modified Cas13 enzymes

[0646] In particular embodiments, it is of interest to make use of an engineered Cas13 protein as defined herein, such as Cas13, wherein the protein complexes with a nucleic acid molecule comprising RNA to form a CRISPR complex, wherein when in the CRISPR complex, the nucleic acid molecule targets one or more target polynucleotide loci, the protein comprises at least one modification compared to unmodified Cas13 protein, and wherein the CRISPR complex comprising the modified protein has altered activity as compared to the complex comprising the unmodified Cas13 protein. It is to be understood that when referring herein to CRISPR "protein", the Cas13 protein preferably is a modified CRISPR-Cas protein (e.g. having increased or decreased (or no) enzymatic activity, such as without limitation including

Cas13. The term "CRISPR protein" may be used interchangeably with "CRISPR-Cas protein", irrespective of whether the CRISPR protein has altered, such as increased or decreased (or no) enzymatic activity, compared to the wild type CRISPR protein.

[0647] Computational analysis of the primary structure of Cas13 nucleases reveals three distinct regions. First a C-terminal RuvC like domain, which is the only functional characterized domain. Second a N-terminal alpha-helical region and thirst a mixed alpha and beta region, located between the RuvC like domain and the alpha-helical region.

[0648] Several small stretches of unstructured regions are predicted within the Cas13 primary structure. Unstructured regions, which are exposed to the solvent and not conserved within different Cas13 orthologs, are preferred sides for splits and insertions of small protein sequences . In addition, these sides can be used to generate chimeric proteins between Cas13 orthologs.

[0649] Based on the above information, mutants can be generated which lead to inactivation of the enzyme or which modify the double strand nuclease to nickase activity. In alternative embodiments, this information is used to develop enzymes with reduced off-target effects (described elsewhere herein)

[0650] In certain of the above-described Cas13 enzymes, the enzyme is modified by mutation of one or more residues (in the RuvC domain) including but not limited to positions R909, R912, R930, R947, K949, R951, R955, K965, K968, K1000, K1002, R1003, K1009, K1017, K1022, K1029, K1035, K1054, K1072, K1086, R1094, K1095, K1109, K1118, K1142, K1150, K1158, K1159, R1220, R1226, R1242, and/or R1252 with reference to amino acid position numbering of AsCas13 (Acidaminococcus sp. BV3L6). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0651] In certain of the above-described non-naturally-occurring CRISPR-Cas proteins, the enzyme is modified by mutation of one or more residues (in the RAD50) domain including but not limited positions K324, K335, K337, R331, K369, K370, R386, R392, R393, K400, K404, K406, K408, K414, K429, K436, K438, K459, K460, K464, R670, K675, R681, K686, K689, R699, K705, R725, K729, K739, K748, and/or K752 with reference to amino acid position numbering of AsCas13 (Acidaminococcus sp. BV3L6). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0652] In certain of the Cas13 enzymes, the enzyme is modified by mutation of one or more residues including but not limited positions R912, T923, R947, K949, R951, R955, K965,

K968, K1000, R1003, K1009, K1017, K1022, K1029, K1072, K1086, F1103, R1226, and/or R1252 with reference to amino acid position numbering of AsCas13 (Acidaminococcus sp. BV3L6). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0653] In certain embodiments, the Cas13 enzyme is modified by mutation of one or more residues including but not limited positions R833, R836, K847, K879, K881, R883, R887, K897, K900, K932, R935, K940, K948, K953, K960, K984, K1003, K1017, R1033, R1138, R1165, and/or R1252 with reference to amino acid position numbering of LbCas13 (Lachnospiraceae bacterium ND2006). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0654] In certain embodiments, the Cas13 enzyme is modified by mutation of one or more residues including but not limited positions K15, R18, K26, Q34, R43, K48, K51, R56, R84, K85, K87, N93, R103, N104, T118, K123, K134, R176, K177, R192, K200, K226, K273, K275, T291, R301, K307, K369, S404, V409, K414, K436, K438, K468, D482, K516, R518, K524, K530, K532, K548, K559, K570, R574, K592, D596, K603, K607, K613, C647, R681, K686, H720, K739, K748, K757, T766, K780, R790, P791, K796, K809, K815, T816, K860, R862, R863, K868, K897, R909, R912, T923, R947, K949, R951, R955, K965, K968, K1000, R1003, K1009, K1017, K1022, K1029, A1053, K1072, K1086, F1103, S1209, R1226, R1252, K1273, K1282, and/or K1288 with reference to amino acid position numbering of AsCas13 (Acidaminococcus sp. BV3L6). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0655] In certain embodiments, the enzyme is modified by mutation of one or more residues including but not limited positions K15, R18, K26, R34, R43, K48, K51, K56, K87, K88, D90, K96, K106, K107, K120, Q125, K143, R186, K187, R202, K210, K235, K296, K298, K314, K320, K326, K397, K444, K449, E454, A483, E491, K527, K541, K581, R583, K589, K595, K597, K613, K624, K635, K639, K656, K660, K667, K671, K677, K719, K725, K730, K763, K782, K791, R800, K809, K823, R833, K834, K839, K852, K858, K859, K869, K871, R872, K877, K905, R918, R921, K932, I960, K962, R964, R968, K978, K981, K1013, R1016, K1021, K1029, K1034, K1041, K1065, K1084, and/or K1098 with reference to amino acid position numbering of FnCas13 (Francisella novicida U112). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0656] In certain embodiments, the enzyme is modified by mutation of one or more residues including but not limited positions K15, R18, K26, K34, R43, K48, K51, R56, K83, K84, R86, K92, R102, K103, K116, K121, R158, E159, R174, R182, K206, K251, K253, K269, K271, K278, P342, K380, R385, K390, K415, K421, K457, K471, A506, R508, K514, K520, K522, K538, Y548, K560, K564, K580, K584, K591, K595, K601, K634, K640, R645, K679, K689, K707, T716, K725, R737, R747, R748, K753, K768, K774, K775, K785, K787, R788, Q793, K821, R833, R836, K847, K879, K881, R883, R887, K897, K900, K932, R935, K940, K948, K953, K960, K984, K1003, K1017, R1033, K1121, R1138, R1165, K1190, K1199, and/or K1208 with reference to amino acid position numbering of LbCas13 (Lachnospiraceae bacterium ND2006). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0657] In certain embodiments, the enzyme is modified by mutation of one or more residues including but not limited positions K14, R17, R25, K33, M42, Q47, K50, D55, K85, N86, K88, K94, R104, K105, K118, K123, K131, R174, K175, R190, R198, I221, K267, Q269, K285, K291, K297, K357, K403, K409, K414, K448, K460, K501, K515, K550, R552, K558, K564, K566, K582, K593, K604, K608, K623, K627, K633, K637, E643, K780, Y787, K792, K830, Q846, K858, K867, K876, K890, R900, K901, M906, K921, K927, K928, K937, K939, R940, K945, Q975, R987, R990, K1001, R1034, I1036, R1038, R1042, K1052, K1055, K1087, R1090, K1095, N1103, K1108, K1115, K1139, K1158, R1172, K1188, K1276, R1293, A1319, K1340, K1349, and/or K1356 with reference to amino acid position numbering of MbCas13 (Moraxella bovoculi 237). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0658] In one embodiment, the Cas13 protein is modified with a mutation at S1228 (e.g., S1228A) with reference to amino acid position numbering of AsCas13. See Yamano et al., Cell 165:949–962 (2016), which is incorporated herein by reference in its entirety.

[0659] In certain embodiments, the Cas13 protein has been modified to recognize a non-natural PAM, such as recognizing a PAM having a sequence or comprising a sequence YCN, YCV, AYV, TYV, RYN, RCN, TGYV, NTTN, TTN, TRTN, TYTV, TYCT, TYCN, TRTN, NTTN, TACT, TYCC, TRTC, TATV, NTTV, TTV, TSTG, TVTS, TYYS, TCYS, TBYS, TCYS, TNYS, TYYS, TNTN, TSTG, TTCC, TCCC, TATC, TGTG, TCTG, TYCV, or TCTC. In particular embodiments, said mutated Cas13 comprises one or more mutated amino acid residue at position 11, 12, 13, 14, 15, 16, 17, 34, 36, 39, 40, 43, 46, 47, 50, 54, 57, 58, 111, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 157, 158, 159, 160, 161, 162, 163, 164,

165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 642, 643, 644, 645, 646, 647, 648, 649, 651, 652, 653, 654, 655, 656, 676, 679, 680, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 707, 711, 714, 715, 716, 717, 718, 719, 720, 721, 722, 739, 765, 768, 769, 773, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, or 1048 of AsCas13 or a position corresponding thereto in a Cas13 ortholog; preferably, one or more mutated amino acid residue at position 130, 131, 132, 133, 134, 135, 136, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 570, 571, 572, 573, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 630, 631, 632, 646, 647, 648, 649, 650, 651, 652, 653, 683, 684, 685, 686, 687, 688, 689, or 690;

[0660] In certain embodiments, the Cas13 protein is modified to have increased activity, i.e. wider PAM specificity. In particular embodiments, the Cas13 protein is modified by mutation of one or more residues including but not limited positions 539, 542, 547, 548, 550, 551, 552, 167, 604, and/or 607 of AsCas13, or the corresponding position of an AsCas13 orthologue, homologue, or variant, preferably mutated amino acid residues at positions 542 or 542 and 607, wherein said mutations preferably are 542R and 607R, such as S542R and K607R; or preferably mutated amino acid residues at positions 542 and 548 (and optionally 552), wherein said mutations preferably are 542R and 548V (and optionally 552R), such as S542R and K548V (and optionally N552R); or at position 532, 538, 542, and/or 595 of LbCas13, or the corresponding position of an AsCas13 orthologue, homologue, or variant, preferably mutated amino acid residues at positions 532 or 532 and 595, wherein said mutations preferably are 532R and 595R, such as G532R and K595R; or preferably mutated amino acid residues at positions 532 and 538 (and optionally 542), wherein said mutations preferably are 532R and 538V (and optionally 542R), such as G532R and K538V (and optionally Y542R), most preferably wherein said mutations are S542R and K607R, S542R and K548V, or S542R, K548V and N552R of AsCas13.

Deactivated / inactivated Cas13 protein

[0661] Where the Cas13 protein has nuclease activity, the Cas13 protein may be modified to have diminished nuclease activity e.g., nuclease inactivation of at least 70%, at least 80%,

at least 90%, at least 95%, at least 97%, or 100% as compared with the wild type enzyme; or to put in another way, a Cas13 enzyme having advantageously about 0% of the nuclease activity of the non-mutated or wild type Cas13 enzyme or CRISPR-Cas protein, or no more than about 3% or about 5% or about 10% of the nuclease activity of the non-mutated or wild type Cas13 enzyme, e.g. of the non-mutated or wild type Francisella novicida U112 (FnCas13), Acidaminococcus sp. BV3L6 (AsCas13), Lachnospiraceae bacterium ND2006 (LbCas13) or Moraxella bovoculi 237 (MbCas13 Cas13 enzyme or CRISPR-Cas protein. This is possible by introducing mutations into the nuclease domains of the Cas13 and orthologs thereof.

[0662] In preferred embodiments of the present invention at least one Cas13 protein is used which is a Cas13 nickase. More particularly, a Cas13 nickase is used which does not cleave the target strand but is capable of cleaving only the strand which is complementary to the target strand, i.e. the non-target DNA strand also referred to herein as the strand which is not complementary to the guide sequence. More particularly the Cas13 nickase is a Cas13 protein which comprises a mutation in the arginine at position 1226A in the Nuc domain of Cas13 from Acidaminococcus sp., or a corresponding position in a Cas13 ortholog. In further particular embodiments, the enzyme comprises an arginine-to-alanine substitution or an R1226A mutation. It will be understood by the skilled person that where the enzyme is not AsCas13, a mutation may be made at a residue in a corresponding position. In particular embodiments, the Cas13 is FnCas13 and the mutation is at the arginine at position R1218. In particular embodiments, the Cas13 is LbCas13 and the mutation is at the arginine at position R1138. In particular embodiments, the Cas13 is MbCas13 and the mutation is at the arginine at position R1293.

[0663] In certain embodiments, use is made additionally or alternatively of a CRISPR-Cas protein which is is engineered and can comprise one or more mutations that reduce or eliminate a nuclease activity. The amino acid positions in the FnCas13p RuvC domain include but are not limited to D917A, E1006A, E1028A, D1227A, D1255A, N1257A, D917A, E1006A, E1028A, D1227A, D1255A and N1257A. Applicants have also identified a putative second nuclease domain which is most similar to PD-(D/E)XK nuclease superfamily and HincII endonuclease like. The point mutations to be generated in this putative nuclease domain to substantially reduce nuclease activity include but are not limited to N580A, N584A, T587A, W609A, D610A, K613A, E614A, D616A, K624A, D625A, K627A and Y629A. In a preferred embodiment, the mutation in the FnCas13p RuvC domain is D917A or E1006A, wherein the D917A or E1006A mutation completely inactivates the DNA cleavage activity of the FnCas13 effector protein. In another embodiment, the mutation in the FnCas13p RuvC domain is

D1255A, wherein the mutated FnCas13 effector protein has significantly reduced nucleolytic activity.

[0664] More particularly, the inactivated Cas13 enzymes include enzymes mutated in amino acid positions As908, As993, As1263 of AsCas13 or corresponding positions in Cas13 orthologs. Additionally, the inactivated Cas13 enzymes include enzymes mutated in amino acid position Lb832, 925, 947 or 1180 of LbCas13 or corresponding positions in Cas13 orthologs. More particularly, the inactivated Cas13 enzymes include enzymes comprising one or more of mutations AsD908A, AsE993A, AsD1263A of AsCas13 or corresponding mutations in Cas13 orthologs. Additionally, the inactivated Cas13 enzymes include enzymes comprising one or more of mutations LbD832A, E925A, D947A or D1180A of LbCas13 or corresponding mutations in Cas13 orthologs.

[0665] Mutations can also be made at neighboring residues, e.g., at amino acids near those indicated above that participate in the nuclease acrivity. In some embodiments, only the RuvC domain is inactivated, and in other embodiments, another putative nuclease domain is inactivated, wherein the effector protein complex functions as a nickase and cleaves only one DNA strand. In a preferred embodiment, the other putative nuclease domain is a HincII-like endonuclease domain.

[0666] The inactivated Cas13 or Cas13 nickase may have associated (e.g., via fusion protein) one or more functional domains, including for example, an adenosine deaminase or catalytic domain thereof. In some cases it is advantageous that additionally at least one heterologous NLS is provided. In some instances, it is advantageous to position the NLS at the N terminus. In general, the positioning of the one or more functional domain on the inactivated Cas13 or Cas13 nickase is one which allows for correct spatial orientation for the functional domain to affect the target with the attributed functional effect. For example, when the functional domain is an adenosine deaminase catalytic domain thereof, the adenosine deaminase catalytic domain is placed in a spatial orientation which allows it to contact and deaminate a target adenine. This may include positions other than the N- / C- terminus of Cas13. In some embodiments, the adenosine deaminase protein or catalytic domain thereof is inserted into an internal loop of Cas13.

Determination of PAM

[0667] Determination of PAM can be ensured as follows. This experiment closely parallels similar work in *E. coli* for the heterologous expression of StCas9 (Sapranauskas, R. et al. Nucleic Acids Res 39, 9275–9282 (2011)). Applicants introduce a plasmid containing both a

PAM and a resistance gene into the heterologous E. coli, and then plate on the corresponding antibiotic. If there is DNA cleavage of the plasmid, Applicants observe no viable colonies.

[0668] In further detail, the assay is as follows for a DNA target. Two E.coli strains are used in this assay. One carries a plasmid that encodes the endogenous effector protein locus from the bacterial strain. The other strain carries an empty plasmid (e.g.pACYC184, control strain). All possible 7 or 8 bp PAM sequences are presented on an antibiotic resistance plasmid (pUC19 with ampicillin resistance gene). The PAM is located next to the sequence of proto-spacer 1 (the DNA target to the first spacer in the endogenous effector protein locus). Two PAM libraries were cloned. One has a 8 random bp 5' of the proto-spacer (e.g. total of 65536 different PAM sequences = complexity). The other library has 7 random bp 3' of the proto-spacer (e.g. total complexity is 16384 different PAMs). Both libraries were cloned to have in average 500 plasmids per possible PAM. Test strain and control strain were transformed with 5'PAM and 3'PAM library in separate transformations and transformed cells were plated separately on ampicillin plates. Recognition and subsequent cutting/interference with the plasmid renders a cell vulnerable to ampicillin and prevents growth. Approximately 12h after transformation, all colonies formed by the test and control strains where harvested and plasmid DNA was isolated. Plasmid DNA was used as template for PCR amplification and subsequent deep sequencing. Representation of all PAMs in the untransfomed libraries showed the expected representation of PAMs in transformed cells. Representation of all PAMs found in control strains showed the actual representation. Representation of all PAMs in test strain showed which PAMs are not recognized by the enzyme and comparison to the control strain allows extracting the sequence of the depleted PAM.

[0669] The following PAMs have been identified for certain wild-type Cas13 orthologues: the Acidaminococcus sp. BV3L6 Cas13 (AsCas13), Lachnospiraceae bacterium ND2006 Cas13 (LbCas13) and Prevotella albensis (PaCas13) can cleave target sites preceded by a TTTV PAM, where V is A/C or G, FnCas13p, can cleave sites preceded by TTN, where N is A/C/G or T. The Moraxella bovoculi AAX08_00205, Moraxella bovoculi AAX11_00205, Butyrivibrio sp. NC3005, Thiomicrospira sp. XS5, or Lachnospiraceae bacterium MA2020 PAM is 5' TTN, where N is A/C/G or T. The natural PAM sequence is TTTV or BTTV, wherein B is T/C or G and V is A/C or G and the effector protein is Moraxella lacunata Cas13.

Codon optimized nucleic acid sequences

[0670] Where the effector protein is to be administered as a nucleic acid, the application envisages the use of codon-optimized CRISPR-Cas type V protein, and more particularly Cas13-encoding nucleic acid sequences (and optionally protein sequences). An example of a

codon optimized sequence, is in this instance a sequence optimized for expression in a eukaryote, e.g., humans (i.e. being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed; see, e.g., SaCas9 human codon optimized sequence in WO 2014/093622 (PCT/US2013/074667) as an example of a codon optimized sequence (from knowledge in the art and this disclosure, codon optimizing coding nucleic acid molecule(s), especially as to effector protein (e.g., Cas13) is within the ambit of the skilled artisan). Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species other than human, or for codon optimization for specific organs is known. In some embodiments, an enzyme coding sequence encoding a DNA/RNA-targeting Cas protein is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a plant or a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In some embodiments, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g., about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. 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 www.kazusa.orjp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. “Codon usage tabulated from the international DNA sequence databases: status for the year 2000” Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. In some embodiments, one or more codons

(e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a DNA/RNA-targeting Cas protein corresponds to the most frequently used codon for a particular amino acid. As to codon usage in yeast, reference is made to the online Yeast Genome database available at http://www.yeastgenome.org/community/codon_usage.shtml, or Codon selection in yeast, Bennetzen and Hall, J Biol Chem. 1982 Mar 25;257(6):3026-31. As to codon usage in plants including algae, reference is made to Codon usage in higher plants, green algae, and cyanobacteria, Campbell and Gowri, Plant Physiol. 1990 Jan; 92(1): 1–11.; as well as Codon usage in plant genes, Murray et al, Nucleic Acids Res. 1989 Jan 25;17(2):477-98; or Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages, Morton BR, J Mol Evol. 1998 Apr;46(4):449-59.

[0671] In certain example embodiments, the CRISPR Cas protein is selected from Table 1.

Table 1

C2c2 orthologue	Code	Multi Letter
Leptotrichia shahii	C2-2	Lsh
L wadei F0279 (Lw2)	C2-3	Lw2
Listeria seeligeri	C2-4	Lse
Lachnospiraceae bacterium MA2020	C2-5	LbM
Lachnospiraceae bacterium NK4A179	C2-6	LbNK179
[Clostridium] aminophilum DSM 10710	C2-7	Ca
Carnobacterium gallinarum DSM 4847	C2-8	Cg
Carnobacterium gallinarum DSM 4847	C2-9	Cg2
Paludibacter propionicigenes WB4	C2-10	Pp
Listeria weihenstephanensis FSL R9-0317	C2-11	Lwei
Listeriaceae bacterium FSL M6-0635	C2-12	LbFSL
Leptotrichia wadei F0279	C2-13	Lw
Rhodobacter capsulatus SB 1003	C2-14	Rc
Rhodobacter capsulatus R121	C2-15	Rc
Rhodobacter capsulatus DE442	C2-16	Rc

[0672] In certain example embodiments, the CRISPR effector protein is a Cas13a protein selected from Table 2

Table 2

c2c2-5	1	Lachnospiraceae bacterium MA2020 (SEQ ID No. 34)	mqiskvnhkhvavgqkdreritgfiyndpvgdeksledvvakrandtkvlfnvfnt kdlydsqesdksekdkeiiskgakfvaksfnsaitilkkqnkiystltsqqvikelkdk fggariydddieealtetlkksfrkenvrnsikvlienaagirsslskdeeeliqeyfvk qlveeytktklqknvvksiknqnmviqpdsdsqvlslsesrrekqssavssdtlvnc kekdvlkafltdyavldedernsllwklrnlvnlyfygsesirdysytkeksvwkeh deqkanktlfideichitkigkngkeqkvldyeenrsrcrkqninyyrsalnyaknnt sgifenedsnhfwihlieneverlyngiengeefkfetgyisekvwkavinhlsikyi algkavynyamkelsspgdiepgkiddsyingitsfdyeiikaeeslqrdismnvvf atnylacatvdtdkdfllfskedirsctkkdgnlcknimqfwggystwknfceeylk ddkdalellyslksmlysmrnssfhfstenvdngswdteligklfeedcnraarieke kfynnnlhmfysssllekvlerlysshherasqvpsfnrvfvrknfpsslseqritpkft dskdeqiwqsavyylckeiyyndflqskeayklfregvknldkndinnqkaadsfk qavvyygkaignatlsqvcqaimteynrqnndglkkksayaekqnsnkykhyplf lkqvlqsafweyldenkeiygfisaqihksnveikaedfianyssqqykklvdkvk ktpelqkwytlgrlinprqanqflgsirnyvqfvkdiqrrakengnpirnyyevlesd siikilemctklngttsndihdyfrdedeyaeyisqfvnfgdvhsgaalnafcnsese gkkngiyydginpivnrnwvlcklygspdliskiisrvnenmihdfhkqedlireyq ikgicsnkkeqqdlrtfqvlknrvelrdiveyseiinelygqlikwcylrerdlmyfql gfhylclnnasskeadyikinvddrnisgailyqiaamyinglpvyykkddmyval ksgkkasdelnsneqtskkinyflkygnnilgdkkdqlylaglelfenvaeheniiifr neidhfhyfydrdrsmldlysevfdrfftydmklrknvvnmlynilldhnivssfvf etgekkvgrgdsevikpsakirlranngvssdvftykvgskdelkiatlpakneeflln varliyypdmeavsenmvregvvkveksndkkgkisrgsntrssnqskynnksk nrmnysmgsifekmdlkfd
c2c2-6	2	Lachnospiraceae bacterium NK4A179 (SEQ ID No. 35)	mkiskvreenrgakltvnaktavvsenrsqegilyndpsrygksrkndedrdryies rlkssgklyrifnedknkretdelqwflseivkkinrrnglvlsdmlsvddrafekafe kyaelsytnrrnkvsgspafetcgvdaataerlkgiisetnfinriknnidnkvsediid riiakylkkslcrervkrglkkllmnafdlpysdpdidvqrdfidyvledfyhvraks qvsrsiknmnmpvqpegdgkfaitvskggtesgnkrsaekeafkkflsdyaslder vrddmlrrmrrlvvlyfygsddsklsdvnekfdvwedhaarrvdnrefiklplenkl angktdkdaerirkntvkelyrnqnigcyrqavkaveednngryfddkmlnmffi

			hrieygvekiyanlkqvtefkartgylsekiwkdlinyisikyiamgkavynyamd elnasdkkeielgkiseeylsgissfdyelikaeemlqretavyvafaarhlssqtveld sensdflllkpkgtmdkndknklasnnilnflkdketlrdtilqyfgghslwtdfpfdk ylaggkddvdfltdlkdviysmrndsfhyatenhnngkwnkelisamfeheterm tvvmkdkfysnnlpmfyknddlkkllidlykdnverasqvpsfnkvfvrknfpalv rdkdnlgieldlkadadkgenelkfynalyymfkeiyynaflndknvrerfitkatkv adnydrnkernlkdriksagsdekkklreqlqnyiaendfgqriknivqvnpdytla qicqlimteynqqnngcmqkksaarkdinkdsyqhykmlllvnlrkaflefikeny afvlkpykhdlcdkadfvpdfakyvkpyaglisrvagsselqkwyivsrflspaqan hmlgflhsykqyvwdiyrrasetgteinhsiaedkiagvditdvdavidlsvklcgti sseisdyfkddevyaeyissyldfeydggnykdslnrfcnsdavndqkvalyydge hpklnrniilsklygerrflekitdrvsrsdiveyyklkketsqyqtkgifdsedeqkni kkfqemknivefrdlmdyseiadelqgqlinwiylrerdlmnfqlgyhyaclnnds nkqatyvtldyqgkknrkingailyqicamyinglplyyvdkdssewtvsdgkest gakigefyryaksfentsdcyasgleifenisehdnitelrnyiehfryyssfdrsflgiy sevfdrfftydlkyrknvptilynillqhfvnvrfefvsgkkmigidkkdrkiakekec aritirekngvyseqftyklkngtvyvdardkrylqsiirllfypekvnmdemievke kkkpsdnntgkgyskrdrqqdrkeydkykekkkkegnflsgmggninwdeina qlkn
c2c2-7	3	[Clostridium] aminophilum DSM 10710 SEQ ID No. 36)	mkfskvdhtrsavgiqkatdsvhgmlytdpkkqevndldkrfdqlnvkakrlynv fnqskaeedddekrfgkvvkklnrelkdllfhrevsrynsignakynyygiksnpee ivsnlgmveslkgerdpqkviskllyylrkglkpgtdglrmileascglrklsgdeke lkvflqtldedfekktfkknlirsienqnmavqpsnegdpiigitqgrfnsqkneeks aiermmsmyadlnedhredvlrklrrlnvlyfnvdtekteeptlpgevdtnpvfev whdhekgkendrqfatfakiltedretrkkeklavkealndlksairdhnimayrcsi kvteqdkdglffedqrinrfwihhiesaverilasinpeklyklrigylgekvwkdlln ylsikyiavgkavfhfamedlgktgqdielgklsnsvsggltsfdyeqiradetlqrql svevafaannlfravvgqtgkkieqskseeneedfllwkaekiaesikkegegntlks ilqffggasswdlnhfcaaygnessalgyetkfaddlrkaiyslrnetfhfttlnkgsfd wnakligdmfsheaatgiavertrfysnnlpmfyresdlkrimdhlyntyhprasqv psfnsvfvrknfrlflsntlntntsfdtevyqkwesgvyylfkeiyynsflpsgdahhlf feglrrirkeadnlpivgkeakkrnavqdfgrrcdelknlslsaicqmimteyneqnn

			gnrkvkstredkrkpdifqhykmlllrtlqeafaiyirreefkfifdlpktlyvmkpve eflpnwksgmfdslvervkqspdlqrwyvlckflngrllnqlsgvirsyiqfagdiqr rakanhnrlymdntqrveyysnvlevvdfcikgtsrfsnvfsdyfrdedayadyldn ylqfkdekiaevssfaalktfcneeevkagiymdgenpvmqrnivmaklfgpdev lknvvpkvtreeieeyyqlekqiapyrqngyckseedqkkllrfqriknrvefqtitef seiinellgqliswsflrerdllyfqlgfhylclhndtekpaeykeisredgtvirnailhq vaamyvgglpvytladkklaafekgeadcklsiskdtagagkkikdffryskyvlik drmltdqnqkytiylaglelfentdehdnitdvrkyvdhfkyyatsdenamsildlys eihdrfftydmkyqknvanmlenillrhfvlirpefftgskkvgegkkitckaraqiei aengmrsedftyklsdgkknistcmiaardqkylntvarllyypheakksivdtrek knnkktnrgdgtfnkqkgtarkekdngprefndtgfsntpfagfdpfrns
c2c2-8	5	Carnobacterium gallinarum DSM 4847 (SEQ ID No. 37)	mritkvkikldnklyqvtmqkeekygtlklneesrkstaeilrlkkasfnksfhsktin sqkenknatikkngdyisqifeklvgvdtnknirkpkmsltdlkdlpkkdlalfikrk fknddiveiknldlislfynalqkvpgehftdeswadfcqemmpyreyknkfierk iillansieqnkgfsinpetfskrkrvlhqwaievqergdfsildeklsklaeiynfkk mckrvqdelndleksmkkgknpekekeaykkqknfkiktiwkdypykthiglie kikeneelnqfnieigkyfehyfpikkerctedepyylnsetiattvnyqlknalisyl mqigkykqfglenqvldskklqeigiyegfqtkfmdacvfatsslkniiepmrsgdi lgkrefkeaiatssfvnyhhffpyfpfelkgmkdreselipfgeqteakqmqniwal rgsvqqirneifhsfdknqkfnlpqldksnfefdasenstgksqsyietdykflfeaek nqleqffierikssgaleyyplkslekIfakkemkfslgsqvvafapsykklvkkghs yqtategtanylglsyynryelkeesfqaqyyllkliyqyvflpnfsqgnspafretvk ailrinkdearkkmkknkkflrkyafeqvremefketpdqymsylqsemreekvr kaekndkgfeknitmnfekllmqifvkgfdvflttfagkelllsseekviketeislsk kinerektlkasiqvehqlvatnsaisywlfckIldsrhlnelrnemikfkqsrikfnht qhaeliqnllpiveltilsndydekndsqnvdvsayfedkslyetapyvqtddrtrvsf rpilklekyhtkslieallkdnpqfrvaatdiqewmhkreeigelvekrknlhtewae gqqtlgaekreeyrdyckkidrfnwkankvtltylsqlhylitdllgrmvgfsalferd lvyfsrsfselggetyhisdyknlsgvlrlnaevkpikiknikvidneenpykgnepe vkpfldrlhaylenvigikavhgkirnqtahlsvlqlelsmiesmnnlrdlmaydrkl knavtksmikildkhgmilklkidenhknfeieslipkeiihlkdkaiktnqvseeyc qlvlallttnpgnqln

c2c2-9	6	Carnobacterium gallinarum DSM 4847 (SEQ ID No. 38)	mrmtkvkingspvsmnrsklnghlvwngttntvniltkkeqsfaasflnktlvkad qvkgykvlaenifiifeqleksnsekpsvylnnirrlkeaglkrffkskyheeikytse knqsvptklnliplffnavdriqedkfdeknwsyfckemspyldykksylnrkkeil ansiqqnrgfsmptaeepnllskrkqlfqqwamkfqespliqqnnfaveqfnkefa nkinelaavynvdelctaiteklmnfdkdksnktrnfeikklwkqhphnkdkalikl fnqegnealnqfnielgkyfehyfpktgkkesaesyylnpqtiiktvgyqlrnafvqy llqvgklhqynkgvldsqtlqeigmyegfqtkfmdacvfassslrniiqattnediltr ekfkkeleknvelkhdlffkteiveerdenpakkiamtpneldlwairgavqrvrnq ifhqqinkrhepnqlkvgsfengdlgnvsyqktiyqklfdaeikdieiyfaekikssg aleqysmkdleklfsnkeltlslggqvvafapsykklykqgyfyqnektieleqftdy dfsndvfkanyylikliyhyvflpqfsqannklfkdtvhyviqqnkelnttekdkkn nkkirkyafeqvklmknespekymqylqremqeertikeakktneekpnynfek lliqifikgfdtflrnfdlnlnpaeelvgtvkekaeglrkrkeriakilnvdeqiktgdeei afwifaklldarhlselrnemikfkqssvkkglikngdlieqmqpilelcilsndses mekesfdkievflekvelaknepymqedkltpvkfrfmkqlekyqtrnfienlvie npefkvsekivlnwheekekiadlvdkrtklheewaskareieeynekikknkskk ldkpaefakfaeykiiceaienfnrldhkvrltylknlhylmidlmgrmvgfsvlfer dfvymgrsysalkkqsiylndydtfanirdwevnenkhlfgtsssdltfqetaefknl kkpmenqlkallgvtnhsfeirnniahlhvlrndgkgegvsllscmndlrklmsydr klknavtkaiikildkhgmilkltnndhtkpfeieslkpkkiihleksnhsfpmdqvs qeycdlvkkmlvftn
c2c2-10	7	Paludibacter propionicigenes WB4 (SEQ ID No. 39)	mrvskvkvkdggkdkmvlvhrkttgaqlvysgqpvsnetsnilpekkrqsfdlstl nktiikfdtakkqklnvdqykivekifkypkqelpkqikaeeilpflnhkfqepvky wkngkeesfnltlliveavqaqdkrklqpyydwktwyiqtksdllkksiennridlte nlskrkkallaweteftasgsidlthyhkvymtdvlckmlqdvkpltddkgkintna yhrglkkalqnhqpaifgtrevpneanradnqlsiyhlevvkylehyfpiktskrrnt addiahylkaqtlkttiekqlvnairaniiqqgktnhhelkadttsndliriktneafvln ltgtcafaannirnmvdneqtndilgkgdfiksllkdntnsqlysfffgeglstnkaek etqlwgirgavqqirnnvnhykkdalktvfnisnfenptitdpkqqtnyadtiykarf inelekipeafaqqlktggavsyytienlkslttfqfslcrstipfapgfkkvfngginy qnakqdesfyelmleqylrkenfaeesynaryfmlkliynnlflpgfttdrkafadsv gfvqmqnkkqaekvnprkkeayafeavrpmtaadsiadymayvqselmqeqn

			kkeekvaeetrinfekfvlqvfikgfdsflrakefdfvqmpqpqltatasnqqkadkl nqleasitadckltpqyakaddathiafyvfcklldaahlsnlrnelikfresvnefkfh hlleiieicllsadvvptdyrdlysseadclarlrpfieqgaditnwsdlfvqsdkhspvi hanielsvkygttklleqiinkdtqfktteanftawntaqksieqlikqredhheqwvk aknaddkekqerkreksnfaqkfiekhgddyldicdyintynwldnkmhfvhlnr lhgltiellgrmagfvalfdrdfqffdeqqiadefklhgfvnlhsidkklnevptkkike iydirnkiiqingnkinesvranliqfisskrnyynnaflhvsndeikekqmydirnh iahfnyltkdaadfslidlinelrellhydrklknavskafidlfdkhgmilklklnadh klkveslepkkiyhlgssakdkpeyqyctnqvmmaycnmcrsllemkk
c2c2-11	9	Listeria weihenstephanensis FSL R9-0317 (SEQ ID No. 40)	mlallhqevpsqklhnlkslntesltklfkpkfqnmisyppskgaehvqfcltdiavp airdldeikpdwgiffeklkpytdwaesyihykqttiqksieqnkiqspdsprklvlq kyvtaflngeplgldlvakkykladlaesfkvvdlnedksanykikaclqqhqrnild elkedpelnqygievkkyiqryfpikrapnrskharadflkkeliestveqqfknavy hyvleqgkmeayeltdpktkdlqdirsgeafsfkfinacafasnnlkmilnpecekd ilgkgdfkknlpnsttqsdvvkkmipffsdeiqnvnfdeaiwairgsiqqirnevyh ckkhswksilkikgfefepnnmkytdsdmqklmdkdiakipdfieeklkssgiirf yshdklqsiwemkqgfsllttnapfvpsfkrvyakghdyqtsknryydlglttfdile ygeedfraryfltklvyyqqfmpwftadnnafrdaanfvlrlnknrqqdakafinire veegemprdymgyvqgqiaihedstedtpnhfekfisqvfikgfdshmrsadlkfi knprnqgleqseieemsfdikvepsflknkddyiafwtfckmldarhlselrnemi kydghltgeqeiiglallgvdsrendwkqffssereyekimkgyvgeelyqrepyrq sdgktpilfrgveqarkygtetviqrlfdaspefkvskcnitewerqketieetierrkel hneweknpkkpqnnaffkeykeccdaidaynwhknkttlvyvnelhhllieilgry vgyvaiadrdfqcmanqyfkhsgiterveywgdnrlksikkldtflkkeglfvsekn arnhiahlnylslksectllylserlreifkydrklknavskslidildrhgmsvvfanlk enkhrlvikslepkklrhlgekkidngyietnqvseeycgivkrllei
c2c2-12	10	Listeriaceae bacterium FSL M6-0635 = Listeria newyorken	mkitkmrvdgrtivmertskegqlgyegidgnktteiifdkkkesfyksilnktvrkp dekeknrrkqainkainkeitelmlavlhqevpsqklhnlkslntesltklfkpkfqn misyppskgaehvqfcltdiavpairdldeikpdwgiffeklkpytdwaesyihyk qttiqksieqnkiqspdsprklvlqkyvtaflngeplgldlvakkykladlaesfklvdl nedksanykikaclqqhqrnildelkedpelnqygievkkyiqryfpikrapnrskh aradflkkeliestveqqfknavyhyvleqgkmeayeltdpktkdlqdirsgeafsfk

		sis FSL M6-0635 (SEQ ID No. 41)	finacafasnnlkmilnpecekdilgkgnfkknlpnsttrsdvvkkmipffsdelqn vnfdeaiwairgsiqqirnevyhckkhswksilkikgfefepnnmkyadsdmqkl mdkdiakipefieeklkssgvvrfyrhdelqsiwemkqgfsllttnapfvpsfkrvy akghdyqtsknryynldlttfdileygeedfraryfltklvyyqqfmpwftadnnafr daanfvlrlnknrqqdakafinireveegemprdymgyvqgqiaihedsiedtpnh fekfisqvfikgfdrhmrsanlkfiknprnqgleqseieemsfdikvepsflknkdd yiafwifckmldarhlselrnemikydghltgeqeiiglallgvdsrendwkqffsse reyekimkgyvveelyqrepyrqsdgktpilfrgveqarkygtetviqrlfdanpefk vskcnlaewerqketieetikrrkelhnewaknpkkpqnnaffkeykeccdaiday nwhknkttlayvnelhhllieilgryvgyvaiadrdfqcmanqyfkhsgiterveyw gdnrlksikkldtflkkeglfvseknarnhiahlnylslksectllylserlreifkydrkl knavskslidildrhgmsvvfanlkenkhrlvikslepkklrhlggkkidggyietnq vseeycgivkrllem
c2c2-13	12	Leptotrichia wadei F0279 (SEQ ID No. 42)	mkvtkvdgishkkyieegklvkstseenrtserlsellsirldiyiknpdnaseeenrir renlkkffsnkvlhlkdsvlylknrkeknavqdknyseediseydlknknsfsvlkk illnedvnseeleifrkdveaklnkinslkysfeenkanyqkinennvekvggkskr niiydyyresakrndyinnvqeafdklykkedieklfflienskkhekykireyyhki igrkndkenfakiiyeeiqnvnnikeliekipdmselkksqvfykyyldkeelndkn ikyafchfveiemsqllknyvykrlsnisndkikrifeyqnlkklienkllnkldtyvr ncgkynyylqvgeiatsdfiarnrqneaflrniigvssvayfslrniletenenditgrm rgktvknnkgeekyvsgevdkiynenkqnevkenlkmfysydfnmdnkneied ffanideaissirhgivhfnlelegkdifafkniapseiskkmfqneinekklklkifkq lnsanvfnyyekdviikylkntkfnfvnknipfvpsftklynkiedlrntlkffwsvp kdkeekdaqiyllkniyygeflnkfvknskvffkitnevikinkqrnqktghykyqk feniektvpveylaiiqsreminnqdkeekntyidfiqqiflkgfidylnknnlkyies nnnndnndifskikikkdnkekydkilknyekhnrnkeipheinefvreiklgkilk ytenlnmfylilkllnhkeltnlkgslekyqsankeetfsdelelinllnldnnrvtedfe leaneigkfldfnenkikdrkelkkfdtnkiyfdgeniikhrafynikkygmlnlleki adkakykislkelkeysnkkneieknytmqqnlhrkyarpkkdekfndedykeye kaigniqkythlknkvefnelnllqglllkilhrlvgytsiwerdlrfrlkgefpenhyie eifnfdnsknvkyksgqivekyinfykelykdnvekrsiysdkkvkklkqekkdly

			irnyiahfnyiphaeisllevlenlrkllsydrklknaimksivdilkeygfvatfkiga dkkieiqtlesekivhlknlkkkklmtdrnseelcelvkvmfeykale
c2c2-14	15	Rhodobacter capsulatus SB 1003 (SEQ ID No. 43)	mqigkvqgrtisefgdpagglkrkistdgknrkelpahlssdpkaligqwisgidkiy rkpdsrksdgkaihsptpskmqfdarddlgeafwklvseaglaqdsdydqfkrrlh pygdkfqpadsgaklkfeadppepqafhgrwygamskrgndakelaaalyehlh vdekridgqpkrnpktdkfapglvvaralgiessvlprgmarlarnwgeeeiqtyfv vdvaasvkevakaavsaaqafdpprqvsgrslspkvgfalaehlervtgskrcsfdp aagpsvlalhdevkktykrlcargknaarafpadktellalmrhthenrvrnqmvr mgrvseyrgqqagdlaqshywtsagqteikeseifvrlwvgafalagrsmkawid pmgkivntekndrdltaavnirqvisnkemvaeamarrgiyfgetpeldrlgaegn egfvfallrylrgcrnqtfhlgaragflkeirkelektrwgkakeaehvvltdktvaair aiidndakalgarlladlsgafvahyaskehfstlyseivkavkdapevssglprlklll kradgvrgyvhglrdtrkhafatklppppaprelddpatkaryiallrlydgpfrayas gitgtalagpaarakeaatalaqsvnvtkaysdvmegrtsrlrppndgetlreylsaltg etatefrvqigyesdsenarkqaefienyrrdmlafmfedyirakgfdwilkiepgat amtrapvlpepidtrgqyehwqaalylvmhfvpasdvsnllhqlrkwealqgkyel vqdgdatdqadarrealdlvkrfrdvlvlflktgearfegraapfdlkpfralfanpatf drlfmatpttarpaeddpegdgasepelrvartlrglrqiarynhmavlsdlfakhkvr deevarlaeiedetqeksqivaaqelrtdlhdkvmkchpktispeerqsyaaaiktie ehrflvgrvylgdhlrlhrlmmdvigrlidyagayerdtgtflinaskqlgagadwav tiagaantdartqtrkdlahfnvldradgtpdltalvnraremmaydrkrknavprsil dmlarlgltlkwqmkdhllqdatitqaaikhldkvrltvggpaavtearfsqdylqm vaavfngsvqnpkprrrddgdawhkppkpataqsqpdqkppnkapsagsrlppp qvgevyegvvvkvidtgslgflavegvagniglhisrlrriredaiivgrryrfrveiyv ppksntsklnaadlvrid
c2c2-15	16	Rhodobacter capsulatus R121 (SEQ ID No. 44)	mqigkvqgrtisefgdpagglkrkistdgknrkelpahlssdpkaligqwisgidkiy rkpdsrksdgkaihsptpskmqfdarddlgeafwklvseaglaqdsdydqfkrrlh pygdkfqpadsgaklkfeadppepqafhgrwygamskrgndakelaaalyehlh vdekridgqpkrnpktdkfapglvvaralgiessvlprgmarlarnwgeeeiqtyfv vdvaasvkevakaavsaaqafdpprqvsgrslspkvgfalaehlervtgskrcsfdp aagpsvlalhdevkktykrlcargknaarafpadktellalmrhthenrvrnqmvr mgrvseyrgqqagdlaqshywtsagqteikeseifvrlwvgafalagrsmkawid

			pmgkivntekndrdltaavnirqvisnkemvaeamarrgiyfgetpeldrlgaegn egfvfallrylrgcrnqtfhlgaragflkeirkelektrwgkakeaehvvltdktvaair aiidndakalgarlladlsgafvahyaskehfstlyseivkavkdapevssglprlklll kradgvrgyvhglrdtrkhafatklppppaprelddpatkaryiallrlydgpfrayas gitgtalagpaarakeaatalaqsvnvtkaysdvmegrssrlrppndgetlreylsalt getatefrvqigyesdsenarkqaefienyrrdmlafmfedyirakgfdwilkiepga tamtrapvlpepidtrgqyehwqaalylvmhfvpasdvsnllhqlrkwealqgkye lvqdgdatdqadarrealdlvkrfrdvlvlflktgearfegraapfdlkpfralfanpatf drlfmatpttarpaeddpegdgasepelrvartlrglrqiarynhmavlsdlfakhkvr deevarlaeiedetqeksqivaaqelrtdlhdkvmkchpktispeerqsyaaaiktie ehrflvgrvylgdhlrlhrlmmdvigrlidyagayerdtgtflinaskqlgagadwav tiagaantdartqtrkdlahfnvldradgtpdltalvnraremmaydrkrknavprsil dmlarlgltlkwqmkdhllqdatitqaaikhldkvrltvggpaavtearfsqdylqm vaavfngsvqnpkprrrddgdawhkppkpataqsqpdqkppnkapsagsrlppp qvgevyegvvvkvidtgslgflavegvagniglhisrlrriredaiivgrryrfrveiyv ppksntsklnaadlvrid
c2c2-16	17	Rhodobacter capsulatus DE442 (SEQ ID No. 45)	mqigkvqgrtisefgdpagglkrkistdgknrkelpahlssdpkaligqwisgidkiy rkpdsrksdgkaihsptpskmqfdarddlgeafwklvseaglaqdsdydqfkrrlh pygdkfqpadsgaklkfeadppepqafhgrwygamskrgndakelaaalyehlh vdekridgqpkrnpktdkfapglvvaralgiessvlprgmarlarnwgeeeiqtyfv vdvaasvkevakaavsaaqafdpprqvsgrslspkvgfalaehlervtgskrcsfdp aagpsvlalhdevkktykrlcargknaarafpadktellalmrhthenrvrnqmvr mgrvseyrgqqagdlaqshywtsagqteikeseifvrlwvgafalagrsmkawid pmgkivntekndrdltaavnirqvisnkemvaeamarrgiyfgetpeldrlgaegn egfvfallrylrgcrnqtfhlgaragflkeirkelektrwgkakeaehvvltdktvaair aiidndakalgarlladlsgafvahyaskehfstlyseivkavkdapevssglprlklll kradgvrgyvhglrdtrkhafatklppppaprelddpatkaryiallrlydgpfrayas gitgtalagpaarakeaatalaqsvnvtkaysdvmegrssrlrppndgetlreylsalt getatefrvqigyesdsenarkqaefienyrrdmlafmfedyirakgfdwilkiepga tamtrapvlpepidtrgqyehwqaalylvmhfvpasdvsnllhqlrkwealqgkye lvqdgdatdqadarrealdlvkrfrdvlvlflktgearfegraapfdlkpfralfanpatf drlfmatpttarpaeddpegdgasepelrvartlrglrqiarynhmavlsdlfakhkvr

			deevarlaeiedetqeksqivaaqelrtdlhdkvmkchpktispeerqsyaaaiktie ehrflvgrvylgdhlrlhrlmmdvigrlidyagayerdtgtflinaskqlgagadwav tiagaantdartqtrkdlahfnvldradgtpdltalvnraremmaydrkrknavprsil dmlarlgltlkwqmkdhllqdatitqaaikhldkvrltvggpaavtearfsqdylqm vaavfngsvqnpkprrrddgdawhkppkpataqsqpdqkppnkapsagsrlppp qvgevyegvvvkvidtgslgflavegvagniglhisrlrriredaiivgrryrfrveiyv ppksntsklnaadlvrid
c2c2-2		(SEQ ID No. 46)	mgnlfghkrwyevrdkkdfkikrkvkvkrnydgnkyilninennnkekidnnkfi rkyinykkndnilkeftrkfhagnilfklkgkegiiriennddfleteevvlyieaygks eklkalgitkkkiideairqgitkddkkieikrqeneeeieidirdeytnktlndcsiilri iendeletkksiyeifkninmslykiiekiienetekvfenryyeehlrekllkddkid viltnfmeirekiksnleilgfvkfylnvggdkkksknkkmlvekilninvdltvedi adfvikelefwnitkriekvkkvnneflekrrnrtyiksyvlldkhekfkierenkkd kivkffveniknnsikekiekilaefkidelikklekelkkgncdteifgifkkhykvn fdskkfskksdeekelykiiyrylkgriekilvneqkvrlkkmekieiekilnesilse kilkrvkqytlehimylgklrhndidmttvntddfsrlhakeeldlelitffastnmeln kifsreninndenidffggdreknyvldkkilnskikiirdldfidnknnitnnfirkftk igtnernrilhaiskerdlqgtqddynkviniiqnlkisdeevskalnldvvfkdkknii tkindikiseennndikylpsfskvlpeilnlyrnnpknepfdtietekivlnaliyvnk elykklileddleeneskniflqelkktlgnideideniienyyknaqisaskgnnkai kkyqkkviecyigylrknyeelfdfsdfkmniqeikkqikdindnktyeritvktsd ktivinddfeyiisifallnsnavinkirnrffatsvwlntseyqniidildeimqlntlrn ecitenwnlnleefiqkmkeiekdfddfkiqtkkeifnnyyediknniltefkdding cdvlekklekivifddetkfeidkksnilqdeqrklsninkkdlkkkvdqyikdkdq eikskilcriifnsdflkkykkeidnliedmesenenkfqeiyypkerknelyiykkn lflnignpnfdkiyglisndikmadakflfnidgknirknkiseidailknlndklngy skeykekyikklkenddffakniqnknyksfekdynrvseykkirdlvefnylnki esylidinwklaiqmarferdmhyivnglrelgiiklsgyntgisraypkrngsdgfy tttayykffdeesykkfekicygfgidlsenseinkpenesirnyishfyivrnpfady siaeqidrvsnllsystrynnstyasvfevfkkdvnldydelkkkfklignndilerlm kpkkvsvlelesynsdyiknliiellkientndtl

c2c2-3		L wadei (Lw2) (SEQ ID No. 47)	mkvtkvdgishkkyieegklvkstseenrtserlsellsirldiyiknpdnaseeenrir renlkkffsnkvlhlkdsvlylknrkeknavqdknyseediseydlknknsfsvlkk illnedvnseeleifrkdveaklnkinslkysfeenkanyqkinennvekvggkskr niiydyyresakrndyinnvqeafdklykkedieklfflienskkhekykireyyhki igrkndkenfakiiyeeiqnvnnikeliekipdmselkksqvfykyyldkeelndkn ikyafchfveiemsqllknyvykrlsnisndkikrifeyqnlkklienkllnkldtyvr ncgkynyylqvgeiatsdfiarnrqneaflrniigvssvayfslrniletenenditgrm rgktvknnkgeekyvsgevdkiynenkqnevkenlkmfysydfnmdnkneied ffanideaissirhgivhfnlelegkdifafkniapseiskkmfqneinekklklkifkq lnsanvfnyyekdviikylkntkfnfvnknipfvpsftklynkiedlrntlkffwsvp kdkeekdaqiyllkniyygeflnkfvknskvffkitnevikinkqrnqktghykyqk feniektvpveylaiiqsreminnqdkeekntyidfiqqiflkgfidylnknnlkyies nnnndnndifskikikkdnkekydkilknyekhnrnkeipheinefvreiklgkilk ytenlnmfylilkllnhkeltnlkgslekyqsankeetfsdelelinllnldnnrvtedfe leaneigkfldfnenkikdrkelkkfdtnkiyfdgeniikhrafynikkygmlnlleki adkakykislkelkeysnkkneieknytmqqnlhrkyarpkkdekfndedykeye kaigniqkythlknkvefnelnllqglllkilhrlvgytsiwerdlrfrlkgefpenhyie eifnfdnsknvkyksgqivekyinfykelykdnvekrsiysdkkvkklkqekkdly irnyiahfnyiphaeisllevlenlrkllsydrklknaimksivdilkeygfvatfkiga dkkieiqtlesekivhlknlkkkklmtdrnseelcelvkvmfeykalekrpaatkka gqakkkkgsypydvpdyaypydvpdyaypydvpdya*
c2c2-4		Listeria seeligeri (SEQ ID No. 48)	mwisiktlihhlgvlffcdymynrrekkiievktmritkvevdrkkvlisrdknggkl vyenemqdnteqimhhkkssfyksvvnkticrpeqkqmkklvhgllqensqeki kvsdvtklnisnflnhrfkkslyyfpenspdkseeyrieinlsqlledslkkqqgtfic wesfskdmelyinwaenyissktklikksirnnriqstesrsgqlmdrymkdilnkn kpfdiqsvsekyqlekltsalkatfkeakkndkeinyklkstlqnherqiieelkense lnqfnieirkhletyfpikktnrkvgdirnleigeiqkivnhrlknkivqrilqegklasy eiestvnsnslqkikieeafalkfinaclfasnnlrnmvypvckkdilmigefknsfk eikhkkfirqwsqffsqeitvddielaswglrgaiapirneiihlkkhswkkffnnptf kvkkskiingktkdvtseflyketlfkdyfyseldsvpeliinkmesskildyyssdql nqvftipnfelslltsavpfapsfkrvylkgfdyqnqdeaqpdynlklniynekafns eafqaqyslfkmvyyqvflpqfttnndlfkssvdfiltlnkerkgyakafqdirkmn

			kdekpseymsyiqsqlmlyqkkqeekekinhfekfinqvfikgfnsfieknrltyic hptkntvpendnieipfhtdmddsniafwlmcklldakqlselrnemikfscslqst eeistftkareviglallngekgcndwkelfddkeawkknmslyvseellqslpytq edgqtpvinrsidlvkkygtetileklfsssddykvsakdiaklheydvtekiaqqesl hkqwiekpglardsawtkkyqnvindisnyqwaktkveltqvrhlhqltidllsrla gymsiadrdfqfssnyilerenseyrvtswillsenknknkyndyelynlknasikv sskndpqlkvdlkqlrltleylelfdnrlkekrnnishfnylngqlgnsilelfddardvl sydrklknavskslkeilsshgmevtfkplyqtnhhlkidklqpkkihhlgekstvss nqvsneycqlvrtlltmk

C2-17		Leptotrichia buccalis C-1013-b (SEQ ID No. 49)	mkvtkvggishkkytsegrlvkseseenrtderlsallnmrldmyiknpsstetken qkrigklkkffsnkmvylkdntlslkngkkenidreysetdilesdvrdkknfavlkk iylnenvnseelevfrndikkklnkinslkysfeknkanyqkinenniekvegkskr niiydyyresakrdayvsnvkeafdklykeediaklvleienltklekykirefyheii grkndkenfakiiyeeiqnvnnmkeliekvpdmselkksqvfykyyldkeelndk nikyafchfveiemsqllknyvykrlsnisndkikrifeyqnlkklienkllnkldtyv rncgkynyylqdgeiatsdfiarnrqneaflrniigvssvayfslrniletenenditgr mrgktvknnkgeekyvsgevdkiynenkknevkenlkmfysydfnmdnknei edffanideaissirhgivhfnlelegkdifafkniapseiskkmfqneinekklklkif rqlnsanvfrylekykilnylkrtrfefvnknipfvpsftklysriddlknslgiywktp ktnddnktkeiidaqiyllkniyygeflnyfmsnngnffeiskeiielnkndkrnlktg fyklqkfediqekipkeylaniqslyminagnqdeeekdtyidfiqkiflkgfmtyla nngrlsliyigsdeetntslaekkqefdkflkkyeqnnnikipyeineflreiklgnilk yterlnmfylilkllnhkeltnlkgslekyqsankeeafsdqlelinllnldnnrvtedfe leadeigkfldfngnkvkdnkelkkfdtnkiyfdgeniikhrafynikkygmlnlle kiadkagykisieelkkysnkkneieknhkmqenlhrkyarprkdekftdedyesy kqaienieeythlknkvefnelnllqglllrilhrlvgytsiwerdlrfrlkgefpenqyi eeifnfenkknvkykggqivekyikfykelhqndevkinkyssanikvlkqekkdl yirnyiahfnyiphaeisllevlenlrkllsydrklknavmksvvdilkeygfvatfki gadkkigiqtlesekivhlknlkkkklmtdrnseelcklvkimfeykmeekksen
C2-18		Herbinix hemicellulo	mkltrrrisgnsvdqkitaafyrdmsqgllyydsedndctdkviesmdferswrgril kngeddknpfymfvkglvgsndkivcepidvdsdpdnldilinknltgfgrnlkap

		silytica (SEQ ID No. 50)	dsndtlenlirkiqagipeeevlpelkkikemiqkdivnrkeqllksiknnripfslegs klvpstkkmkwlfklidvpnktfnekmlekyweiydydklkanitnrldktdkkar sisravseelreyhknlrtnynrfvsgdrpaagldnggsakynpdkeefllflkeveq yfkkyfpvkskhsnkskdkslvdkyknycsykvvkkevnrsiinqlvagliqqgkl lyyfyyndtwqedflnsyglsyiqveeafkksvmtslswginrltsffiddsntvkfd dittkkakeaiesnyfnklrtcsrmqdhfkeklaffypvyvkdkkdrpdddienlivl vknaiesvsylrnrtfhfkessllellkelddknsgqnkidysvaaefikrdienlydvf reqirslgiaeyykadmisdcfktcglefalyspknslmpafknvykrganlnkayir dkgpketgdqgqnsykaleeyreltwyievknndqsynayknllqliyyhaflpev renealitdfinrtkewnrketeerlntknnkkhknfdendditvntyryesipdyqg eslddylkvlqrkqmarakevnekeegnnnyiqfirdvvvwafgaylenklknyk nelqpplskeniglndtlkelfpeekvkspfnikcrfsistfidnkgkstdntsaeavkt dgkedekdkknikrkdllcfylflrlldeneicklqhqfikyrcslkerrfpgnrtklek etellaeleelmelvrftmpsipeisakaesgydtmikkyfkdfiekkvfknpktsnl yyhsdsktpvtrkymallmrsaplhlykdifkgyylitkkecleyiklsniikdyqns lnelheqleriklksekqngkdslyldkkdfykvkeyvenleqvarykhlqhkinfe slyrifrihvdiaarmvgytqdwerdmhflfkalvyngvleerrfeaifnnnddnnd grivkkiqnnlnnknrelvsmlcwnkklnknefgaiiwkrnpiahlnhftqteqns kssleslinslrillaydrkrqnavtktindlllndyhirikwegrvdegqiyfnikeked ienepiihlkhlhkkdcyiyknsymfdkqkewicngikeevydksilkcignlfkf dyedknkssanpkht
C2-19		[Eubacterium] rectale (SEQ ID No. 51)	mlrrdkevkklynvfnqiqvgtkpkkwnndeklspeenerraqqknikmknyk wreacskyvessqriindvifysyrkaknklrymrknedilkkmqeaeklskfsgg kledfvaytlrkslvvskydtqefdslaamvvflecigknnisdhereivckllelirkd fskldpnvkgsqganivrsvrnqnmivqpqgdrflfpqvyakenetvtnknveke glnefllnyanlddekraeslrklrrildvyfsapnhyekdmditlsdniekekfnvw ekhecgkketglfvdipdvlmeaeaenikldavvekrerkvlndrvrkqniicyryt ravvekynsneplffennainqywihhienaverilknckagklfklrkgylaekv wkdainlisikyialgkavynfalddiwkdkknkelgivderirngitsfdyemika henlqrelavdiafsvnnlaravcdmsnlgnkesdfllwkrndiadklknkddmas vsavlqffggksswdinifkdaykgkkkynyevrfiddlrkaiycarnenfhfktal vndekwntelfgkiferetefclnvekdrfysnnlymfyqvselrnmldhlysrsvsr

			aaqvpsynsvivrtafpeyitnvlgyqkpsydadtlgkwysacyyllkeiyynsflq sdralqlfeksvktlswddkkqqravdnfkdhfsdiksactslaqvcqiymteynqq nnqikkvrssndsifdqpvyqhykvllkkaianafadylknnkdlfgfigkpfkane ireidkeqflpdwtsrkyealcievsgsqelqkwyivgkflnarslnlmvgsmrsyi qyvtdikrraasignelhvsvhdvekvekwvqvievcsllasrtsnqfedyfndkdd yarylksyvdfsnvdmpseysalvdfsneeqsdlyvdpknpkvnrnivhsklfaa dhilrdivepvskdnieefysqkaeiayckikgkeitaeeqkavlkyqklknrvelrd iveygeiinellgqlinwsfmrerdllyfqlgfhydclrndskkpegyknikvdensi kdailyqiigmyvngvtvyapekdgdklkeqcvkggvgvkvsafhryskylglne ktlynagleifevvaehediinlrngidhfkyylgdyrsmlsiysevfdrfftydikyq knvlnllqnillrhnvivepilesgfktigeqtkpgaklsirsiksdtfqykvkggtlitd akderyletirkilyyaeneednlkksvvvtnadkyeknkesddqnkqkekknkd nkgkkneetksdaeknnnerlsynpfanlnfklsn
C2-20		Eubacteriaceae bacterium CHKCI004 (SEQ ID No. 52)	mkiskeshkrtavavmedrvggvvyvpggsgidlsnnlkkrsmdtkslynvfnqi qagtapseyewkdylseaenkkreaqkmiqkanyelrrecedyakkanlavsriif skkpkkifsdddiishmkkqrlskfkgrmedfvlialrkslvvstynqevfdsrkaat vflknigkknisadderqikqlmaliredydkwnpdkdssdkkessgtkvirsiehq nmviqpeknklslskisnvgkktktkqkekagldaflkeyaqidensrmeylkklrr lldtyfaapssyikgaavslpeninfsselnvwerheaakkvninfveipesllnaeq nnnkinkveqehsleqlrtdirrrnitcyhfanalaaderyhtlffenmamnqfwihh menaverilkkcnvgtlfklrigylsekvwkdmlnllsikyialgkavyhfalddiw kadiwkdasdknsgkindltlkgissfdyemvkaqedlqremavgvafstnnlarv tckmddlsdaesdfllwnkeairrhvkytekgeilsailqffggrslwdeslfekaysd snyelkflddlkraiyaarnetfhfktaaidggswntrlfgslfekeaglclnveknkfy snnlvlfykqedlrvfldklygkecsraaqipsyntilprksfsdfmkqllglkepvyg saildqwysacyylfkevyynlflqdssakalfekavkalkgadkkqekavesfrkr yweisknaslaeicqsyiteynqqnnkerkvrsandgmfnepiyqhykmllkeal kmafasyikndkelkfvykpteklfevsqdnflpnwnsekyntlisevknspdlqk wyivgkfmnarmlnlllgsmrsylqyvsdiqkraaglgenqlhlsaenvgqvkkw iqvlevclllsvrisdkftdyfkdeeeyasylkeyvdfedsampsdysallafsnegki dlyvdasnpkvnrniiqaklyapdmvlkkvvkkisqdeckefnekkeqimqfkn kgdevsweeqqkileyqklknrvelrdlseygelinellgqlinwsylrerdllyfqlg

			fhysclmneskkpdayktirrgtvsienavlyqiiamyingfpvyapekgelkpqc ktgsagqkirafcqwasmvekkkyelynaglelfevvkehdniidlrnkidhfkyy qgndsilalygeifdrfftydmkyrnnvlnhlqnillrhnviikpiiskdkkevgrgk mkdraaflleevssdrftykvkegerkidaknrlyletvrdilyfpnravndkgedvii cskkaqdlnekkadrdknhdkskdtnqkkegknqeeksenkepysdrmtwkpf agikle
C2-21		Blautia sp. Marseille-P2398 (SEQ ID No. 53)	mkiskvdhvksgidqklssqrgmlykqpqkkyegkqleehvrnlsrkakalyqvf pvsgnskmekelqiinsfiknillrldsgktseeivgyintysvasqisgdhiqelvdq hlkeslrkytcvgdkriyvpdiivallkskfnsetlqydnselkilidfiredylkekqik qivhsiennstplriaeingqkrlipanvdnpkksyifeflkeyaqsdpkgqesllqh mrylillylygpdkitddyceeieawnfgsivmdneqlfseeasmliqdriyvnqqi eegrqskdtakvkknkskyrmlgdkiehsinesvvkhyqeackaveekdipwik yisdhvmsvyssknrvdldklslpylakntwntwisfiamkyvdmgkgvyhfa msdvdkvgkqdnliigqidpkfsdgissfdyerikaeddlhrsmsgyiafavnnfar aicsdefrkknrkedvltvgldeiplydnvkrkllqyfggasnwddsiidiiddkdlv acikenlyvarnvnfhfagsekvqkkqddileeivrketrdigkhyrkvfysnnvav fycdediiklmnhlyqrekpyqaqipsynkvisktylpdlifmllkgknrtkisdpsi mnmfrgtfyfllkeiyyndflqasnlkemfceglknnvknkksekpyqnfmrrfe elenmgmdfgeicqqimtdyeqqnkqkkktatavmsekdkkirtldndtqkykh frtllyiglreafiiylkdeknkewyeflrepvkreqpeekefvnkwklnqysdcseli lkdslaaawyvvahfinqaqlnhligdiknyiqfisdidrrakstgnpvsesteiqier yrkilrvlefakffcgqitnvltdyyqdendfsthvghyvkfekknmepahalqafs nslyacgkekkkagfyydgmnpivnrnitlasmygnkkllenamnpvteqdirk yyslmaeldsvlkngavcksedeqknlrhfqnlknrielvdvltlselvndlvaqlig wvyirerdmmylqlglhyiklyftdsvaedsylrtldleegsiadgavlyqiaslysfn lpmyvkpnkssvyckkhvnsvatkfdifekeycngdetvienglrlfeninlhkdm vkfrdylahfkyfakldesilelyskaydfffsyniklkksvsyvltnvllsyfinaklsf stykssgnktvqhrttkisvvaqtdyftyklrsivknkngvesienddrrcevvniaar dkefvdevcnvinynsdk
C2-22		Leptotrichia sp. oral taxon 879	mgnlfghkrwyevrdkkdfkikrkvkvkrnydgnkyilninennnkekidnnkfi gefvnykknnnvlkefkrkfhagnilfklkgkeeiirienndfleteevvlyievyg kseklkaleitkkkiideairqgitkddkkieikrqeneeeieidirdeytnktlndcsiil

		str. F0557 (SEQ ID No. 54)	riiendeletkksiyeifkninmslykiiekiienetekvfenryyeehlrekllkdnki dviltnfmeirekiksnleimgfvkfylnvsgdkkksenkkmfvekilntnvdltve divdfivkelkfwnitkriekvkkfnneflenrrnrtyiksyvlldkhekfkierenkk dkivkffveniknnsikekiekilaefkinelikklekelkkgncdteifgifkkhykv nfdskkfsnksdeekelykiiyrylkgriekilvneqkvrlkkmekieiekilnesils ekilkrvkqytlehimylgklrhndivkmtvntddfsrlhakeeldlelitffastnme lnkifngkekvtdffgfnlngqkitlkekvpsfklnilkklnfinnennideklshfysf qkegyllrnkilhnsygniqetknlkgeyenveklikelkvsdeeiskslsldvifegk vdiinkinslkigeykdkkylpsfskivleitrkfreinkdklfdiesekiilnavkyvn kilyekitsneeneflktlpdklvkksnnkkenknllsieeyyknaqvssskgdkkai kkyqnkvtnayleylentfteiidfskfnlnydeiktkieerkdnkskiiidsistninit ndieyiisifallnsntyinkirnrffatsvwlekqngtkeydyeniisildevllinllre nnitdildlknaiidakivendetyiknyifesneeklkkrlfceelvdkedirkifede nfkfksfikkneignfkinfgilsnlecnseveakkiigknskklesfiqniideyksni rtlfsseflekykeeidnlvedtesenknkfekiyypkehknelyiykknlflnignpn fdkiygliskdiknvdtkilfdddikknkiseidailknlndklngysndykakyvnk lkenddffakniqnenyssfgefekdynkvseykkirdlvefnylnkiesylidinw klaiqmarferdmhyivnglrelgiiklsgyntgisraypkrngsdgfytttayykffd eesykkfekicygfgidlsenseinkpenesirnyishfyivrnpfadysiaeqidrvs nllsystrynnstyasvfevfkkdvnldydelkkkfrlignndilerlmkpkkvsvlel esynsdyiknliielltkientndtl
C2-23		Lachnospiraceae bacterium NK4A144 (SEQ ID No. 55)	mkiskvdhtrmavakgnqhrrdeisgilykdptktgsidfderfkklncsakilyhv fngiaegsnkyknivdkvnnnldrvlftgksydrksiididtvlrnvekinafdriste ereqiiddlleiqlrkglrkgkaglrevlligagvivrtdkkqeiadfleildedfnktnq akniklsienqglvvspvsrgeerifdvsgaqkgksskkaqekealsaflldyadldk nvrfeylrkirrlinlyfyvknddvmslteipaevnlekdfdiwrdheqrkeengdfv gcpdilladrdvkksnskqvkiaerqlresireknikryrfsiktiekddgtyffankqi svfwihrienaverilgsindkklyrlrlgylgekvwkdilnflsikyiavgkavfnfa mddlqekdrdiepgkisenavngltsfdyeqikademlqrevavnvafaannlarv tvdipqngekedillwnksdikkykknskkgilksilqffggastwnmkmfeiayh dqpgdyeenylydiiqiiyslrnksfhfktydhgdknwnreligkmiehdaervisv erekfhsnnlpmfykdadlkkildllysdyagrasqvpafntvlvrknfpeflrkdm

			gykvhfnnpevenqwhsavyylykeiyynlflrdkevknlfytslknirsevsdkk qklasddfasrceeiedrslpeicqiimteynaqnfgnrkvksqrvieknkdifrhyk mlliktlagafslylkqerfafigkatpipyettdvknflpewksgmyasfveeiknnl dlqewyivgrflngrmlnqlagslrsyiqyaedierraaenrnklfskpdekieackk avrvldlcikistrisaeftdyfdseddyadylekylkyqddaikelsgssyaaldhfc nkddlkfdiyvnagqkpilqrnivmaklfgpdnilsevmekvtesaireyydylkk vsgyrvrgkcstekeqedllkfqrlknavefrdvteyaevinellgqliswsylrerdll yfqlgfhymclknksfkpaeyvdirrnngtiihnailyqivsmyingldfyscdkeg ktlkpietgkgvgskigqfikysqylyndpsykleiynaglevfenidehdnitdlrk yvdhfkyyaygnkmslldlyseffdrfftydmkyqknvvnvlenillrhfvifypkf gsgkkdvgirdckkeraqieiseqsltsedfmfklddkageeakkfparderylqtia kllyypneiedmnrfmkkgetinkkvqfnrkkkitrkqknnssnevlsstmgylfk nikl
C2-24		Chloroflexus aggregans (SEQ ID No. 56)	mtdqvrreevaageladtplaaaqtpaadaavaatpapaeavaptpeqavdqpattg eseapvttaqaaaheaepaeatgasftpvseqqpqkprrlkdlqpgmelegkvtsial ygifvdvgvgrdglvhisemsdrridtpselvqigdtvkvwvksvdldarrisltml npsrgekprrsrqsqpaqpqprrqevdreklaslkvgeivegvitgfapfgafadigv gkdglihiselsegrvekpedavkvgeryqfkvleidgegtrislslrraqrtqrmqql epgqiiegtvsgiatfgafvdigvgrdglvhisalaphrvakvedvvkvgdkvkvk vlgvdpqskrisltmrleeeqpattagdeaaepaeevtptrrgnlerfaaaaqtarerse rgersergerrerrerrpaqsspdtyivgedddesfegnatiedlltkfggsssrrdrdrr rrhedddeemerpsnrrqreairrtlqqigyde
C2-25		Demequina aurantiaca (SEQ ID No. 57)	mdltwhallilfivallagfldtlaggggllvpallltgipplqalgtnklqssfgtgmat yqvirkkrvhwrdvrwpmvwaflgsaagavavqfidtdalliiipvvlalvaayflf vpkshlpppeprmsdpayeatlvpiigaydgafgpgtgslyalsgvalraktlvqsta iaktlnfatnfaallvfafaghmlwtvgavmiagqligayagshmlfrvnplvlrvli vvmslgmlirvlld
C2-26		Thalassospira sp. TSL5-1	mriikpygrshvegvatqeprrklrlnsspdisrdipgfaqshdaliiaqwisaidkiat kpkpdkkptqaqinlrttlgdaawqhvmaenllpaatdpaireklhliwqskiapw gtarpqaekdgkptpkggwyerfcgvlspeaitqnvarqiakdiydhllhvaakrkg repakqgessnkpgkfkpdrkrglieeraesiaknalrpgshapcpwgpddqatye

		(SEQ ID No. 58)	qagdvagqiyaaardcleekkrrsgnrntssvqylprdlaakilyaqygrvfgpdtti kaaldeqpslfalhkaikdcyhrlindarkrdilrilprnmaalfrlvraqydnrdinali rlgkvihyhaseqgksehhgirdywpsqqdiqnsrfwgsdgqadikrheafsriwr hiialasrtlhdwadphsqkfsgenddilllakdaieddvfkaghyerkcdvlfgaqa slfcgaedfekailkqaitgtgnlrnatfhfkgkvrfekelqeltkdvpvevqsaiaal wqkdaegrtrqiaetlqavlaghflteeqnrhifaaltaamaqpgdvplprlrrvlarh dsicqrgrilplspcpdrakleespaltcqytvlkmlydgpfrawlaqqnstilnhyid stiartdkaardmngrklaqaekdlitsraadlprlsvdekmgdflarltaatatemrv qrgyqsdgenaqkqaafigqfecdvigrafadflnqsgfdfvlklkadtpqpdaaqc dvtaliapddisvsppqawqqvlyfilhlvpvddashllhqirkwqvlegkekpaqi ahdvqsvlmlyldmhdakftggaalhgiekfaeffahaadfravfppqslqdqdrsi prrglreivrfghlpllqhmsgtvqithdnvvawqaartagatgmspiarrqkqreel halavertarfrnadlqnymhalvdvikhrqlsaqvtlsdqvrlhrlmmgvlgrlvd yaglwerdlyfvvlallyhhgatpddvfkgqgkknladgqvvaalkpknrkaaap vgvfddldhygiyqddrqsirnglshfnmlrggkapdlshwvnqtrslvahdrklk navaksviemlaregfdldwgiqtdrgqhilshgkirtrqaqhfqksrlhivkksakp dkndtvkirenlhgdamvervvqlfaaqvqkryditvekrldhlflkpqdqkgkng ihthngwsktekkrrpsrenrkgnhen
C2-27		SAMN04487830_13920 [Pseudobutyrivibrio sp. OR37] (SEQ ID No. 59)	mkfskeshrktavgvtesngiigllykdplnekekiedvvnqranstkrlfnlfgteat skdisraskdlakvvnkaignlkgnkkfnkkeqitkglntkiiveelknvlkdekkli vnkdiideacsrllktsfrtaktkqavkmiltavlientnlskedeafvheyfvkklvne ynktsvkkqipvalsnqnmviqpnsvngtleisetkksketkttekdafraflrdyatl denrrhkmrlclrnlvnlyfygetsvskddfdewrdhedkkqndelfvkkivsiktd rkgnvkevldvdatidairtnniacyrralayanenpdvffsdtmlnkfwihhvene veriyghinnntgdykyqlgylsekvwkgiinylsikyiaegkavynyamnalak dnnsnafgkldekfvngitsfeyerikaeetlqrecavniafaanhlanatvdlnekds dflllkhednkdtlgavarpnilrnilqffggksrwndfdfsgideiqllddlrkmiysl rnssfhfktenidndswntkligdmfaydfnmagnvqkdkmysnnvpmfysts diekmldrlyaevherasqvpsfnsvfvrknfpdylkndlkitsafgvddalkwqsa vyyvckeiyyndflqnpetftmlkdyvqclpididksmdqklksernahknfkeaf atyckecdslsaicqmimteynnqnkgnrkvisartkdgdkliykhykmilfealk nvftiyleknintygflkkpklinnvpaieeflpnyngrqyetlvnriteetelqkwyi

			vgrllnpkqvnqlignfrsyvqyvndvarrakqtgnnlsndniawdvkniiqifdvc tklngvtsniledyfddgddyarylknfvdytnknndhsatllgdfcakeidgikigi yhdgtnpivnrniiqcklygatgiisdltkdgsilsvdyeiikkymqmqkeikvyqq kgicktkeeqqnlkkyqelknivelrniidyseildelqgqlinwgylrerdlmyfql gfhylclhneskkpvgynnagdisgavlyqivamytnglslidangkskknakasa gakvgsfcsyskeirgvdkdtkedddpiylagvelfeninehqqcinlrnyiehfhy yakhdrsmldlysevfdrfftydmkytknvpnmmynillqhlvvpafefgssekrl ddndeqtkpramftlreknglsseqftyrlgdgnstvklsargddylravasllyypdr apeglirdaeaedkfakinhsnpksdnrnnrgnfknpkvqwynnktkrk
C2-28		SAMN02910398_00008 [Butyrivibrio sp. YAB3001] (SEQ ID No. 60)	mkiskvdhrktavkitdnkgaegfiyqdptrdsstmeqiisnrarsskvlfnifgdtk kskdlnkytesliiyvnkaikslkgdkrnnkyeeiteslktervlnaliqagneftcsen niedalnkylkksfrvgntksalkkllmaaycgyklsieekeeiqnyfvdklvkeyn kdtvlkytakslkhqnmvvqpdtdnhvflpsriagatqnkmsekealteflkayavl deekrhnlriilrklvnlyfyespdfiypennewkehddrknktetfvspvkvneek ngktfvkidvpatkdlirlkniecyrrsvaetagnpityftdhniskfwihhienevek ifallksnwkdyqfsvgyisekvwkeiinylsikyiaigkavynyaledikkndgtl nfgvidpsfydginsfeyekikaeetfqrevavyvsfavnhlssatvklseaqsdmlv lnkndiekiaygntkrnilqffggqskwkefdfdryinpvnytdidflfdikkmvysl rnesfhftttdtesdwnknlisamfeyecrristvqknkffsnnlplfygenslervlhk lyddyvdrmsqvpsfgnvfvrkkfpdymkeigikhnlssednlklqgalyflykei yynafissekamkifvdlvnkldtnarddkgritheamahknfkdaishymthdcs ladicqkimteynqqntghrkkqttysseknpeifrhykmilfmllqkamteyisse eifdfimkpnspktdikeeeflpqykscaydnlikliadnvelqkwyitarllsprev nqligsfrsykqfvsdierraketnnslsksgmtvdvenitkvldlctklngrfsneltd yfdskddyavyvskfldfgfkidekfpaallgefcnkeengkkigiyhngtepilns niiksklygitdvvsravkpvseklireylqqevkikpylengvcknkeeqaalrky qelknriefrdiveyseiinelmgqlinfsylrerdlmyfqlgfhylclnnygakpegy ysivndkrtikgailyqivamytyglpiyhyvdgtisdrrknkktvldtlnssetvgak ikyfiyysdelfndslilynaglelfeninehenivnlrkyidhfkyyvsqdrslldiys evfdryftydrkykknvmnlfsnimlkhfiitdfefstgektigekntakkecakvri krgglssdkftykfkdakpielsakntefldgvarilyypenvvltdlvrnsevedekr

			iekydrnhnssptrkdktykqdvkknynkktskafdsskldtksvgnnlsdnpvlk qflseskkkr
C2-29		Blautia sp. Marseille-P2398 (SEQ ID No. 61)	mkiskvdhvksgidqklssqrgmlykqpqkkyegkqleehvrnlsrkakalyqvf pvsgnskmekelqiinsfiknillrldsgktseeivgyintysvasqisgdhiqelvdq hlkeslrkytcvgdkriyvpdiivallkskfnsetlqydnselkilidfiredylkekqik qivhsiennstplriaeingqkrlipanvdnpkksyifeflkeyaqsdpkgqesllqh mrylillylygpdkitddyceeieawnfgsivmdneqlfseeasmliqdriyvnqqi eegrqskdtakvkknkskyrmlgdkiehsinesvvkhyqeackaveekdipwik yisdhvmsvyssknrvdldklslpylakntwntwisfiamkyvdmgkgvyhfa msdvdkvgkqdnliigqidpkfsdgissfdyerikaeddlhrsmsgyiafavnnfar aicsdefrkknrkedvltvgldeiplydnvkrkllqyfggasnwddsiidiiddkdlv acikenlyvarnvnfhfagsekvqkkqddileeivrketrdigkhyrkvfysnnvav fycdediiklmnhlyqrekpyqaqipsynkvisktylpdlifmllkgknrtkisdpsi mnmfrgtfyfllkeiyyndflqasnlkemfceglknnvknkksekpyqnfmrrfe elenmgmdfgeicqqimtdyeqqnkqkkktatavmsekdkkirtldndtqkykh frtllyiglreafiiylkdeknkewyeflrepvkreqpeekefvnkwklnqysdcseli lkdslaaawyvvahfinqaqlnhligdiknyiqfisdidrrakstgnpvsesteiqier yrkilrvlefakffcgqitnvltdyyqdendfsthvghyvkfekknmepahalqafs nslyacgkekkkagfyydgmnpivnrnitlasmygnkkllenamnpvteqdirk yyslmaeldsvlkngavcksedeqknlrhfqnlknrielvdvltlselvndlvaqlig wvyirerdmmylqlglhyiklyftdsvaedsylrtldleegsiadgavlyqiaslysfn lpmyvkpnkssvyckkhvnsvatkfdifekeycngdetvienglrlfeninlhkdm vkfrdylahfkyfakldesilelyskaydfffsyniklkksvsyvltnvllsyfinaklsf stykssgnktvqhrttkisvvaqtdyftyklrsivknkngvesienddrrcevvniaar dkefvdevcnvinynsdk
C2-30		Leptotrichia sp. Marseille-P3007 (SEQ ID No. 62)	mkitkidgishkkyikegklvkstseenktderlselltirldtyiknpdnaseeenrirr enlkeffsnkvlylkdgilylkdrreknqlqnknyseediseydlknknnflvlkkill nedinseeleifrndfekkldkinslkysleenkanyqkinennikkvegkskrnify nyykdsakrndyinniqeafdklykkedienlfflienskkhekykirecyhkiigrk ndkenfatiiyeeiqnvnnmkeliekvpnvselkksqvfykyylnkeklndeniky vfchfveiemskllknyvykkpsnisndkvkrifeyqslkklienkllnkldtyvrnc gkysfylqdgeiatsdfivgnrqneaflrniigvsstayfslrniletenenditgrmrg

			ktvknnkgeekyisgeidklydnnkqnevkknlkmfysydfnmnskkeiedffs nideaissirhgivhfnlelegkdiftfknivpsqiskkmfhdeinekklklkifkqlns anvfrylekykilnylnrtrfefvnknipfvpsftklysriddlknslgiywktpktndd nktkeitdaqiyllkniyygeflnyfmsnngnffeitkeiielnkndkrnlktgfyklq kfenlqektpkeylaniqslyminagnqdeeekdtyidfiqkiflkgfmtylanngrl sliyigsdeetntslaekkqefdkflkkyeqnnnieipyeinefvreiklgkilkyterl nmfylilkllnhkeltnlkgslekyqsankeeafsdqlelinllnldnnrvtedfelead eigkfldfngnkvkdnkelkkfdtnkiyfdgeniikhrafynikkygmlnllekisd eakykisieelknyskkkneieenhttqenlhrkyarprkdekftdedykkyekair niqqythlknkvefnelnllqslllrilhrlvgytsiwerdlrfrlkgefpenqyieeifnf dnsknvkykngqivekyinfykelykddtekisiysdkkvkelkkekkdlyirnyi ahfnyipnaeisllemlenlrkllsydrklknaimksivdilkeygfvvtfkiekdkki rieslkseevvhlkklklkdndkkkepiktyrnskelcklvkvmfeykmkekksen
C2-31		Bacteroides ihuae (SEQ ID No. 63)	mritkvkvkessdqkdkmvlihrkvgegtlvldenladltapiidkykdksfelsllk qtlvsekemnipkcdkctakerclsckqrekrlkevrgaiektigaviagrdiiprlnif nedeicwlikpklrneftfkdvnkqvvklnlpkvlveyskkndptlflayqqwiaay lknkkghikksilnnrvvidysdesklskrkqalelwgeeyetnqrialesyhtsyni gelvtllpnpeeyvsdkgeirpafhyklknvlqmhqstvfgtneilcinpifnenrani qlsaynlevvkyfehyfpikkkkknlslnqaiyylkvetlkerlslqlenalrmnllqk gkikkhefdkntcsntlsqikrdeffvlnlvemcafaannirnivdkeqvneilskkd lcnslskntidkelctkfygadfsqipvaiwamrgsvqqirneivhykaeaidkifal ktfeyddmekdysdtpfkqylelsiekidsffieqlssndvlnyyctedvnkllnkck lslrrtsipfapgfktiyelgchlqdssntyrighylmliggrvanstvtkaskaypayrf mlkliynhlflnkfldnhnkrffmkavafvlkdnrenarnkfqyafkeirmmnnde siasymsyihslsvqeqekkgdkndkvryntekfiekvfvkgfddflswlgvefils pnqeerdktvtreeyenlmikdrvehsinsnqeshiafftfckklldanhlsdlrnewik frssgdkegfsynfaidiielclltvdrveqrrdgykeqtelkeylsffikgnesentvw kgfyfqqdnytpvlyspielirkygtlellkliivdedkitqgefeewqtlkkvvedkv trrnelhqewedmknkssfsqekcsiyqklcrdidrynwldnklhlvhlrklhnlvi qilsrmarfialwdrdfvlldasranddykllsffnfrdfinakktktddellaefgskie kknapfikaedvplmvecieakrsfyqkvffrnnlqvladrnfiahynyisktakcsl

			femiiklrtlmyydrklrnavvksianvfdqngmvlqlslddshelkvdkviskriv hlknnnimtdqvpeeyykicrrllemkk
C2-32		SAMN05216357_1045 [Porphyromonadaceae bacterium KH3CP3RA] (SEQ ID No. 64)	mefrdsifksllqkeiekaplcfaeklisggvfsyypserlkefvgnhpfslfrktmpf spgfkrvmksggnyqnanrdgrfydldigvylpkdgfgdeewnaryflmkliyn qlflpyfadaenhlfrecvdfvkrvnrdyncknnnseeqafidirsmredesiadyla fiqsniiieenkkketnkegqinfnkfllqvfvkgfdsflkdrtelnflqlpelqgdgtrg ddlesldklgavvavdlkldatgidadlnenisfytfckllldsnhlsrlrneiikyqsans dfshnedfdydriisiielcmlsadhvstndnesifpnndkdfsgirpylstdakvetf edlyvhsdaktpitnatmvlnwkygtdklferlmisdqdflvtekdyfvwkelkkd ieekiklreelhslwvntpkgkkgakkkngrettgefseenkkeylevcreidryvnl dnklhfvhllkrmhslliellgrfvgftylferdyqyyhleirsrrnkdagvvdkleynk ikdqnkydkddffactflyekankvrnfiahfnyltmwnspqeeehnsnlsgakns sgrqnlkcsltelinelrevmsydrklknavtkavidlfdkhgmvikfrivnnnnnd nknkhhlelddivpkkimhlrgiklkrqdgkpipiqtdsvdplycrmwkklldlkp tpf
C2-33		Listeria riparia (SEQ ID No. 65)	mhdawaenpkkpqsdaflkeykacceaidtynwhknkatlvyvnelhhllidilg rlvgyvaiadrdfqcmanqylkssghtervdswintirknrpdyiekldifmnkagl fvsekngrnyiahlnylspkhkysllylfeklremlkydrklknavtkslidlldkhg mcvvfanlknnkhrlviaslkpkkietfkwkkik
C2-34		Insolitispirillum peregrinum (SEQ ID No. 66)	mriirpygsstvaspspqdaqplrslqrqngtfdvaefsrrhpelvlaqwvamldkii rkpapgknstalprptaeqrrlrqqvgaalwaemqrhtpvppelkavwdskvhpy skdnapataktpshrgrwydrfgdpetsaatvaegvrrhlldsaqpfranggqpkgk gviehraltiqngtllhhhqsekagplpedwstyradelvstigkdarwikvaaslyq hygrifgpttpiseaqtrpefvlhtavkayyrrlfkerklpaerlerllprtgealrhavtv qhgnrsladavrigkilhygwlqngepdpwpddaalyssrywgsdgqtdikhsea vsrvwrraltaaqrtltswlypagtdagdilligqkpdsidrnrlpllygdstrhwtrsp gdvwlflkqtlenlrnssfhfktlsaftshldgtcesepaeqqaaqalwqddrqqdhq qvflslraldattylptgplhrivnavqstdatlplprfrrvvtraantrlkgfpvepvnrrt meddpllrcrygvlkllyergfrawletrpsiascldqslkrstkaaqtingknspqgv eilsratkllqaeggggghgihdlfdrlyaataremrvqvgyhhdaeaarqqaefiedl kcevvarafcaylktlgiqgdtfrrqpeplptwpdlpdlpsstigtaqaalysvlhlmp

			vedvgsllhqlrrwlvalqarggedgtaitatipllelylnrhdakfsgggagtglrwd dwqvffdcqatfdrvfppgpaldshrlplrglrevlrfgrvndlaaligqdkitaaevd rwhtaeqtiaaqqqrrealheqlsrkkgtdaevdeyralvtaiadhrhltahvtlsnvv rlhrlmttvlgrlvdygglwerdltfvtlyeahrlgglrnllsesrvnkfldgqtpaalsk knnaeengmiskvlgdkarrqirndfahfnmlqqgkktinltdeinnarklmahdr klknaitrsvttllqqdgldivwtmdashrltdakidsrnaihlhkthnranireplhgk sycrwvaalfgatstpsatkksdkir

[0673] In certain example embodiments, the CRISPR effector protein is a Cas13b protein selected from Table 3.

Table 3

Bergeyella zoohelcum (SEQ ID No. 67)	1	menktslgnniyynpfkpqdksyfagyfnaamentdsvfrelgkrlkgkeytsenf fdaifkenislveyeryvkllsdyfpmarlldkkevpikerkenfkknfkgiikavrd lrnfythkehgeveitdeifgvldemlkstvltvkkkkvktdktkeilkksiekqldil cqkkleylrdtarkieekrrnqrergekelvapfkysdkrddliaaiyndafdvyidk kkdslkesskakyntksdpqqeegdlkipiskngvvfllslfltkqeihafkskiagfk atvideatvseatvshgknsicfmatheifshlaykklkrkvrtaeinygeaenaeqls vyaketlmmqmldelskvpdvvyqnlsedvqktfiedwneylkenngdvgtme eeqvihpvirkryedkfnyfairfldefaqfptlrfqvhlgnylhdsrpkenlisdrrik ekitvfgrlselehkkalfikntetnedrehyweifpnpnydfpkenisvndkdfpia gsildrekqpvagkigikvkllnqqyvsevdkavkahqlkqrkaskpsiqniieeiv pinesnpkeaivfggqptaylsmndihsilyeffdkwekkkeklekkgekelrkei gkelekkivgkiqaqiqqiidkdtnakilkpyqdgnstaidkeklikdlkqeqnilqk lkdeqtvrekeyndfiayqdknreinkvrdrnhkqylkdnlkrkypeaparkevly yrekgkvavwlandikrfmptdfknewkgeqhsllqkslayyeqckeelknllpe kvfqhlpfklggyfqqkylyqfytcyldkrleyisglvqqaenfksenkvfkkvene cfkflkkqnythkeldarvqsilgypiflergfmdekptiikgktfkgnealfadwfr yykeyqnfqtfydtenyplvelekkqadrkrktkiyqqkkndvftllmakhifksvf kqdsidqfsledlyqsreerlgnqerarqtgerntnyiwnktvdlklcdgkitvenvkl knvgdfikyeydqrvqaflkyeeniewqaflikeskeeenypyvvereieqyekvr reellkevhlieeyilekvkdkeilkkgdnqnfkyyilngllkqlknedvesykvfnl ntepedvninqlkqeatdleqkafvltyirnkfahnqlpkkefwdycqekygkiek ektyaeyfaevfkkekealik

Prevotella intermedia (SEQ ID No. 68)	2	meddkkttdsiryelkdkhfwaaflnlarhnvyitvnhinkileegeinrdgyettlk ntwneikdinkkdrlskliikhfpfleaatyrlnptdttkqkeekqaeaqsleslrksff vfiyklrdlrnhyshykhskslerpkfeegllekmynifnasirlvkedyqynkdin pdedfkhldrteeefnyyftkdnegnitesgllffvslflekkdaiwmqqklrgfkdn renkkkmtnevfcrsrmllpklrlqstqtqdwilldmlnelircpkslyerlreedrek frvpieiadedydaeqepfkntlvrhqdrfpyfalryfdyneiftnlrfqidlgtyhfsi ykkqigdykeshhlthklygferiqeftkqnrpdewrkfvktfnsfetskepyipett phyhlenqkigirfrndndkiwpslktnsekneksky kldksfqaeaflsvhellpm mfyylllktentdndneietkkkenkndkqekhkieeiienkiteiyalydtfangei ksideleeyckgkdieighlpkqmiailkdehkvmateaerkqeemlvdvqksle sldnqineeienverknsslksgkiaswlvndmmrfqpvqkdnegkplnnskans teyqllqrtlaffgseherlapyfkqtkliessnphpflkdtewekcnnilsfyrsylea kknfleslkpedweknqyflklkepktkpktlvqgwkngfnlprgiftepirkwfm khrenitvaelkrvglvakviplffseeykdsvqpfynyhfnvgninkpdeknflnc eerrellrkkkdefkkmtdkekeenpsylefkswnkferelrlvrnqdivtwllcme lfnkkkikelnvekiylknintnttkkeknteekngeeknikeknnilnrimpmrlpi kvygrenfsknkkkkirrntfftvyieekgtkllkqgnfkalerdrrlgglfsfvktpsk aesksntisklrveyelgeyqkarieiikdmlalektlidkynsldtdnfnkmltdwle lkgepdkasfqndvdlliavrnafshnqypmrnriafaninpfslssantseekglgi anqlkdkthktiekiieiekpietke
Prevotella buccae (SEQ ID No. 69)	3	mqkqdklfvdrkknaifafpkyitimenkekpepiyyeltdkhfwaaflnlarhnv yttinhinrrleiaelkddgymmgikgswneqakkldkkvrlrdlimkhfpfleaaa yemtnskspnnkeqrekeqsealslnnlknvlfiflekl qvlrnyyshykyseespk pifetsllknmykvfdanvrlvkrdymhhenidmqrdfthlnrkkqvgrtkniids pnfhyhfadkegnmtiagllffvslfldkkdaiwmqkklkgfkdgrnlreqmtnev fcrsrislpklklenvqtkdwmqldmlnelvrcpkslyerlrekdresfkvpfdifsd dynaeeepfkntlvrhqdrfpyfvlryfdlneifeqlrfqidlgtyhfsiynkrigdede vrhlthhlygfariqdfapqnqpeewrklvkdldhfetsqepyisktaphyhleneki gikfcsahnnlfpslqtdktcngrskfnlgtqftaeaflsvhellpmmfyylllt kdysr kesadkvegiirkeisniyaiydafanneinsiadltrrlqntnilqghlpkqmisilkg rqkdmgkeaerkigemiddtqrrldllckqtnqkirigkrnagllksgkiadwlvnd mmrfqpvqkdqnnipinnskansteyrmlqralalfgsenfrlkayfnqmnlvgn dnphpflaetqwehqtnilsfyrnylearkkylkglkpqnwkqyqhflilkvqktnr

		ntlvtgwknsfnlprgiftqpirewfekhnnskriydqilsfdrvgfvakaiplyfaee ykdnvqpfydypfnignrlkpkkrqfldkkervelwqknkelfknypsekkktdl ayldflswkkferelrliknqdivtwlmfkelfnmatveglkigeihlrdidtntanee snnilnrimpmklpvktyetdnkgnilkerplatfyieetetkvlkqgnfkalvkdrrl nglfsfaettdlnleehpisklsvdlelikyqttrisifemtlglekklidkystlptdsfrn mlerwlqckanrpelknyvnsliavrnafshnqypmydatlfaevkkftlfpsvdtk kielniapqlleivgkaikeieksenkn
Porphyromonas gingivalis (SEQ ID No. 70)	4	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfavffkpddfvlakn rkeqlisvadgkecltvsgfafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls ensldeesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnpqsmgfisvhdlrklllmellcegsfsr mqsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykq eikgrkdklnsqllsafdmdqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlr lrkfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqya geenrrqfraivaelrlldpssghpflsatmetahrytegfykcylekkrewlakifyr peqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlf dskvmellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyi psdgkkfadcythlmektvrdkkrelrtagkpvppdlaadikrsfhravnerefmlr lvqeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegeg gdnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeyd rcrikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildpenrffgkllnnmsqpindl
Bacteroides pyogenes (SEQ ID No. 71)	5	mesiknsqkstgktlqkdppyfglylnmallnvrkvenhirkwlgdvallpeksgf hsllttdnlssakwtrfyyksrkflpflemfdsdkksyenrretaecldtidrqkissllk evygklqdirnafshyhiddqsvkhtaliissemhrfienaysfalqktrarftgvfvet dflqaeekgdnkkffaiggnegiklkdnalifliclfldreeafkflsratgfkstkekgf lavretfcalccrqpherllsvnpreallmdmlnelnrcpdilfemldekdqksflpll geeeqahilenslndelceaiddpfemiaslskrvryknrfpylmlryieeknllpfir

		fridlgclelasypkkmgeennyersvtdhamafgrltdfhnedavlqqitkgitdev rfslyapryaiynnkigfvrtsgsdkisfptlkkkggeghcvaytlqntksfgfisiydl rkilllsfldkdkaknivsglleqcekhwkdlsenlfdairtelqkefpvplirytlprsk ggklvsskladkqekyeseferrkeklteilsekdfdlsqiprrmidewlnvlptsrek klkgyvetlkldcrerlrvfekrekgehplpprigematdlakdiirmvidqgvkqri tsayyseiqrclaqyagddnrrhldsiirelrlkdtknghpflgkvlrpglghteklyqr yfeekkewleatfypaaspkrvprfvnpptgkqkelpliirnlmkerpewrdwkqr knshpidlpsqlfeneicrllkdkigkepsgklkwnemfklywdkefpngmqrfy rckrrvevfdkvveyeyseeggnykkyyealidevvrqkissskeksklqvedltls vrrvfkrainekeyqlrllceddrllfmavrdlydwkeaqldldkidnmlgepvsvs qviqleggqpdavikaecklkdvsklmrycydgrvkglmpyfanheatqeqvem elrhyedhrrrvfnwvfaleksvlkneklrrfyeesqggcehrrcidalrkaslvseee yeflvhirnksahnqfpdleigklppnvtsgfceciwskykaiicriipfidperrffgk lleqk
Alistipes sp. ZOR0009 (SEQ ID No. 72)	6	msneigafrehqfayapgnekqeeatfatyfnlalsnvegmmfgevesnpdkiek sldtlppailrqiasfiwlskedhpdkaysteevkvivtdlvrrlcfyrnyfshcfyldtq yfysdelvdttaigeklpynfhhfitnrlfryslpeitlfrwnegerkyeilrdgliffcclf lkrgqaerflnelrffkrtdeegrikrtiftkyctreshkhigieeqdflifqdiigdlnrvp kvcdgvvdlskeneryiknretsnesdenkaryrllirekdkfpyylmryivdfgvl pcitfkqndystkegrgqfhyqdaavaqeercynfvvrngnvyysympqaqnvv riselqgtisveelrnmvyasingkdvnksveqylyhlhllyekiltisgqtikegrvd vedyrplldkllrpasngeelrrelrkllpkrvcdllsnrfdcsegvsavekrlkaillrh eqlllsqnpalhidkiksvidylylffsddekfrqqptekahrglkdeefqmyhylvg dydshplalwkeleasgrlkpemrkltsatslhglymlclkgtvewcrkqlmsigk gtakveaiadrvglklydklkeytpeqlerevklvvmhgyaaaatpkpkaqaaips kltelrfysflgkremsfaafirqdkkaqklwlrnfytveniktlqkrqaaadaackkl ynlvgevervhtndkvlvlvaqryrerllnvgskcavtldnperqqkladvyevqna wlsirfddldftlthvnlsnlrkaynliprkhilafkeyldnrvkqklceecrnvrrkedl ctccsprysnltswlkenhsessiereaatmmlldverkllsflderrkaiieygkfip fsalvkecrladaglcgirndvlhdnvisyadaigklsayfpkeaseaveyirrtkevr eqrreelmanssq
Prevotella sp.	7a	mskeckkqrqekkrrlqkanfsisltgkhvfgayfnmartnfvktinyilpiagvrg nysenqinkmlhalfliqagrneeltteqkqwekklrlnpeqqtkfqkllfkhfpvlg

MA2016 (SEQ ID No. 73)		pmmadvadhkaylnkkkstvqtedetfamlkgvsladcldiiclmadtltecrnfy thkdpynkpsqladqylhqemiakkldkvvvasrrilkdreglsvnevefltgidhl hqevlkdefgnakvkdgkvmktfveyddfyfkisgkrlvngytvttkddkpvnvn tmlpalsdfgllyfcvlflskpyaklfidevrlfeyspfddkenmimsemlsiyrirtp rlhkidshdskatlamdifgelrrcpmelynlldknagqpffhdevkhpnshtpdvs krlryddrfptlalryidetelfkrirfqlqlgsfrykfydkencidgrvrvrriqkeingy grmqevadkrmdkwgdliqkreersvkleheelyinldqfledtadstpyvtdrrp aynihanriglywedsqnpkqykvfdengmyipelvvtedkkapikmpaprcal svydlpamlfyeylreqqdnefpsaeqviieyeddyrkffkavaegklkpfkrpkef rdflkkeypklrmadipkklqlflcshglcynnkpetvyerldrltlqhleerelhiqnr lehyqkdrdmignkdnqygkksfsdvrhgalarylaqsmmewqptklkdkekg hdkltglnynvltaylatyghpqvpeegftprtleqvlinahliggsnphpfinkvlal gnrnieelylhyleeelkhirsriqslssnpsdkalsalpfihhdrmryhertseemm alaaryttiqlpdglftpyileilqkhytensdlqnalsqdvpvklnptcnaaylitlfyq tvlkdnaqpfylsdktytrnkdgekaesfsfkrayelfsvlnnnkkdtfpfemiplflt sdeiqerlsaklldgdgnpvpevgekgkpatdsqgntiwkrriysevddyaekltdr dmkisfkgeweklprwkqdkiikrrdetrrqmrdellqrmpryirdikdnertlrry ktqdmvlfllaekmftniiseqssefnwkqmrlskvcneaflrqtltfrvpvtvgetti yveqenmslknygefyrfltddrlmsllnnivetlkpnengdlvirhtdlmselaay dqyrstifmliqsienliitnnavlddpdadgfwvredlpkrnnfasllelinqlnnvel tdderkllvairnafshnsynidfslikdvkhlpevakgilqhlqsmlgveitk
Prevotella sp. MA2016 (SEQ ID No. 74)	7b	mskeckkqrqekkrrlqkanfsisltgkhvfgayfnmartnfvktinyilpiagvrg nysenqinkmlhalfliqagrneeltteqkqwekklrlnpeqqtkfqkllfkhfpvlg pmmadvadhkaylnkkkstvqtedetfamlkgvsladcldiiclmadtltecrnfy thkdpynkpsqladqylhqemiakkldkvvvasrrilkdreglsvnevefltgidhl hqevlkdefgnakvkdgkvmktfveyddfyfkisgkrlvngytvttkddkpvnvn tmlpalsdfgllyfcvlflskpyaklfidevrlfeyspfddkenmimsemlsiyrirtp rlhkidshdskatlamdifgelrrcpmelynlldknagqpffhdevkhpnshtpdvs krlryddrfptlalryidetelfkrirfqlqlgsfrykfydkencidgrvrvrriqkeingy grmqevadkrmdkwgdliqkreersvkleheelyinldqfledtadstpyvtdrrp aynihanriglywedsqnpkqykvfdengmyipelvvtedkkapikmpaprcal svydlpamlfyeylreqqdnefpsaeqviieyeddyrkffkavaegklkpfkrpkef rdflkkeypklrmadipkklqlflcshglcynnkpetvyerldrltlqhleerelhiqnr

		lehyqkdrdmignkdnqygkksfsdvrhgalarylaqsmmewqptklkdkekg hdkltglnynvltaylatyghpqvpeegftprtleqvlinahliggsnphpfinkvlal gnrnieelylhyleeelkhirsriqslssnpsdkalsalpfihhdrmryhertseemm alaaryttiqlpdglftpyileilqkhytensdlqnalsqdvpvklnptcnaaylitlfyq tvlkdnaqpfylsdktytrnkdgekaesfsfkrayelfsvlnnnkkdtfpfemiplflt sdeiqerlsaklldgdgnpvpevgekgkpatdsqgntiwkrriysevddyaekltdr dmkisfkgeweklprwkqdkiikrrdetrrqmrdellqrmpryirdikdnertlrry ktqdmvlfllaekmftniiseqssefnwkqmrlskvcneaflrqtltfrvpvtvgetti yveqenmslknygefyrfltddrlmsllnnivetlkpnengdlvirhtdlmselaay dqyrstifmliqsienliitnnavlddpdadgfwvredlpkrnnfasllelinqlnnvel tdderkllvairnafshnsynidfslikdvkhlpevakgilqhlqsmlgveitk
Riemerella anatipestifer (SEQ ID No. 75)	8	mekpllpnvytlkhkffwgaflniarhnafitichineqlglktpsnddkivdvvcet wnnilnndhdllkksqltelilkhfpfltamcyhppkkegkkkghqkeqqkekese aqsqaealnpskliealeilvnqlhslrnyyshykhkkpdaekdifkhlykafdaslr mvkedykahftvnltrdfahlnrkgknkqdnpdfnryrfekdgfftesgllfftnlfld krdaywmlkkvsgfkashkqrekmttevfcrsrillpklrlesrydhnqmlldmls elsrcpkllyeklseenkkhfqveadgfldeieeeqnpfkdtlirhqdrfpyfalryldl nesfksirfqvdlgtyhyciydkkigdeqekrhltrtllsfgrlqdfteinrpqewkalt kdldyketsnqpfiskttphyhitdnkigfrlgtskelypsleikdganriakypynsg fvahafisvhellplmfyqhltgksedllketvrhiqriykdfeeerintiedlekanqg rlplgafpkqmlgllqnkqpdlsekakikiekliaetkllshrlntklksspklgkrrek liktgvladwlvkdfmrfqpvaydaqnqpiksskanstefwfirralalyggeknrl egyfkqtnligntnphpflnkfnwkacrnlvdfyqqyleqrekfleaiknqpwepy qyclllkipkenrknlvkgweqggislprglfteairetlsedlmlskpirkeikkhgr vgfisraitlyfkekyqdkhqsfynlsykleakapllkreehyeywqqnkpqsptes qrlelhtsdrwkdyllykrwqhlekklrlyrnqdvmlwlmtleltknhfkelnlnyh qlklenlavnvqeadaklnplnqtlpmvlpvkvypatafgevqyhktpirtvyiree htkalkmgnfkalvkdrrlnglfsfikeendtqkhpisqlrlrreleiyqslrvdafketl sleekllnkhtslsslenefralleewkkeyaassmvtdehiafiasvrnafchnqypf ykealhapiplftvaqptteekdglgiaeallkvlreyceivksqi
Prevotella aurantiaca	9	meddkkttgsisyelkdkhfwaaflnlarhnvyitinhinkllleireidndekvldikt lwqkgnkdlnqkarlrelmtkhfpfletaiytknkedkkevkqekqaeaqsleslkd clflfldklqearnyyshykysefskepefeegllekmynifgnniqlvindyqhnk

(SEQ ID No. 76)		dinpdedfkhldrkgqfkysfadnegnitesgllffvslflekkdaiwmqqklngfk dnlenkkkmthevfcrsrilmpklrlestqtqdwilldmlnelircpkslyerlqgdd rekfkvpfdpadedynaeqepfkntlirhqdrfpyfvlryfdyneifknlrfqidlgty hfsiykkliggqkedrhlthklygferiqefakqnrpdewkaivkdldtyetsnkryi settphyhlenqkigirfrngnkeiwpslktndennekskykldkqyqaeaflsvhe llpmmfyylllkkekpnndeinasivegfikreirnifklydafangeinniddleky cadkgipkrhlpkqmvailydehkdmvkeakrkqkemvkdtkkllatlekqtqk ekeddgrnvkllksgeiarwlvndmmrfqpvqkdnegkplnnskansteyqml qrslalynneekptryfrqvnliesnnphpflkwtkweecnniltfyysyltkkiefln klkpedwkknqyflklkepktnretlvqgwkngfnlprgiftepirewfkrhqnns keyekvealdrvglvtkviplffkeeyfkdkeenfkedtqkeindcvqpfynfpyn vgnihkpkekdflhreerielwdkkkdkfkgykekikskkltekdkeefrsylefqs wnkferelrlvrnqdivtwllckelidklkidelnieelkklrlnnidtdtakkeknnil nrvmpmelpvtvyeiddshkivkdkplhtiyikeaetkllkqgnfkalvkdrrlngl fsfvktnseaeskrnpisklrveyelgeyqearieiiqdmlaleeklinkykdlptnkf semlnswlegkdeadkarfqndvdfliavrnafshnqypmhnkiefanikpfslyt annseekglgianqlkdktkettdkikkiekpietke
Prevotella saccharolytica (SEQ ID No. 77)	10	medkpfwaaffnlarhnvyltvnhinkllldleklydegkhkeiferedifnisddvm ndansngkkrkldikkiwddldtdltrkyqlrelilkhfpfiqpaiigaqtkerttidkd krststsndslkqtgegdindllslsnvksmffrllqileqlrnyyshvkhsksatmpn fdedllnwmryifidsvnkvkedyssnsvidpntsfshliykdeqgkikpcrypfts kdgsinafgllffvslflekqdsiwmqkkipgfkkasenymkmtnevfcrnhillp kirletvydkdwmlldmlnevvrcplslykrltpaaqnkfkvpekssdnanrqedd npfsrilvrhqnrfpyfvlrffdlnevfttlrfqinlgcyhfaickkqigdkkevhhlirtl ygfsrlqnftqntrpeewntlvkttepssgndgktvqgvplpyisytiphyqieneki gikifdgdtavdtdiwpsvstekqlnkpdkytltpgfkadvflsvhellpmmfyyql llcegmlktdagnavekvlidtrnaifnlydafvqekintitdlenylqdkpilighlpk qmidllkghqrdmlkaveqkkamlikdterrlklldkqlkqetdvaakntgtllkng qiadwlvndmmrfqpvkrdkegnpincskansteyqmlqrafafyatdscrlsryf tqlhlihsdnshlflsrfeydkqpnliafyaaylkakleflnelqpqnwasdnyflllra pkndrqklaegwkngfnlprglftekiktwfnehktivdisdcdifknrvgqvarlip vffdkkfkdhsqpfyrydfnvgnvskpteanylskgkreelfksyqnkfknnipae ktkeyreyknfslwkkferelrliknqdiliwlmcknlfdekikpkkdilepriavsyi

		kldslqtntstagslnalakvvpmtlaihidspkpkgkagnnekenkeftvyikeegt kllkwgnfktlladrrikglfsyiehddidlkqhpltkrrvdleldlyqtcridifqqtlgl eaqlldkysdlntdnfyqmligwrkkegiprnikedtdflkdvrnafshnqypdsk kiafrrirkfnpkelileeeeglgiatqmykevekvvnrikrielfd
HMPREF9712_03108 [Myroides odoratimimus CCUG 10230] (SEQ ID No. 78)	11	mkdilttdttekqnrfyshkiadkyffggyfnlasnniyevfeevnkrntfgklakrd ngnlknyiihvfkdelsisdfekrvaifasyfpiletvdkksikernrtidltlsqrirqfr emlislvtavdqlrnfythyhhsdivienkvldflnssfvstalhvkdkylktdktkefl ketiaaeldilieaykkkqiekkntrfkankredilnaiyneafwsfindkdkdkdke tvvakgadayfeknhhksndpdfalnisekgivyllsffltnkemdslkanltgfkg kvdresgnsikymatqriysfhtyrglkqkirtseegvketllmqmidelskvpnvv yqhlsttqqnsfiedwneyykdyeddvetddlsrvihpvirkryedrfnyfairflde ffdfptlrfqvhlgdyvhdrrtkqlgkvesdriikekvtvfarlkdinsakasyfhslee qdkeeldnkwtlfpnpsydfpkehtlqhqgeqknagkigiyvklrdtqykekaale earkslnpkersatkaskydiitqiieandnvksekplvftgqpiaylsmndihsmlf slltdnaelkktpeeveaklidqigkqineilskdtdtkilkkykdndlketdtdkitrdl ardkeeieklileqkqraddynytsstkfnidksrkrkhllfnaekgkigvwlandikr fmfkeskskwkgyqhtelqklfayfdtsksdlelilsnmvmvkdypielidlvkks rtlvdflnkylearleyienvitrvknsigtpqfktvrkecftflkksnytvvsldkqver ilsmplfiergfmddkptmlegksykqhkekfadwfvhykensnyqnfydtevy eittedkrekakvtkkikqqqkndvftlmmvnymleevlklssndrlslnelyqtke erivnkqvakdtqernknyiwnkvvdlqlcdglvhidnvklkdignfrkyendsrv kefltyqsdivwsaylsnevdsnklyvierqldnyesirskellkevqeiecsvynqv ankeslkqsgnenfkqyvlqgllpigmdvremlilstdvkfkkeeiiqlgqageveq dlysliyirnkfahnqlpikeffdfcennyrsisdneyyaeyymeifrsikekyan
Prevotella intermedia (SEQ ID No. 79)	12	meddkkttdsiryelkdkhfwaaflnlarhnvyitvnhinkileedeinrdgyentle nswneikdinkkdrlskliikhfpfleattyrqnptdttkqkeekqaeaqsleslkksff vfiyklrdlrnhyshykhskslerpkfeedlqnkmynifdvsiqfvkedykhntdin pkkdfkhldrkrkgkfhysfadnegnitesgllffvslflekkdaiwvqkklegfkcs nksyqkmtnevfcrsrmllpklrlestqtqdwilldmlnelircpkslyerlqgvnrk kfyvsfdpadedydaeqepfkntlvrhqdrfpyfalryfdynevfanlrfqidlgtyh fsiykkliggqkedrhlthklygferiqefdkqnrpdewkaivkdsdtfkkkeekee ekpyisettphyhlenkkigiafknhniwpstqteltnnkrkkynlgtsikaeaflsvh ellpmmfyylllktentkndnkvggkketkkqgkhkieaiieskikdiyalydafan

		geinsedelkeylkgkdikivhlpkqmiailknehkdmaekaeakqekmklaten rlktldkqlkgkiqngkrynsapksgeiaswlvndmmrfqpvqkdengeslnnsk ansteyqllqrtlaffgseherlapyfkqtkliessnphpflndtewekcsnilsfyrsyl karknfleslkpedweknqyflmlkepktnretlvqgwkngfnlprgfftepirkwf mehwksikvddlkrvglvakvtplffsekykdsvqpfynypfnvgdvnkpkeed flhreerielwdkkkdkfkgykakkkfkemtdkekeehrsylefqswnkferelrl vrnqdivtwllctelidklkidelnikelkklrlkdintdtakkeknnilnrvmpmelp vtvykvnkggyiiknkplhtiyikeaetkllkqgnfkalvkdrrlnglfsfvktpseae sesnpisklrveyelgkyqnarldiiedmlalekklidkynsldtdnfhnmltgwlel kgeakkarfqndvkltavrnafshnqypmydenlfgnierfslsssniieskgldia aklkeevskaakkiqneednkkeket
Capnocytophaga canimorsus (SEQ ID No. 80)	13	mkniqrlgkgnefspfkkedkfyfggflnlannniedffkeiitrfgivitdenkkpk etfgekilneifkkdisivdyekwvnifadyfpftkylslyleemqfknrvicfrdvm kellktvealrnfythydhepikiedrvfyfldkvlldvsltvknkylktdktkeflnqh igeelkelckqrkdylvgkgkridkeseiingiynnafkdfickrekqddkenhnsv ekilcnkepqnkkqkssatvwelcskssskyteksfpnrendkhclevpisqkgivf llsfflnkgeiyaltsnikgfkakitkeepvtydknsirymathrmfsflaykglkrkir tseinynedgqasstyeketlmlqmldelnkvpdvvyqnlsedvqktfiedwneyl kenngdvgtmeeeqvihpvirkryedkfnyfairfldefaqfptlrfqvhlgnylcd krtkqicdttterevkkkitvfgrlselenkkaiflnereeikgwevfpnpsydfpken isvnykdfpivgsildrekqpvsnkigirvkiadelqreidkaikekklrnpknrkan qdekqkerlvneivstnsneqgepvvfigqptaylsmndihsvlyeflinkisgeale tkivekietqikqiigkdattkilkpytnansnsinrekllrdleqeqqilktlleeqqqre kdkkdkkskrkhelypsekgkvavwlandikrfmpkafkeqwrgyhhsllqkyl ayyeqskeelknllpkevfkhfpfklkgyfqqqylnqfytdylkrrlsyvnelllniq nfkndkdalkatekecfkffrkqnyiinpiniqiqsilvypiflkrgfldekptmidre kfkenkdteladwfmhyknykednyqkfyayplekveekekfkrnkqinkqkk ndvytlmmveyiiqkifgdkfveenplvlkgifqskaerqqnnthaattqernlngil nqpkdikiqgkitvkgvklkdignfrkyeidqrvntfldyeprkewmaylpndwk ekekqgqlppnnvidrqiskyetvrskillkdvqelekiisdeikeehrhdlkqgkyy nfkyyilngllrqlknenvenykvfklntnpekvnitqlkqeatdleqkafvltyirnk fahnqlpkkefwdycqekygkiekektyaeyfaevfkrekealik

Porphyromonas gulae (SEQ ID No. 81)	14	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndqdvlsfkalwknfdndlerksrlrslilkhfsflegaaygkklfeskssgnkssknkeltkkekeelqanalsldnlksilfdflqklkdfrnyyshyrhsgsselplfdgnmlqrlynvfdvsvqrvkidhehndevdphyhfnhlvrkgkkdryghndnpsfkhhfvdgegmvteagllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrmddwmlldmlnelvrcpkplydrlreddracfrvpvdilpdeddtdgggedpfkntlvrhqdrfpyfalryfdlkkvftslrfhidlgtyhfaiykkmigeqpedrhltrnlygfgriqdfaeehrpeewkrlvrdldyfetgdkpyisqtsphyhiekgkiglrfmpegqhlwpspevgttrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaervqgrikrviedvyavydafardeintrdeldacladkgirrghlprqmiailsqehkdmeekirkklqemmadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpvakdasgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetrweshtnilsfyrsylrarkaflerigrsdrvenrpflllkepktdrqtlvagwkgefhlprgifteavrdcliemghdevasykevgfmakavplyferacedrvqpfydspfnvgnslkpkkgrflskeeraeewergkerfrdleawsysaarriedafagieyaspgnkkkieqllrdlslweafesklkvradrinlaklkkeileaqehpyhdfkswqkferelrlvknqdiitwmmcrdlmeenkvegldtgtlylkdirpnvqeqgslnvlnrvkpmrlpvvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtgglameqypisklrveyelakyqtarvcvfeltlrleeslltryphlpdesfremleswsdpllakwpelhgkvrlliavrnafshnqypmydeavfssirkydpsspdaieermglniahrlseevkqaketveriiqa
Prevotella sp. P5-125 (SEQ ID No. 82)	15	mnipalvenqkkyfgtysvmamlnaqtvldhiqkvadiegeqnennenlwfhpvmshlynakngydkqpektmfiierlqsyfpflkimaenqreysngkykqnrvevnsndifevlkrafgvlkmyrdltnhyktyeeklndgcefltsteqplsgminnyytvalrnmnerygyktedlafiqdkrfkfvkdaygkkksqvntgfflslqdyngdtqkklhlsgvgialliclfldkqyiniflsrlpifssynaqseerriiirsfginsiklpkdrihseksnksvamdmlnevkrcpdelfttlsaekqsrfriisddhnevlmkrssdrfvplllqyidygklfdhirfhvnmgklryllkadktcidgqtrvrvieqplngfgrleeaetmrkqengtfgnsgirirdfenmkrddanpanypyivdtythyilennkvemfindkedsapllpvieddryvvktipscrmstleipamafhmflfgskkteklivdvhnrykrlfqamqkeevtaeniasfgiaesdlpqkildlisgnahgkdvdafirltvddmltdterrikrfkddrksirsadnkmgkrgfkqistgkladflakdivlfqpsvndgenkitglnyrimqsaiavydsgddyeakqqfklmfekarligkgttephpflykvfarsipanavefyer

		ylierkfyltglsneikkgnrvdvpfirrdqnkwktpamktlgriysedlpvelprqm fdneikshlkslpqmegidfnnanvtyliaeymkrvldddfqtfyqwnrnyrymd mlkgeydrkgslqhcftsveereglwkerasrteryrkqasnkirsnrqmrnassee ietildkrlsnsrneyqksekvirryrvqdallfllakktlteladfdgerfklkeimpda ekgilseimpmsftfekggkkytitsegmklknygdffvlasdkrignllelvgsdiv skedimeefnkydqcrpeissivfnlekwafdtypelsarvdreekvdfksilkilln nkninkeqsdilrkirnafdhnnypdkgvveikalpeiamsikkafgeyaimk
Flavobacterium branchiophilum (SEQ ID No. 83)	16	menlnkildkeneiciskifntkgiaapitekaldnikskqkndlnkearlhyfsighs fkqidtkkvfdyvlieelkdekplkfitlqkdfftkefsiklqklinsirninnhyvhnf ndinlnkidsnvfhflkesfelaiiekyykvnkkypldneivlflkelfikdentallny ftnlskdeaieyiltftitenkiwninnehnilniekgkyltfeamlflitiflykneanhl lpklydfknnkskqelftffskkftsqdidaeeghlikfrdmiqylnhyptawnndlk lesenknkimttklidsiiefelnsnypsfatdiqfkkeakaflfasnkkrnqtsfsnks yneeirhnphikqyrdeiasaltpisfnvkedkfkifvkkhvleeyfpnsigyekfle yndftekekedfglklysnpktnklieridnhklvkshgrnqdrfmdfsmrflaenn yfgkdaffkcykfydtqeqdeflqsnennddvkfhkgkvttyikyeehlknysyw dcpfveennsmsvkisigseekilkiqrnlmiyflenalynenvenqgyklvnnyy relkkdveesiasldliksnpdfkskykkilpkrllhnyapakqdkapenafetllkk adfreeqykkllkkaeheknkedfvkrnkgkqfklhfirkacqmmyfkekyntlk egnaafekkdpviekrknkehefghhknlnitreefndyckwmfafngndsykk ylrdlfsekhffdnqeyknlfessvnleafyaktkelfkkwietnkptnnenrytleny knlilqkqvfinvyhfskylidknllnsennviqykslenveylisdfyfqsklsidqy ktcgklfnklksnkledcllyeiaynyidkknvhkidiqkiltskiiltindantpykis vpfnklerytemiaiknqnnlkarflidlplylsknkikkgkdsagyeiiikndleied intinnkiindsvkftevlmelekyfilkdkcilsknyidnseipslkqfskvwikene neiinyrniachfhlplletfdnlllnveqkfikeelqnvstindlskpqeylillfikfkh nnfylnlfnknesktikndkevkknrvlqkfinqvilkkk
Myroides odoratimimus (SEQ ID No. 84)	17	mkdilttdttekqnrfyshkiadkyffggyfnlasnniyevfeevnkrntfgklakrd ngnlknyiihvfkdelsisdfekrvaifasyfpiletvdkksikernrtidltlsqrirqfr emlislvtavdqlrnfythyhhsdivienkvldflnssfvstalhvkdkylktdktkefl ketiaaeldilieaykkkqiekkntrfkankredilnaiyneafwsfindkdkdkdke tvvakgadayfeknhhksndpdfalnisekgivyllsffltnkemdslkanltgfkg kvdresgnsikymatqriysfhtyrglkqkirtseegvketllmqmidelskvpnvv

		yqhlsttqqnsfiedwneyykdyeddvetddlsrvthpvirkryedrfnyfairflde ffdfptlrfqvhlgdyvhdrrtkqlgkvesdriikekvtvfarlkdinsakasyfhslee qdkeeldnkwtlfpnpsydfpkehtlqhqgeqknagkigiyvklrdtqykekaale earkslnpkersatkaskydiitqiieandnvksekplvftgqpiaylsmndihsmlf slltdnaelkktpeeveaklidqigkqineilskdtdtkilkkykdndlketdtdkitrdl ardkeeieklileqkqraddynytsstkfnidksrkrkhllfnaekgkigvwlandikr fmfkeskskwkgyqhielqklfayfdtsksdlelilsnmvmvkdypielidlvkks rtlvdflnkylearleyienvitrvknsigtpqfktvrkecftflkksnytvvsldkqver ilsmplfiergfmddkptmlegksykqhkekfadwfvhykensnyqnfydtevy eittedkrekakvtkkikqqqkndvftlmmvnymleevlklssndrlslnelyqtke erivnkqvakdtqernknyiwnkvvdlqlcdglvhidnvklkdignfrkyendsrv keftyqsdivwsaylsnevdsnklyvierqldnyesirskellkevqeiecsvynqv ankeslkqsgnenfkqyvlqgllpigmdvremlilstdvkfkkeeiiqlgqageveq dlysliyirnkfahnqlpikeffdfcennyrsisdneyyaeyymeifrsikekyan
Flavobacterium columnare (SEQ ID No. 85)	18	mssknesynkqktfnhykqedkyffggflnnaddnlrqvgkefktrinfnhnnnel asvfkdyfnkeksvakrehalnllsnyfpvleriqkhtnhnfeqtreifellldtikklrd yythhyhkpitinpkiydflddtlldvlitikkkkvkndtsrellkeklrpeltqlknqk reelikkgkklleenlenavfnhclipfleenktddkqnktvslrkyrkskpneetsitl tqsglvflmsfflhrkefqvftsglerfkakvntikeeeislnknnivymithwsysyy nfkglkhriktdqgvstleqnntthsltntntkealltqivdylskvpneiyetlsekqq kefeedineymrenpenedstfssivshkvirkryenkfnyfamrfldeyaelptlrf mvnfgdyikdrqkkilesiqfdseriikkeihlfeklslvteykknvylketsnidlsrf plfpnpsyvmannnipfyidsrsnnldeylnqkkkaqsqnkkrnltfekynkeqsk daiiamlqkeigvkdlqqrstigllscnelpsmlyevivkdikgaelenkiaqkireq yqsirdftldspqkdnipttliktintdssvtfenqpidiprlknalqkeltltqekllnvk eheievdnynrnkntykfknqpknkvddkklqrkyvfyrneirqeanwlasdlihf mknkslwkgymhnelqsflaffedkkndcialletvfnlkedciltkglknlflkhg nfidfykeylklkedflstestflengfiglppkilkkelskrlkyifivfqkrqfiikelee kknnlyadainlsrgifdekptmipfkkpnpdefaswfvasyqynnyqsfyeltpd iverdkkkkyknlrainkvkiqdyylklmvdtlyqdlfnqpldkslsdfyvskaere kikadakayqklndsslwnkvihlslqnnritanpklkdigkykralqdekiatllty dartwtyalqkpekenendykelhytalnmelqeyekvrskellkqvqelekkild kfydfsnnashpedleiedkkgkrhpnfklyitkallkneseiinlenidieillkyyd

		ynteelkekiknmdedekakiintkenynkitnvlikkalvliiirnkmahnqyppk fiydlanrfvpkkeeeyfatyfnrvfetitkelwenkekkdktqv
Porphyromonas gingivalis (SEQ ID No. 86)	19	mteqnekpyngtyytledkhfwaaflnlarhnayitlahidrqlayskaditndedil ffkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnksskkkelskke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdkygnndnpffkhhfvdregtvte agllffvslflekrdaiwmqkkirgfkggteayqqmtnevfcrsrislpklkleslrtd dwmlldmlnelvrcpkslydrlreedrarfrvpvdilsdeddtdgteedpfkntlvrh qdrfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriqdf aeehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqhlwpspe vgatrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqgrikr viedvyavydafardeintrdeldacladkgirrghlprqmiailsqehkdmeekvr kklqemiadtdhrldmldrqtdrkirigrknaglpksgvvadwlvrdmmrfqpva kdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetr weshtnilsfyrsylearkaflqsigrsdrvenhrflllkepktdrqtlvagwkgefhlp rgifteavrdcliemgydevgsykevgfmakavplyferaskdrvqpfydypfnv gnslkpkkgrflskekraeewesgkerfrlaklkkeileakehpyhdfkswqkfere lrlvknqdiitwmmcrdlmeenkvegldtgtlylkdirtdvqeqgslnvlnrvkpm rlpvvvyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtg alameqypisklrveyelakyqtarvcafeqtleleeslltryphlpdknfrkmlesw sdplldkwpdlhgnvrlliavrnafshnqypmydetlfssirkydpsspdaieermg lniahrlseevkqakemveriiqa
Porphyromonas sp. COT-052 OH4946 (SEQ ID No. 87)	20	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndqdvls fkalwknfdndlerksrlrslilkhfsflegaaygkklfeskssgnkssknkeltkkek eelqanalsldnlksilfdflqklkdfrnyyshyrhsesselplfdgnmlqrlynvfdv svqrvkrdhehndkvdphrhfnhlvrkgkkdryghndnpsfkhhfvdsegmvte agllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrtdd wmlldmlnelvrcpkplydrlreddracfrvpvdilpdeddtdgggedpfkntlvr hqdrfpyfalryfdlkkvftslrfhidlgtyhfaiykkmigeqpedrhltrnlygfgriq dfaeehrpeewkrlvrdldyfetgdkpyisqttphyhiekgkiglrfvpegqhlwps pevgttrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqgri krviedvyaiydafardeintlkeldacladkgirrghlpkqmigilsqerkdmeek vrkklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpv

		akdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhet rweshtnilsfyrsylrarkaflerigrsdrvencpfllkepktdrqtlvagwkgefhl prgifteavrdcliemgydevgsyrevgfmakavplyferacedrvqpfydspfnv gnslkpkkgrflskedraeewergkerfrdleawshsaarrikdafagieyaspgnk kkieqllrdlslweafesklkvradkinlaklkkeileaqehpyhdfkswqkferelrl vknqdiitwmmcrdlmeenkvegldtgtlylkdirpnvqeqgslnvlnrvkpmrl pvvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtggl ameqypisklrveyelakyqtarvcvfeltlrleesllsryphlpdesfremleswsdp llakwpelhgkvrlliavrnafshnqypmydeavfssirkydpsspdaieermglni ahrlseevkqaketveriiqa
Prevotella intermedia (SEQ ID No. 88)	21	meddkktkestnmldnkhfwaaflnlarhnvyitvnhinkvlelknkkdqdiiidn dqdilaikthwekvngdlnkterlrelmtkhfpfletaiytknkedkeevkqekqak aqsfdslkhclflfleklqearnyyshykysestkepmlekellkkmynifddniqlv ikdyqhnkdinpdedfkhldrteeefnyyfttnkkgnitasgllffvslflekkdaiw mqqklrgfkdnreskkkmthevfcrsrmllpklrlestqtqdwilldmlnelircpk slyerlqgeyrkkfnvpfdsadedydaeqepfkntlvrhqdrfpyfalryfdyneift nlrfqidlgtyhfsiykkliggqkedrhlthklygferiqefakqnrtdewkaivkdfd tyetseepyisetaphyhlenqkigirfrndndeiwpslktngennekrkykldkqy qaeaflsvhellpmmfyylllkkeepnndkknasivegfikreirdiyklydafang einniddlekycedkgipkrhlpkqmvailydehkdmaeeakrkqkemvkdtk kllatlekqtqgeiedggrnirllksgeiarwlvndmmrfqpvqkdnegnplnnsk ansteyqmlqrslalynkeekptryfrqvnlinssnphpflkwtkweecnnilsfyrs yltkkieflnklkpedweknqyflklkepktnretlvqgwkngfnlprgiftepirew fkrhqndseeyekvetldrvglvtkviplffkkedskdkeeylkkdaqkeinncvq pfygfpynvgnihkpdekdflpseerkklwgdkkykfkgykakvkskkltdkek eeyrsylefqswnkferelrlvrnqdivtwllctelidklkveglnveelkklrlkdidt dtakqeknnilnrvmpmqlpvtvyeiddshnivkdrplhtvyieetktkllkqgnfk alvkdrrlnglfsfvdtssetelksnpiskslveyelgeyqnarietikdmllleetliek yktlptdnfsdmlngwlegkdeadkarfqndvkllvavrnafshnqypmrnriafa ninpfslssadtseekkldianqlkdkthkiikriieiekpietke
PIN17_0200 [Prevotella	AFJ07523	mkmeddkktkestnmldnkhfwaaflnlarhnvyitvnhinkvlelknkkdqdii idndqdilaikthwekvngdlnkterlrelmtkhfpfletaiytknkedkeevkqek qakaqsfdslkhclflfleklqearnyyshykysestkepmlekellkkmynifddn

intermedia 17] (SEQ ID No. 89)		iqlvikdyqhnkdinpdedfkhldrteeefnyyfttnkkgnitasgllffvslflekkd aiwmqqklrgfkdnreskkkmthevfcrsrmllpklrlestqtqdwilldmlnelir cpkslyerlqgeyrkkfnvpfdsadedydaeqepfkntlvrhqdrfpyfalryfdyn eiftnlrfqidlgtyhfsiykkliggqkedrhlthklygferiqefakqnrtdewkaivk dfdtyetseepyisetaphyhlenqkigirfrndndeiwpslktngennekrkykld kqyqaeaflsvhellpmmfyylllkkeepnndkknasivegfikreirdiyklydaf angeinniddlekycedkgipkrhlpkqmvailydehkdmaeeakrkqkemvk dtkkllatlekqtqgeiedggrnirllksgeiarwlvndmmrfqpvqkdnegnplnn skansteyqmlqrslalynkeekptryfrqvnlinssnphpflkwtkweecnnilsf yrsyltkkieflnklkpedweknqyflklkepktnretlvqgwkngfnlprgiftepir ewfkrhqndseeyekvetldrvglvtkviplffkkedskdkeeylkkdaqkeinnc vqpfygfpynvgnihkpdekdflpseerkklwgdkkykfkgykakvkskkltdk ekeeyrsylefqswnkferelrlvrnqdivtwllctelidklkveglnveelkklrlkdi dtdtakqeknnilnrvmpmqlpvtvyeiddshnivkdrplhtvyieetktkllkqgn fkalvkdrrlnglfsfvdtssetelksnpiskslveyelgeyqnarietikdmllleetlie kyktlptdnfsdmlngwlegkdeadkarfqndvkllvavrnafshnqypmrnriaf aninpfslssadtseekkldianqlkdkthkiikriieiekpietke
Prevotella intermedia (SEQ ID No. 90)	BAU18623	meddkkttdsisyelkdkhfwaaflnlarhnvyitvnhinkvlelknkkdqdiiidn dqdilaikthwekvngdlnkterlrelmtkhfpfletaiysknkedkeevkqekqak aqsfdslkhclflfleklqetrnyyshykysestkepmlekellkkmynifddniqlv ikdyqhnkdinpdedfkhldrteedfnyyftrnkkgnitesgllffvslflekkdaiw mqqklrgfkdnreskkkmthevfcrsrmllpklrlestqtqdwilldmlnelircpk slyerlqgedrekfkvpfdpadedydaeqepfkntlvrhqdrfpyfalryfdyneift nlrfqidlgtfhfsiykkliggqkedrhlthklygferiqefakqnrpdewkaivkdld tyetsneryisettphyhlenqkigirfrndndeiwpslktngennekskykldkqyq aeaflsvhellpmmfyylllkkeepnndkknasivegfikreirdmyklydafang einniddlekycedkgipkrhlpkqmvailydehkdmvkeakrkqrkmvkdtek llaalekqtqektedggrnirllksgeiarwlvndmmrfqpvqkdnegnplnnska nsteyqmlqrslalynkeekptryfrqvnlinssnphpflkwtkweecnnilsfyrsy ltkkieflnklkpedweknqyflklkepktnretlvqgwkngfnlprgiftepirewf krhqndskeyekvealdrvglvtkviplffkkedskdkeedlkkdaqkeinncvqp fysfpynvgnihkpdekdflhreerielwdkkkdkfkgykakvkskkltdkekee yrsylefqswnkferelrlvrnqdivtwllctelidklkveglnveelkklrlkdidtdta

		kqeknnilnrvmpmqlpvtvyeiddshnivkdrplhtvyieetktkllkqgnfkalv kdrrlnglfsfvdtsseaelksnpiskslveyelgeyqnarietikdmllleetliekyk nlptdnfsdmlngwlegkdeadkarfqndvkllvavrnafshnqypmrnriafani npfslssadtseekkldianqlkdkthkiikriieiekpietke
HMPREF6485_0083 [Prevotella buccae ATCC 33574] (SEQ ID No. 91)	EFU31981	mqkqdklfvdrkknaifafpkyitimenkekpepiyyeltdkhfwaaflnlarhnv yttinhinrrleiaelkddgymmgikgswneqakkldkkvrlrdlimkhfpfleaaa yemtnskspnnkeqrekeqsealslnnlknvlfifleklqvlrnyyshykyseespk pifetsllknmykvfdanvrlvkrdymhhenidmqrdfthlnrkkqvgrtkniids pnfhyhfadkegnmtiagllffvslfldkkdaiwmqkklkgfkdgrnlreqmtnev fcrsrislpklklenvqtkdwmqldmlnelvrcpkslyerlrekdresfkvpfdifsd dynaeeepfkntlvrhqdrfpyfvlryfdlneifeqlrfqidlgtyhfsiynkrigdede vrhlthhlygfariqdfapqnqpeewrklvkdldhfetsqepyisktaphyhleneki gikfcsahnnlfpslqtdktcngrskfnlgtqftaeaflsvhellpmmfyyllltkdysr kesadkvegiirkeisniyaiydafanneinsiadltrrlqntnilqghlpkqmisilkg rqkdmgkeaerkigemiddtqrrldllckqtnqkirigkrnagllksgkiadwlvnd mmrfqpvqkdqnnipinnskansteyrmlqralalfgsenfrlkayfnqmnlvgn dnphpflaetqwehqtnilsfyrnylearkkylkglkpqnwkqyqhflilkvqktnr ntlvtgwknsfnlprgiftqpirewfekhnnskriydqilsfdrvgfvakaiplyfaee ykdnvqpfydypfnignrlkpkkrqfldkkervelwqknkelfknypsekkktdl ayldflswkkferelrliknqdivtwlmfkelfnmatveglkigeihlrdidtntanee snnilnrimpmklpvktyetdnkgnilkerplatfyieetetkvlkqgnfkalvkdrrl nglfsfaettdlnleehpisklsvdlelikyqttrisifemtlglekklidkystlptdsfrn mlerwlqckanrpelknyvnsliavrnafshnqypmydatlfaevkkftlfpsvdtk kielniapqlleivgkaikeieksenkn
HMPREF9144_1146 [Prevotella pallens ATCC 700821] (SEQ ID No. 92)	EGQ18444	mkeeekgktpvvstynkddkhfwaaflnlarhnvyitvnhinkilgegeinrdgye ntlekswneikdinkkdrlskliikhfpflevttyqrnsadttkqkeekqaeaqslesl kksffvfiyklrdlrnhyshykhskslerpkfeedlqekmynifdasiqlvkedykh ntdikteedfkhldrkgqfkysfadnegnitesgllffvslflekkdaiwvqkklegfk csnesyqkmtnevfcrsrmllpklrlqstqtqdwilldmlnelircpkslyerlreedr kkfrvpieiadedydaeqepfknalvrhqdrfpyfalryfdyneiftnlrfqidlgtyh fsiykkqigdykeshhlthklygferiqeftkqnrpdewrkfvktfnsfetskepyip ettphyhlenqkigirfrndndkiwpslktnseknekskykldksfqaeaflsvhell pmmfyylllktentdndneietkkkenkndkqekhkieeiienkiteiyalydafan

		gkinsidkleeyckgkdieighlpkqmiailksehkdmateakrkqeemladvqk slesldnqineeienverknsslksgeiaswlvndmmrfqpvqkdnegnplnnsk ansteyqmlqrslalynkeekptryfrqvnliessnphpflnntewekcnnilsfyrs yleakknfleslkpedweknqyflmlkepktncetlvqgwkngfnlprgiftepirk wfmehrknitvaelkrvglvakviplffseeykdsvqpfynylfnvgninkpdekn flnceerrellrkkkdefkkmtdkekeenpsylefqswnkferelrlvrnqdivtwll cmelfnkkkikelnvekiylknintnttkkeknteekngeekiikeknnilnrimp mrlpikvygrenfsknkkkkirrntfftvyieekgtkllkqgnfkalerdrrlgglfsfv kthskaesksntisksrveyelgeyqkarieiikdmlaleetlidkynsldtdnfhnml tgwlklkdepdkasfqndvdlliavrnafshnqypmrnriafaninpfslssantsee kglgianqlkdkthktiekiieiekpietke
HMPREF9714_02132 [Myroides odoratimimus CCUG 12901] (SEQ ID No. 93)	EHO08761	mkdilttdttekqnrfyshkiadkyffggyfnlasnniyevfeevnkrntfgklakrd ngnlknyiihvfkdelsisdfekrvaifasyfpiletvdkksikernrtidltlsqrirqfr emlislvtavdqlrnfythyhhseivienkvldflnsslvstalhvkdkylktdktkefl ketiaaeldilieaykkkqiekkntrfkankredilnaiyneafwsfindkdkdketv vakgadayfeknhhksndpdfalnisekgivyllsffltnkemdslkanltgfkgkv dresgnsikymatqriysfhtyrglkqkirtseegvketllmqmidelskvpnvvyq hlsttqqnsfiedwneyykdyeddvetddlsrvihpvirkryedrfnyfairfldeffd fptlrfqvhlgdyvhdrrtkqlgkvesdriikekvtvfarlkdinsakanyfhsleeqd keeldnkwtlfpnpsydfpkehtlqhqgeqknagkigiyvklrdtqykekaaleea rkslnpkersatkaskydiitqiieandnvksekplvftgqpiaylsmndihsmlfsll tdnaelkktpeeveaklidqigkqineilskdtdtkilkkykdndlketdtdkitrdlar dkeeieklileqkqraddynytsstkfnidksrkrkhllfnaekgkigvwlandikrf mteefkskwkgyqhtelqklfayydtsksdldlilsdmvmvkdypielialvkksrt lvdflnkylearlgymenvitrvknsigtpqfktvrkecftflkksnytvvsldkqver ilsmplfiergfmddkptmlegksyqqhkekfadwfvhykensnyqnfydtevy eittedkrekakvtkkikqqqkndvftlmmvnymleevlklssndrlslnelyqtke erivnkqvakdtqernknyiwnkvvdlqlceglvridkvklkdignfrkyendsrv kefltyqsdivwsaylsnevdsnklyvierqldnyesirskellkevqeiecsvynqv ankeslkqsgnenfkqyvlqglvpigmdvremlilstdvkfikeeiiqlgqageveq dlysliyirnkfahnqlpikeffdfcennyrsisdneyyaeyymeifrsikekyts
HMPREF9711_00870	EKB06014	mkdilttdttekqnrfyshkiadkyffggyfnlasnniyevfeevnkrntfgklakrd ngnlknyiihvfkdelsisdfekrvaifasyfpiletvdkksikernrtidltlsqrirqfr

[Myroides odoratimimus CCUG 3837] (SEQ ID No. 94)		emlislvtavdqlrnfythyhhseivienkvldflnsslvstalhvkdkylktdktkefl ketiaaeldilieaykkkqiekkntrfkankredilnaiyneafwsfindkdkdketv vakgadayfeknhhksndpdfalnisekgivyllsffltnkemdslkanltgfkgkv dresgnsikymatqriysfhtyrglkqkirtseegvketllmqmidelskvpnvvyq hlsttqqnsfiedwneyykdyeddvetddlsrvihpvirkryedrfnyfairfldeffd fptlrfqvhlgdyvhdrrtkqlgkvesdriikekvtvfarlkdinsakasyfhsleeqd keeldnkwtlfpnpsydfpkehtlqhqgeqknagkigiyvklrdtqykekaaleea rkslnpkersatkaskydiitqiieandnvksekplvftgqpiaylsmndihsmlfsll tdnaelkktpeeveaklidqigkqineilskdtdtkilkkykdndlketdtdkitrdlar dkeeieklileqkqraddynytsstkfnidksrkrkhllfnaekgkigvwlandikrf mfkeskskwkgyqhtelqklfayfdtsksdlelilsdmvmvkdypielidlvrksrt lvdflnkylearlgyienvitrvknsigtpqfktvrkecfaflkesnytvasldkqieril smplfiergfmdskptmlegksyqqhkedfadwfvhykensnyqnfydtevyei itedkreqakvtkkikqqqkndvftlmmvnymleevlklpsndrlslnelyqtkee rivnkqvakdtqernknyiwnkvvdlqlceglvridkvklkdignfrkyendsrvk efltyqsdivwsgylsnevdsnklyvierqldnyesirskellkevqeiecivynqva nkeslkqsgnenfkqyvlqgllprgtdvremlilstdvkfkkeeimqlgqvreveqd lysliyirnkfahnqlpikeffdfcennyrpisdneyyaeyymeifrsikekyas
HMPREF9699_02005 [Bergeyella zoohelcum ATCC 43767] (SEQ ID No. 95)	EKB54193	menktslgnniyynpfkpqdksyfagyfnaamentdsvfrelgkrlkgkeytsenf fdaifkenislveyeryvkllsdyfpmarlldkkevpikerkenfkknfkgiikavrd lrnfythkehgeveitdeifgvldemlkstvltvkkkkvktdktkeilkksiekqldil cqkkleylrdtarkieekrrnqrergekelvapfkysdkrddliaaiyndafdvyidk kkdslkesskakyntksdpqqeegdlkipiskngvvfllslfltkqeihafkskiagfk atvideatvseatvshgknsicfmatheifshlaykklkrkvrtaeinygeaenaeqls vyaketlmmqmldelskvpdvvyqnlsedvqktfiedwneylkenngdvgtme eeqvihpvirkryedkfnyfairfldefaqfptlrfqvhlgnylhdsrpkenlisdrrik ekitvfgrlselehkkalfikntetnedrehyweifpnpnydfpkenisvndkdfpia gsildrekqpvagkigikvkllnqqyvsevdkavkahqlkqrkaskpsiqniieeiv pinesnpkeaivfggqptaylsmndihsilyeffdkwekkkeklekkgekelrkei gkelekkivgkiqaqiqqiidkdtnakilkpyqdgnstaidkeklikdlkqeqnilqk lkdeqtvrekeyndfiayqdknreinkvrdrnhkqylkdnlkrkypeaparkevly yrekgkvavwlandikrfmptdfknewkgeqhsllqkslayyeqckeelknllpe kvfqhlpfklggyfqqkylyqfytcyldkrleyisglvqqaenfksenkvfkkvene

		cfkflkkqnythkeldarvqsilgypiflergfmdekptiikgktfkgnealfadwfr yykeyqnfqtfydtenyplvelekkqadrkrktkiyqqkkndvftllmakhifksvf kqdsidqfsledlyqsreerlgnqerarqtgerntnyiwnktvdlklcdgkitvenvkl knvgdfikyeydqrvqaflkyeeniewqaflikeskeeenypyvvereieqyekvr reellkevhlieeyilekvkdkeilkkgdnqnfkyyilngllkqlknedvesykvfnl ntepedvninqlkqeatdleqkafvltyirnkfahnqlpkkefwdycqekygkiek ektyaeyfaevfkkekealik
HMPREF9151_01387 [Prevotella saccharolytica F0055] (SEQ ID No. 96)	EKY00089	mmekenvqgshiyyeptdkcfwaafynlarhnayltiahinsfvnskkginnddk vldiiddwskfdndllmgarlnklilkhfpflkaplyqlakrktrkqqgkeqqdyek kgdedpeviqeaianafkmanvrktlhaflkqledlrnhfshynynspakkmevk fddgfcnklyyvfdaalqmvkddnrmnpeinmqtdfehlvrlgrnrkipntfkyn ftnsdgtinnngllffvslflekrdaiwmqkkikgfkggtenymrmtnevfcrnrm vipklrletdydnhqlmfdmlnelvrcplslykrlkqedqdkfrvpiefldednead npyqenansdenpteetdplkntlvrhqhrfpyfvlryfdlnevfkqlrfqinlgcyh fsiydktigertekrhltrtlfgfdrlqnfsvklqpehwknmvkhldteessdkpylsd amphyqienekigihflktdtekketvwpsleveevssnrnkykseknltadaflst hellpmmfyyqllsseektraaagdkvqgvlqsyrkkifdiyddfangtinsmqkl derlakdnllrgnmpqqmlailehqepdmeqkakekldrlitetkkrigkledqfkq kvrigkrradlpkvgsiadwlvndmmrfqpakrnadntgvpdskansteyrllqea lafysaykdrlepyfrqvnliggtnphpflhrvdwkkcnhllsfyhdyleakeqyls hlspadwqkhqhflllkvrkdiqnekkdwkkslvagwkngfnlprglftesiktwf stdadkvqitdtklfenrvgliakliplyydkvyndkpqpfyqypfnindrykpedtr krftaassklwnekkmlyknaqpdssdkieypqyldflswkklerelrmlrnqdm mvwlmckdlfaqctvegvefadlklsqlevdvnvqdnlnvlnnvssmilplsvyp sdaqgnvlrnskplhtvyvqenntkllkqgnfksllkdrrlnglfsfiaaegedlqqhp ltknrleyelsiyqtmrisvfeqtlqlekailtrnktlcgnnfnnllnswsehrtdkktlq pdidfliavrnafshnqypmstntvmqgiekfniqtpkleekdglgiasqlakktkd aasrlqniinggtn
A343_1752 [Porphyromonas gingivalis	EOA10535	mteqnekpyngtyytledkhfwaaffnlarhnayitlthidrqlayskaditndedilf fkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnksskkkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdrcgnndnpffkhhfvdreekvte agllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrtdd

JCVI SC001] (SEQ ID No. 97)		wmlldmlnelvrcpkslydrlreedrarfrvpvdilsdeddtdgteedpfkntlvrhq drfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriqdfa eehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqllwpspev gatrtgrskyaqdkrftaeaflsvhelmpmmfyyfllrekyseeasaervqgrikrvi edvyavydafargeidtldrldacladkgirrghlprqmiailsqehkdmeekvrkk lqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpvakdt sgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetrwe shtnilsfyrsylkarkaflqsigrsdrvenhrfllllkepktdrqtlvagwkgefhlprgi fteavrdcliemgldevgsykevgfmakavplyferackdrvqpfydypfnvgnsl kpkkgrflskekraeewesgkerfrdleawshsaarriedafagienasrenkkkie qllqdlslwetfesklkvkadkiniaklkkeileakehpyldfkswqkferelrlvkn qdiitwmmcrdlmeenkvegldtgtlylkdirtdvheqgslnvlnrvkpmrlpvv vyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtgalam eqypisklrveyelakyqtarvcafeqtleleeslltryphlpdknfrkmleswsdpll dkwpdlhgnvrlliavrnafshnqypmydetlfssirkydpsspdaieermglnia hrlseevkqakemveriiqa
HMPREF1 981_03090 [Bacteroides pyogenes F0041] (SEQ ID No. 98)	ERI81700	mesiknsqkstgktlqkdppyfglylnmallnvrkvenhirkwlgdvallpeksgf hsllttdnlssakwtrfyyksrkflpflemfdsdkksyenrretteclдtidrqkissllk evygklqdirnafshyhiddqsvkhtaliissemhrfienaysfalqktrarftgvfvet dflqaeekgdnkkffaiggnegiklkdnalifliclfldreeafkflsratgfkstkekgf lavretfcalccrqpherllsvnpreallmdmlnelnrcpdilfemldekdqksflpll geeeqahilenslndelceaiddpfemiaslskrvryknrfpylmlryieeknllpfir fridlgclelasypkkmgeennyersvtdhamafgrltdfhnedavlqqitkgitdev rfslyapryaiynnkigfvrtggsdkisfptlkkkggeghcvaytlqntksfgfisiydl rkilllsfldkdkaknivsglleqcekhwkdlsenlfdairtelqkefpvplirytlprsk ggklvsskladkqekyeseferrkeklteilsekdfdlsqiprrmidewlnvlptsrek klkgyvetlkldcrerlrvfekrekgehpvpprigematdlakdiirmvidqgvkqri tsayyseiqrclaqyagddnrrhldsiirelrlkdtknghpflgkvlrpglghteklyqr yfeekkewleatfypaaspkrvprfvnpptgkqkelpliirnlmkerpewrdwkqr knshpidlpsqlfeneicrllkdkigkepsgklkwnemfklywdkefpngmqrfy rckrrvevfdkvveyeyseeggnykkyyealidevvrqkissskeksklqvedltls vrrvfkrainekeyqlrllceddrllfmavrdlydwkeaqldldkidnmlgepvsvs qviqleggqpdavikaecklkdvsklmrycydgrvkglmpyfanheatqeqvem

		elrhyedhrrrvfnwvfaleksvlkneklrrfyeesqggcehrrcidalrkaslvseee yeflvhirnksahnqfpdleigklppnvtsgfceciwskykaiicriipfidperrffgk lleqk
HMPREF1553_02065 [Porphyromonas gingivalis F0568] (SEQ ID No. 99)	ERJ65637	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnprsmgfisvhdlrklllmellcegsfsrm qsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykqei kgrkdklnsqllsafdmdqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlrlq kfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqyag eenrhqfraivaelrlldpssghpflsatmetahrytedfykcylekkrewlaktfyrp eqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlfds kimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyips dgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlrlv qeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegegg dnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeydr crikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildpenrffgkllnnmsqpindl
HMPREF1988_01768 [Porphyromonas gingivalis F0185] (SEQ ID No. 100)	ERJ81987	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnpqsmgfisvhdlrklllmellcegsfsr mqsgflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykq eikgrkdklnsqllsafdmnqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlr

		lrkfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqya geenrrqfraivaelhlldpssghpflsatmetahrytedfykcylekkrewlaktfyr peqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlf dskimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyi psdgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlr lvqeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegeg gdnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeyd rcrikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildhenrffgkllnnmsqpindl
HMPREF1990_01800 [Porphyromonas gingivalis W4087] (SEQ ID No. 101)	ERJ87335	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnprsmgfisvhdlrklllmellcegsfsrm qsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykqei kgrkdklnsqllsafdmdqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlrlq kfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqyag eenrhqfraivaelrlldpssghpflsatmetahrytedfykcylekkrewlaktfyrp eqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlfds kvmellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyips dgkkfadcythlmektvrdkkrelrtagkpvppdlaayikrsfhravnerefmlrlv qeddrlmlmainkimtdreedilpglknidsildkenqfslavhakvlekegeggd nslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeydrcr ikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpdesq ylilirnkaahnqfpcaaeipliyrdvsakvgsiegssakdlpegsslvdslwkkye miirkilpildpenrffgkllnnmsqpindl
M573_117042 [Prevotella	KJJ86756	mkmeddkkttestnmldnkhfwaaflnlarhnvyitvnhinkvlelknkkdqdiii dndqdilaikthwekvngdlnkterlrelmtkhfpfletaiytknkedkeevkqekq aeaqsleslkdclflfleklqearnyyshykysestkepmleegllekmynifddniq

intermedia ZT] (SEQ ID No. 102)		lvikdyqhnkdinpdedfkhldrkgqfkysfadnegnitesgllffvslflekkdaiw mqqkltgfkdnreskkkmthevfcrrrmllpklrlestqtqdwilldmlnelircpks lyerlqgeyrkkfnvpfdsadedydaeqepfkntlvrhqdrfpyfalryfdyneiftn lrfqidlgtyhfsiykkliggqkedrhlthklygferiqefakqnrpdewkalvkdldt yetsneryisettphyhlenqkigirfrngnkeiwpslktngennekskykldkpyq aeaflsvhellpmmfyylllkkeepnndkknasivegfikreirdmyklydafang einnigdlekycedkgipkrhlpkqmvailydepkdmvkeakrkqkemvkdtk kllatlekqtqeeiedggrnirllksgeiarwlvndmmrfqpvqkdnegnplnnska nsteyqmlqrslalynkeekptryfrqvnlinssnphpflkwtkweecnnilsfyrn yltkkieflnklkpedweknqyflklkepktnretlvqgwkngfnlprgiftepirew fkrhqndskeyekvealkrvglvtkviplffkeeyfkedaqkeinncvqpfysfpyn vgnihkpdekdflpseerkklwgdkkdkfkgykakvkskkltdkekeeyrsylef qswnkferelrlvrnqdivtwllctelidkmkveglnveelqklrlkdidtdtakqek nnilnrimpmqlpvtvyeiddshnivkdrplhtvyieetktkllkqgnfkalvkdrrl nglfsfvdtsskaelkdkpisksvveyelgeyqnarietikdmlllektlikkyeklpt dnfsdmlngwlegkdesdkarfqndvkllvavrnafshnqypmrnriafaninpf slssadiseekkldianqlkdkthkiikkiieiekpietke
A2033_10205 [Bacteroidetes bacterium GWA2_31_9] (SEQ ID No. 103)	OFX18020.1	menqtqkgkgiyyyytknedkhyfgsflnlannnieqiieefrirlslkdeknikeii nnyftdkksytdwerginilkeylpvidyldlaitdkefekidlkqketakrkyfrtnf sllidtiidlrnfythyfhkpisinpdvakfldknllnvcldikkqkmktdktkqalkd gldkelkklielkkaelkekkiktwnitenvegavyndafnhmvyknnagvtilkd yhksilpddkidselklnfsisglvfllsmflskkeieqfksnlegfkgkvigengeye iskfnnslkymathwifsyltfkglkqrvkntfdketllmqmidelnkvphevyqtl skeqqnefledineyvqdneenkksmensivvhpvirkryddkfnyfairfldefa nfptlkffvtagnfvhdkrekqiqgsmltsdrmikekinvfgklteiakyksdyfsne ntletsewelfpnpsylliqnnipvhidlihnteeakqcqiaidrikcttnpakkrntrk skeeiikiiyqknknikygdptallssnelpaliyellvnkksgkeleniivekivnqy ktiagfekgqnlsnslitkklkksepnedkinaekiilainreleitenklniiknnraef rtgakrkhifyskelgqeatwiaydlkrfmpeasrkewkgfhhselqkflafydrnk ndakallnmfwnfdndqligndlnsafrefhfdkfyekylikrdeilegfksfisnfk depkllkkgikdiyrvfdkryyiikstnaqkeqllskpiclprgifdnkptyiegvkve snsalfadwyqytysdkhefqsfydmprdykeqfekfelnniksiqnkknlnksd kfiyfrykqdlkikqiksqdlfiklmvdelfnvvfknnielnlkklyqtsderfknqli

		advqknrekgdtsdnkmnenfiwnmtiplslcngqieepkvklkdigkfrkletdd kviqlleydkskvwkkleiedelenmpnsyerirrekllkgiqefehfllekekfdgi nhpkhfeqdlnpnfktyvingvlrknsklnyteidklldlehisikdietsakeihlayf lihvrnkfghnqlpkleafelmkkyykknneetyaeyfhkvssqivnefknslekh s
SAMN05421542_0666 [Chryseobacterium jejuense] (SEQ ID No. 104)	SDI27289.1	mektqtglgiyydhtklqdkyffggffnlaqnnidnvikafiikffperkdkdiniaq fldicfkdndadsdfqkknkflrihfpvigfltsdndkagfkkkfalllktiselrnfyth yyhksiefpselfellddifvkttseikklkkkddktqqllnknlseeydiryqqqierl kelkaqgkrvsltdetairngvfnaafnhliyrdgenvkpsrlyqssysepdpaengi slsqnsilfllsmflerketedlksrvkgfkakiikqgeeqisglkfmathwvfsylcf kgikqklstefheetlliqiidelskvpdevysafdsktkekfledineymkegnadls ledskvihpvirkryenkfnyfairfldeylsstslkfqvhvgnyvhdrrvkhingtgf qterivkdrikvfgrlsnisnlkadyikeqlelpndsngweifpnpsyifidnnvpih vladeatkkgielfkdkrrkeqpeelqkrkgkiskynivsmiykeakgkdklridep lallslneipallyqilekgatpkdieliiknklterfekiknydpetpapasqiskrlrnn ttakgqealnaeklsllereientetklssieekrlkakkeqrrntpqrsifsnsdlgriaa wladdikrfmpaeqrknwkgyqhsqlqqslayfekrpqeaflllkegwdtsdgss ywnnwvmnsflennhfekfyknylmkrvkyfselagnikqhthntkflrkfikqq mpadlfpkrhyilkdleteknkvlskplvfsrglfdnnptfikgvkvtenpelfaewy sygyktehvfqhfygwerdynelldselqkgnsfaknsiyynresqldliklkqdlki kkikiqdlflkriaeklfenvfnypttlsldefyltqeeraekerialaqslreegdnspni ikddfiwsktiafrskqiyepaiklkdigkfnrfvlddeeskaskllsydknkiwnke qlerelsigensyevirreklfkeiqnlelqilsnwswdginhprefemedqkntrhp nfkmylvngilrkninlykededfwleslkendfktlpsevletksemvqllflvilir nqfahnqlpeiqfynfirknypeiqnntvaelylnliklavqklkdns
SAMN05444360_11366 [Chryseobacterium carnipullorum] (SEQ	SHM52812.1	mntrvtgmgvsydhtkkedkhffggflnlaqdnitavikafcikfdknpmssvqfa escftdkdsdtdfqnkvryvrthlpvigylnyggdrntfrqklstllkavdslrnfythy yhsplalstelfelldtvfasvavevkqhkmkddktrqllskslaeeldirykqqlerlk elkeqgknidlrdeagirngvlnaafnhliykegeiakptlsyssfyygadsaengiti sqsgllfllsmflgkkeiedlksrirgfkakivrdgeenisglkfmathwifsylsfkg mkqrlstdfheetlliqiidelskvpdevyhdfdtatrekfvedineyiregnedfslg dstiihpvirkryenkfnyfavrfldefikfpslrfqvhlgnfvhdrrikdihgtgfqter vvkdrikvfgklseisslkteyiekeldldsdtgweifpnpsyvfidnnipiyistnktf

ID No. 105)		kngssefiklrrkekpeemkmrgedkkekrdiasmignagslnsktplamlslne mpallyeilvkkttpeeieliikekldshfeniknydpekplpasqiskrlrnnttdkg kkvinpeklihlinkeidateakfallaknrkelkekfrgkplrqtifsnmelgreatwl addikrfmpdilrknwkgyqhnqlqqslaffnsrpkeaftilqdgwdfadgssfwn gwiinsfvknrsfeyfyeayfegrkeyfsslaenikqhtsnhrnlrrfidqqmpkglf enrhyllenleteknkilskplvfprglfdtkptfikgikvdeqpelfaewyqygyste hvfqnfygwerdyndlleselekdndfsknsihysrtsqleliklkqdlkikkikiqdl flkliaghifenifkypasfsldelyltqeerlnkeqealiqsqrkegdhsdniikdnfig sktvtyeskqisepnvklkdigkfnrfllddkvktllsynedkvwnkndldlelsige nsyevirreklfkkiqnfelqtltdwpwngtdhpeefgttdnkgvnhpnfkmyvv ngilrkhtdwfkegednwlenlnethfknlsfqeletksksiqtafliimirnqfahnq lpavqffefiqkkypeiqgsttselylnfinlavvellellek
SAMN05421786_1011119 [Chryseobacterium ureilyticum] (SEQ ID No. 106)	SIS70481.1	metqilgngisydhtktedkhffggflntaqnnidllikayiskfessprklnsvqfpd vcfkkndsdadfqhklqfirkhlpviqylkyggnrevlkekfrlllqavdslrnfythf yhkpiqlpnelltlldtifgeignevrqnkmkddktrhllkknlseeldfryqeqlerlr klksegkkvdlrdteairngvlnaafnhlifkdaedfkptvsyssyyydsdtaengisi sqsgllfllsmflgrremedlksrvrgfkariikheeqhvsglkfmathwvfsefcfk giktrlnadyheetlliqlidelskvpdelyrsfdvatrerfiedineyirdgkedkslies kivhpvirkryeskfnyfairfldefvnfptlrfqvhagnyvhdrriksiegtgfkterl vkdrikvfgklstisslkaeylakavnitddtgwellphpsyvfidnnipihltvdpsf kngvkeyqekrklqkpeemknrqggdkmhkpaisskigkskdinpespvalls mneipallyeilvkkaspeeveakirqkltavferirdydpkvplpasqvskrlrnnt dtlsynkeklvelankeveqterklalitknrrecrekvkgkfkrqkvfknaelgteat wlandikrfmpeeqkknwkgyqhsqlqqslaffesrpgearsllqagwdfsdgssf wngwvmnsfardntfdgfyesylngrmkyflrladniaqqsstnklisnfikqqm pkglfdrrlymledlateknkilskplifprgifddkptfkkgvqvseepeafadwys ygydvkhkfqefyawdrdyeellreelekdtaftknsihysresqiellakkqdlkvk kvriqdlylklmaeflfenvfghelalpldqfyltqeerlkqeqeaivqsqrpkgdds pnivkenfiwsktipfksgrvfepnvklkdigkfrnlltdekvdillsynnteigkqvi eneliigagsyefirreqlfkeiqqmkrlslrsvrgmgvpirlnlk
Prevotella buccae	WP_004343581	mqkqdklfvdrkknaifafpkyitimenqekpepiyyeltdkhfwaaflnlarhnv yttinhinrrleiaelkddgymmdikgswneqakkldkkvrlrdlimkhfpfleaaa yeitnskspnnkeqrekeqsealslnnlknvlfifleklqvlrnyyshykyseespkp

(SEQ ID No. 107)		ifetsllknmykvfdanvrlvkrdymhhenidmqrdfthlnrkkqvgrtkniidsp nfhyhfadkegnmtiagllffvslfldkkdaiwmqkklkgfkdgrnlreqmtnevf crsrislpklklenvqtkdwmqldmlnelvrcpkslyerlrekdresfkvpfdifsdd ydaeeepfkntlvrhqdrfpyfvlryfdlneifeqlrfqidlgtyhfsiynkrigdedev rhlthhlygfariqdfaqqnqpevwrklvkdldyfeasqepyipktaphyhleneki gikfcsthnnlfpslktektcngrskfnlgtqftaeaflsvhellpmmfyyllltkdysr kesadkvegiirkeisniyaiydafangeinsiadltcrlqktnilqghlpkqmisileg rqkdmekeaerkigemiddtqrrldllckqtnqkirigkrnagllksgkiadwlvnd mmrfqpvqkdqnnipinnskansteyrmlqralalfgsenfrlkayfnqmnlvgn dnphpflaetqwehqtnilsfyrnylearkkylkglkpqnwkqyqhflilkvqktnr ntlvtgwknsfnlprgiftqpirewfekhnnskriydqilsfdrvgfvakaiplyfaee ykdnvqpfydypfnignklkpqkgqfldkkervelwqknkelfknypsekkktdl ayldflswkkferelrliknqdivtwlmfkelfnmatveglkigeihlrdidtntanee snnilnrimpmklpvktyetdnkgnilkerplatfyieetetkvlkqgnfkvlakdrrl ngllsfaettdidleknpitklsvdhelikyqttrisifemtlglekklinkyptlptdsfrn mlerwlqckanrpelknyvnsliavrnafshnqypmydatlfaevkkftlfpsvdtk kielniapqlleivgkaikeieksenkn
Porphyromonas gingivalis (SEQ ID No. 108)	WP_005873511	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnpqsmgfisvhnlrklllmellcegsfsr mqsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykq eikgrkdklnsqllsafdmnqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlr lrkfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqya geenrrqfraivaelhlldpssghpflsatmetahrytedfykcylekkrewlaktfyr peqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlf dskimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyi psdgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlr lvqeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegeg

		gdnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeyd rcrikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildpenrffgkllnnmsqpindl
Porphyromonas gingivalis (SEQ ID No. 109)	WP_005874195	mteqnekpyngtyytledkhfwaaffnlarhnayitlahidrqlayskaditndedil ffkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnksskkkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdkygnndnpffkhhfvdreekvt eagllffvslflekrdaiwmqkkirgfkggteayqqmtnevfcrsrislpklkleslrtd dwmlldmlnelvrcpkslydrlreedrarfrvpvdilsdeddtdgteedpfkntlvrh qdrfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriqdf aeehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqllwpspe vgatrtgrskyaqdkrftaeaflsvhelmpmmfyyfllrekyseeasaekvqgrikr viedvyavydafardeintrdeldacladkgirrghlprqmiailsqehkdmeekvr kklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpva kdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetr weshtnilsfyrsylkarkaflqsigrsdreenhrfllllkepktdrqtlvagwksefhlp rgifteavrdcliemgydevgsykevgfmakavplyferackdrvqpfydypfnv gnslkpkkgrflskekraeewesgkerfrdleawshsaarriedafvgieyaswenk kkieqllqdlslwetfesklkvkadkiniaklkkeileakehpyhdfkswqkferelrl vknqdiitwmmcrdlmeenkvegldtgtlylkdirtdvqeqgslnvlnhvkpmrl pvvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtgal ameqypisklrveyelakyqtarvcafeqtleleeslltryphlpdesfremleswsd plldkwpdlqrevrlliavrnafshnqypmydetifssirkydpssldaieermglni ahrlseevklakemveriiqa
Prevotella pallens (SEQ ID No. 110)	WP_006044833	mkeeekgktpvvstynkddkhfwaaflnlarhnvyitvnhinkilgegeinrdgye ntlekswneikdinkkdrlskliikhfpflevttyqrnsadttkqkeekqaeaqslesl kksffvfiyklrdlrnhyshykhskslerpkfeedlqekmynifdasiqlvkedykh ntdikteedfkhldrkgqfkysfadnegnitesgllffvslflekkdaiwvqkklegfk csnesyqkmtnevfcrsrmllpklrlqstqtqdwilldmlnelircpkslyerlreedr kkfrvpieiadedydaeqepfknalvrhqdrfpyfalryfdyneiftnlrfqidlgtyh fsiykkqigdykeshhlthklygferiqeftkqnrpdewrkfvktfnsfetskepyip ettphyhlenqkigirfrndndkiwpslktnseknekskykldksfqaeaflsvhell

		pmmfyylllktentdndneietkkkenkndkqekhkieeiienkiteiyalydafan gkinsidkleeyckgkdieighlpkqmiailksehkdmateakrkqeemladvqk slesldnqineeienverknsslksgeiaswlvndmmrfqpvqkdnegnplnnsk ansteyqmlqrslalynkeekptryfrqvnliessnphpflnntewekcnnilsfyrs yleakknfleslkpedweknqyflmlkepktncetlvqgwkngfnlprgiftepirk wfmehrknitvaelkrvglvakviplffseeykdsvqpfynylfnvgninkpdekn flnceerrellrkkkdefkkmtdkekeenpsylefqswnkferelrlvrnqdivtwll cmelfnkkkikelnvekiylknintnttkkeknteekngeekiikeknnilnrimp mrlpikvygrenfsknkkkkirrntfftvyieekgtkllkqgnfkalerdrrlgglfsfv kthskaesksntisksrveyelgeyqkarieiikdmlaleetlidkynsldtdnfhnml tgwlklkdepdkasfqndvdlliavrnafshnqypmrnriafaninpfslssantsee kglgianqlkdkthktiekiieiekpietke
Myroides odoratimimus (SEQ ID No. 111)	WP_006261414	mkdilttdttekqnrfyshkiadkyffggyfnlasnniyevfeevnkrntfgklakrd ngnlknyiihvfkdelsisdfekrvaifasyfpiletvdkksikernrtidltlsqrirqfr emlislvtavdqlrnfythyhhseivienkvldflnsslvstalhvkdkylktdktkefl ketiaaeldilieaykkkqiekkntrfkankredilnaiyneafwsfindkdkdketv vakgadayfeknhhksndpdfalnisekgivyllsffltnkemdslkanltgfkgkv dresgnsikymatqriysfhtyrglkqkirtseegvketllmqmidelskvpnvvyq hlsttqqnsfiedwneyykdyeddvetddlsrvihpvirkryedrfnyfairfldeffd fptlrfqvhlgdyvhdrrtkqlgkvesdriikekvtvfarlkdinsakanyfhsleeqd keeldnkwtlfpnpsydfpkehtlqhqgeqknagkigiyvklrdtqykekaaleea rkslnpkersatkaskydiitqiieandnvksekplvftgqpiaylsmndihsmlfsll tdnaelkktpeeveaklidqigkqineilskdtdtkilkkykdndlketdtdkitrdlar dkeeieklileqkqraddynytsstkfnidksrkrkhllfnaekgkigvwlandikrf mteefkskwkgyqhtelqklfayydtsksdldlilsdmvmvkdypielialvkksrt lvdflnkylearlgymenvitrvknsigtpqfktvrkecftflkksnytvvsldkqver ilsmplfiergfmddkptmlegksyqqhkekfadwfvhykensnyqnfydtevy eittedkrekakvtkkikqqqkndvftlmmvnymleevlklssndrlslnelyqtke erivnkqvakdtqernknyiwnkvvdlqlceglvridkvklkdignfrkyendsrv kefltyqsdivwsaylsnevdsnklyvierqldnyesirskellkevqeiecsvynqv ankeslkqsgnenfkqyvlqglvpigmdvremlilstdvkfikeeiiqlgqageveq dlysliyirnkfahnqlpikeffdfcennyrsisdneyyaeyymeifrsikekyts

Myroides odoratimimus (SEQ ID No. 112)	WP_006265509	mkdilttdttekqnrfyshkiadkyffggyfnlasnniyevfeevnkrntfgklakrdngnlknyiihvfkdelsisdfekrvaifasyfpiletvdkksikernrtidltlsqrirqfremlislvtavdqlrnfythyhhseivienkvldflnsslvstalhvkdkylktdktkeflketiaaeldilieaykkkqiekkntrfkankredilnaiyneafwsfindkdkdketvvakgadayfeknhhksndpdfalnisekgivyllsffltnkemdslkanltgfkgkvdresgnsikymatqriysfhtyrglkqkirtseegvketllmqmidelskvpnvvyqhlsttqqnsfiedwneyykdyeddvetddlsrvihpvirkryedrfnyfairfldeffdfptlrfqvhlgdyvhdrrtkqlgkvesdriikekvtvfarlkdinsakasyfhsleeqdkeeldnkwtlfpnpsydfpkehtlqhqgeqknagkigiyvklrdtqykekaaleearkslnpkersatkaskydiitqiieandnvksekplvftgqpiaylsmndihsmlfslltdnaelkktpeeveaklidqigkqineilskdtdtkilkkykdndlketdtdkitrdlardkeeieklileqkqraddynytsstkfnidksrkrkhllfnaekgkigvwlandikrfmfkeskskwkgyqhtelqklfayfdtsksdlelilsdmvmvkdypielidlvrksrtlvdflnkylearlgyienvitrvknsigtpqfktvrkecfaflkesnytvasldkqierilsmplfiergfmdskptmlegksyqqhkedfadwfvhykensnyqnfydtevyeiitedkreqakvtkkikqqqkndvftlmmvnymleevlklpsndrlslnelyqtkeerivnkqvakdtqernknyiwnkvvdlqlceglvridkvklkdignfrkyendsrvkefltyqsdivwsgylsnevdsnklyvierqldnyesirskellkevqeiecivynqvankeslkqsgnenfkqyvlqgllprgtdvremlilstdvkfkkeeimqlgqvreveqdlysliyirnkfahnqlpikeffdfcennyrpisdneyyaeyymeifrsikekyas
Prevotella sp. MSX73 (SEQ ID No. 113)	WP_007412163	mqkqdklfvdrkknaifafpkyitimenqekpepiyyeltdkhfwaaflnlarhnvyttinhinrrleiaelkddgymmgikgswneqakkldkkvrlrdlimkhfpfleaaayeitnskspnnkeqrekeqsealslnnlknvlfifleklqvlrnyyshykyseespkpifetsllknmykvfdanvrlvkrdymhhenidmqrdfthlnrkkqvgrtkniidspnfhyhfadkegnmtiagllffvslfldkkdaiwmqkklkgfkdgrnlreqmtnevfcrsrislpklklenvqtkdwmqldmlnelvrcpkslyerlrekdresfkvpfdifsddydaeeepfkntlvrhqdrfpyfvlryfdlneifeqlrfqidlgtyhfsiynkrigdedevrhlthhlygfariqdfapqnqpeewrklvkdldhfetsqepyisktaphyhlenekigikfcsthnnlfpslkrektcngrskfnlgtqftaeaflsvhellpmmfyyllltkdysrkesadkvegiirkeisniyaiydafanneinsiadltcrlqktnilqghlpkqmisilegrqkdmekeaerkigemiddtqrrldllckqtnqkirigkrnagllksgkiadwlvsdmmrfqpvqkdtnnapinnskansteyrmlqhalalfgsessrlkayfrqmnlvgnanphpflaetqwehqtnilsfyrnylearkkylkglkpqnwkqyqhflilkvqktnr

		ntlvtgwknsfnlprgiftqpirewfekhnnskriydqilsfdrvgfvakaiplyfaee ykdnvqpfydypfnignklkpqkgqfldkkervelwqknkelfknypseknktdl ayldflswkkferelrliknqdivtwlmfkelfktttveglkigeihlrdidtntanees nnilnrimpmklpvktyetdnkgnilkerplatfyieetetkvlkqgnfkvlakdrrl ngllsfaettdidleknpitklsvdyelikyqttrisifemtlglekklidkystlptdsfrn mlerwlqckanrpelknyvnsliavrnafshnqypmydatlfaevkkftlfpsvdtk kielniapqlleivgkaikeieksenkn
Porphyromonas gingivalis (SEQ ID No. 114)	WP_012458414	mteqnerpyngtyytledkhfwaaffnlarhnayitlahidrqlayskaditndedilf fkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnksskkkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdrygnndnpffkhhfvdreekvte agllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrtdd wmlldmlnelvrcpkslydrlreedrarfrvpvdilsdeddtdgteedpfkntlvrhq drfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriqdfa eehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqhlwpspev gatrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekysdeasaervqgrikrvi edvyavydafargeintrdeldacladkgirrghlprqmigilsqehkdmeekvrk klqemivdtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpvak dtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetrw eshtnilsfyrsylkarkaflqsigrsdrvenhrflllkepktdrqtlvagwkgefhlprg ifteavrdcliemgldevgsykevgfmakavplyferackdrvqpfydypfnvgns lkpkkgrflskekraeewesgkerfrlaklkkeileakehpyldfkswqkferelrlv knqdiitwmicrdlmeenkvegldtgtlylkdirtdvqeqgnlnvlnrvkpmrlpv vvyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtgala meqypisklrveyelakyqtarvcafeqtleleeslltryphlpdknfrkmleswsdp lldkwpdlhgnvrlliavrnafshnqypmydeavfssirkydpsspdaieermgln iahrlseevkqakemaeriiqa
Paludibacter propionicigenes (SEQ ID No. 115)	WP_013446107	mktsanniyfnginsfkkifdskgaiapiaekscrnfdikaqndvnkeqrihyfavg htfkqldtenlfeyvldenlrakrptrfislqqfdkefienikrlisdirninshyihrfdpl kidavptniidflkesfelaviqiylkekginylqfsenphadqklvaflhdkflplde kktsmlqnetpqlkeykeyrkyfktlskqaaidqllfaeketdyiwnlfdshpvltisa gkylsfysclfllsmflykseanqliskikgfkknttеeekskreiftffskrfnsmdid seenqlvkfrdlilylnhypvawnkdleldssnpamtdklkskiieleinrsfplyeg

		nerfatfakyqiwgkkhlgksiekeyinasftdeeitaytyetdtcpelkdahkkladl kaakglfgkrkeknesdikktetsirelqhepnpikdkliqrieknlltvsygrnqdrf mdfsarflaeinyfgqdasfkmyhfyatdeqnselekyelpkdkkkydslkfhqg klvhfisykehlkryeswddafviennaiqlklsfdgventvtiqralliylledalrni qnntaenagkqllqeyyshnkadlsafkqiltqqdsiepqqktefkkllprrllnnysp ainhlqtphsslplilekallaekrycslvvkakaegnyddfikrnkgkqfklqfirka wnlmyfrnsylqnvqaaghhksfhierdefndfsrymfafeelsqykyylnemfe kkgffennefkilfqsgtslenlyektkqkfeiwlasntaktnkpdnyhlnnyeqqfs nqlffinlshfinylkstgklqtdangqiiyealnnvqylipeyyytdkpersesksgn klynklkatkledallyemamcylkadkqiadkakhpitkllltsdvefnitnkegiql yhllvpfkkidafiglkmhkeqqdkkhptsflanivnylelvkndkdirktyeafstn pvkrtltyddlakidghlisksikftnvtleleryfifkeslivkkgnnidfkyikglrny ynnekkknegirnkafhfgipdsksydqlirdaevmfianevkpthatkytdlnkql htvcdklmetvhndyfskegdgkkkreaagqkyfeniisak
Porphyromonas gingivalis (SEQ ID No. 116)	WP_013816155	mteqnekpyngtyytledkhfwaaffnlarhnayitlahidrqlayskaditndedil ffkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnkssknkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdrygnndnpffkhhfvdregtvte agllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrtdd wmlldmlnelvrcpkslydrlreedrarfrvpvdilsdeedtdgaeedpfkntlvrhq drfpyfalryfdlkkvftslrfqidlgtyhfaiykknigeqpedrhltrnlygfgriqdfa eehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqhlwpspev gatrtgrskyaqdkrftaeaflsahelmpmmfyyfllrekyseeasaervqgrikrvi edvyavydafardeintrdeldacladkgirrghlprqmigilsqehkdmeekirkk lqemmadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpvak dtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetrw eshtnilsfyrsylkarkaflqsigrsdrvenhrfllllkepktdrqtlvagwkgefhlprg ifteavrdcliemgldevgsykevgfmakavplyferackdwvqpfynypfnvgn slkpkkgrflskekraeewesgkerfrlaklkkeileakehpyldfkswqkferelrl vknqdiitwmicgdlmeenkvegldtgtlylkdirtdvqeqgslnvlnrvkpmrlp vvvyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtgala meqypisklrveyelakyqtarvcafeqtleleeslltrcphlpdknfrkmleswsdp

		lldkwpdlhrkvrlliavrnafshnqypmydeavfssirkydpsfpdaieermglni ahrlseevkqaketveriiqa
Flavobacterium columnare (SEQ ID No. 117)	WP_014165541	mssknesynkqktfnhykqedkyffggflnnaddnlrqvgkefktrinfnhnnnel asvfkdyfnkeksvakrehalnllsnyfpvleriqkhtnhnfeqtreifellldtikklrd yythhyhkpitinpkiydflddtlldvlitikkkkvkndtsrellkeklrpeltqlknqk reelikkgkklleenlenavfnhclrpfleenktddkqnktvslrkyrkskpneetsitl tqsglvflmsfflhrkefqvftsglegfkakvntikeeeislnknnivymithwsysy ynfkglkhriktdqgvstleqnntthsltntntkealltqivdylskvpneiyetlsekq qkefeedineymrenpenedstfssivshkvirkryenkfnyfamrfldeyaelptlr fmvnfgdyikdrqkkilesiqfdseriikkeihlfeklslvteykknvylketsnidlsr fplfpnpsyvmannnipfyidsrsnnldeylnqkkkaqsqnkkrnltfekynkeqs kdaiiamlqkeigvkdlqqrstigllscnelpsmlyevivkdikgaelenkiaqkire qyqsirdftldspqkdnipttliktintdssvtfenqpidiprlknaiqkeltltqekllnv keheievdnynrnkntykfknqpknkvddkklqrkyvfyrneirqeanwlasdli hfmknkslwkgymhnelqsflaffedkkndcialletvfnlkedciltkglknlflkh gnfidfykeylklkedflntestflengliglppkilkkelskrfkyifivfqkrqfiikel eekknnlyadainlsrgifdekptmipfkkpnpdefaswfvasyqynnyqsfyelt pdiverdkkkkyknlrainkvkiqdyylklmvdtlyqdlfnqpldkslsdfyvska erekikadakayqkrndsslwnkvihlslqnnritanpklkdigkykralqdekiatl ltyddrtwtyalqkpekenendykelhytalnmelqeyekvrskellkqvqelekqi leeytdflstqihpadferegnpnfkkylahsileneddldklpekveamreldetitn piikkaivliiirnkmahnqyppkfiydlanrfvpkkeeeyfatyfnrvfetitkelwe nkekkdktqv
Psychroflexus torquis (SEQ ID No. 118)	WP_015024765	mesiiglglsfnpyktadkhyfgsflnlvennlnavfaefkerisykakdenissliek hfidnmsivdyekkisilngylpiidflddelennlntrvknfkknfiilaeaieklrdy ythfyhdpitfednkepllelldevllktildvkkkylktdktkeilkdslreemdllvir ktdelrekkktnpkiqhtdssqiknsifndafqgllyedkgnnkktqvshraktrlnp kdihkqeerdfeiplstsglvflmslflskkeiedfksnikgfkgkvvkdenhnslky mathrvysilafkglkyriktdtfsketlmmqmidelskvpdcvyqnlsetkqkdfi edwneyfkdneentenlensrvvhpvirkryedkfnyfairfldefanfktlkfqvf mgyyihdqrtktigttnittertvkekinvfgklskmdnlkkhffsqlsddentdwef fpnpsynfltqadnspannipiylelknqqiikekdaikaevnqtqnrnpnkpskrd llnkilktyedfhqgdptailslneipallhlflvkpnnktgqqieniirikiekqfkain

		hpsknnkgipkslfadtnvrvnaiklkkdleaeldmlnkkhiafkenqkassnydk llkehqftpknkrpelrkyvfyksekgeeatwlandikrfmpkdfktkwkgcqhse lqrklafydrhtkqdikellsgcefdhslldinayfqkdnfedffskylenrietlegvlk klhdfkneptplkgvfkncfkflkrqnyvtespeiikkrilakptflprgvfderptmk kgknplkdknefaewfveylenkdyqkfynaeeyrmrdadfkknavikkqklkd fytlqmvnyllkevfgkdemnlqlselfqtrqerlklqgiakkqmnketgdssentr nqtyiwnkdvpvsffngkvtidkvklknigkykryerdervktfigyevdekwm mylphnwkdrysvkpinvidlqiqeyeeirshellkeiqnleqyiydhttdknillqd gnpnfkmyvlnglligikqvnipdfivlkqntnfdkidftgiascselekktiiliairn kfahnqlpnkmiydlaneflkieknetyanyylkvlkkmisdla
Riemerella anatipestifer (SEQ ID No. 119)	WP_015345620	mffsfhnaqrvifkhlykafdaslrmvkedykahftvnltrdfahlnrkgknkqdn pdfnryrfekdgfftesgllfftnlfldkrdaywmlkkvsgfkashkqrekmttevfc rsrillpklrlesrydhnqmlldmlselsrcpkllyeklseenkkhfqveadgfldeie eeqnpfkdtlirhqdrfpyfalryldlnesfksirfqvdlgtyhyciydkkigdeqekr hltrtllsfgrlqdfteinrpqewkaltkdldyketsnqpfiskttphyhitdnkigfrlgt skelypsleikdganriakypynsgfvahafisvhellplmfyqhltgksedllketvr hiqriykdfeeerintiedlekanqgrlplgafpkqmlgllqnkqpdlsekakikiekl iaetkllshrlntklksspklgkrreklktgvladwlvkdfmrfqpvaydaqnqpik sskanstefwfirralalyggeknrlegyfkqtnligntnphpflnkfnwkacrnlvdf yqqyleqrekfleaikhqpwepyqyclllkvpkenrknlvkgweqggislprglfte airetlskdltlskpirkeikkhgrvgfisraitlyfkekyqdkhqsfynlsykleakapl lkkeehyeywqqnkpqsptesqrlelhtsdrwkdyllykrwqhlekklrlyrnqdi mlwlmtleltknhfkelnlnyhqlklenlavnvqeadaklnplnqtlpmvlpvkvy pttafgevqyhetpirtvyireeqtkalkmgnfkalvkdrrlnglfsfikeendtqkhp isqlrlrreleiyqslrvdafketlsleekllnkhaslsslenefrtlleewkkkyaassm vtdkhiafiasvrnafchnqypfyketlhapillftvaqpteekdglgiaeallkvlre yceivksqi
Prevotella pleuritidis (SEQ ID No. 120)	WP_021584635	mendkrleesacytlndkhfwaaflnlarhnvyitvnhinktlelknkknqeiiidnd qdilaikthwakvngdlnktdrlrelmikhfpfleaaiysnnkedkeevkeekqaka qsfkslkdclflfleklqearnyyshykysesskepefeegllekmyntfdasirlvke dyqynkdidpekdfkhlerkedfnylftdkdnkgkitknglffvslflekkdaiwm qqkfrgfkdnrgnkekmthevfcrsrmllpkirlestqtqdwilldmlnelircpksl yerlqgayrekfkvpfdsidedydaeqepfrntlvrhqdrfpyfalryfdyneifknlr

		fqidlgtyhfsiykkliggkkedrhlthklygferiqeftkqnrpdkwqaiikdldtye tsneryisettphyhlenqkigirfrndnndiwpslktngeknekskynldkpyqae aflsvhellpmmfyylllkmentdndkednevgtkkkgnknnkqekhkieeiien kikdiyalydaftngeinsidelaeqregkdieighlpkqlivilknkskdmaekanr kqkemikdtkkrlatldkqvkgeiedggrnirllksgeiarwlvndmmrfqpvqk dnegkplnnskansteyqmlqrslalynkeekptryfrqvnlikssnphpfledtkw eecynilsfyrnylkakikflnklkpedwkknqyflmlkepktnrktlvqgwkngf nlprgiftepikewfkrhqndseeykkvealdrvglvakviplffkeeyfkedaqke inncvqpfysfpynvgnihkpeeknflhceerrklwdkkkdkfkgykakekskk mtdkekeehrsylefqswnkferelrlvrnqdiltwllctklidklkidelnieelqklrl kdidtdtakkeknnilnrvmpmrlpvtvyeidksfnivkdkplhtvyieetgtkllk qgnfkalvkdrrlnglfsfvktsseaeskskpisklrveyelgayqkaridiikdmlal ektlidndenlptnkfsdmlkswlkgkgeankarlqndvgllvavrnafshnqyp mynsevfkgmkllslssdipekeglgiakqlkdkiketieriieiekeirn
Porphyromonas gingivalis (SEQ ID No. 121)	WP_021663197	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnprsmgfisvhdlrklllmellcegsfsrm qsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykqei kgrkdklnsqllsafdmdqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlrlq kfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqyag eenrhqfraivaelrlldpssghpflsatmetahrytedfykcylekkrewlaktfyrp eqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlfds kimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyips dgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlrlv qeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegegg dnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeydr crikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde

		sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildpenrffgkllnnmsqpindl
Porphyromonas gingivalis (SEQ ID No. 122)	WP_021665475	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtnenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnpqsmgfisvhdlrklllmellcegsfsr mqsgflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykq eikgrkdklnsqllsafdmnqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlr lrkfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqya geenrrqfraivaelhlldpssghpflsatmetahrytedfykcylekkrewlaktfyr peqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlf dskimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyi psdgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlr lvqeddrlmlmainkmmtdreedilpglknidsildkenqfslavhakvlekegeg gdnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeyd rcrikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildhenrffgkllnnmsqpindl
Porphyromonas gingivalis (SEQ ID No. 123)	WP_021677657	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnpqsmgfisvhdlrklllmellcegsfsr mqsgflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykq eikgrkdklnsqllsafdmnqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlr lrkfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqya

		geenrrqfraivaelhlldpssghpflsatmetahrytedfykcylekkrewlaktfyr peqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlf dskimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyi psdgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlr lvqeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegeg gdnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeyd rcrikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildhenrffgkllnnmsqpindl
Porphyromonas gingivalis (SEQ ID No. 124)	WP_021680012	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnprsmgfisvhdlrklllmellcegsfsrm qsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykqei kgrkdklnsqllsafdmdqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlrlq kfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqyag eenrhqfraivaelrlldpssghpflsatmetahrytedfykcylekkrewlaktfyrp eqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlfds kvmellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyips dgkkfadcythlmektvrdkkrelrtagkpvppdlaayikrsfhravnerefmlrlv qeddrlmlmainkimtdreedilpglknidsildkenqfslavhakvlekegeggd nslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeydrcr ikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpdesq ylilirnkaahnqfpcaaeipliyrdvsakvgsiegssakdlpegsslvdslwkkye miirkilpildpenrffgkllnnmsqpindl
Porphyromonas gingivalis	WP_023846767	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd

(SEQ ID No. 125)		lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnprsmgfisvhdlrkllmellcegsfsrm qsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykqei kgrkdklnsqllsafdmnqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlrlr kfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqyag eenrrqfraivaelhlldpssghpflsatmetahrytedfykcylekkrewlaktfyrp eqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlfds kimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyips dgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlrlv qeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegegg dnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeydr crikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildpenrffgkllnnmsqpindl
Prevotella falsenii (SEQ ID No. 126)	WP_036884929	mkndnnstkstdytlgdkhfwaaflnlarhnvyitvnhinkvlelknkkdqeiiidn dqdilaiktlwgkvdtdinkkdrlrelimkhfpfleaatyqqsstnntkqkeeeqaka qsfeslkdclflfleklrearnyyshykhsksleepkleekllenmynifdtnvqlvik dyehnkdinpeedfkhlgraegefnyyftrnkkgnitesgllffvslflekkdaiwaq tkikgfkdnrenkqkmthevfcrsrmllpklrlestqtqdwilldmlnelircpksly krlqgekrekfrvpfdpadedydaeqepfkntlvrhqdrfpyfalryfdyneiftnlrf qidlgtyhfsiykkqigdkkedrhlthklygferiqefakenrpdewkalvkdldtfe esnepyisettphyhlenqkigirnknkkkkktiwpsletkttvnerskynlgksfka eaflsvhellpmmfyylllnkeepnngkinaskvegiiekkirdiyklygafaneei nneeelkeycegkdiairhlpkqmiailkneykdmakkaedkqkkmikdtkkrla aldkqvkgevedggrnikplksgriaswlvndmmrfqpvqrdrdgyplnnskan steyqllqrtlalfgsererlapyfrqmnligkdnphpflkdtkwkehnnilsfyrsyle akknflgslkpedwkknqyflklkepktnretlvqgwkngfnlprgiftepirewfir hqneseeykkvkdfdriglvakviplffkedyqkeiedyvqpfygypfnvgnihns qegtflnkkereelwkgnktkfkdyktkeknkektnkdkfkkktdeekeefrsyld fqswkkferelrlvrnqdivtwllcmelidklkidelnieelqklrlkdidtdtakkek nnilnrimpmelpvtvyetddsnniikdkplhtiyikeaetkllkqgnfkalvkdrrl

		nglfsfvetsseaelkskpiskslveyelgeyqrarveiikdmlrleetligndeklptn kfrqmldkwlehkketddtdlkndvkllteyrnafshnqypmrdriafanikpfsls santsneeglgiakklkdktketidriieieeqtatkr
Prevotella pleuritidis (SEQ ID No. 127)	WP_036931485	mendkrleestcytlndkhfwaaflnlarhnvyitinhinkllleirqidndekvldika lwqkvdkdinqkarlrelmikhfpfleaaiysnnkedkeevkeekqakaqsfkslk dclflfleklqearnyyshykssesskepefeegllekmyntfgvsirlvkedyqynk didpekdfkhlerkedfnylftdkdnkgkitknglffvslflekkdaiwmqqklrgf kdnrgnkekmthevfcrsrmllpkirlestqtqdwilldmlnelircpkslyerlqga yrekfkvpfdsidedydaeqepfrntlvrhqdrfpyfalryfdyneifknlrfqidlgt yhfsiykkligdnkedrhlthklygferiqefakqkrpnewqalvkdldiyetsneq yisettphyhlenqkigirfknkkdkiwpsletngkenekskynldksfqaeaflsih ellpmmfydlllkkeepnndeknasivegfikkeikrmyaiydafaneeinskegl eeycknkgfqerhlpkqmiailtnksknmaekakrkqkemikdtkkrlatldkqv kgeiedggrnirllksgeiarwlvndmmrfqsvqkdkegkplnnskansteyqml qrslalynkeqkptpyfiqvnlikssnphpfleetkweecnnilsfyrsyleakknfle slkpedwkknqyflmlkepktnrktlvqgwkngfnlprgiftepikewfkrhqnd seeykkvealdrvglvakviplffkeeyfkedaqkeinncvqpfysfpynvgnihk peeknflhceerrklwdkkkdkfkgykakekskkmtdkekeehrsylefqswnk ferelrlvrnqdivtwllctelidklkidelnieelqklrlkdidtdtakkeknnilnrim pmqlpvtvyeidksfnivkdkplhtiyieetgtkllkqgnfkalvkdrrlnglfsfvkt sseaeskskpisklrveyelgayqkaridiikdmlalektlidndenlptnkfsdmlk swlkgkgeankarlqndvdllvairnafshnqypmynsevfkgmkllslssdipek eglgiakqlkdkiketieriieiekeirn
[Porphyromonas gingivalis (SEQ ID No. 128)	WP_039417390	mteqnerpyngtyytledkhfwaaffnlarhnayitlahidrqlayskaditndedilf fkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnksskkkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdrygnndnpffkhhfvdregtvte agllffvslflekrdaiwmqkkirgfkggteayqqmtnevfcrsrislpklkleslrtd dwmlldmlnelvrcpkslydrlreedrarfrvpidilsdeddtdgteedpfkntlvrh qdrfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriqdf aeehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqhlwpspe vgatrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqgrikr viedvyavydafargeidtldrldacladkgirrghlprqmiailsqehkdmeekvr

		kklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpva kdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetr weshtnilsfyrsylkarkaflqsigrsdreenhrflllkepktdrqtlvagwksefhlp rgifteavrdcliemgydevgsykevgfmakavplyferackdrvqpfydypfnv gnslkpkkgrflskekraeewesgkerfrlaklkkeileakehpyldfkswqkferel rlvknqdiitwmmcrdlmeenkvegldtgtlylkdirtdvheqgslnvlnrvkpmr lpvvvyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtga lameqypisklrveyelakyqtarvcafeqtleleeslltryphlpdknfrkmlesws dplldkwpdlhrkvrlliavrnafshnqypmydeavfssirkydpsspdaieermg lniahrlseevkqakemaeriiqv
Porphyromonas gulae (SEQ ID No. 129)	WP_039418912	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndqdvls fkalwknldndlerksrlrslilkhfsflegaaygkklfeskssgnkssknkeltkkek eelqanalsldnlksilfdflqklkdfrnyyshyrhsgsselplfdgnmlqrlynvfdv svqrvkrdhehndkvdphrhfnhlvrkgkkdryghndnpsfkhhfvdsegmvte agllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrmd dwmlldmlnelvrcpkplydrlreddracfrvpvdilpdeddtdgggedpfkntlv rhqdrfpyfalryfdlkkvftslrfhidlgtyhfaiykkmigeqpedrhltrnlygfgri qdfaeehrpeewkrlvrdldyfetgdkpyisqtsphyhiekgkiglrfmpegqhlw pspevgttrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqg rikrviedvyaiydafardeintlkeldacladkgirrghlpkqmiailsqehknmee kvrkklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqp vakdasgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflh dtrweshtnilsfyrsylrarkaflerigrsdrmenrpflllkepktdrqtlvagwksef hlprgifteavrdcliemgydevgsyrevgfmakavplyferacedrvqpfydspf nvgnslkpkkgrflskeeraeewergkerfrdleawshsaarriedafagieyaspg nkkkieqllrdlslweafesklkvradkinlaklkkeileaqehpyhdfkswqkfere lrlvknqdiitwmmcrdlmeenkvegldtgtlylkdirtnvqeqgslnvlnhvkp mrlpvvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvd tgglameqypisklrveyelakyqtarvcafeqtleleeslltryphlpdknfrkmles wsdpllakwpelhgkvrlliavrnafshnqypmydeavfssirkydpsspdaieer mglniahrlseevkqaketveriiqa
Porphyromonas gulae	WP_039419792	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndqdvls fkalwknldndlerksrlrslilkhfsflegaaygkklfeskssgnkssknkeltkkek

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME	1	DE	3
CONTENANT LES PAGES	1	À	246

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME	1	OF	3
CONTAINING PAGES	1	TO	246

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

CLAIMS

1. An engineered composition for site directed base editing comprising a targeting domain and an adenosine deaminase, or catalytic domain thereof.

2. The composition of claim 1, wherein the targeting domain is an oligonucleotide binding domain.

3. The composition of claim 1 or 2, wherein the adenosine deaminase, or catalytic domain thereof, comprises one or more mutations that increase activity or specificity of the adenosine deaminase relative to wild type.

4. The composition of claim 1 or 2, wherein the adenosine deaminase comprises one or more mutations that changes the functionality of the adenosine deaminase relative to wild type, preferably an ability of the adenosine deaminase to deaminate cytodine.

5. The composition of any one of the preceding claims, wherein the targeting domain is a CRISPR system comprising a CRISPR effector protein, or fragment thereof which retains DNA and/or RNA binding ability, and a guide molecule.

6. The composition of claim 5, wherein the CRISPR system is catalytically inactive.

7. The composition of claim 5 or 6, wherein the CRISPR system comprises an RNA-binding protein, preferably Cas13, preferably the Cas13 protein is Cas13a, Cas13b or Cas13c, preferably wherein said Cas13 a Cas13 listed in any of Tables 1, 2, 3, 4, or 6 or is from a bacterial species listed in any of Tables 1, 2, 3, 4, or 6, preferably wherein said Cas13 protein is Prevotella sp.P5-125 Cas13b, Porphyromas gulae Cas13b, or Riemerella anatipestifer Cas13b; preferably Prevotella sp.P5-125 Cas13b.

8. The composition of claim 5, 6 or 7, wherein said guide molecule comprises a guide sequence is capable of hybridizing with a target RNA sequence comprising an Adenine to form an RNA duplex, wherein said guide sequence comprises a non-pairing Cytosine at a position corresponding to said Adenine resulting in an A-C mismatch in the RNA duplex formed.

9. The composition of claim 7, wherein said Cas13 protein is a Cas13a protein and said Cas13a comprises one or more mutations the two HEPN domains, particularly at position R474 and R1046 of Cas13a protein originating from *Leptotrichia wadei* or amino acid positions corresponding thereto of a Cas13a ortholog, or wherein said Cas13 protein is a Cas13b protein and said Cas13b comprises a mutation in one or more of positions R116, H121, R1177, H1182, preferably R116A, H121A, R1177A, H1182A of Cas13b protein

originating from *Bergeyella zoohelcum* ATCC 43767 or amino acid positions corresponding thereto of a Cas13b ortholog, or wherein said Cas13 protein is a Cas13b protein and said Cas13b comprises a mutation in one or more of positions R128, H133, R1053, H1058, preferably H133 and H1058, preferably H133A and H1058A, of a Cas13b protein originating from Prevotella sp. P5-125 or amino acid positions corresponding thereto of a Cas13b ortholog.

10. The composition of claim 7, wherein said Cas13, preferably Cas13b, is truncated, preferably C-terminally truncated, preferably wherein said Cas13 is a truncated functional variant of the corresponding wild type Cas13, optionally wherein said truncated Cas13b is encoded by nt 1-984 of Prevotella sp.P5-125 Cas13b or the corresponding nt of a Cas13b orthologue or homologue.

11. The composition of claim 7, wherein said Cas13 is a catalytically inactive Cas13, preferably Cas13b6.

12. The composition of claim 10, wherein said guide sequence has a length of about 20-53 nt, preferably 25-53 nt, more preferably 29-53 nt or 40-50 nt capable of forming said RNA duplex with said target sequence, and/or wherein the distance between said non-pairing C and the 5' end of said guide sequence is 20-30 nucleotides..

13. The composition of claims 12, wherein the guide sequence comprises more than one mismatch corresponding to different adenosine sites in the target RNA sequence or wherein two guide molecules are used, each comprising a mismatch corresponding to a different adenosine sites in the target RNA sequence.

14. The composition of any one of the preceding claims, wherein adenosine deaminase protein or catalytic domain thereof is fused to a N- or C-terminus of said oligonucleotide targeting protein, optionally by a linker, preferably where said linker is $(GGGGS)_{3-11}$, GSG_5 or LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR, or wherein said linker is an XTEN linker.

15. The composition of any one of claims 7 to 13, wherein said adenosine deaminase protein or catalytic domain thereof is inserted into an internal loop of said dead Cas13 protein.

16. The composition of any one of claims 7 to 13, wherein said adenosine deaminase protein or catalytic domain thereof is linked to an adaptor protein and said guide molecule or said dead Cas13 protein comprises an aptamer sequence capable of binding to said adaptor protein, preferably wherein said adaptor sequence is selected from MS2, PP7,

Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1.

17. The composition of any one of the preceding claims, wherein said adenosine deaminase protein or catalytic domain thereof capable of deaminating adenosine or cytodine in RNA or is an RNA specific adenosine deaminase and/or is a bacterial, human, cephalopod, or Drosophila adenosine deaminase protein or catalytic domain thereof, preferably TadA, more preferably ADAR, optionally huADAR, optionally (hu)ADAR1 or (hu)ADAR2, preferably huADAR2 or catalytic domain thereof.

18. The composition of claim 17, wherein the ADAR protein is a mutated hADAR2d comprising mutation E488Q or a mutated hADAR1d comprising mutation E1008Q.

19. The composition of any one of the preceding claims, wherein said targeting domain and optionally said adenosine protein or catalytic domain thereof comprise one or more heterologous nuclear export signal(s) (NES(s)) or nuclear localization signal(s) (NLS(s)), preferably an HIV Rev NES or MAPK NES, preferably C-terminal.

20. The composition of any one of the preceding claims, wherein said target RNA sequence of interest is within a cell, preferably a eukaryotic cell, most preferably a human or non-human animal cell, or a plant cell.

21. The composition of any one of the preceding claims for use in prophylactic or therapeutic treatment, preferably wherein said target locus of interest is within a human or animal.

22. A method of modifying an Adenine or Cytidine in a target RNA sequence of interest, comprising delivering to said target RNA, the composition according to any one of claims 1 to 21.

23. The method of claim 22, wherein the targeting domain comprises the CRISPR system of any one of claims 5 to 7, wherein said guide molecule forms a complex with said CRISPR effector protein and directs said complex to bind said target RNA sequence of interest, wherein said guide sequence is capable of hybridizing with a target sequence comprising said Adenine or Cytosine to form an RNA duplex; wherein said adenosine deaminase protein or catalytic domain thereof deaminates said Adenine or Cytodine in said RNA duplex.

24. The method of claim 22, wherein the CRISPR system comprises the Cas13 of any one of claims 7 to 21.

25. The method of claims 22 or 23, wherein the CRISPR system and the adenonsine deaminase, or catalytic domain thereof, are delivered as one or more polynucleotide molecules, as a ribonucleoprotein complex, optionally via particles, vesicles, or one or more viral vectors.

26. The method of anyone of claims 22 to 24 or the composition of any one of claims 1 to 21 is for use in the treatment or prevention of a disease caused by transcripts containing a pathogenic G→A or C→T point mutation.

27. An isolated cell obtained from the method of any one of claims 22 to 25 and/or comprising the composition of any one of claims 1-21, or progeny of said modified cell, preferably wherein said cell comprises a hypoxanthine or a guanine in replace of said Adenine in said target RNA of interest compared to a corresponding cell not subjected to the method.

28. The cell or progeny thereof of claim 27, wherein said cell is a eukaryotic cell, preferably a human or non-human animal cell, optionally a therapeutic T cell or an antibody-producing B-cell or wherein said cell is a plant cell.

29. A non-human animal comprising said modified cell or progeny thereof of claim 27 or 28.

30. A plant comprising said modified cell or progeny thereof of claim 27.

31. A modified cell according to claim 27 or 28 for use in therapy, preferably cell therapy.

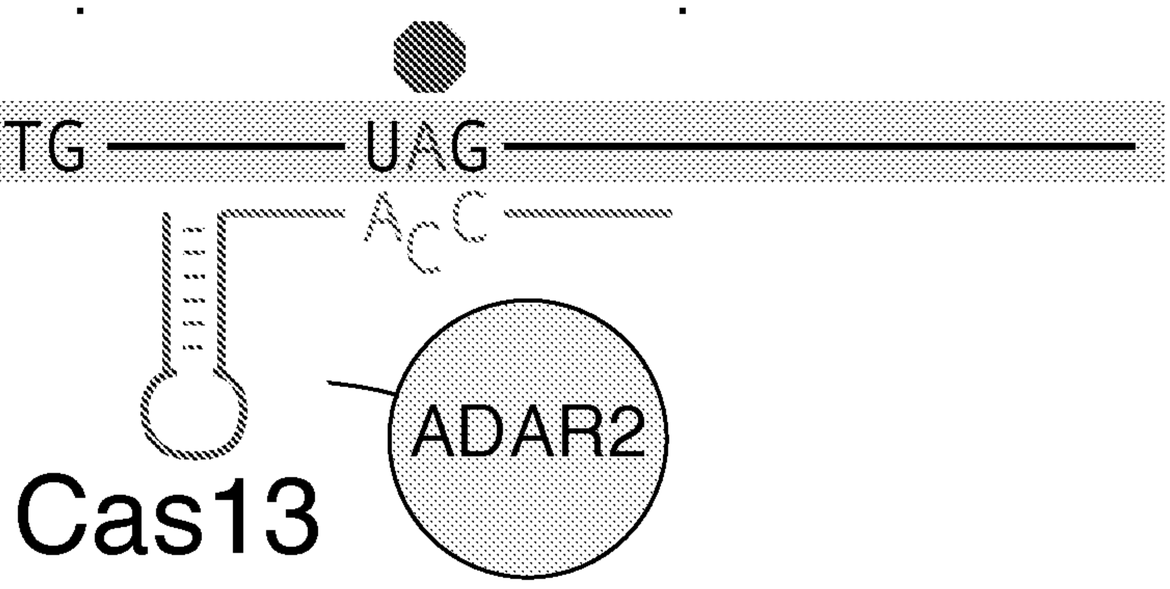


FIG. 1

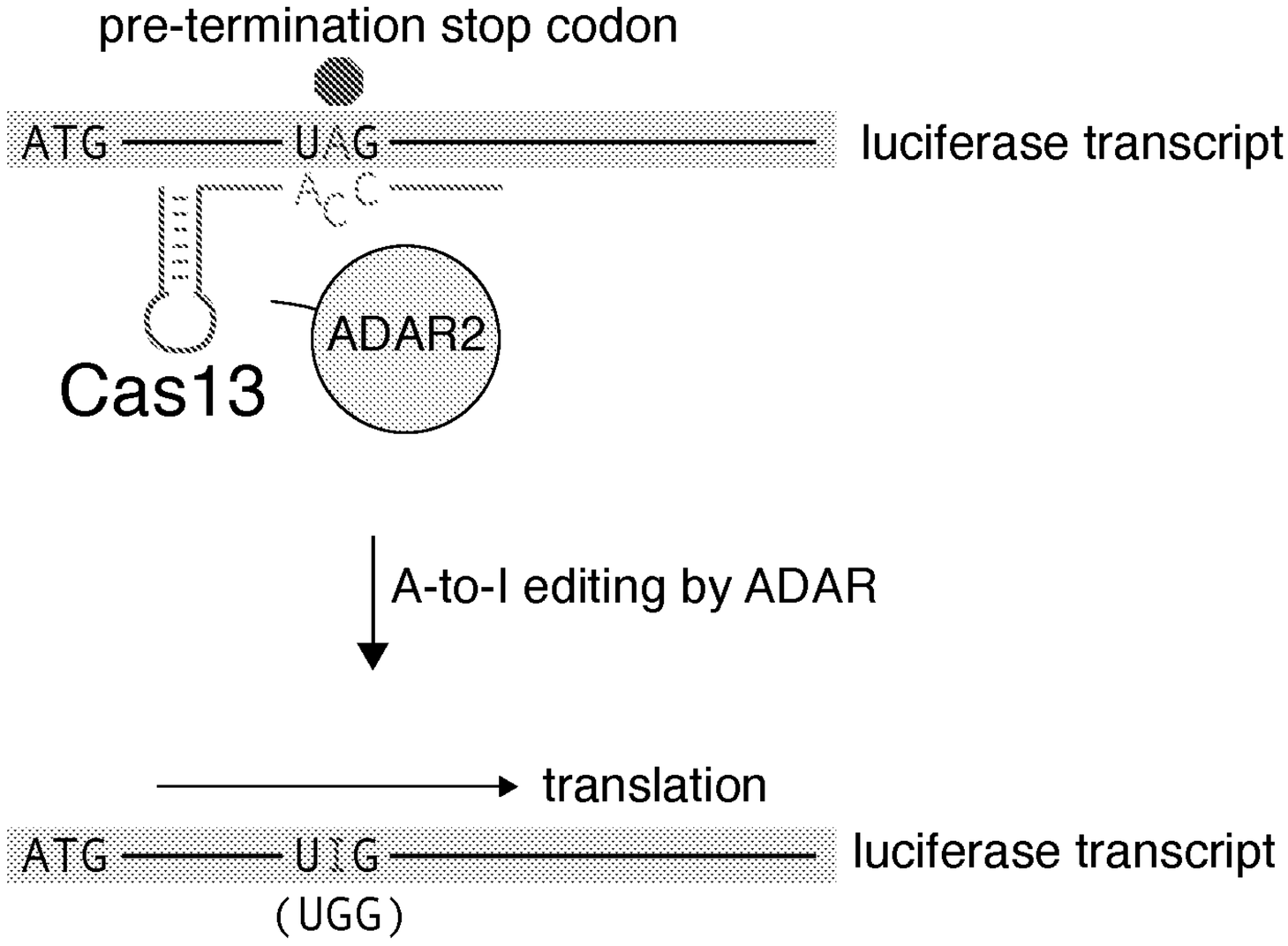

pre-termination stop codon
ATG
UAG
luciferase transcript
Cas13
ADAR2
A-to-I editing by ADAR
translation
ATG
UIG
(UGG)
luciferase transcript

FIG. 2

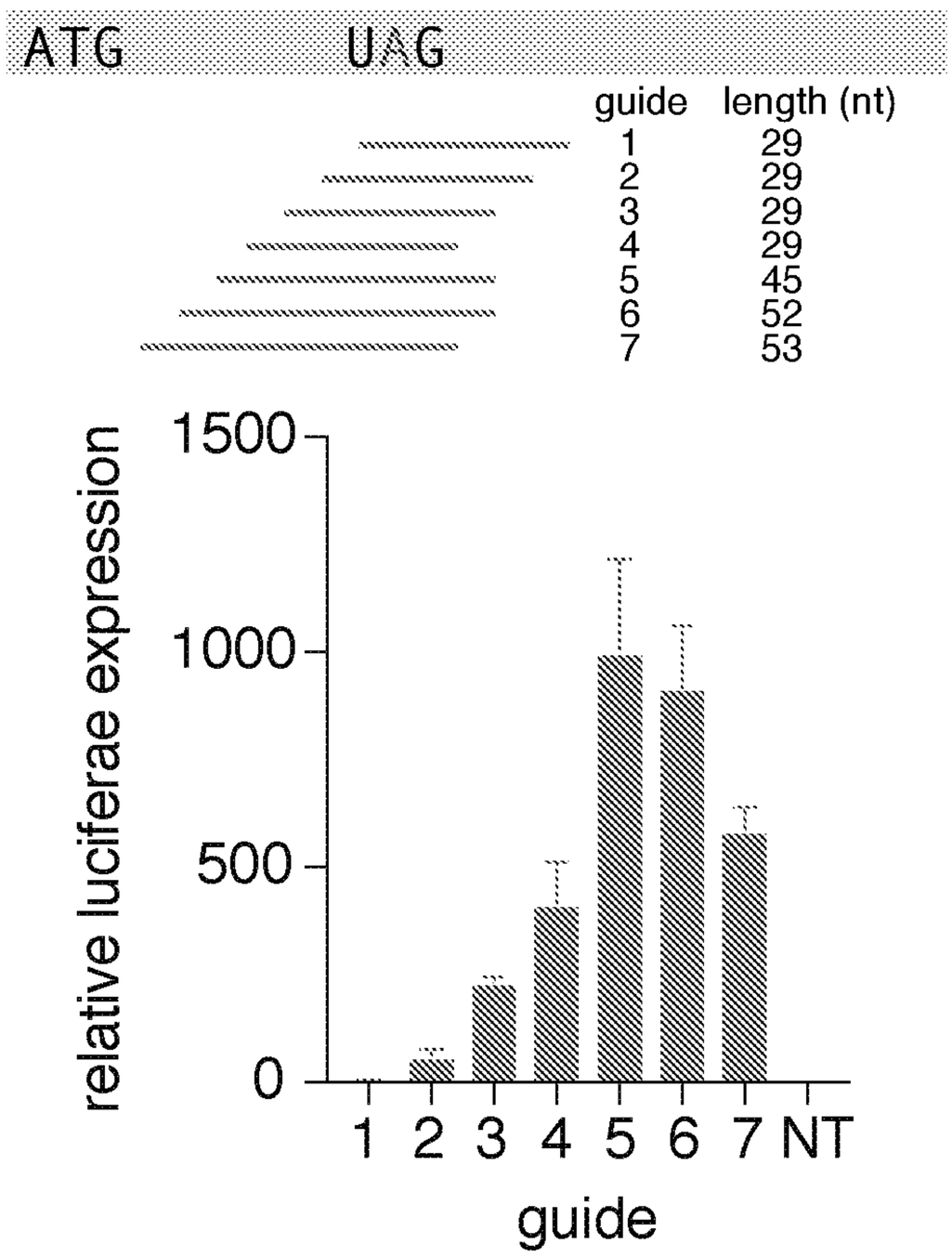

ATG UAG
guide length (nt)
1 29
2 29
3 29
4 29
5 45
6 52
7 53
1500
1000
500
0
relative luciferae expression
1 2 3 4 5 6 7 NT
guide

FIG. 3

ADAR Name	Protein Sequence
Homo sapiens_ADAR2_E_Q_Mutant (SEQ ID NO :650)	QLHLPQVLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMT TGTDVKDAKVISVSTGTKCINGEYMSDRGLALNDCHAEIISRRSL LRFLYTQLELYLNNKDDQKRSIFQKSERGGFRLKENVQFHLYIST SPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKIESGQGTIPVR SNASIQTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEP IYFSSIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPLLSGISN AEARQPGKAPNFSVNWTVGDSAIEVINATTGKDELGRASRLCKH ALYCRWMRVHGKVPSHLLRSKITKPNVYHESKLAAKEYQAAKA RLFTAFIKAGLGAWVEKPTEQDQFSLT*
Homo sapiens_ADAR1_E_Q_Mutant (SEQ ID NO :651)	SLGTGNRCVKGDSLSLKGETVNDCHAEIISRRGFIRFLYSELMKY NSQTAKDSIFEPAKGGEKLQIKKTVSFHLYISTAPCGDGALFDKS CSDRAMESTESRHYPVFENPKQGKLRTKVENGQGTIPVESSDIVP TWDGIRLGERLRTMSCSDKILRWNVLGLQGALLTHFLQPIYLKS VTLGYLFSQGHLTRAICCRVTRDGSAFEDGLRHPFIVNHPKVGR VSIYDSKRQSGKTKETSVNWCLADGYDLEILDGTRGTVDGPRNE LSRVSKKNIFLLFKKLCSFRYRRDLLRLSYGEAKKAARDYETAK NYFKKGLKDMGYGNWISKPQEEKNF
Octopus vulgaris_ADAR1_E_Q_Mutant (SEQ ID NO :652)	SVGTGNRCLTGDHLSLEGNSVNDSHAEMITRRGFLRYLYRHLLE YDAEVPNDLFEKGERSICRIKTNITFHLYISTAPCGDGALFSPRDT DSSNAKMEEENKHIHNPTFSSSVQGLLRTKVEGGQGTIPIDADFT EQTWDGIQRGERLRTMSCSDKICRWNVVGLQGALLSHFIEPIYL DSLTLGYLYDHGHLARAVCCRIERGEASVNQLLPEGYRLNHPW LGRVTACDPPRETQKTKSLSINWCYDDEKSEVLDGTAGICYTAIE KNLFSRLTKHNLYEEFKRVCRKFDRNDLLTAPSYNKAKMMATP FQTAKNVMLKKLKENNCGTWVSKPIEEEMF
Sepia_ADAR1_E_Q_Mutant (SEQ ID NO :653)	SVGTGNRCLTGDRLSLEGNSVNDSHAEMVTRRGFLRYLYKHLL EYDPEKPHDLFEKGERSLCRIKTNITFHLYISTAPCGDGALFSPRD TDSSNVKVDEENKHVHNPTFSSSVQGLLRTKVEGGQGTIPIDADF TEQTWDGIQRGERLRTMSCSDKICRWNVVGLQGALLSHFVEPIY LESLTLGYLYDHGHLARAVCCRIERGEASVNQLLPEGYRLNHPW LGRVTACDPPRETQKTKSLSINWCYDDEKSEVLDGTAGICYTAIE KNLFSRLTKHSLYEEFKKVCQKFEREDLLNVTSYNKAKMMAIPF QTAKNVMLKKLKENNCGTWVSKPIEEEMF

FIG. 4

Octopus vulgaris_ADAR2_E_Q_Mutant (SEQ ID NO :654)	GIGTGTKCINGEHMSDRGFGVNDCHAEIIARRCFLRYIYDQLELH LSDNSDVRNSSIFELRDKGGYQLKENIQFHLYISTAPCGDARIFSP HGQDVETGDRHPNRKARGQLRTKIESGQGTIPVRTSGVIQTWDG VLEGERLLTMSCSDKIARWNVLGIQGSLLSHFMNPIYLESIILGSL YHSDHLSRAMYSRISIIENLPEPFHLNRPFLSGISSPESRQPGKAPN FGINWRKEDETFEVINAMTGRVEGGSVSRICKQALFGRFMSLYG KLSSLTGQSVTTRPTHYSDAKAAVMEYQLAKQCVFQAFQKAGL GNWVQKPIEQDQF
Sepia_ADAR2_E_Q_Mutant(SEQ ID NO :655)	GIGTGTKCINGEYMNDRGFAVNDCHAEIIARRCFLRFIYDQLEMH LSEDPEVRGQSVFELRDGGGYKLKPNIHFHLYISTAPCGDARIFSP HGQDVETGDRHPNRKARGQLRTKIESGQGTIPVRSSGFIQTWDG VLEGERLLTMSCSDKIARWNVLGIQGALLCHFMHPIYLESIILGSL YHSDHLSRAVYCRIASIENLPDLFQLNRPFLSGISSPESRQPGKAP NFGINWRRNDDTFEVINAMTGRVEGGNMSRICKQALFDRFMNL YGRLSSLTGQSVTTRPTLYSEAKAAVMEYQLAKQCVFQAFQKA GLGNWVQKPIEQDQF
Doryteusthis opalescens_ADAR2_E_Q_Mutant(SEQ ID NO :656)	GIGTGTKCINGEYMNDRGFAVNDCHAEIIARRCFLRFIYDQLELH LSDNAEVRGQSIFELRDAGGYKLKPNIQFHLYISTAPCGDARIFSP HGQDVETGDRHPNRKARGQLRTKIESGQGTIPVRSSGFIQTWDG VLEGERLLTMSCSDKIARWNVLGVQGALLCHFMHPIYLESIILGS LYHSDHLSRAVYCRIAAIENLPDLFRLNRPFLSGISSPESRQPGKA PNFGINWRRNDDSFEVINAMTGRVEGGSMSRICKQALFDRFMNL YGKLSSLTGQSVTTRPALYSEAKATVMEYQLAKQCVFQAFQKA GLGNWVQKPIEQDQF

FIG. 4
(continued)

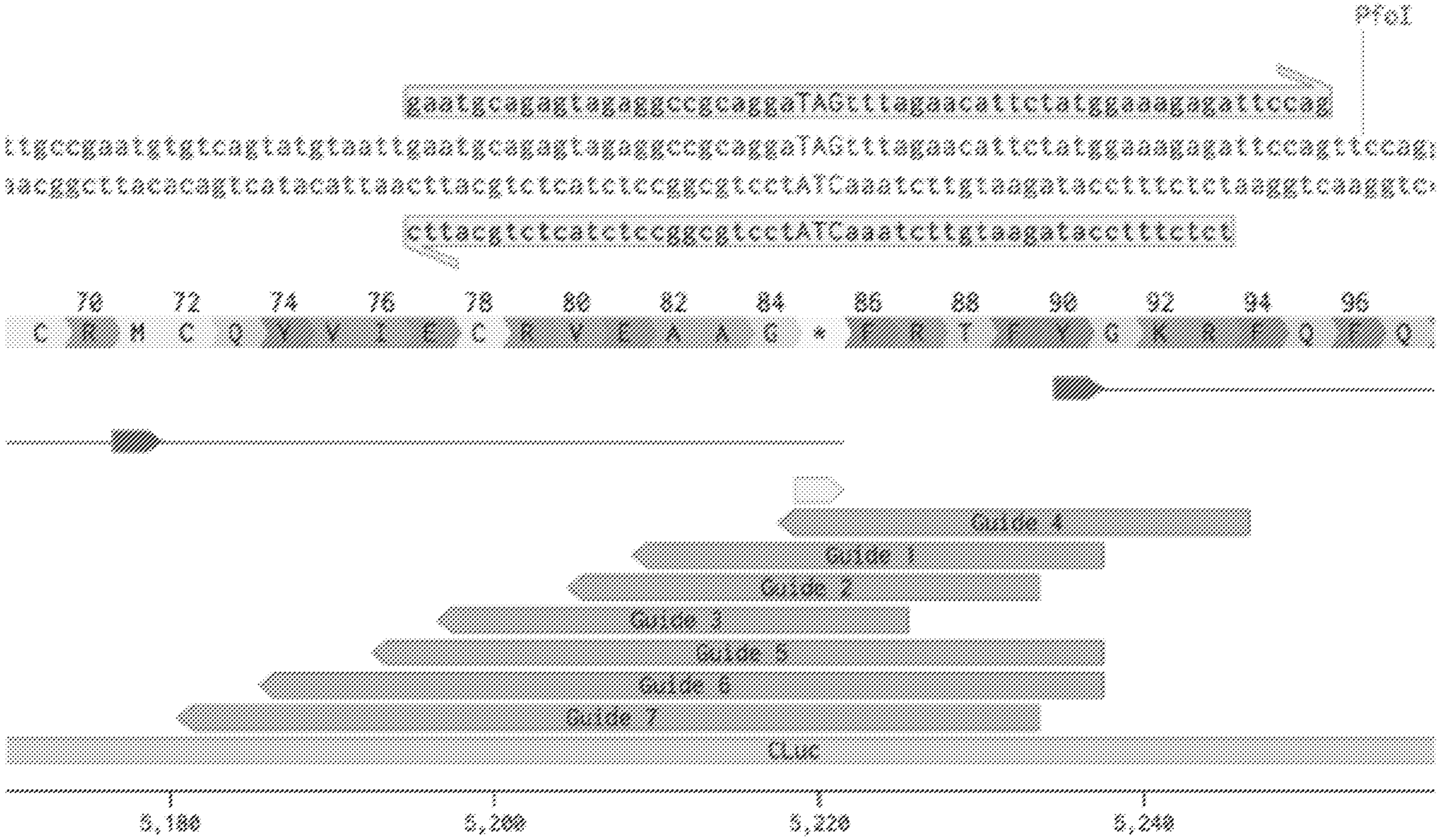

PfoI
gaatgcagagtagaggccgcaggaTAGtttagaacattctatggaaagagattccag
ttgccgaatgtgtcagtatgtaattgaatgcagagtagaggccgcaggaTAGtttagaacattctatggaaagagattccagttccag
aacggcttacacagtcatacattaacttacgtctcatctccggcgtcctATCaaatcttgtaagataccttttctctaaggtcaaggtc
cttacgtctcatctccggcgtcctATCaaatcttgtaagatacctttctct
70 72 74 76 78 80 82 84 86 88 90 92 94 96
C M C Q V T C V E A G * D T G K N Q D Q
Guide 4
Guide 1
Guide 2
Guide 3
Guide 5
Guide 6
Guide 7
Cas
5,180 5,200 5,220 5,240

FIG. 5

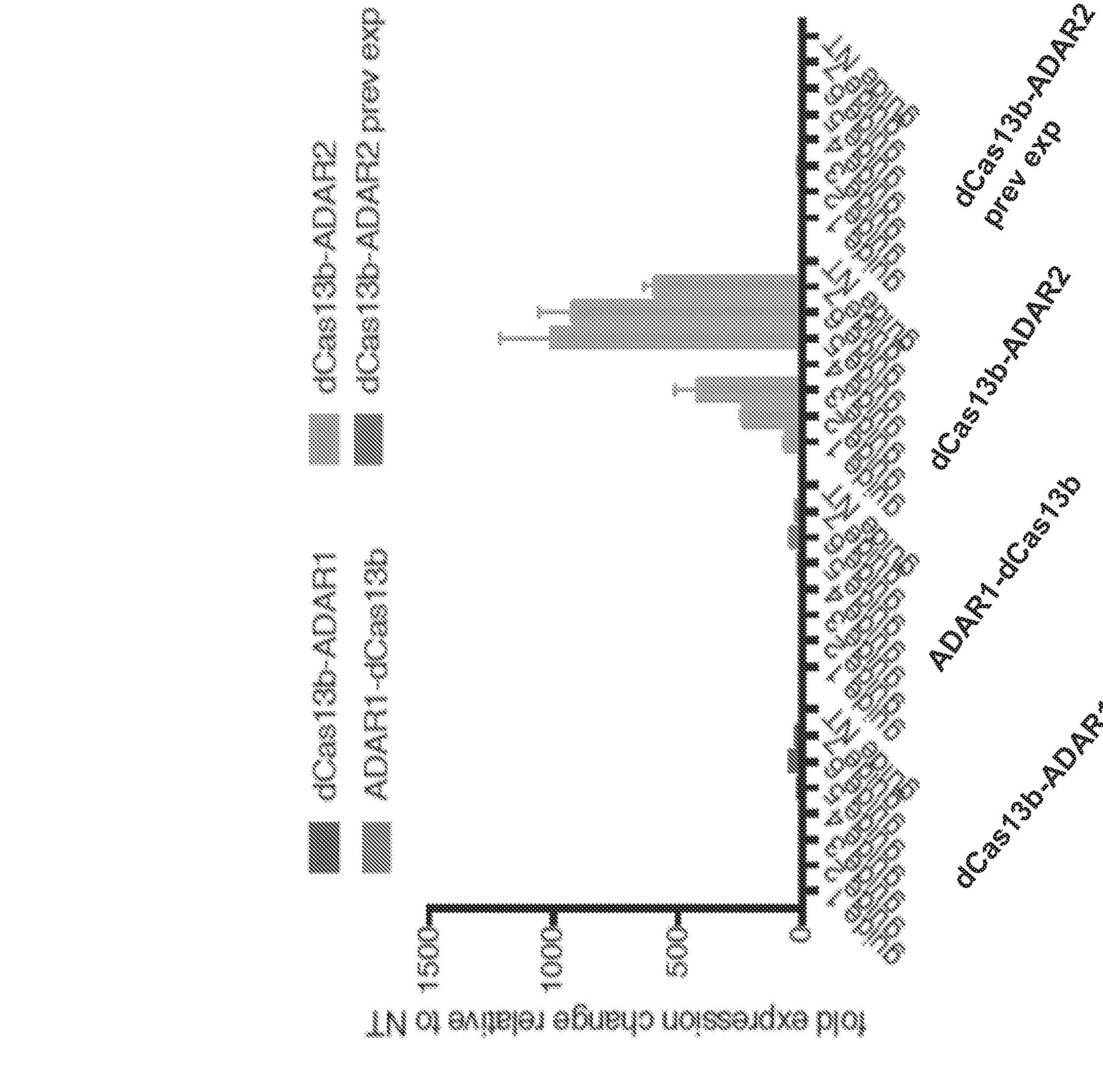

dCas13b-ADAR2
prev exp
dCas13b-ADAR2
ADAR1-dCas13b
dCas13b-ADAR1
fold expression change relative to NT
1500
1000
500
0
dCas13b-ADAR1
ADAR1-dCas13b
dCas13b-ADAR2
dCas13b-ADAR2 prev exp

FIG. 6

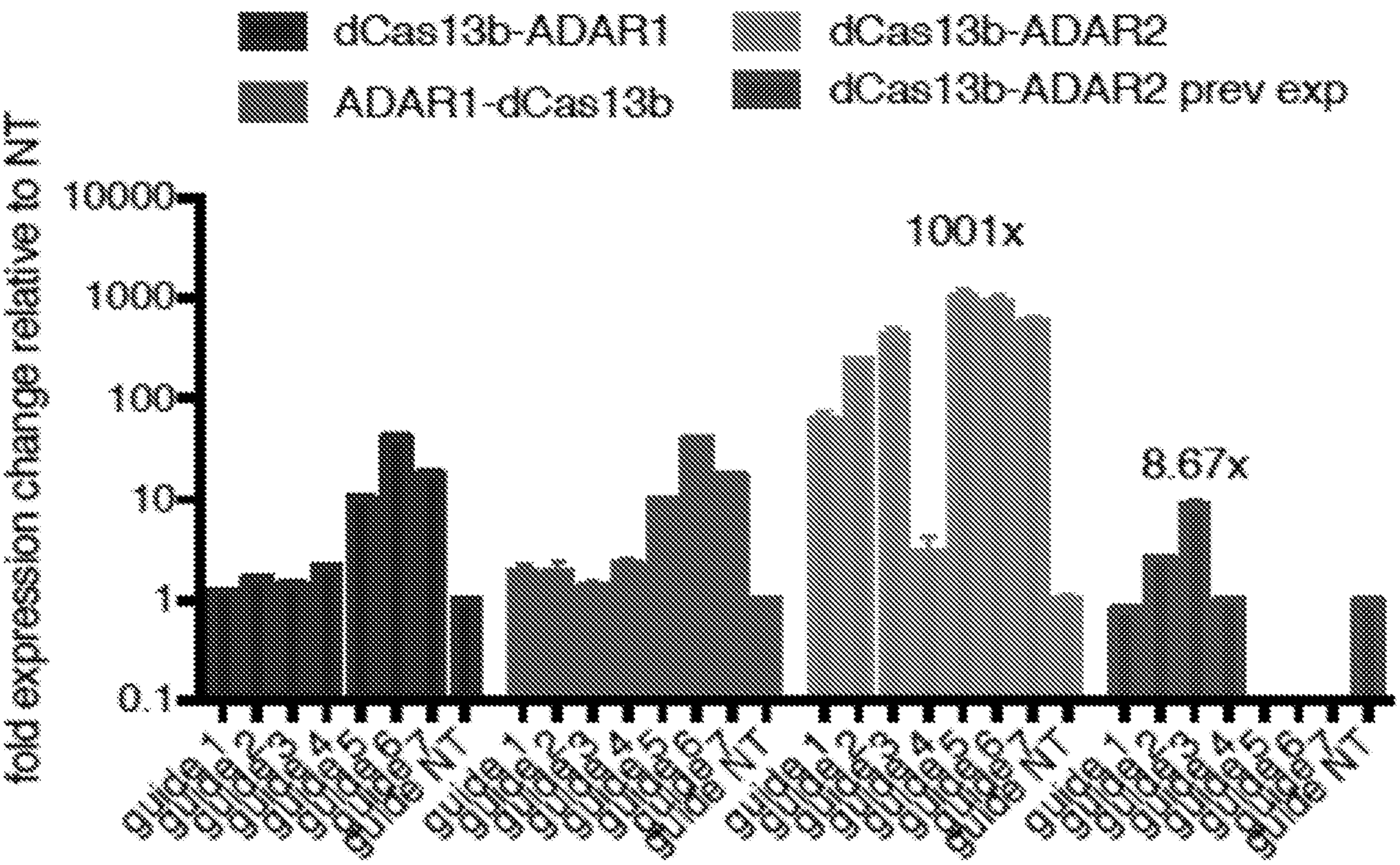

dCas13b-ADAR1
ADAR1-dCas13b
dCas13b-ADAR2
dCas13b-ADAR2 prev exp
fold expression change relative to NT
10000
1000
100
10
1
0.1
1001x
8.67x
guide 1
guide 2
guide 3
guide 4
guide 5
guide 6
guide 7
guide NT

FIG. 7

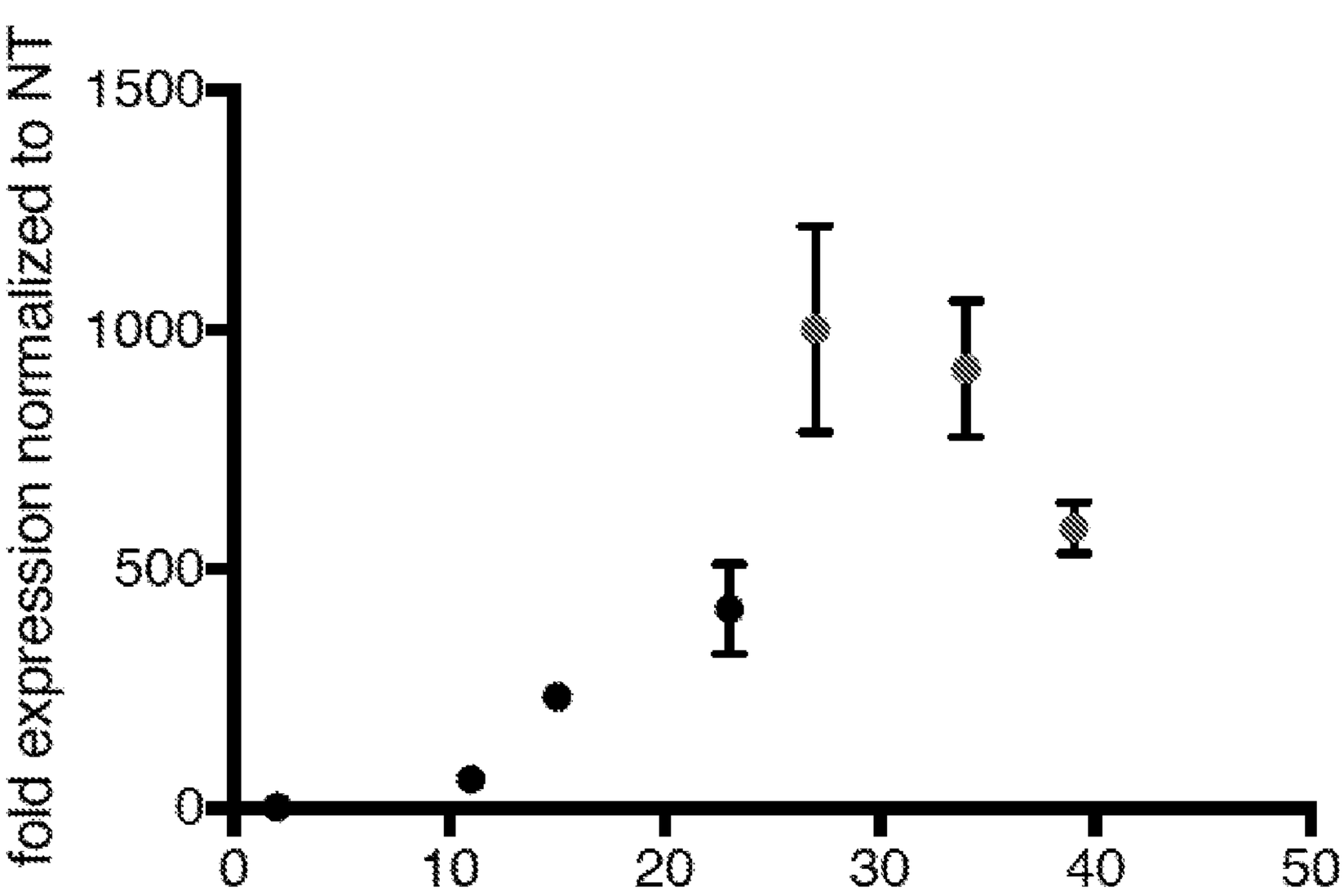

distance from DR
fold expression normalized to NT
1500
1000
500
0
0
10
20
30
40
50

FIG. 8

A

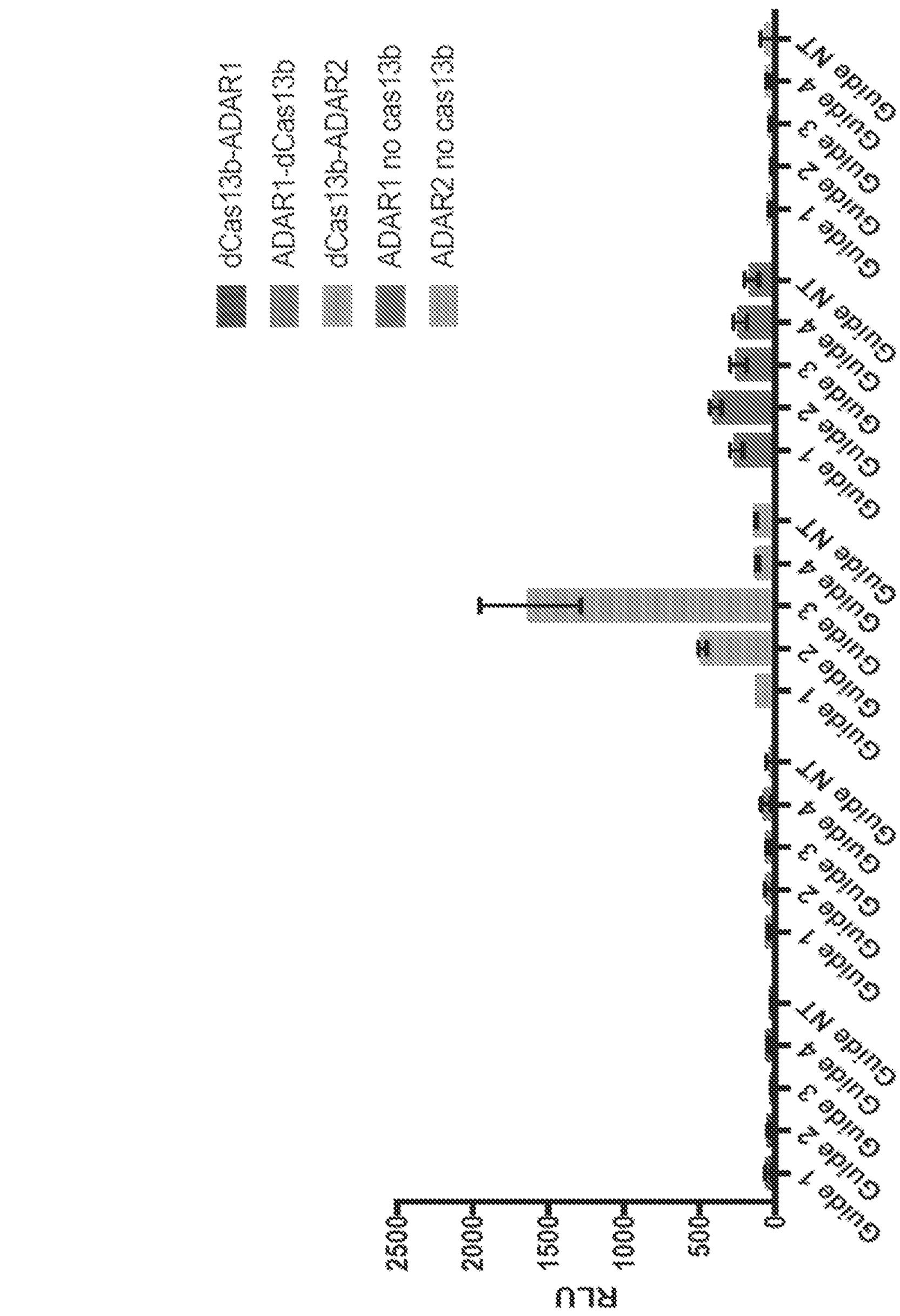

dCas13b-ADAR1
ADAR1-dCas13b
dCas13b-ADAR2
ADAR1 no cas13b
ADAR2 no cas13b
RLU
2500
2000
1500
1000
500
0
1 guide
2 guides
3 guides
4 guides
NT
1 guide
2 guides
3 guides
4 guides
NT
1 guide
2 guides
3 guides
4 guides
NT
1 guide
2 guides
3 guides
4 guides
NT

FIG. 9A

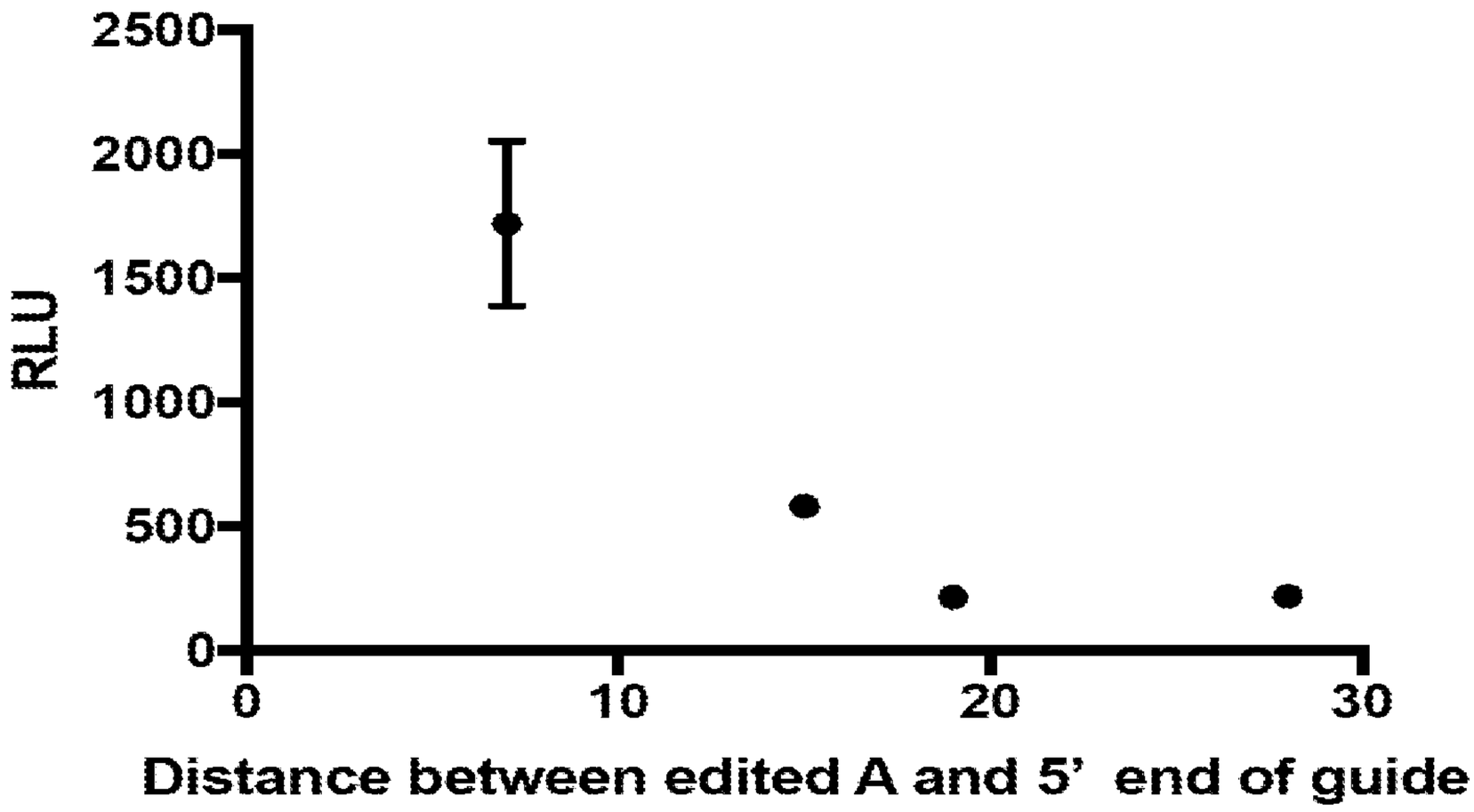
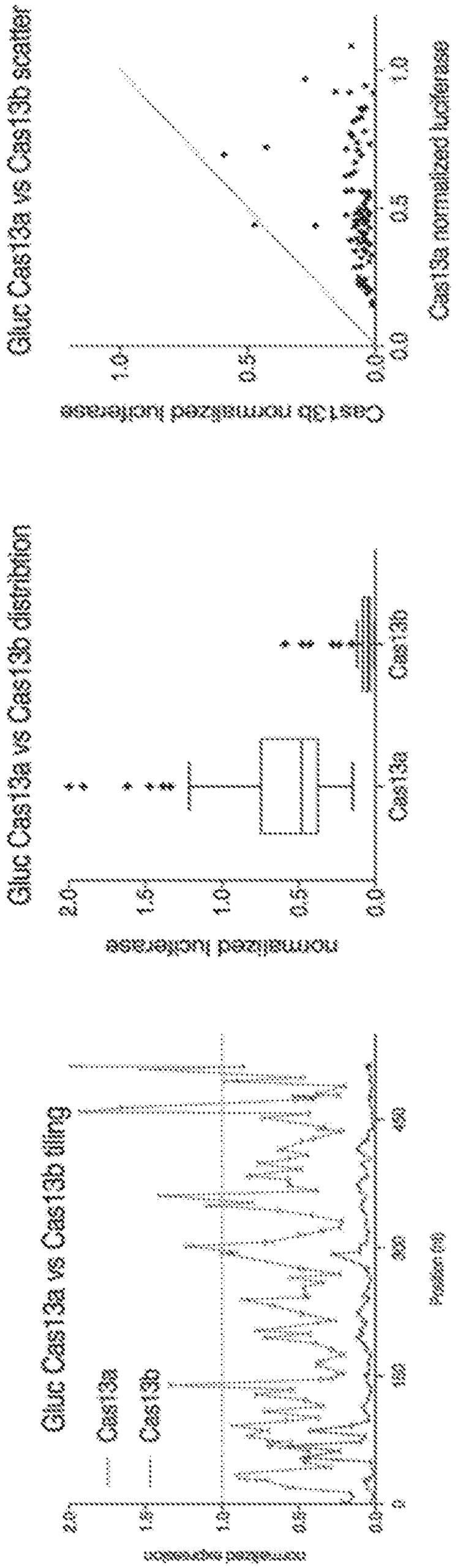

2500
2000
1500
1000
500
0
RLU
0
10
20
30
Distance between edited A and 5' end of guide

FIG. 9B


Gluc Cas13a vs Cas13b scatter
Cas13b normalized luciferase
Cas13a normalized luciferase
0.0
0.5
1.0
Gluc Cas13a vs Cas13b distribution
normalized luciferase
Cas13a
Cas13b
Gluc Cas13a vs Cas13b tiling
Cas13a
Cas13b
normalized luciferase
position (nt)

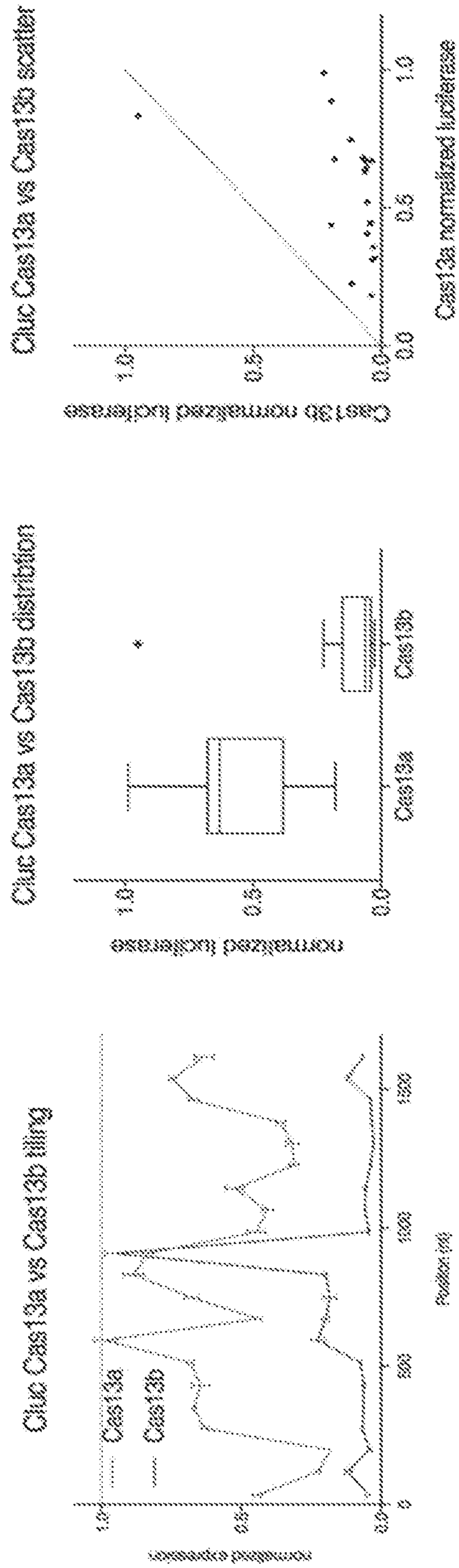
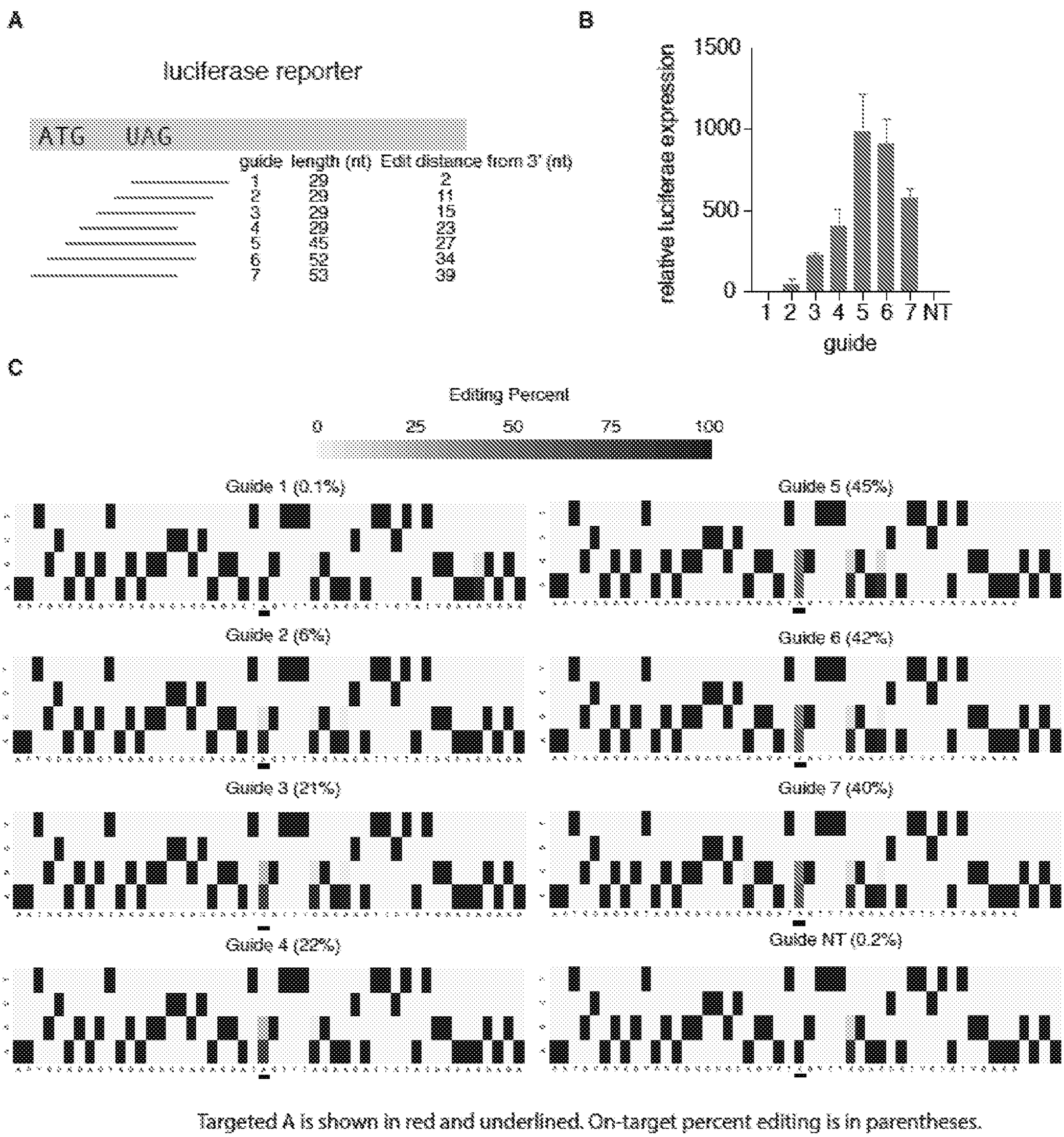
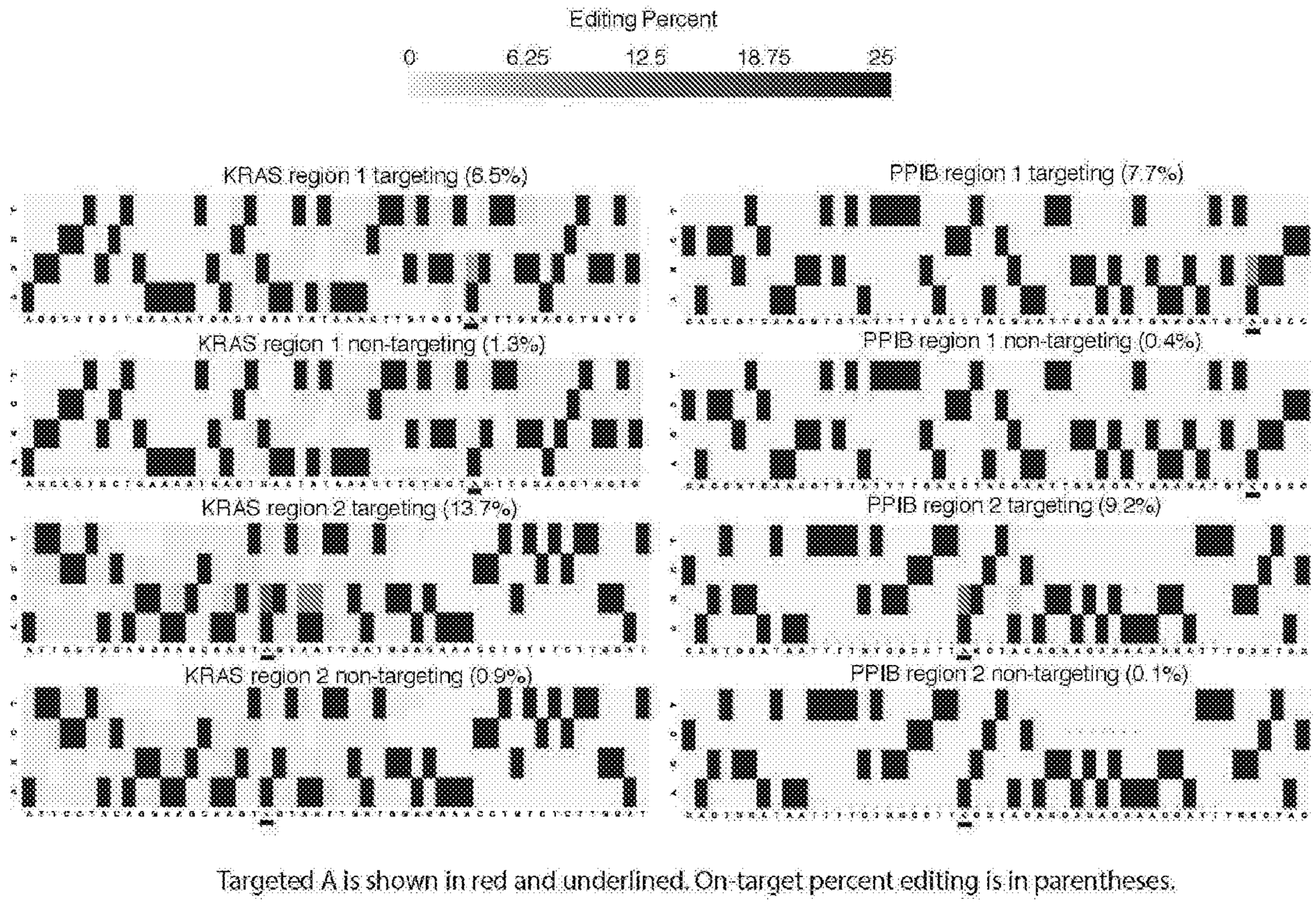

Gluc Cas13a vs Cas13b scatter
Cas13b normalized luciferase
Cas13a normalized luciferase
0.0
0.5
1.0
Gluc Cas13a vs Cas13b distribution
normalized luciferase
Cas13a
Cas13b
Gluc Cas13a vs Cas13b tiling
Cas13a
Cas13b
normalized luciferase
position (nt)

FIG. 10



Targeted A is shown in red and underlined. On-target percent editing is in parentheses.

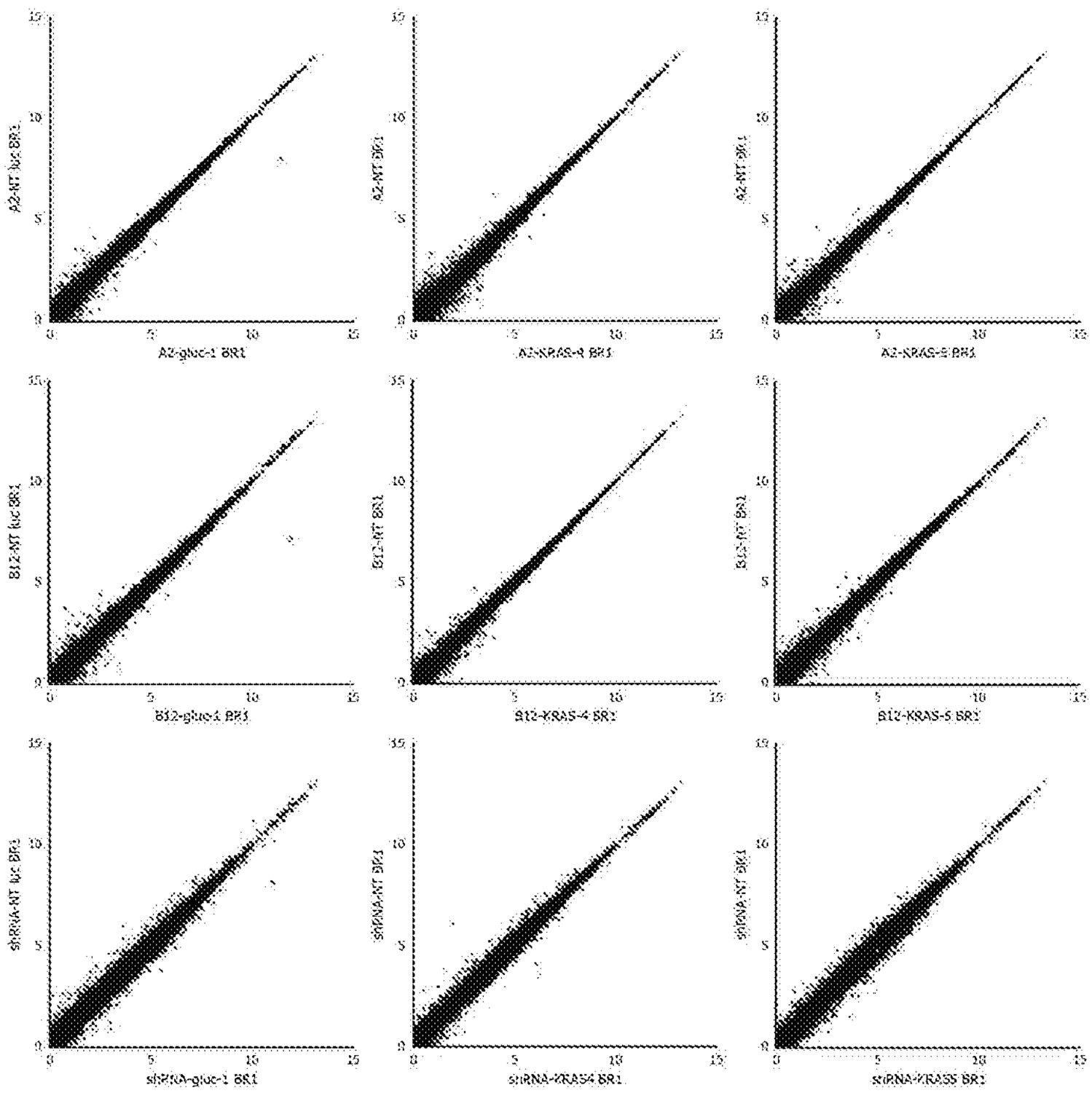
FIG. 11



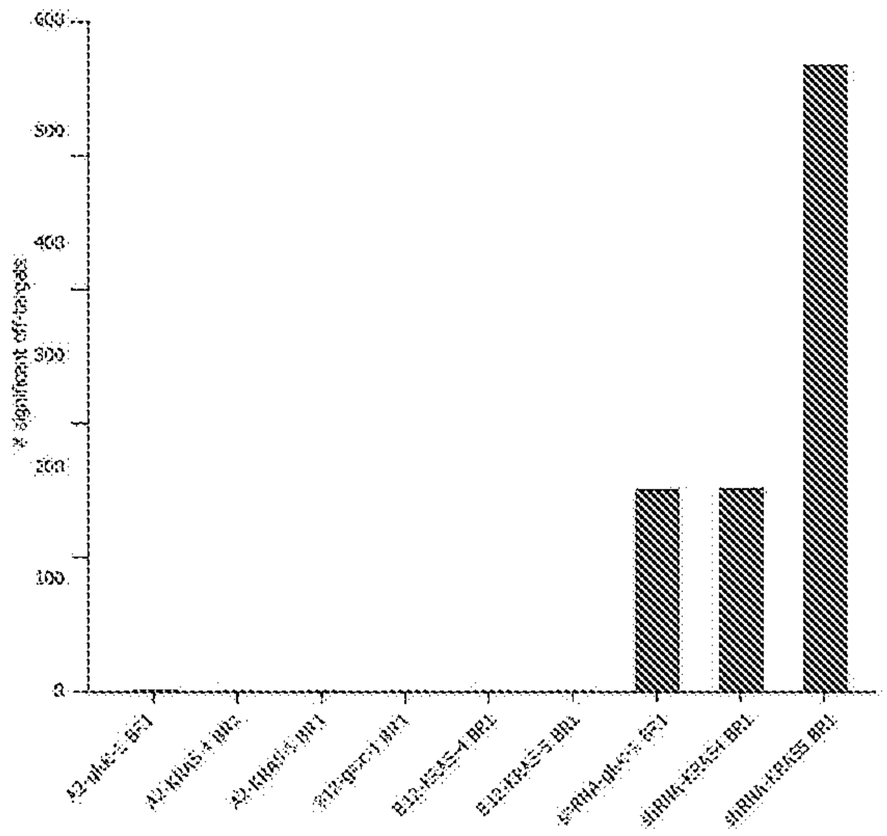
Targeted A is shown in red and underlined. On-target percent editing is in parentheses.

FIG. 12

A


A2-NT luc BR1
A2-gluc-1 BR1
A2-NT BR1
A2-KRAS-4 BR1
A2-KRAS-5 BR1
B12-NT luc BR1
B12-gluc-1 BR1
B12-NT BR1
B12-KRAS-4 BR1
B12-KRAS-5 BR1
shRNA-NT luc BR1
shRNA-gluc-1 BR1
shRNA-NT BR1
shRNA-KRAS4 BR1
shRNA-NT BR1
shRNA-KRAS5 BR1

B


significant off-targets
A2-gluc-1 BR1
A2-KRAS-4 BR1
A2-KRAS-5 BR1
B12-gluc-1 BR1
B12-KRAS-4 BR1
B12-KRAS-5 BR1
shRNA-gluc-1 BR1
shRNA-KRAS4 BR1
shRNA-KRAS5 BR1

C

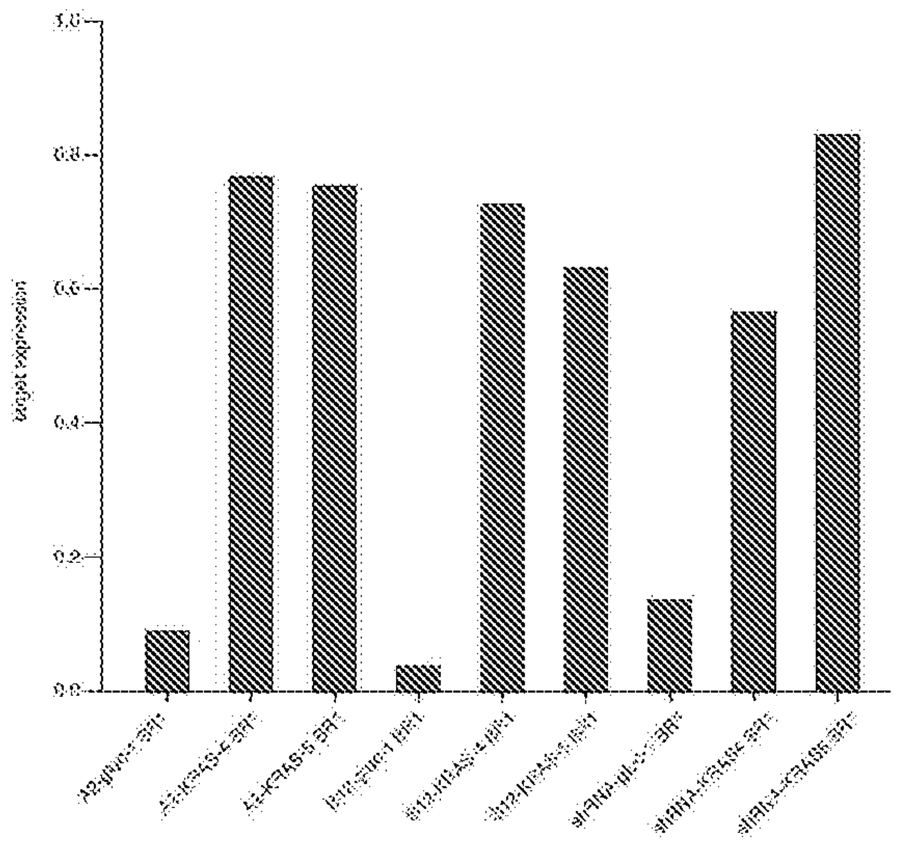

target expression
A2-gluc-1 BR1
A2-KRAS-4 BR1
A2-KRAS-5 BR1
B12-gluc-1 BR1
B12-KRAS-4 BR1
B12-KRAS-5 BR1
shRNA-gluc-1 BR1
shRNA-KRAS4 BR1
shRNA-KRAS5 BR1

FIG. 13

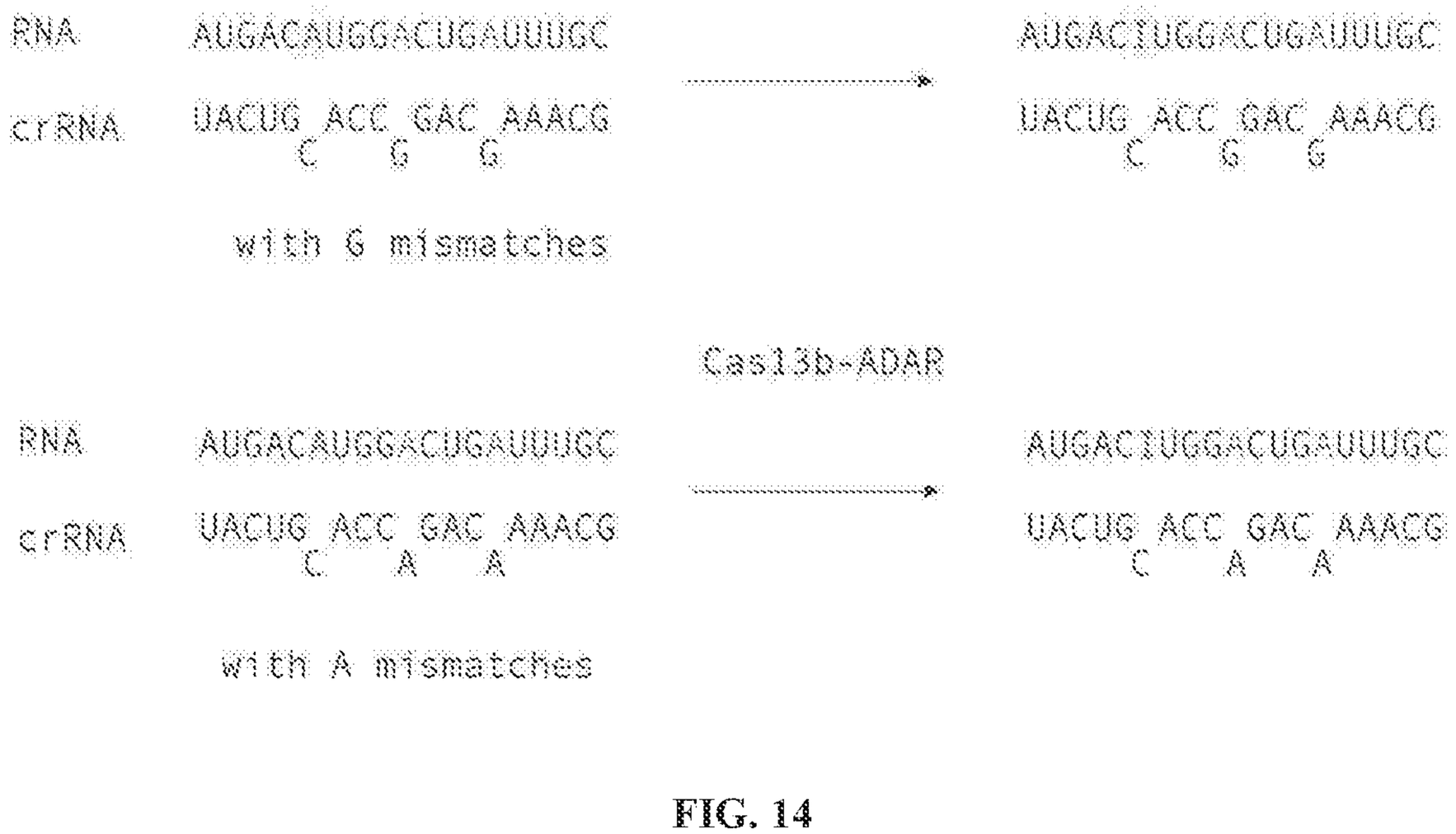

RNA
AUGACAUGGACUGAUUUGC
AUGACTUGGACUGAUUUGC
crRNA
UACUG ACC GAC AAACG
C G G
UACUG ACC GAC AAACG
C G G
with G mismatches
Cas13b-ADAR
RNA
AUGACAUGGACUGAUUUGC
AUGACTUGGACUGAUUUGC
crRNA
UACUG ACC GAC AAACG
C A A
UACUG ACC GAC AAACG
C A A
with A mismatches

FIG. 14

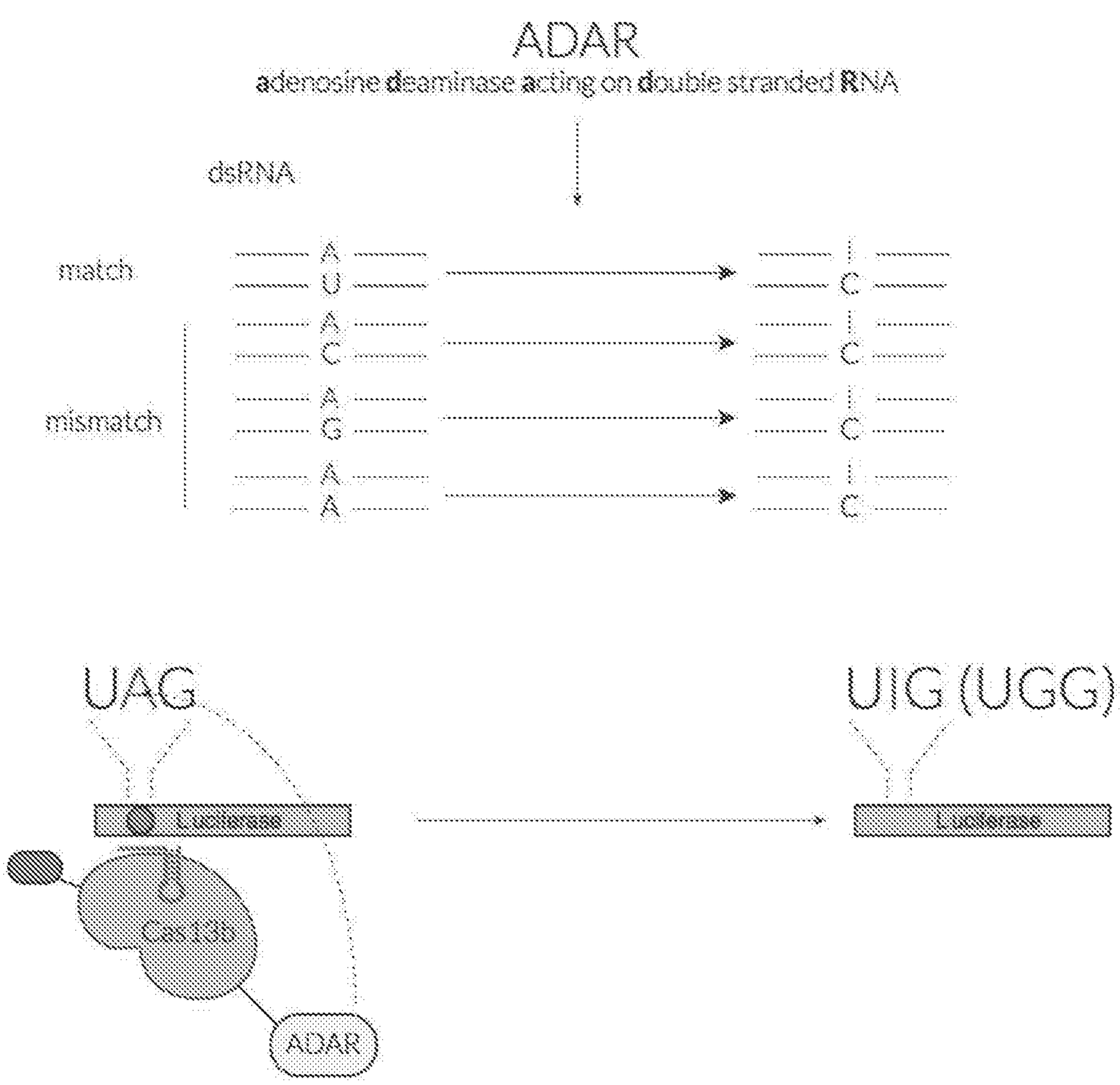

ADAR
adenosine deaminase acting on double stranded RNA
dsRNA
match
mismatch
A
U
A
C
A
G
A
A
I
C
I
C
I
C
I
C
UAG
Luciferase
U
Cas13b
ADAR
UIG (UGG)
Luciferase

FIG. 15

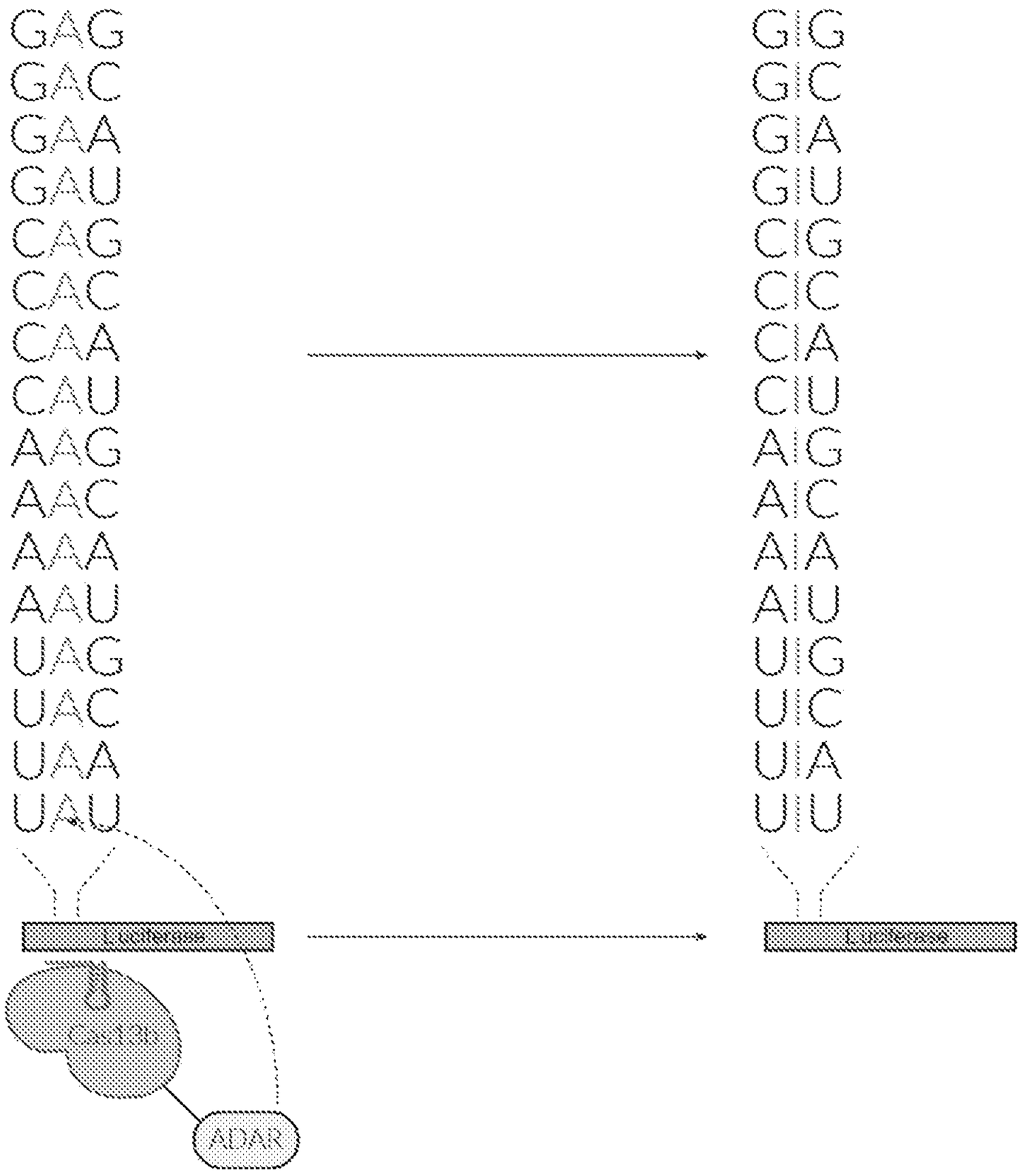

GAG
GAC
GAA
GAU
CAG
CAC
CAA
CAU
AAG
AAC
AAA
AAU
UAG
UAC
UAA
UAU
GIG
GIC
GIA
GIU
CIG
CIC
CIA
CIU
AIG
AIC
AIA
AIU
UIG
UIC
UIA
UIU
ADAR

FIG. 16

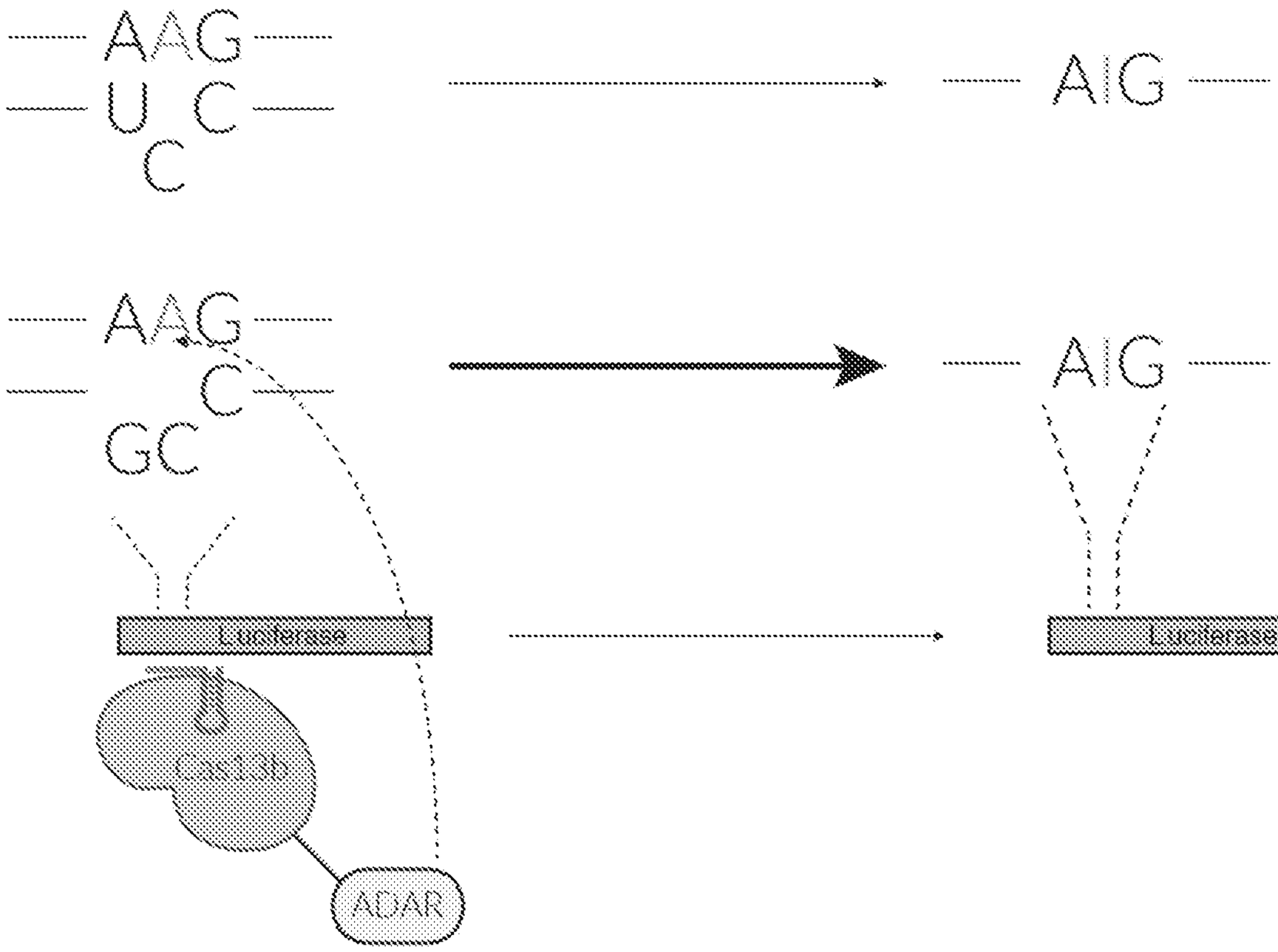


FIG. 17

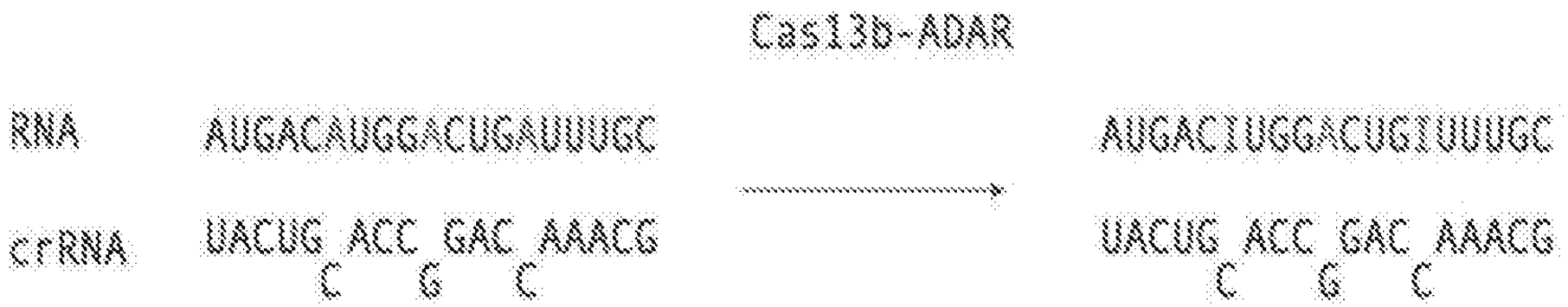

Cas13b-ADAR
RNA
AUGACAUGGACUGAUUUGC
AUGACTUGGACUGTUUUGC
crRNA
UACUG ACC GAC AAACG
C G C
UACUG ACC GAC AAACG
C G C

FIG. 18

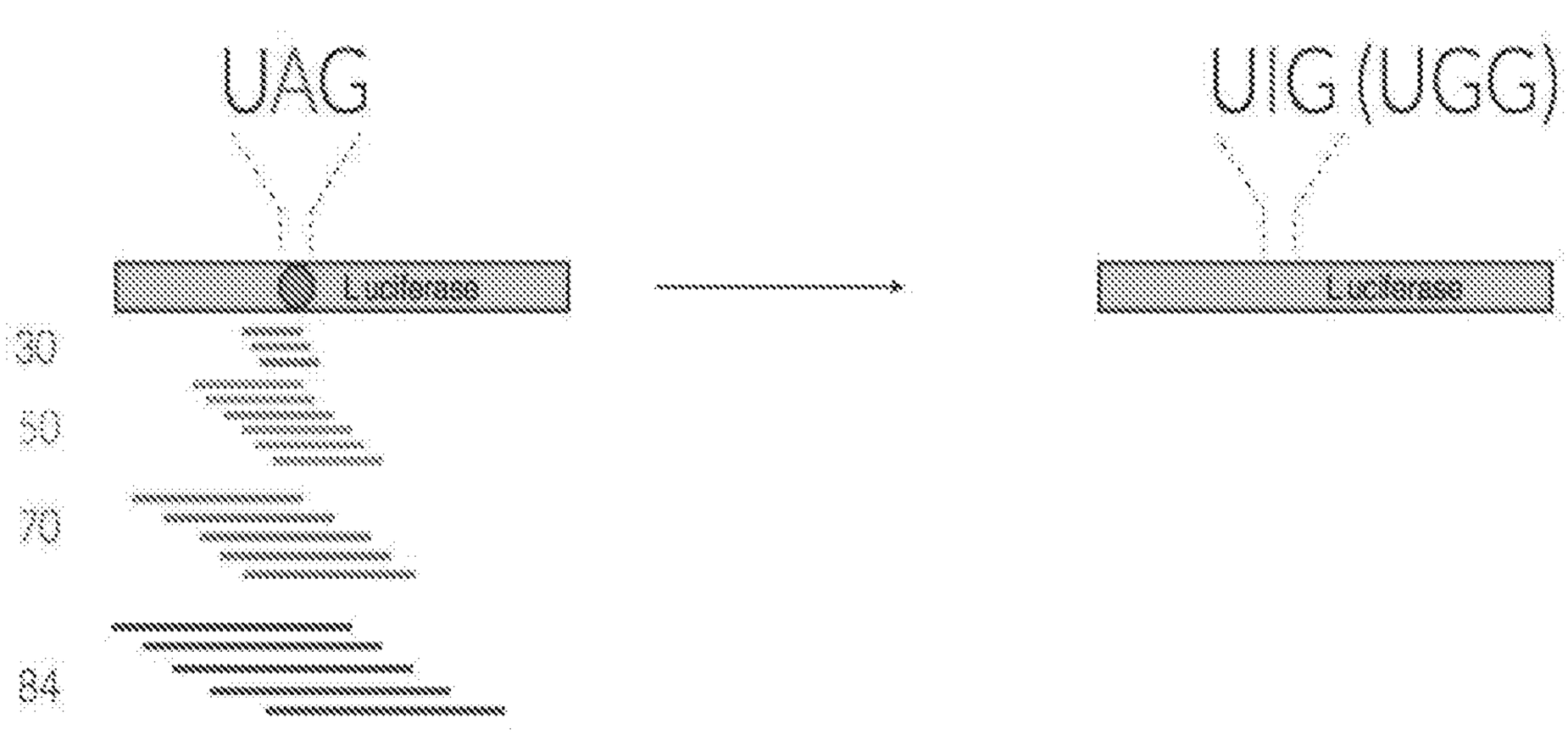
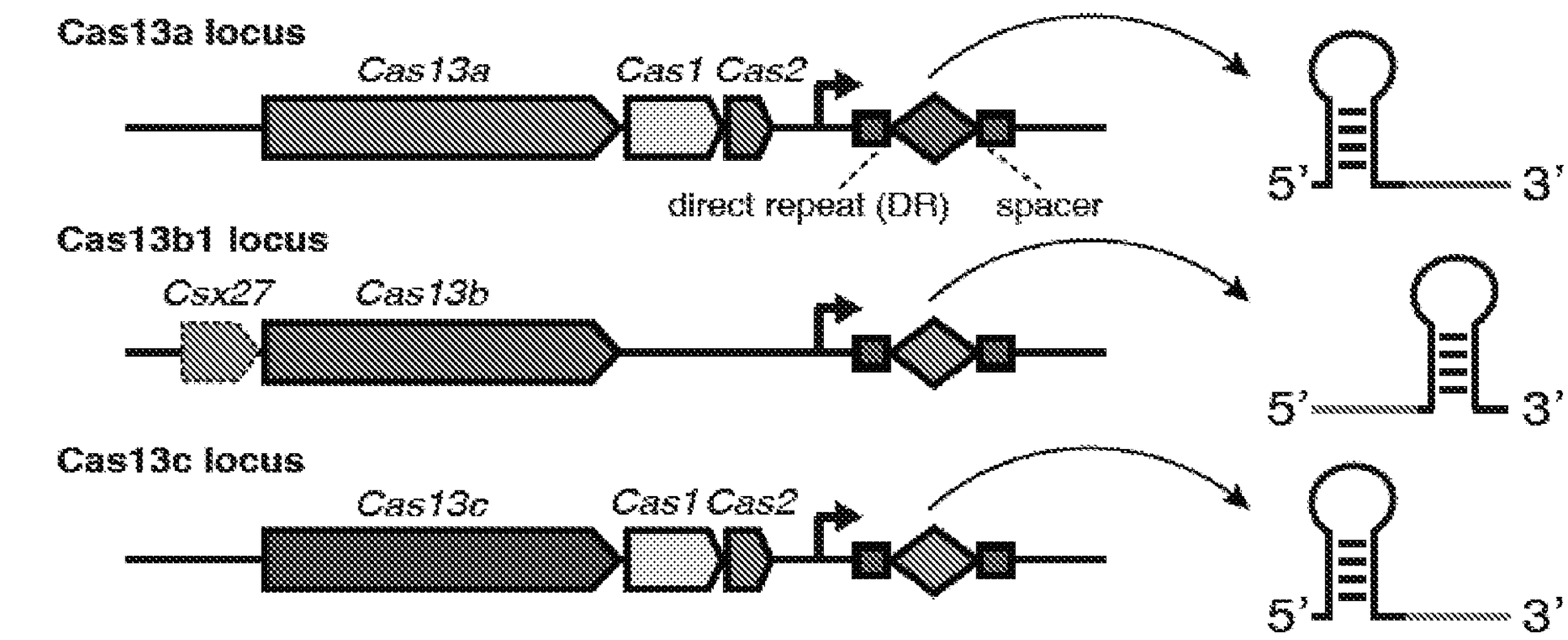

UAG
UIG (UGG)
Luciferase
Luciferase
30
50
70
84

FIG. 19


A
Cas13a locus
Cas13a
Cas1 Cas2
direct repeat (DR)
spacer
5'
3'
Cas13b1 locus
Csx27
Cas13b
5'
3'
Cas13c locus
Cas13c
Cas1 Cas2
5'
3'

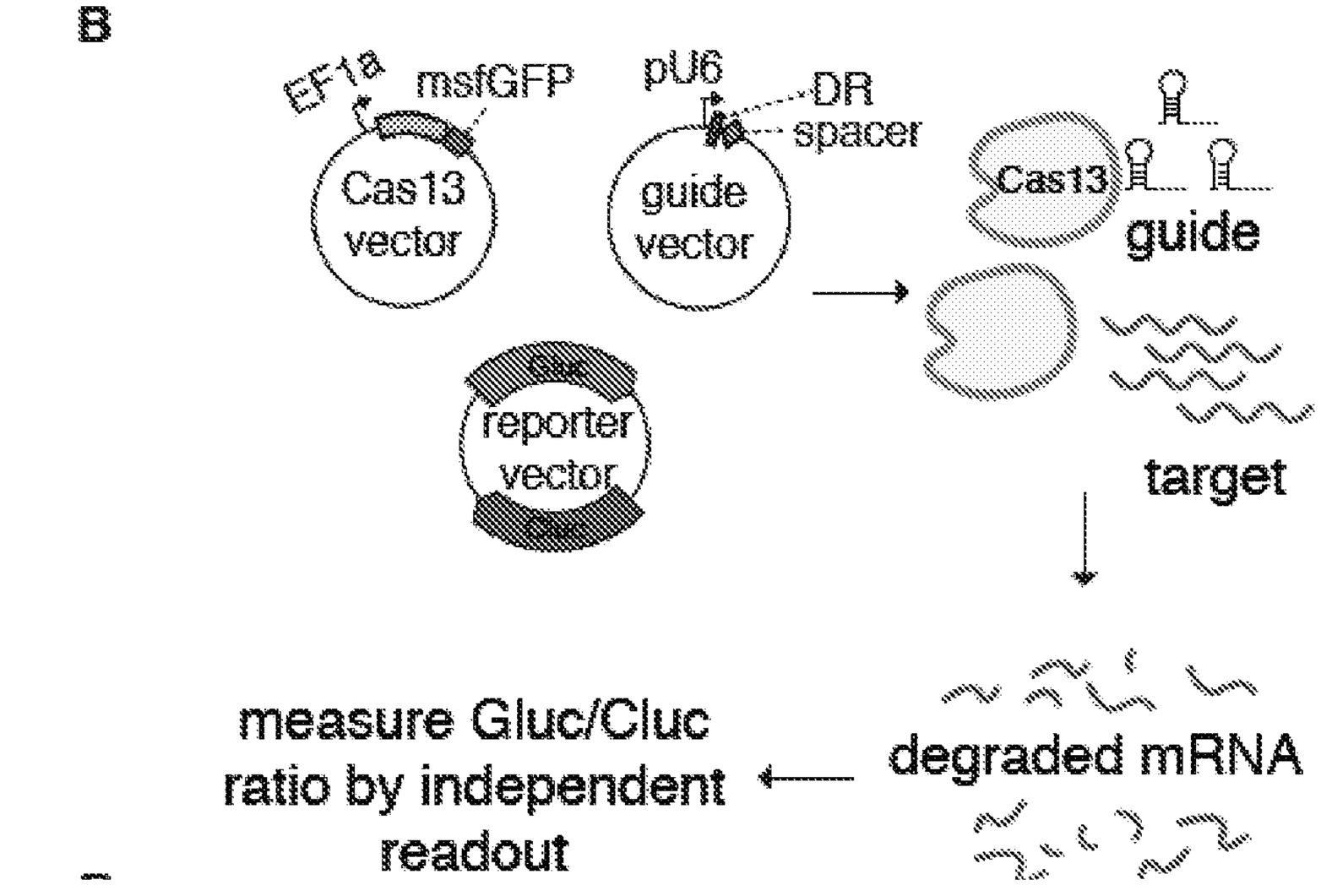
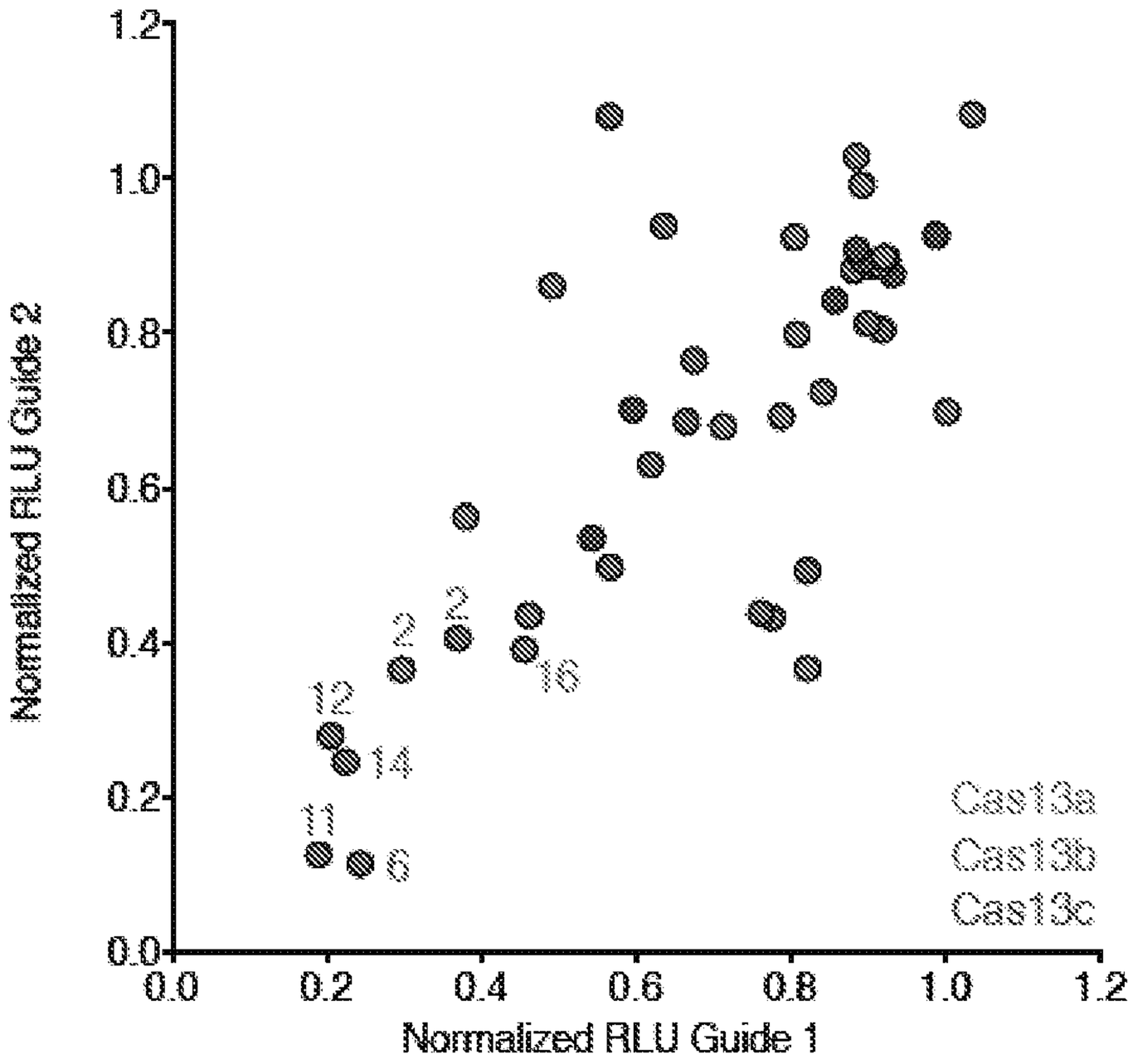

B
EF1α
msfGFP
Cas13 vector
pU6
DR
spacer
guide vector
reporter vector
Cas13
guide
target
degraded mRNA
measure Gluc/Cluc ratio by independent readout

FIG. 20

C



D

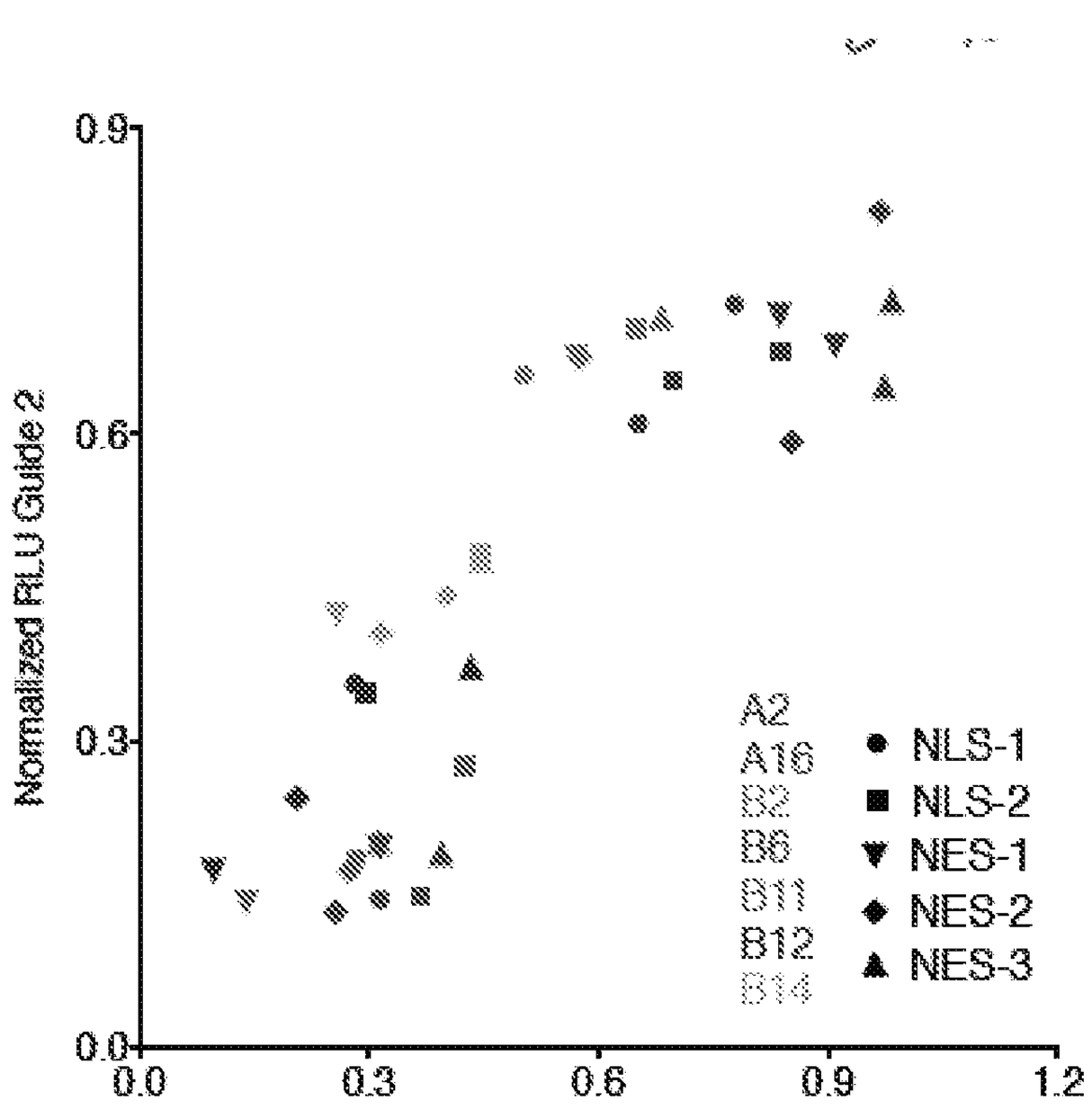


FIG. 20 (continued)

E

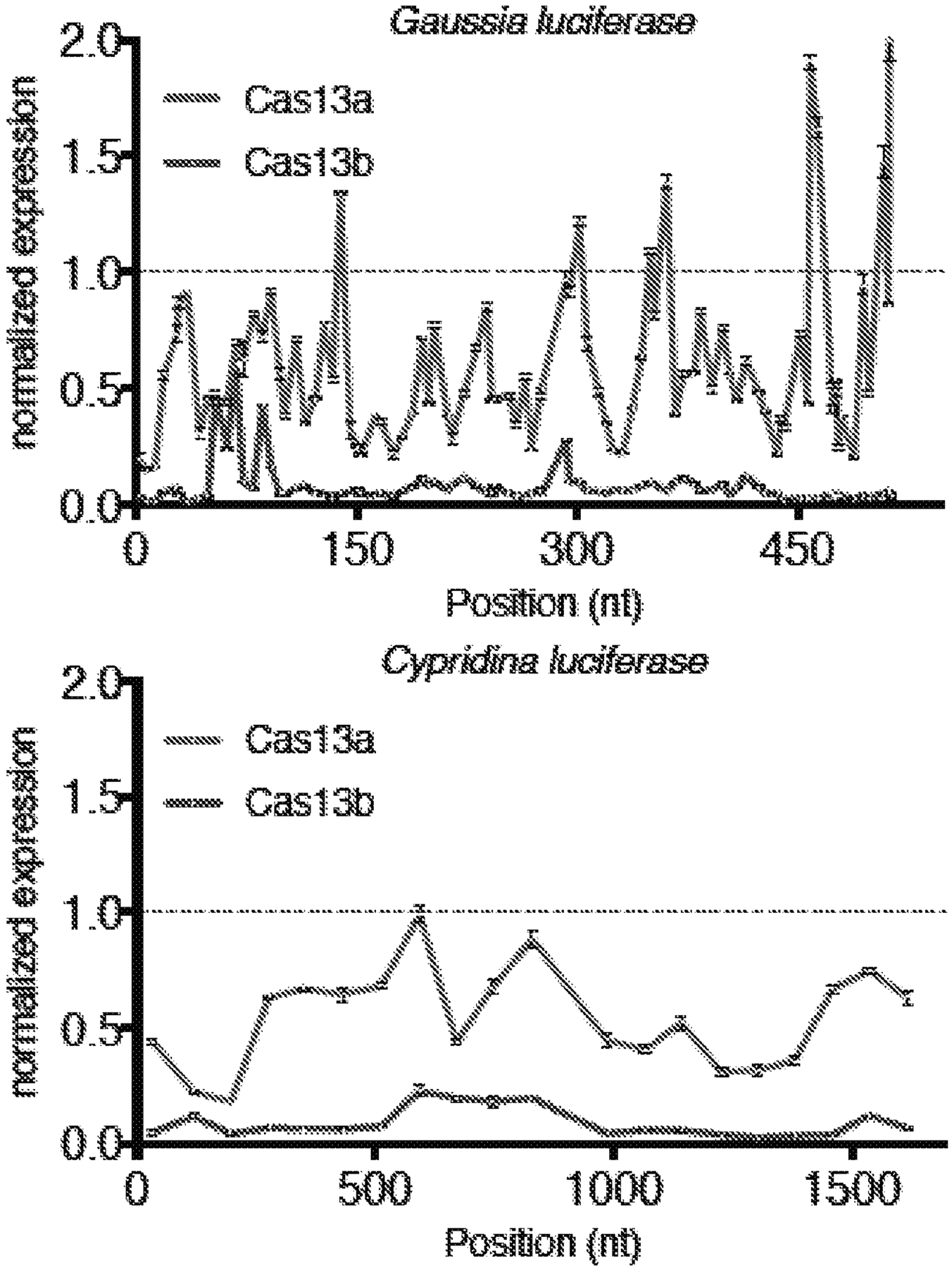


FIG. 20 (continued)

F

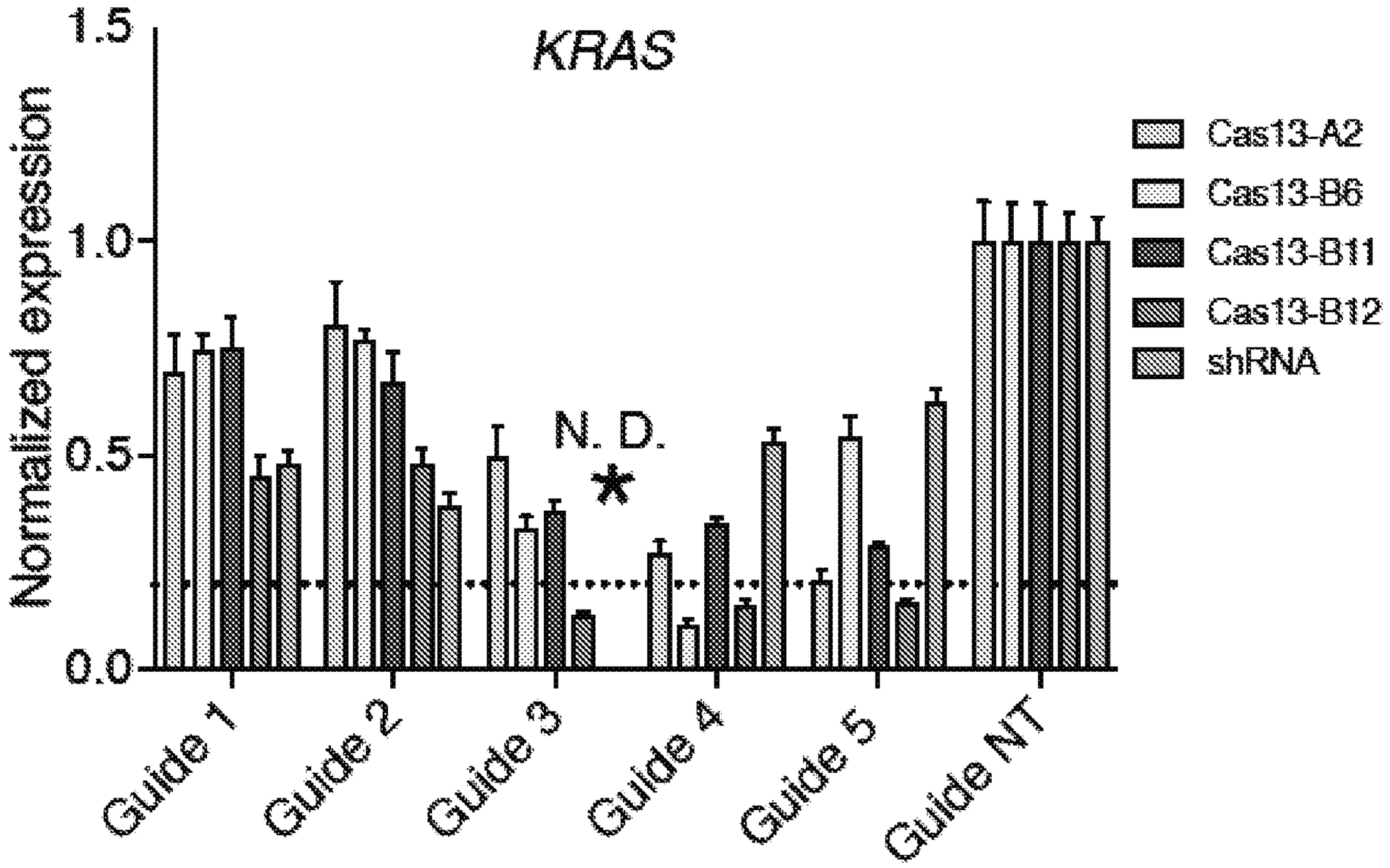


FIG. 20 (continued)

Cas13a target

AGAGATGGAAGCCAATGCCCGGAAAGCTGGCTGC

all single mismatches

NGAGATGGAAGCCAATGCCCGGAAAGCTGGCTGC
ANAGATGGAAGCCAATGCCCGGAAAGCTGGCTGC
AGNGATGGAAGCCAATGCCCGGAAAGCTGGCTGC
AGANATGGAAGCCAATGCCCGGAAAGCTGGCTGC
⋮
AGAGATGGAAGCCAATGCCCGGAAAGCTGGCTGN

all double mismatches

NNAGATGGAAGCCAATGCCCGGAAAGCTGGCTGC
NGNGATGGAAGCCAATGCCCGGAAAGCTGGCTGC
NGANATGGAAGCCAATGCCCGGAAAGCTGGCTGC
NGAGNTGGAAGCCAATGCCCGGAAAGCTGGCTGC
⋮
AGAGATGGAAGCCAATGCCCGGAAAGCTGGCTNN

WT Cas13b target

TGGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA

all single mismatches

NGGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
TNGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
TGNAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
TGGNGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
⋮
TGGNGGTGCTCAAAGAGATGGAAGCCAATGCCCGGN

all double mismatches

NNGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
NGNAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
NGGNGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
NGGANGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
⋮
TGGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGNN

FIG. 21

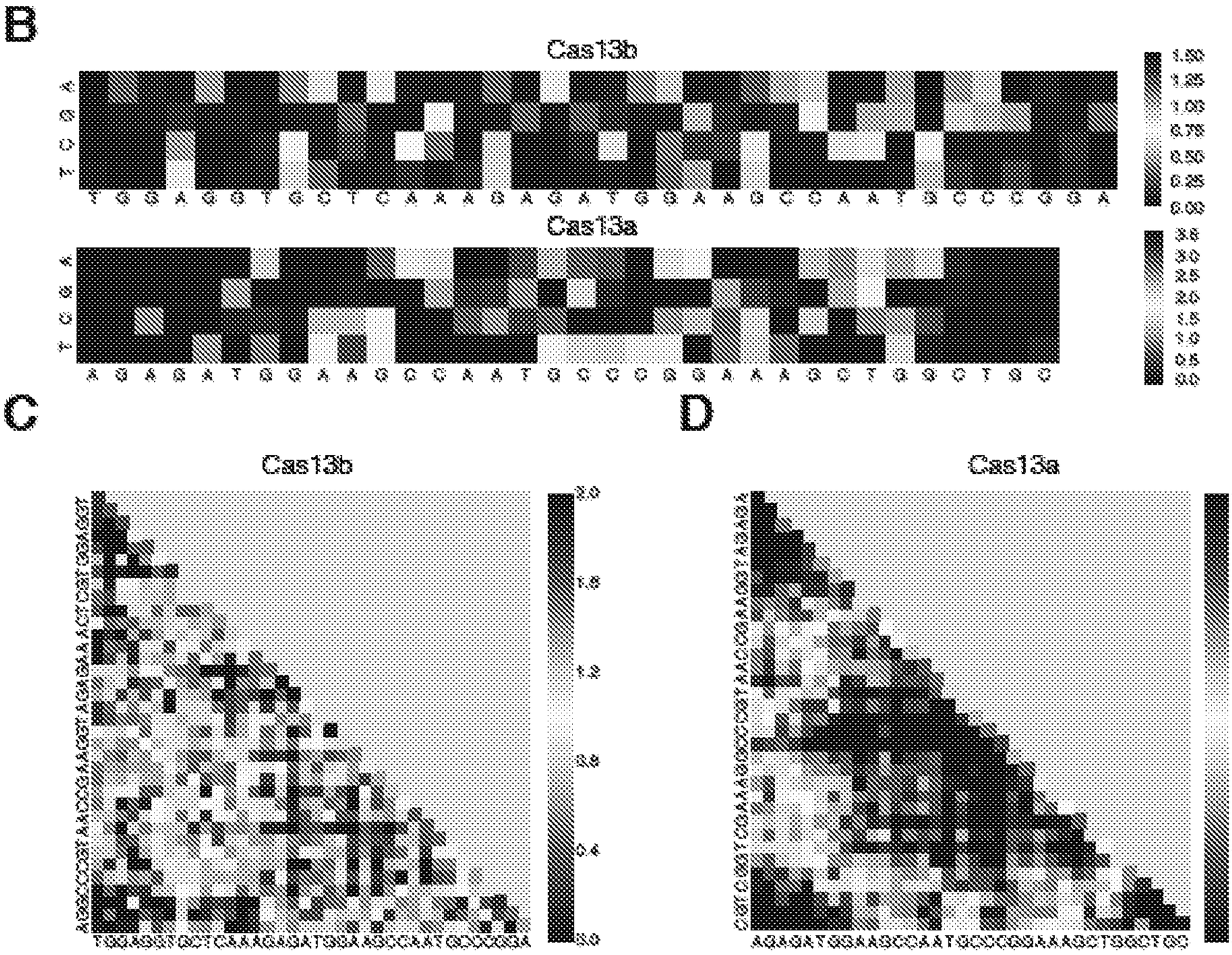
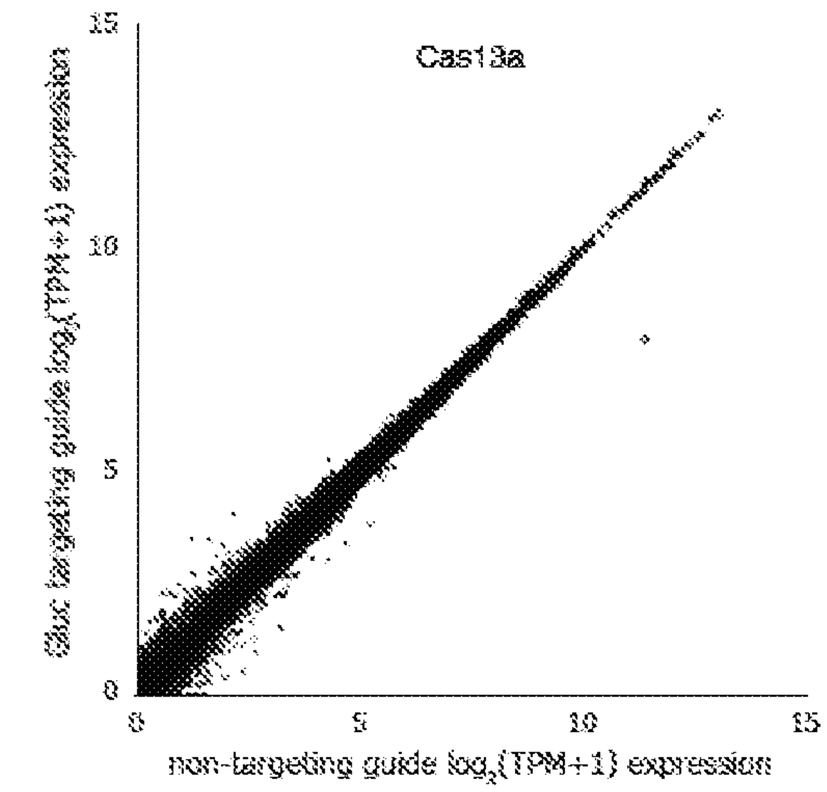
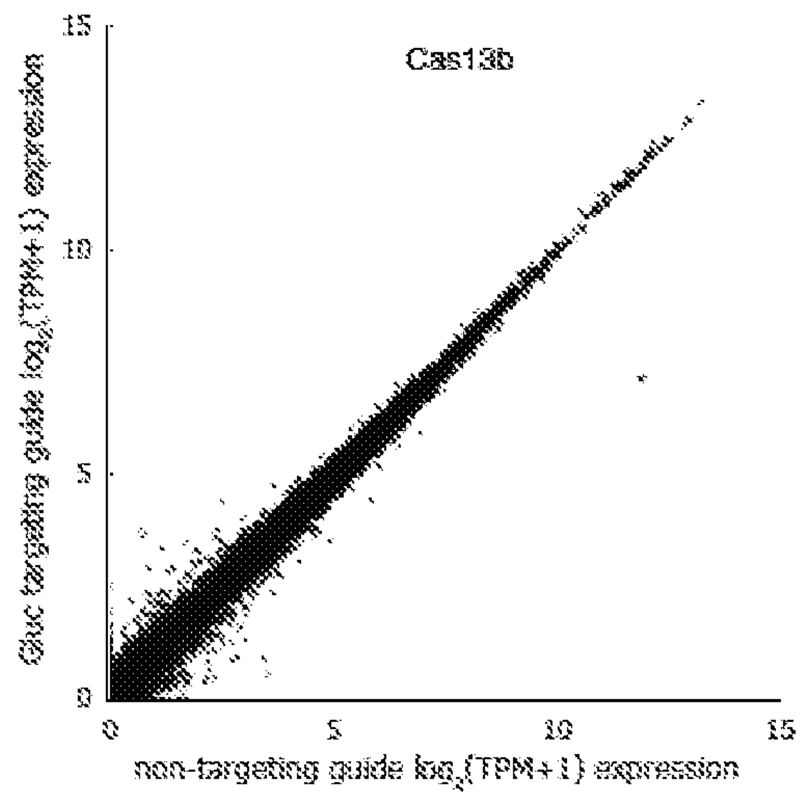


FIG. 21 (continued)

E





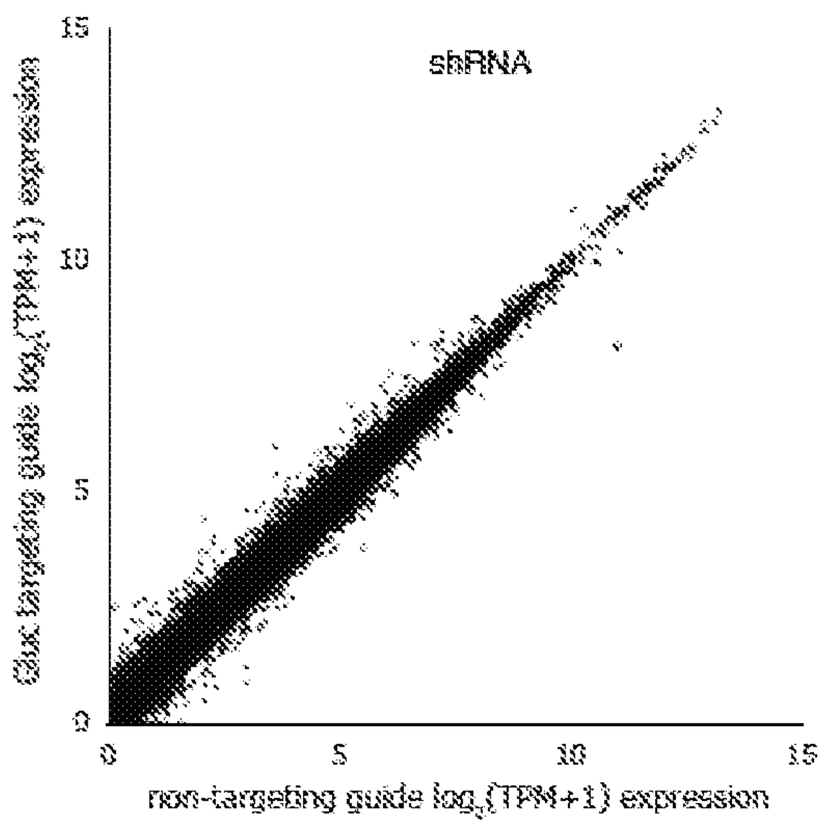
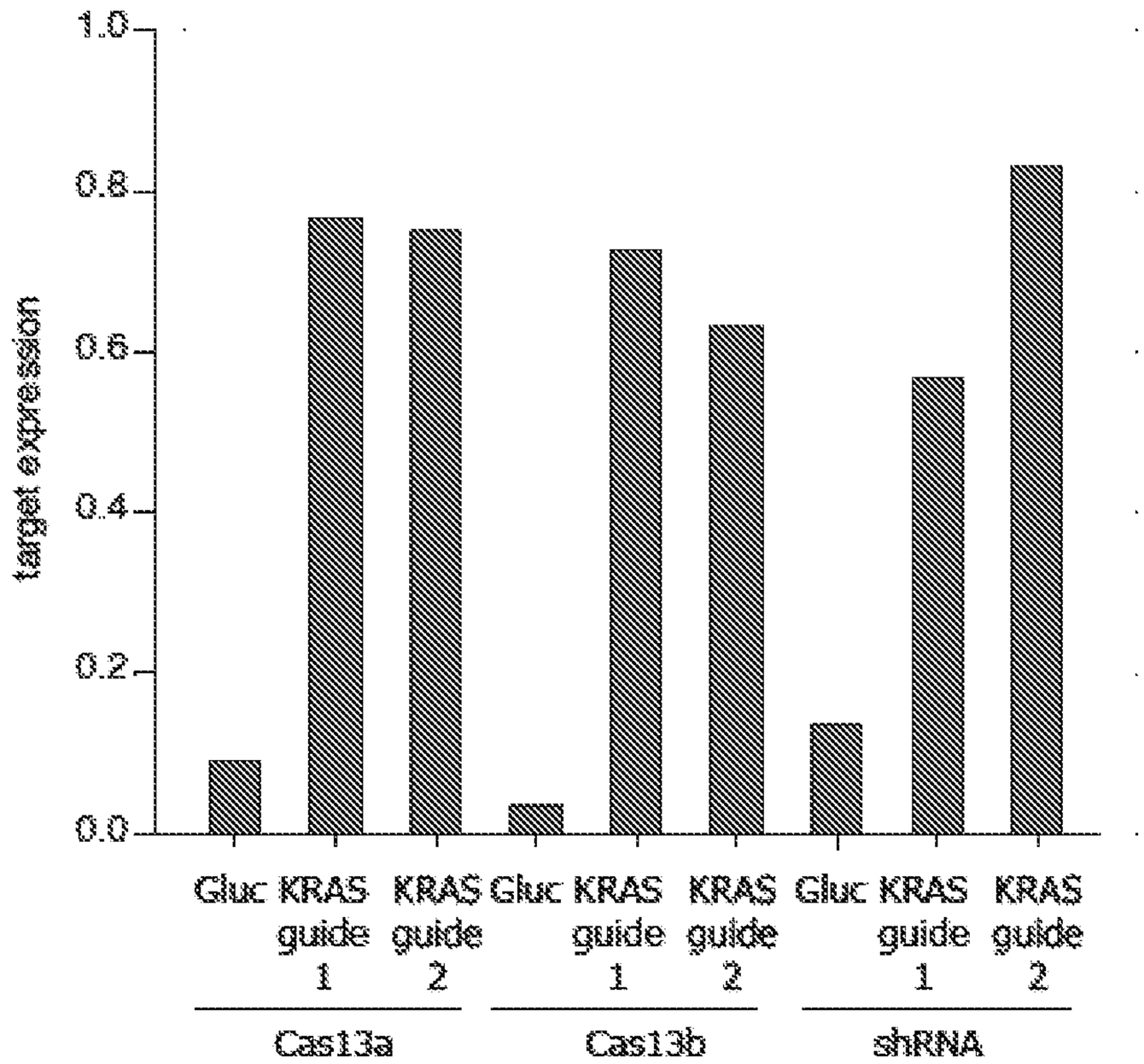


FIG. 21 (continued)

F



G

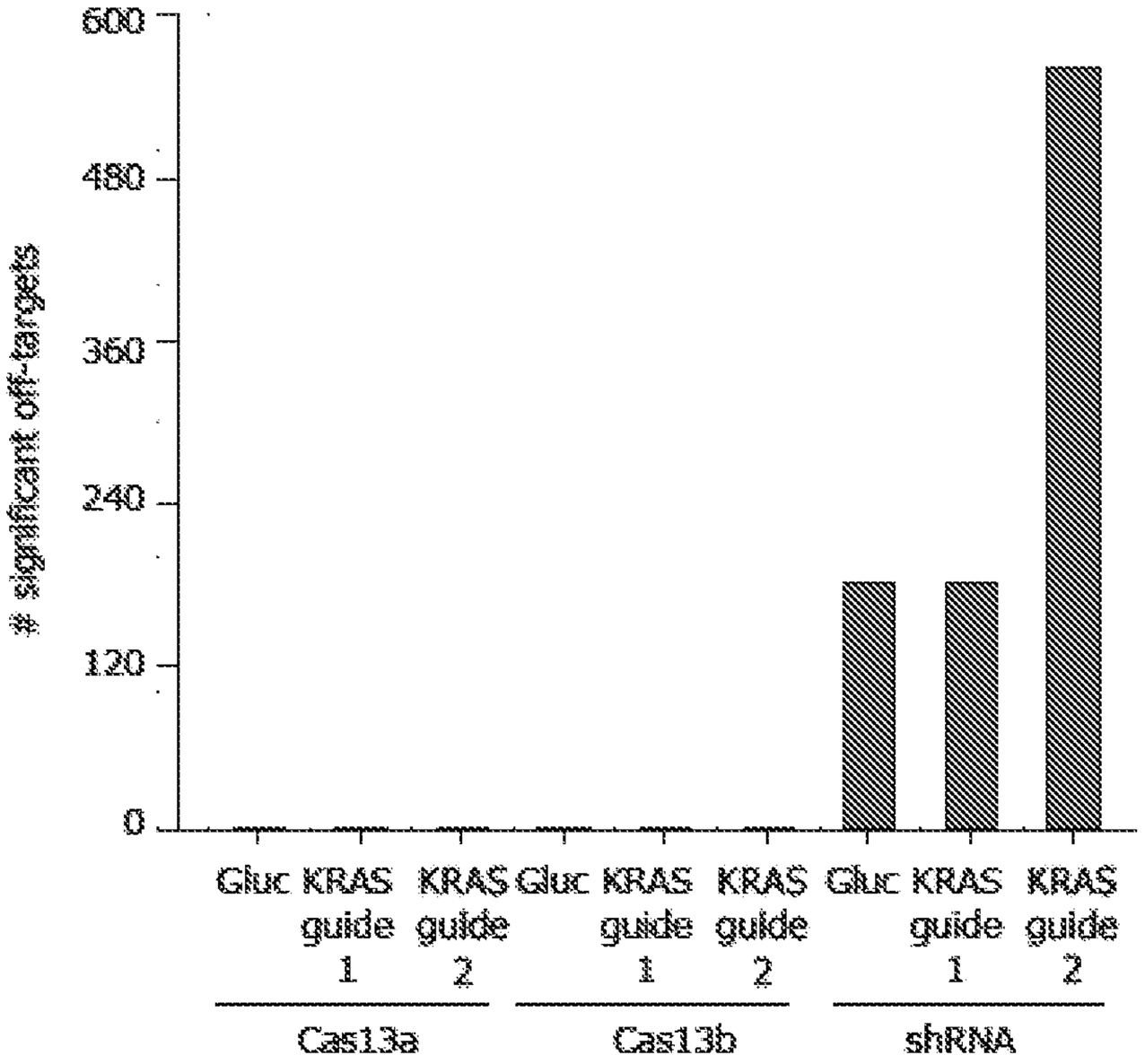
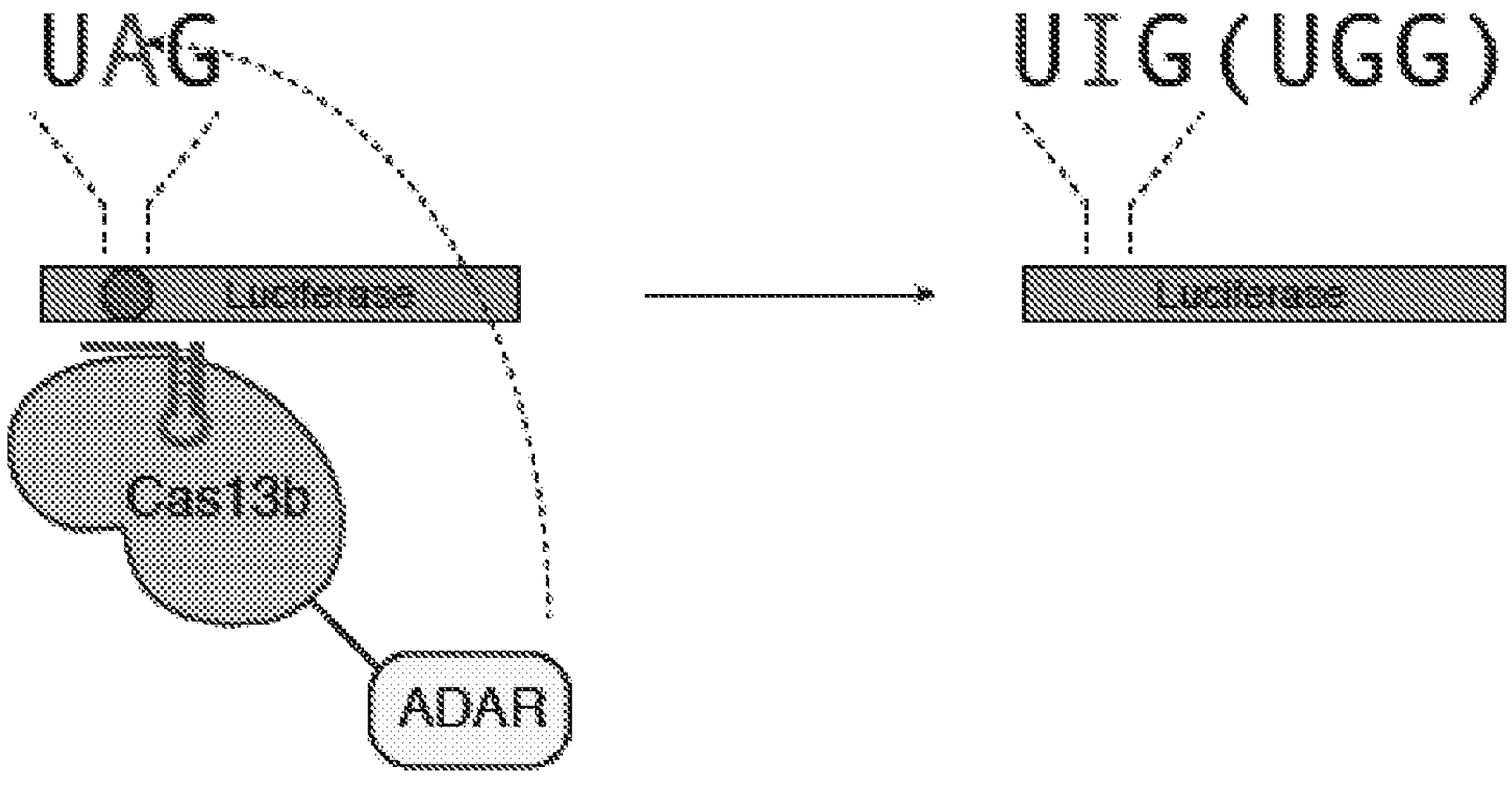


FIG. 21 (continued)

A


UAG
Luciferase
Cas13b
ADAR
UIG (UGG)
Luciferase

B

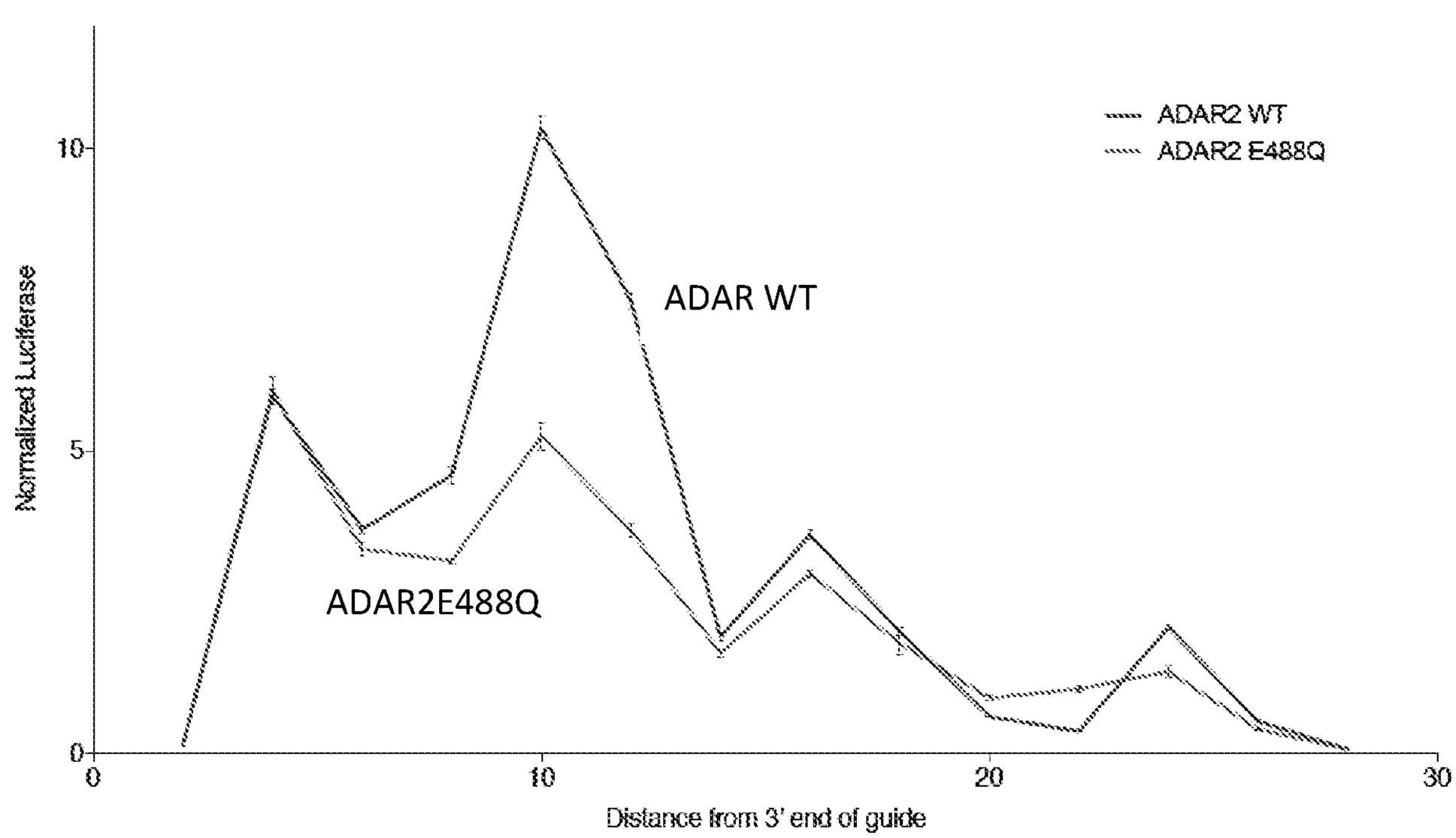

ADAR2 WT
ADAR2 E488Q
ADAR WT
ADAR2E488Q
Normalized Luciferase
0
5
10
0
10
20
30
Distance from 3' end of guide

FIG. 22

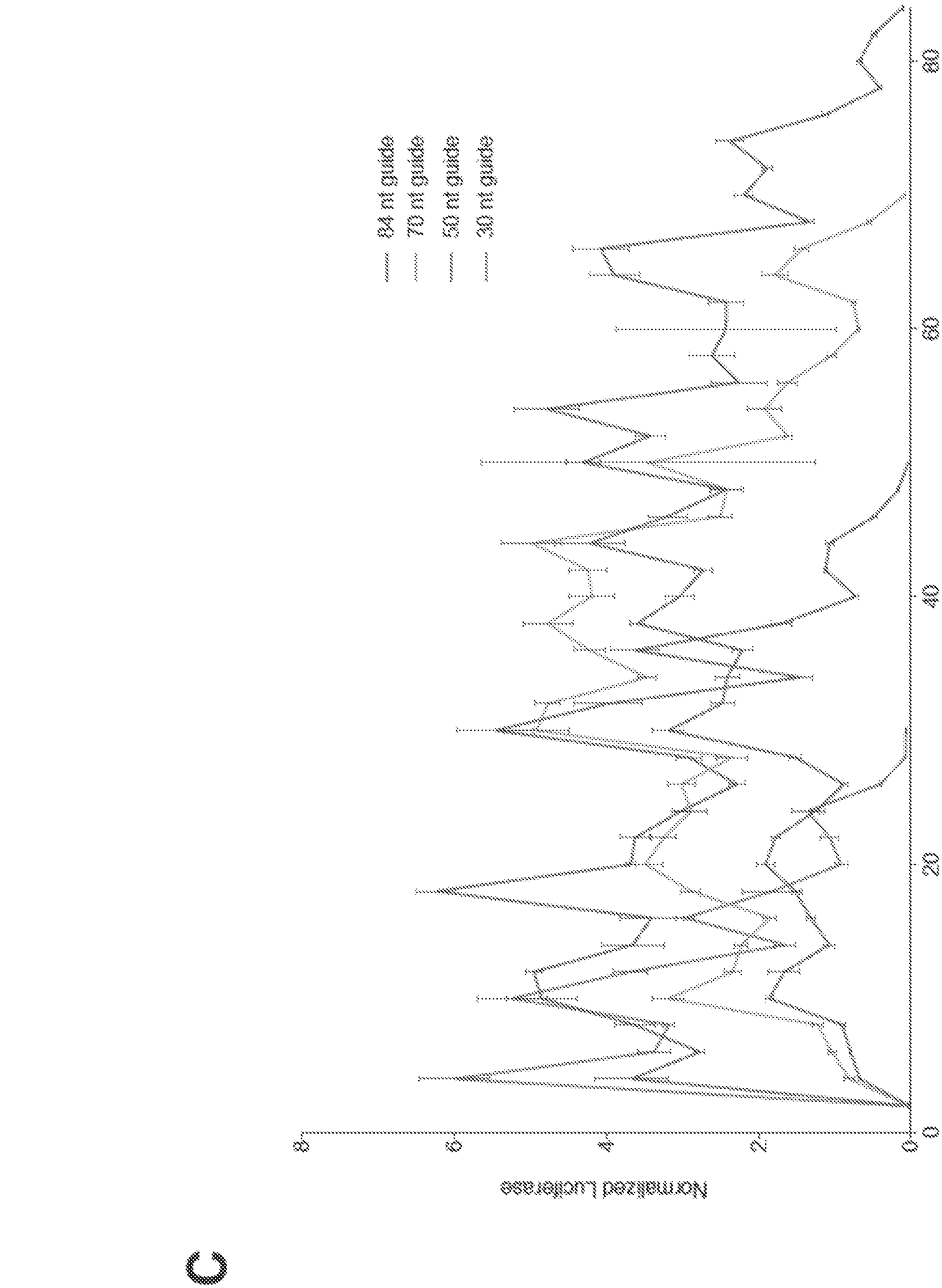
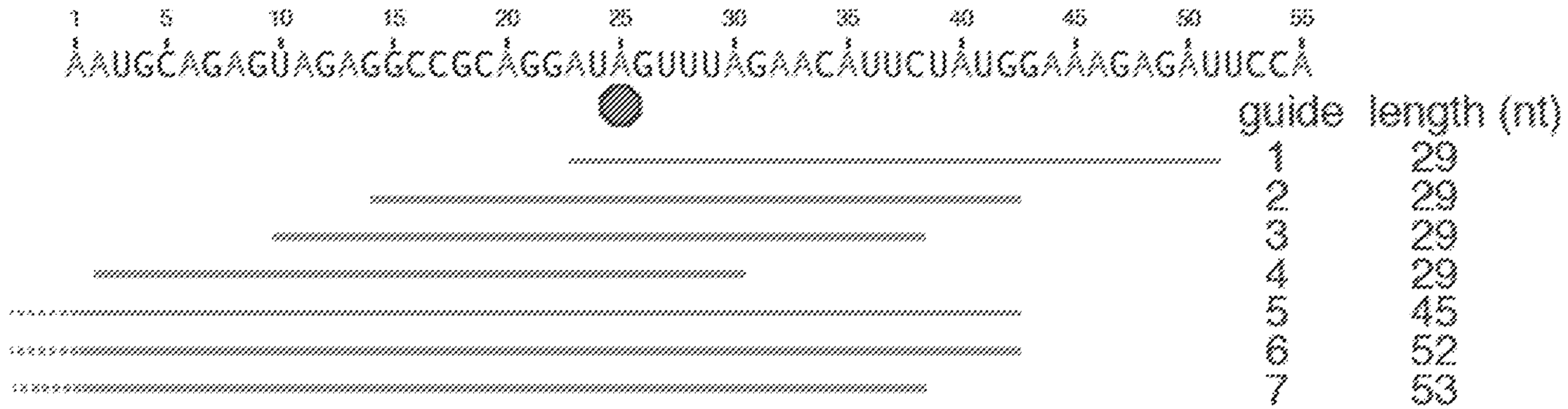

84 nt guide
70 nt guide
50 nt guide
30 nt guide
Normalized Luciferase
0
2
4
6
8
20
40
60
80
Distance from 3' end of guide

FIG. 22

D


1 5 10 15 20 25 30 35 40 45 50 55
AAUGCAGAGUAGAGGCCGCAGGAUAGUUUAGAACAUUCUAUGGAAAGAGAUUCCA
guide length (nt)
1 29
2 29
3 29
4 29
5 45
6 52
7 53

E

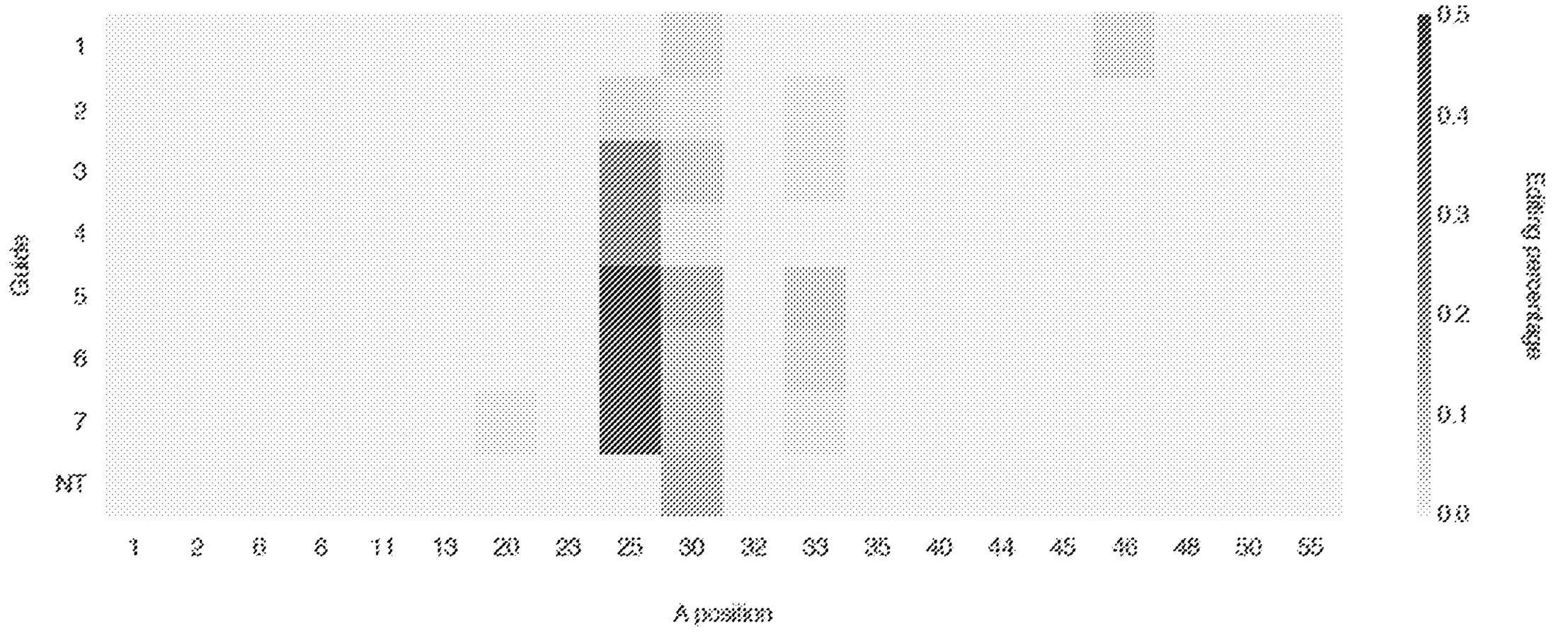

Guide
1
2
3
4
5
6
7
NT
1 2 6 6 11 13 20 23 25 30 32 33 35 40 44 45 46 48 50 55
A position
Editing percentage
0.5
0.4
0.3
0.2
0.1
0.0

FIG. 22

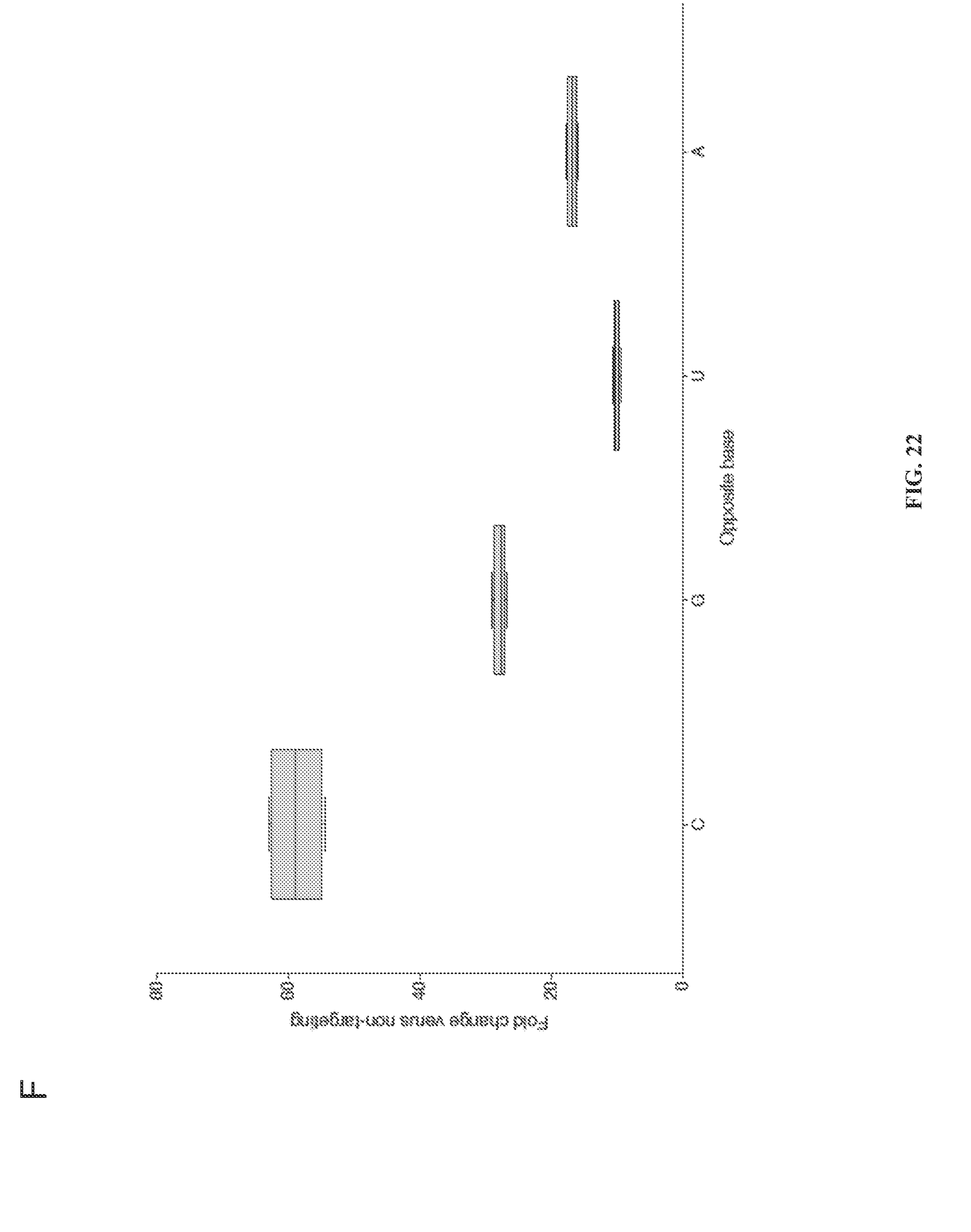


FIG. 22

A

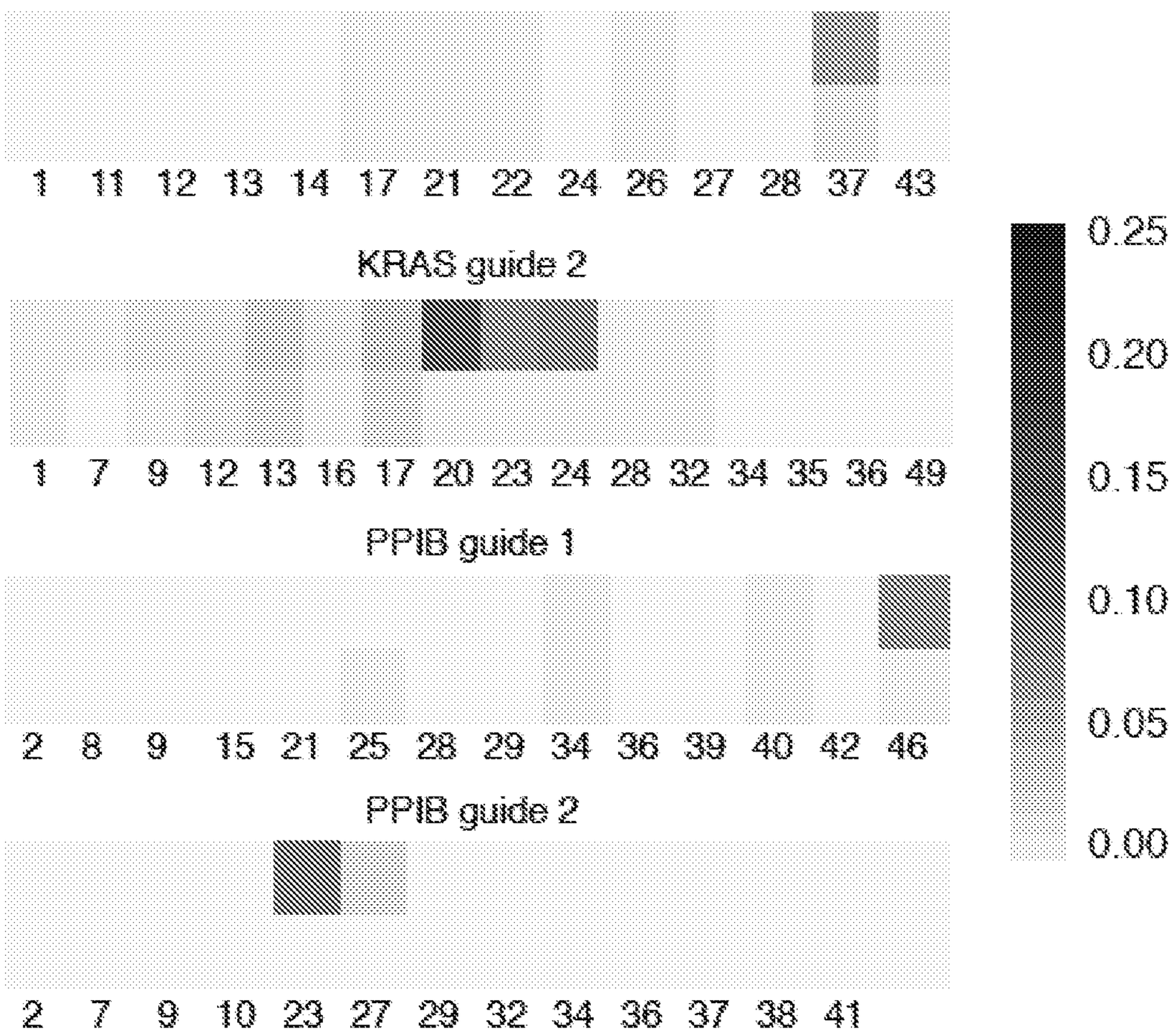

KRAS guide 1
1 11 12 13 14 17 21 22 24 26 27 28 37 43
KRAS guide 2
1 7 9 12 13 16 17 20 23 24 28 32 34 35 36 49
PPIB guide 1
2 8 9 15 21 25 28 29 34 36 39 40 42 46
PPIB guide 2
2 7 9 10 23 27 29 32 34 36 37 38 41
0.25
0.20
0.15
0.10
0.05
0.00

FIG. 23



FIG. 24

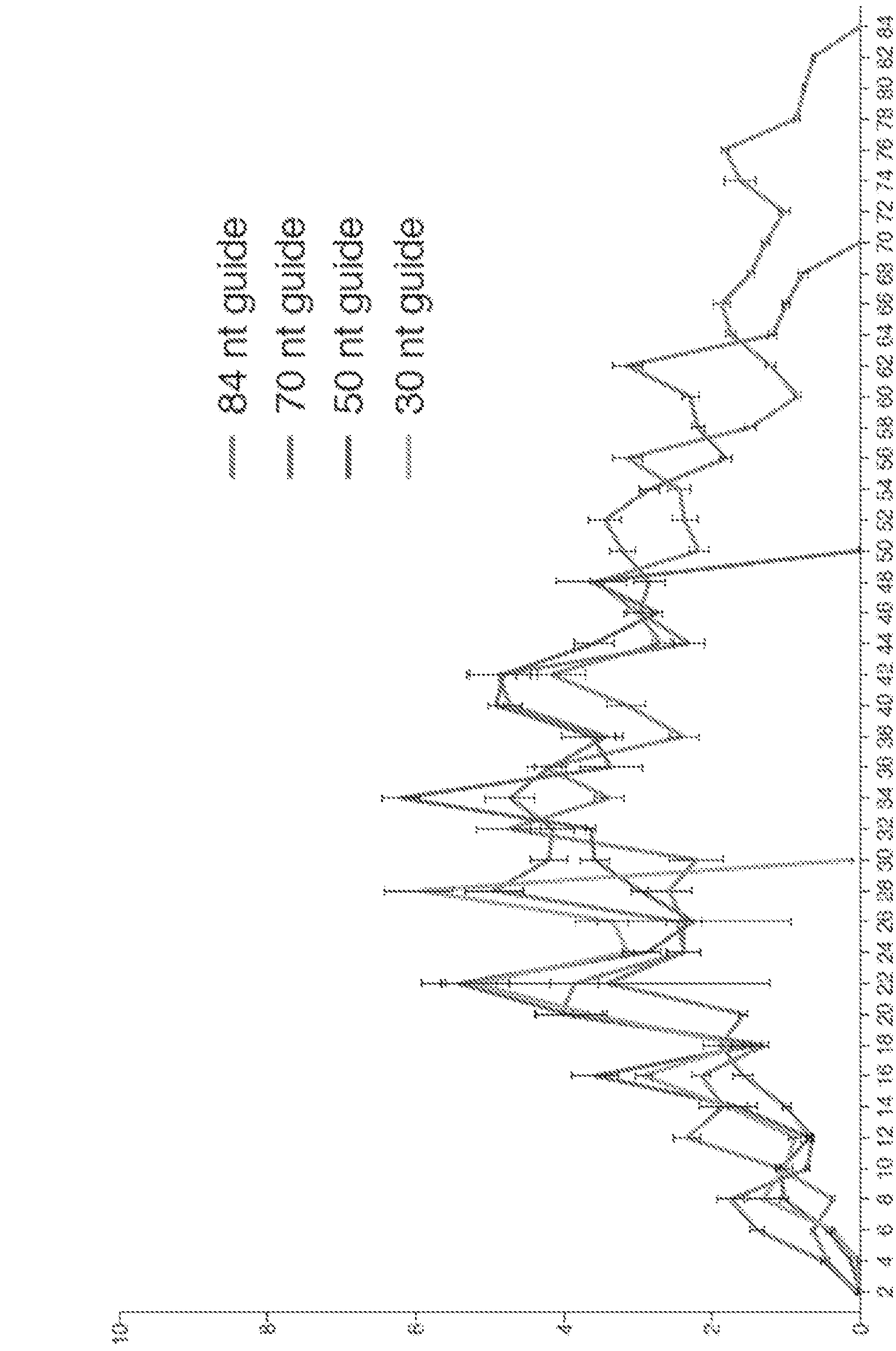

A
84 nt guide
70 nt guide
50 nt guide
30 nt guide
Distance from 3' end of guide
Background subtracted RLU

FIG. 25

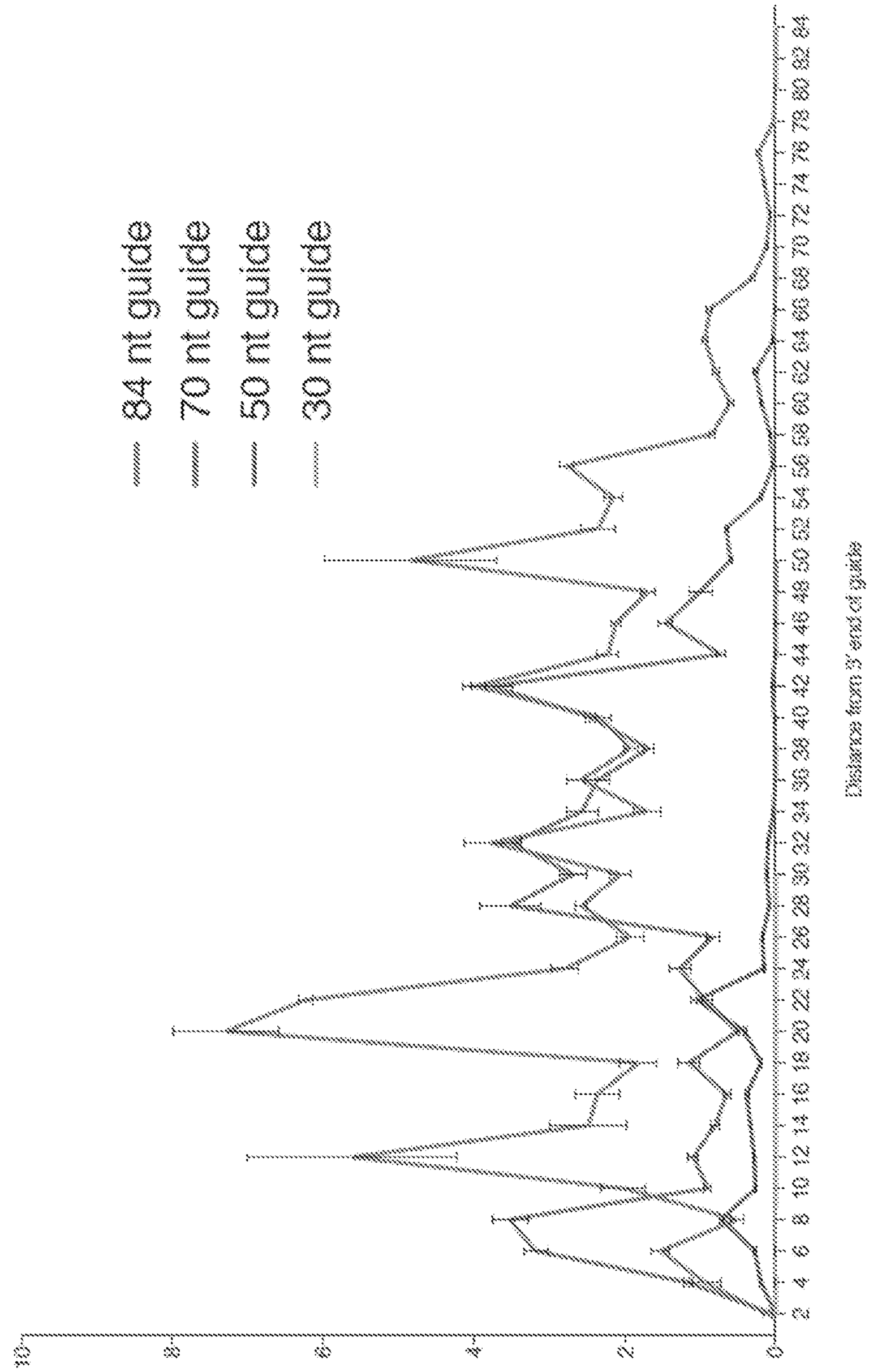
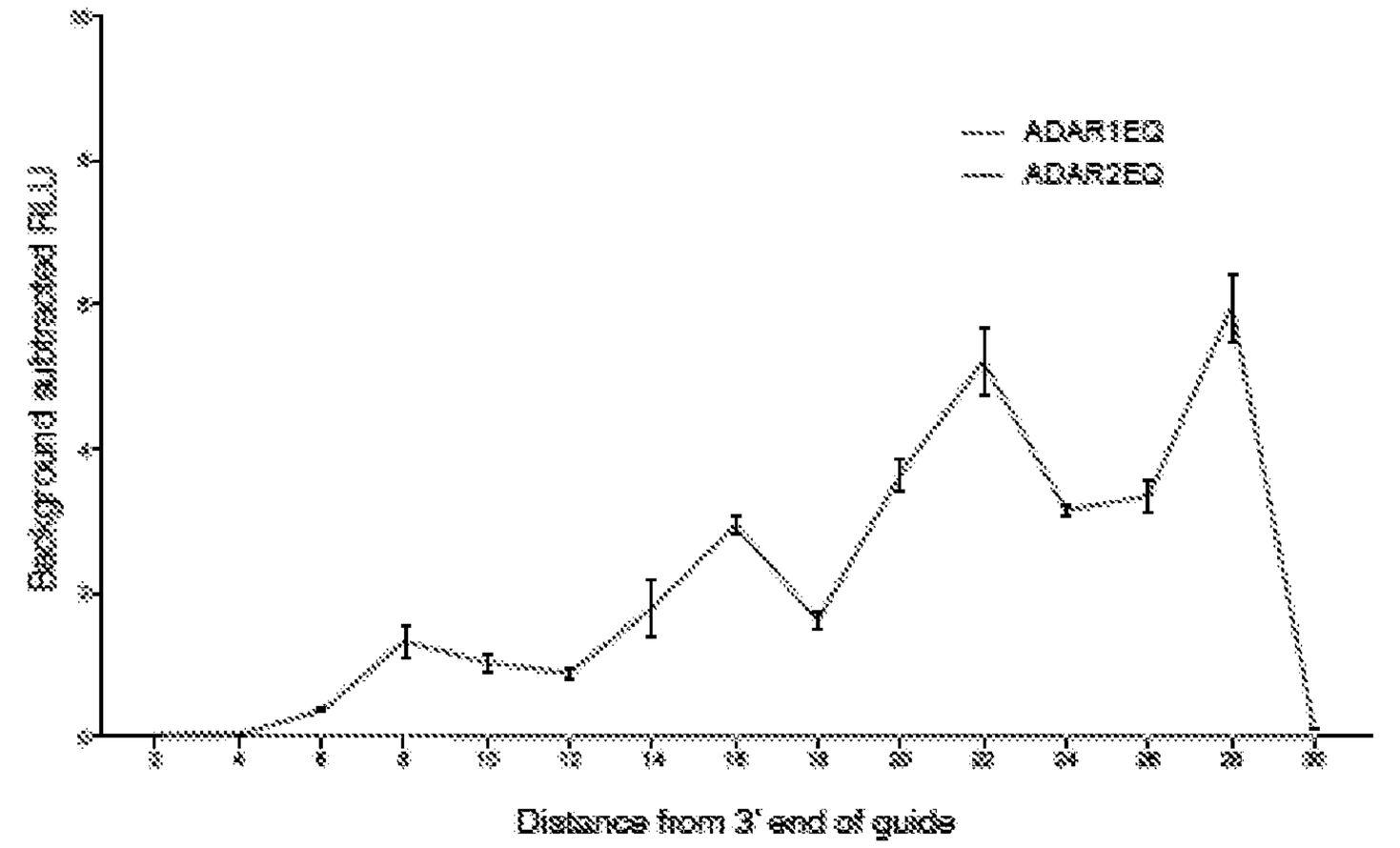

84 nt guide
70 nt guide
50 nt guide
30 nt guide
Background subtracted RLU
Distance between guide 5' ends (nt)

FIG. 25

C



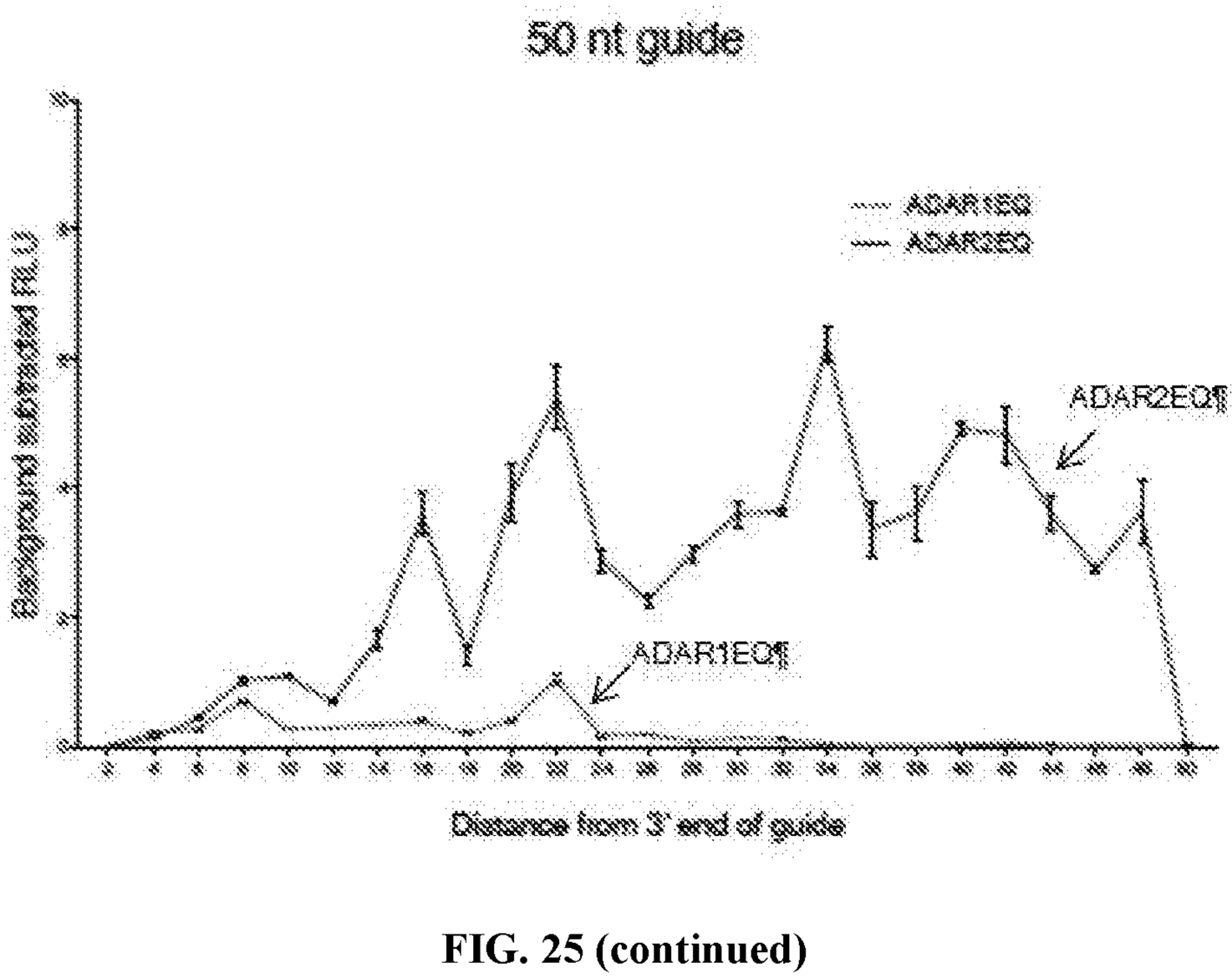


FIG. 25 (continued)

C

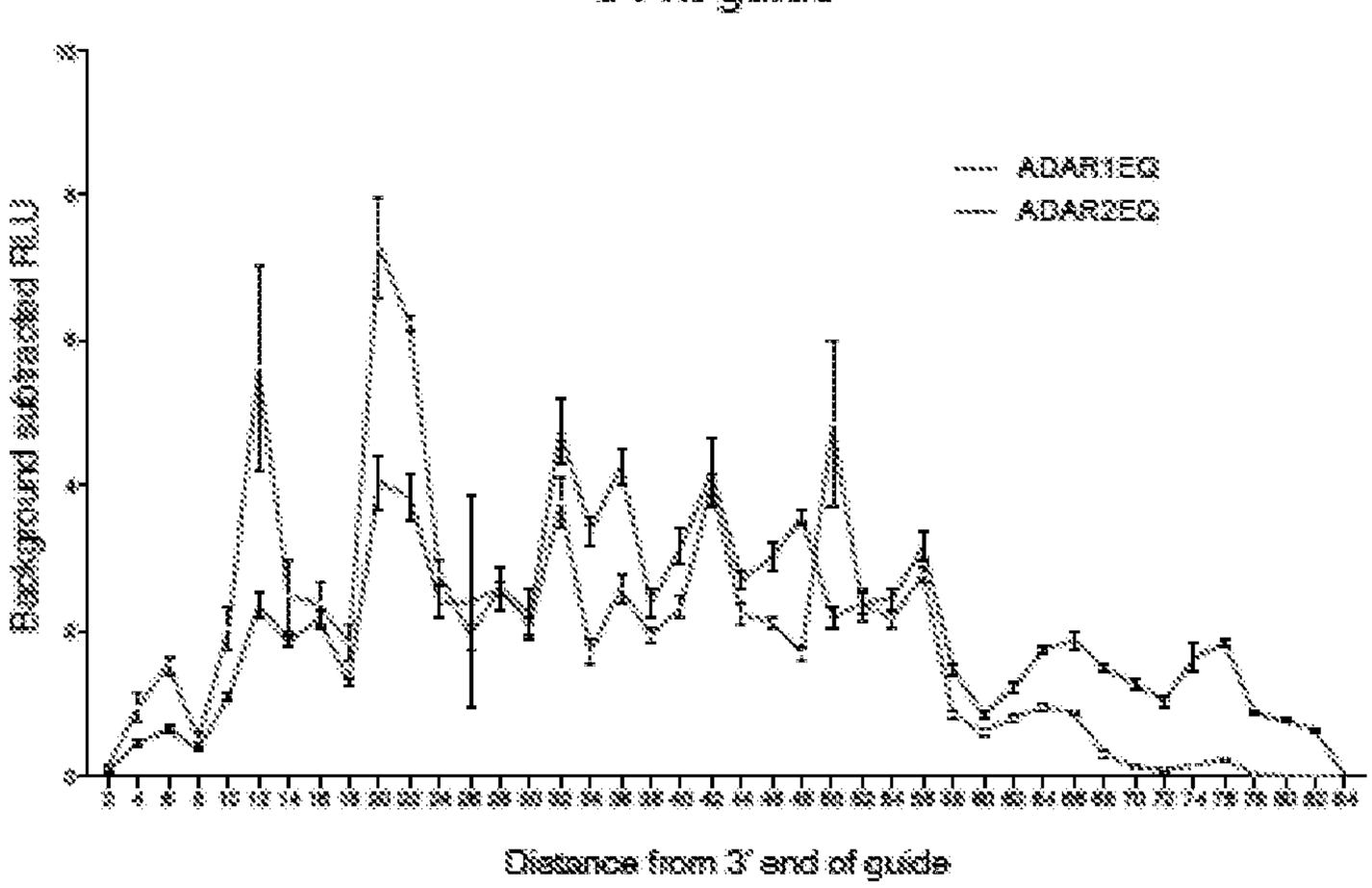


FIG. 25 (continued)

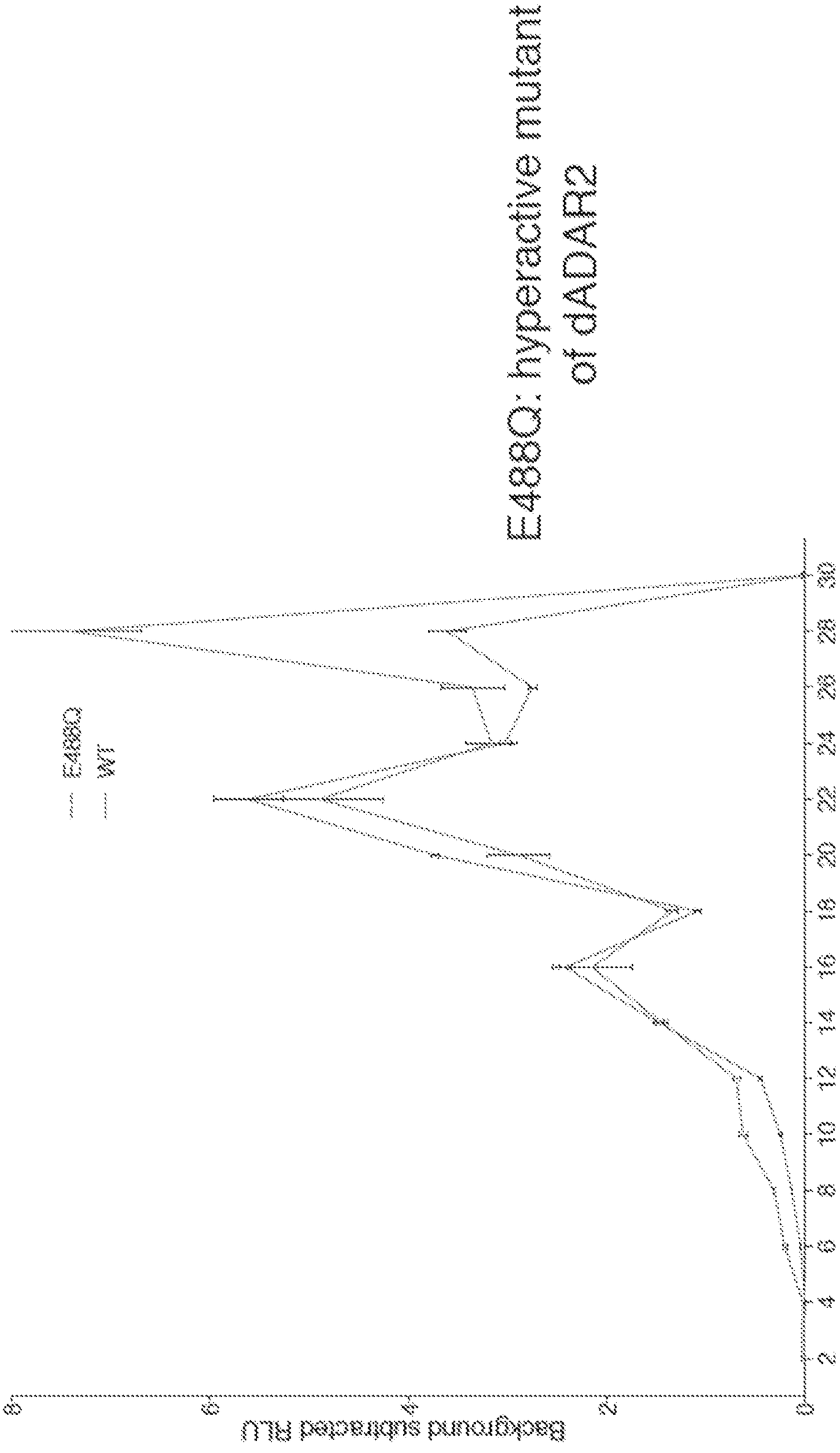

E480Q: hyperactive mutant of dAR2
WT
E480Q
background subtracted RLU
background subtracted counts
0
2
4
6
8
10
12
14
16
18
20
22
24

FIG. 26

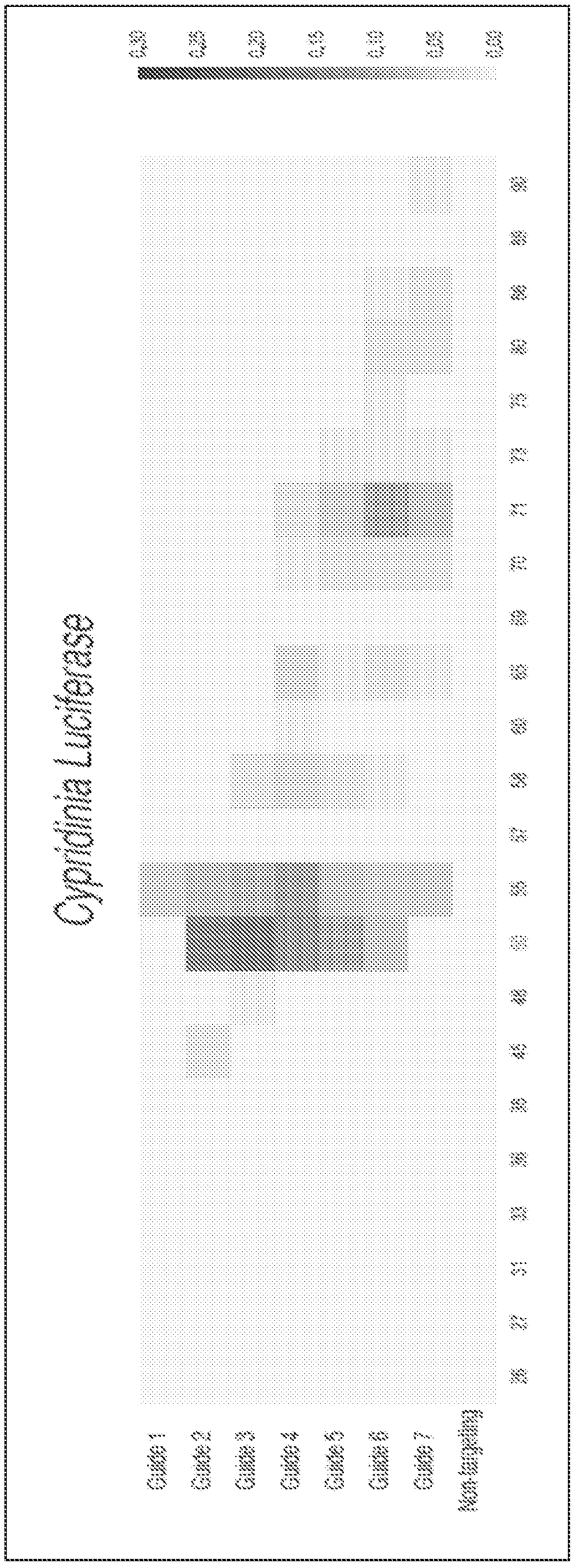

Cypridina Luciferase
Well Number
Cube 1
Cube 2
Cube 3
Cube 4
Cube 5
Cube 6
Cube 7
Well Number
A

FIG. 27

B

Guide	3' start pos.	5' start pos.	A-C pos.
1	1	50	50
2	9	59	50
3	17	67	50
4	25	75	50
5	33	83	50
6	41	91	50
7	49	99	50
NT			

FIG. 27 Cont

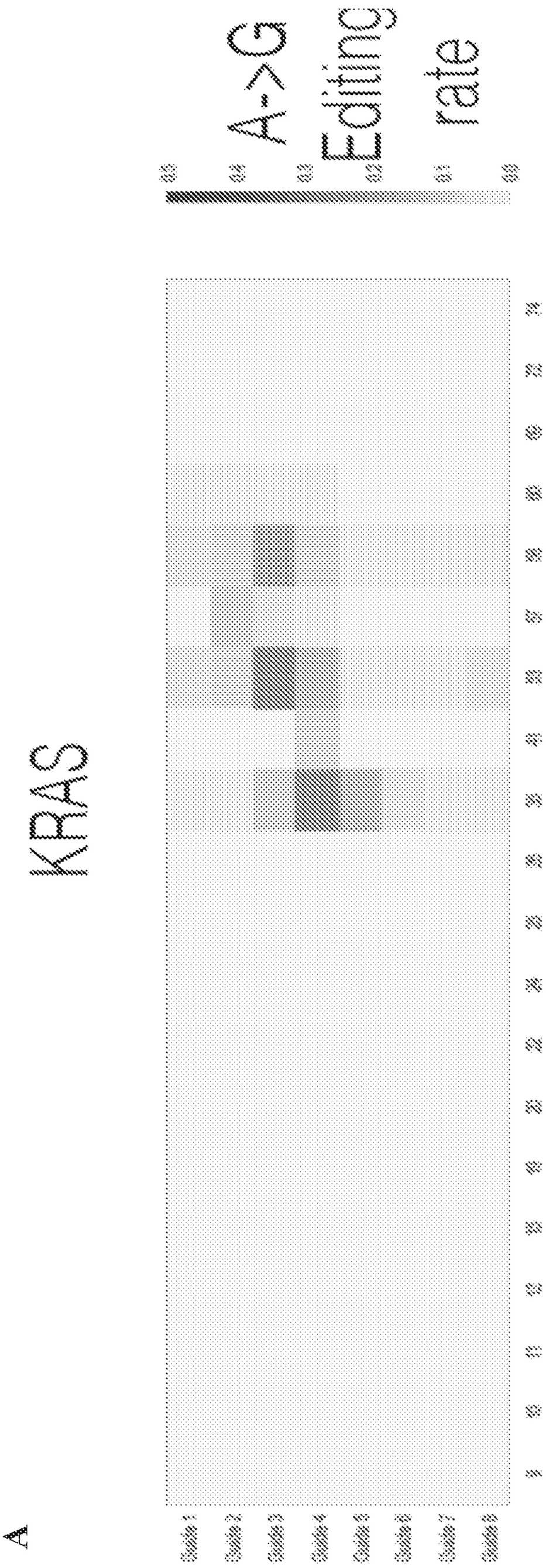


FIG. 28

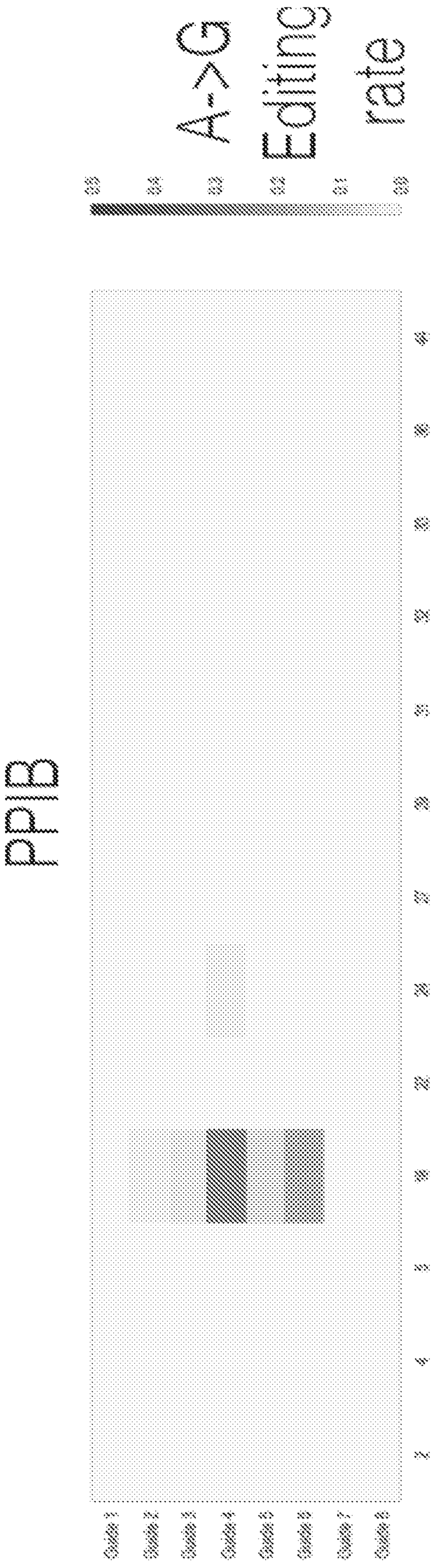
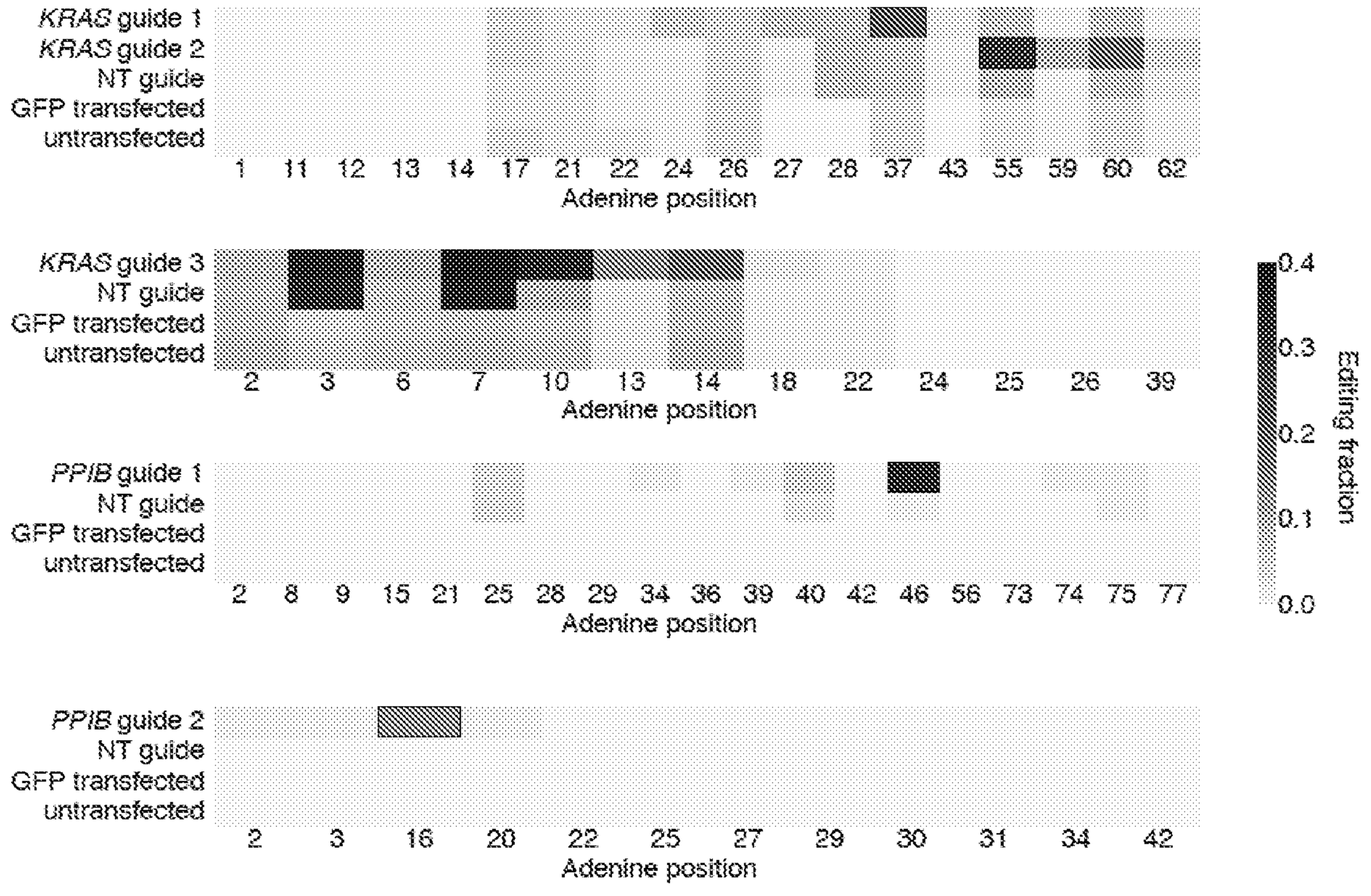

A-to-G Editing rate
Position
Guide
Guide 1
Guide 2
Guide 3
Guide 4
Guide 5
Guide 6
Guide 7
Guide 8
Accumulated mutations
Guide
Guide 1
Guide 2
Guide 3
Guide 4
Guide 5
Guide 6
Guide 7
Guide 8
Guide combination
Position
Guide
Guide 1
Guide 2
Guide 3
Guide 4
Guide 5
Guide 6
Guide 7
Guide 8
A
B
C
D

FIG. 28 cont.


KRAS guide 1
KRAS guide 2
NT guide
GFP transfected
untransfected
1 11 12 13 14 17 21 22 24 26 27 28 37 43 55 59 60 62
Adenine position
KRAS guide 3
NT guide
GFP transfected
untransfected
2 3 6 7 10 13 14 18 22 24 25 26 39
Adenine position
PPIB guide 1
NT guide
GFP transfected
untransfected
2 8 9 15 21 25 28 29 34 36 39 40 42 46 56 73 74 75 77
Adenine position
PPIB guide 2
NT guide
GFP transfected
untransfected
2 3 16 20 22 25 27 29 30 31 34 42
Adenine position
0.4
0.3
0.2
0.1
0.0
Editing fraction

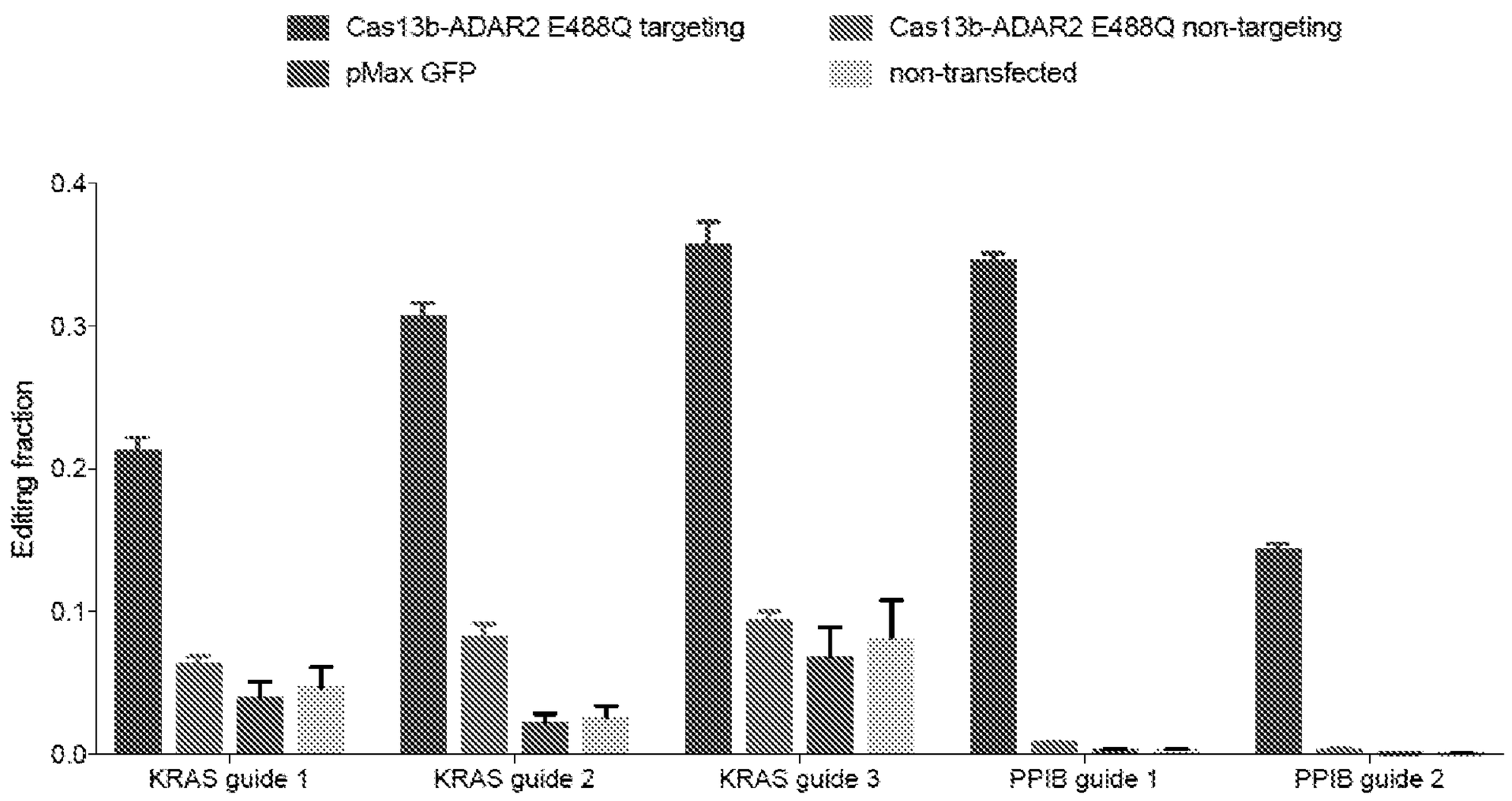

Cas13b-ADAR2 E488Q targeting
Cas13b-ADAR2 E488Q non-targeting
pMax GFP
non-transfected
0.4
0.3
0.2
0.1
0.0
Editing fraction
KRAS guide 1
KRAS guide 2
KRAS guide 3
PPIB guide 1
PPIB guide 2

FIG. 29

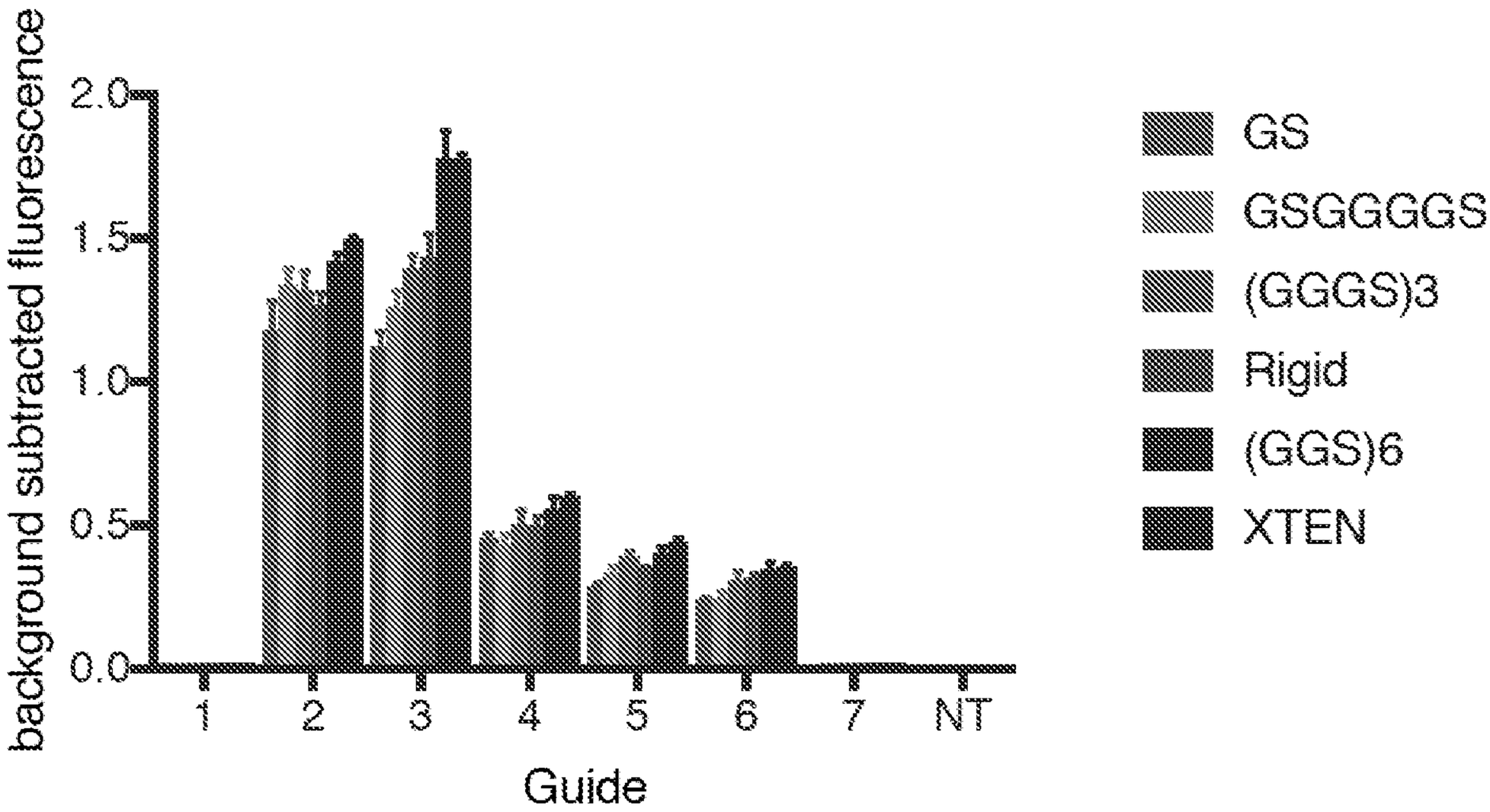

background subtracted fluorescence
2.0
1.5
1.0
0.5
0.0
1
2
3
4
5
6
7
NT
Guide
GS
GSGGGGS
(GGGS)3
Rigid
(GGS)6
XTEN

FIG. 30

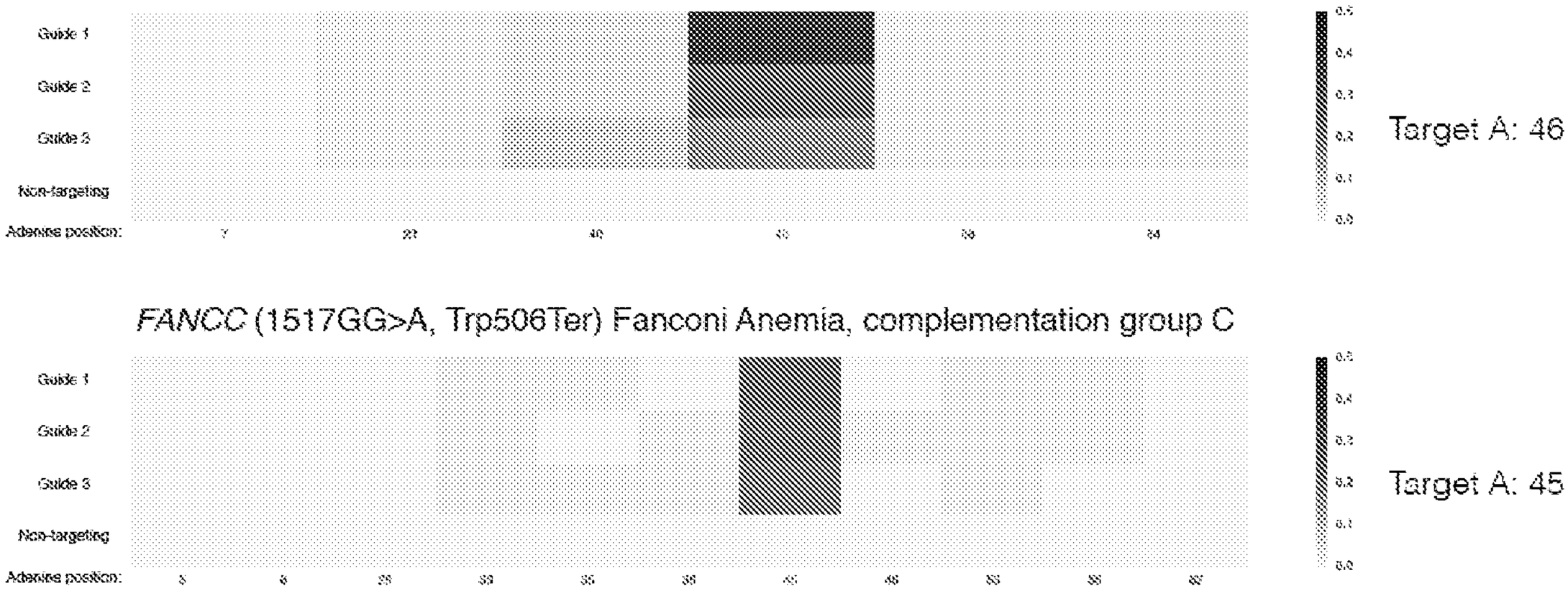

AVPR2 (878G>A, Trp293Ter) X-linked Nephrogenic diabetes insipidus
Guide 1
Guide 2
Guide 3
Non-targeting
Adenine position:
Target A: 46
FANCC (1517GG>A, Trp506Ter) Fanconi Anemia, complementation group C
Guide 1
Guide 2
Guide 3
Non-targeting
Adenine position:
Target A: 45

FIG. 31

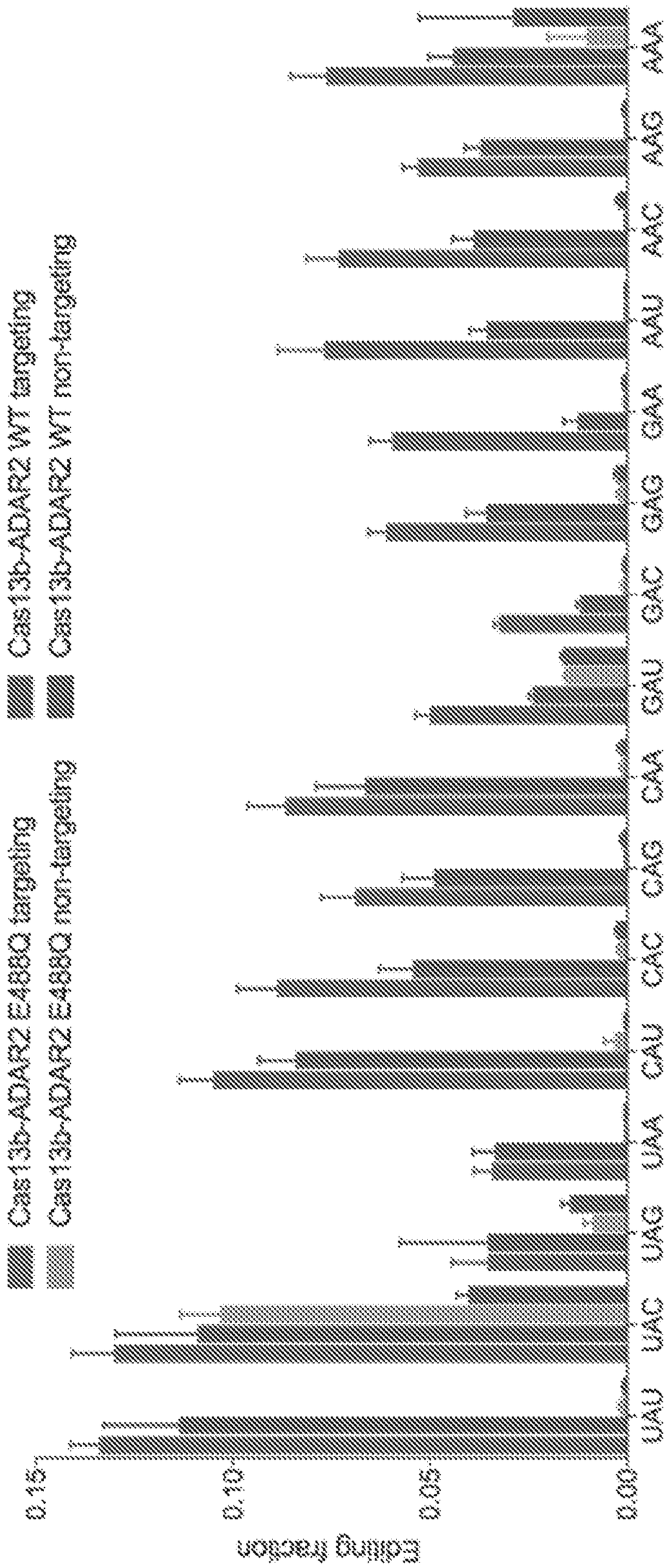
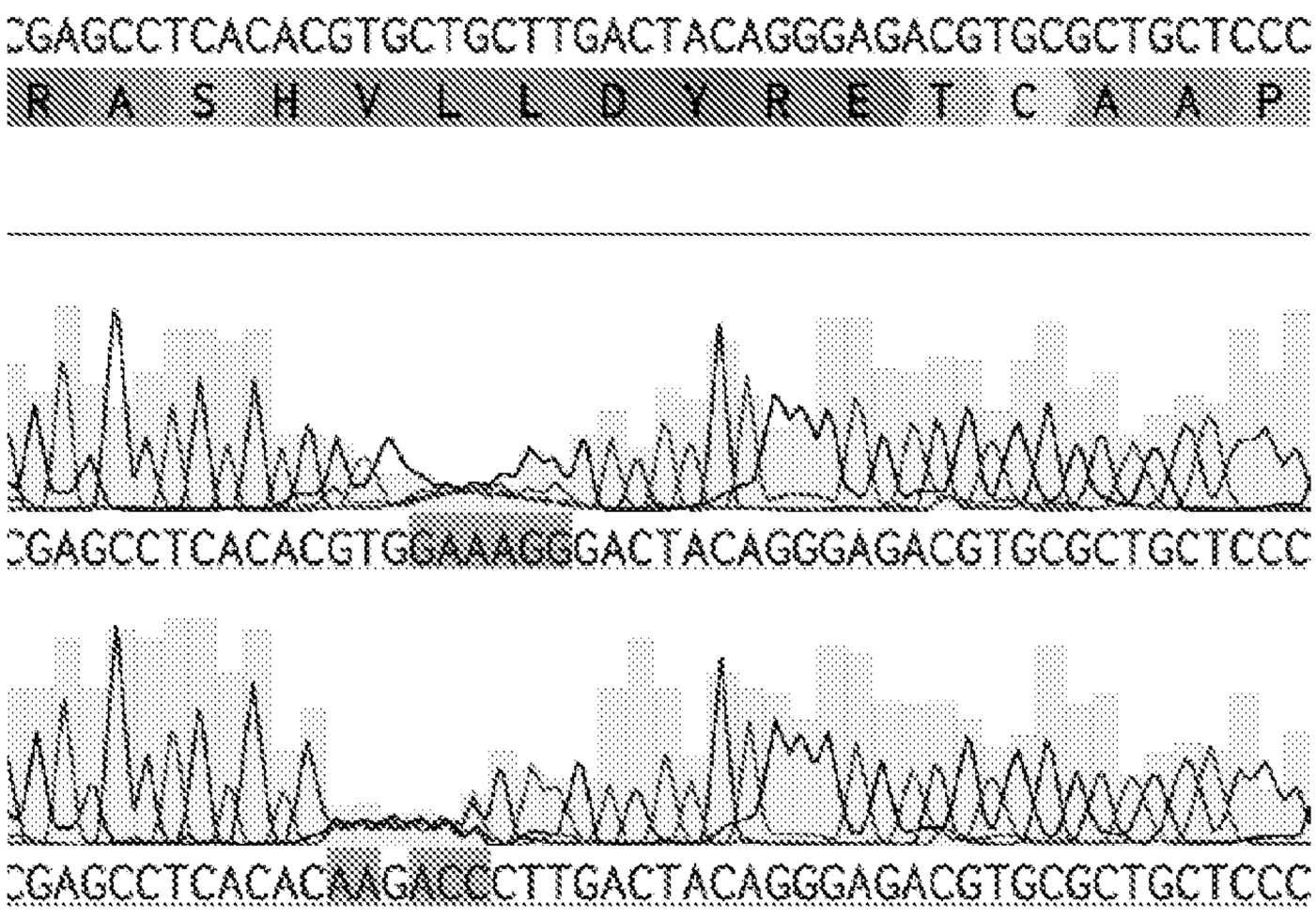

AAA
AAG
Cas13b-ADAR2 WT targeting
Cas13b-ADAR2 WT non-targeting
Cas13b-ADAR2 E488Q targeting
Cas13b-ADAR2 E488Q non-targeting
UAC
UAG
UAA
CAU
CAC
CAG
CAA
AAU
AAC
GAU
GAC
GAG
GAA
Motif identity
editing fraction
0.15
0.10
0.05
0.00
UAU

FIG. 32

A


CGAGCCTCACACGTGCTGCTTGACTACAGGGAGACGTGCGCTGCTCCC
R A S H V L L D Y R E T C A A P
CGAGCCTCACACGTGCAAACGGACTACAGGGAGACGTGCGCTGCTCCC
CGAGCCTCACACAAGACCCTTGACTACAGGGAGACGTGCGCTGCTCCC

B

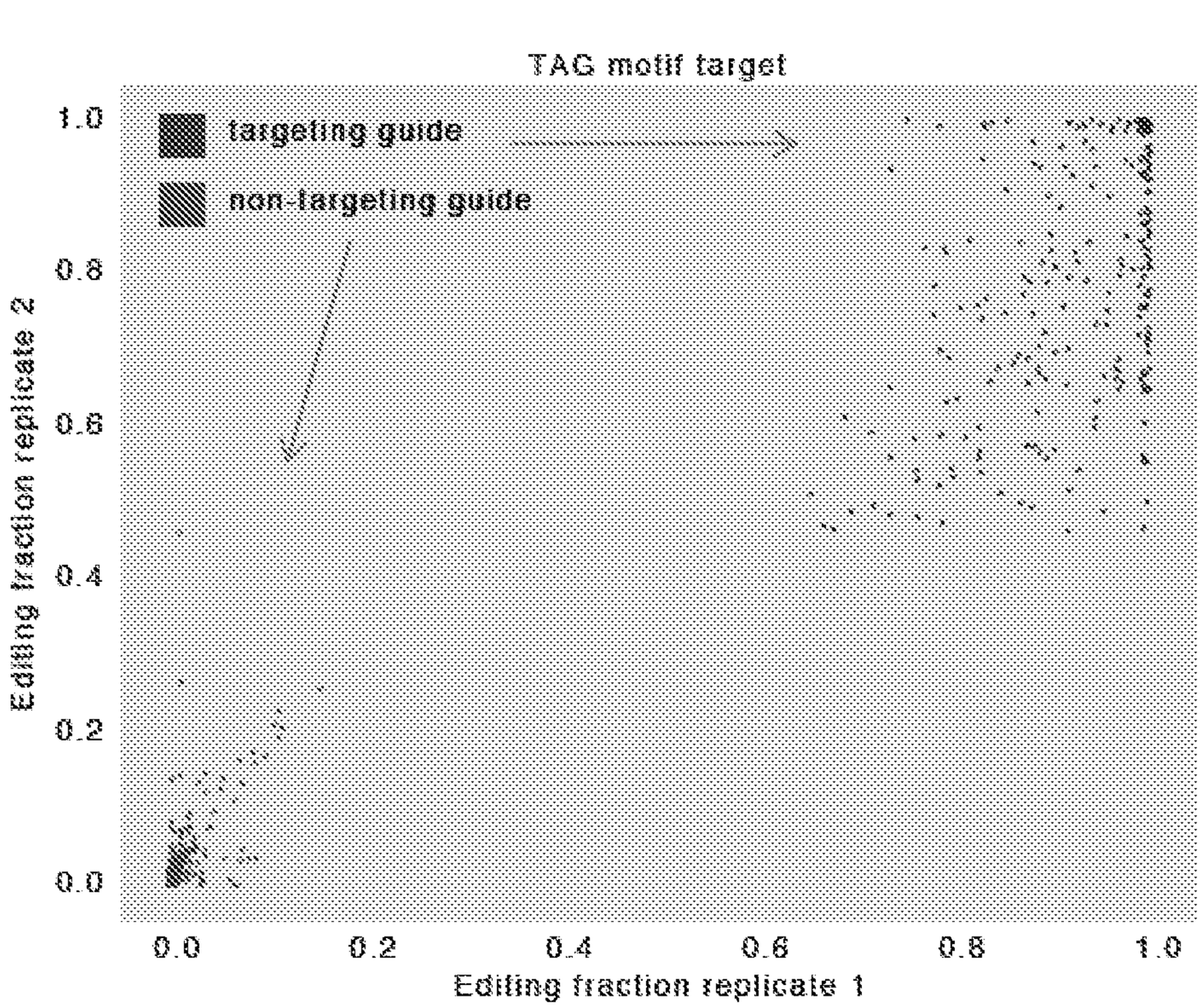
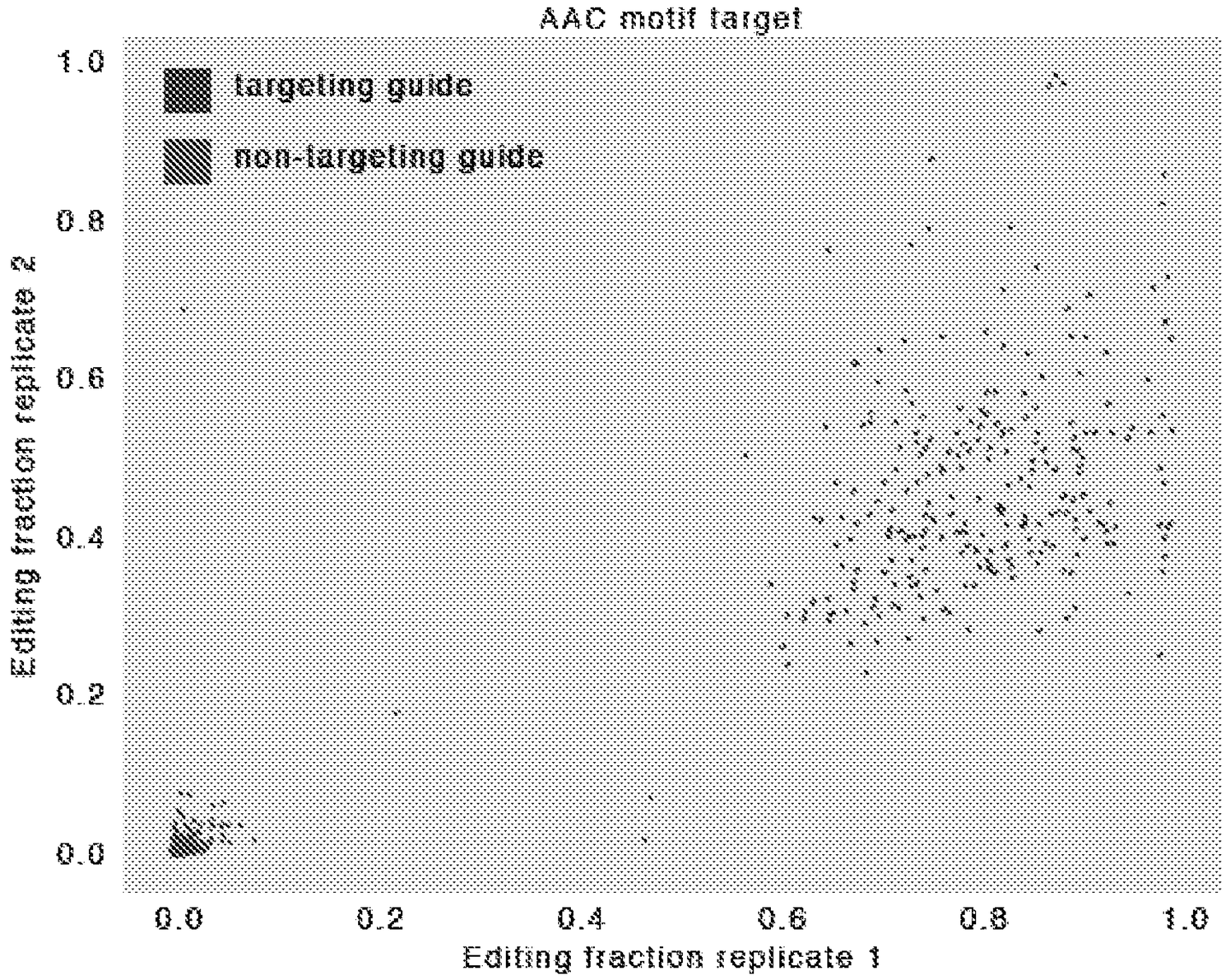

TAG motif target
targeting guide
non-targeting guide
Editing fraction replicate 2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Editing fraction replicate 1

FIG. 33

C



D

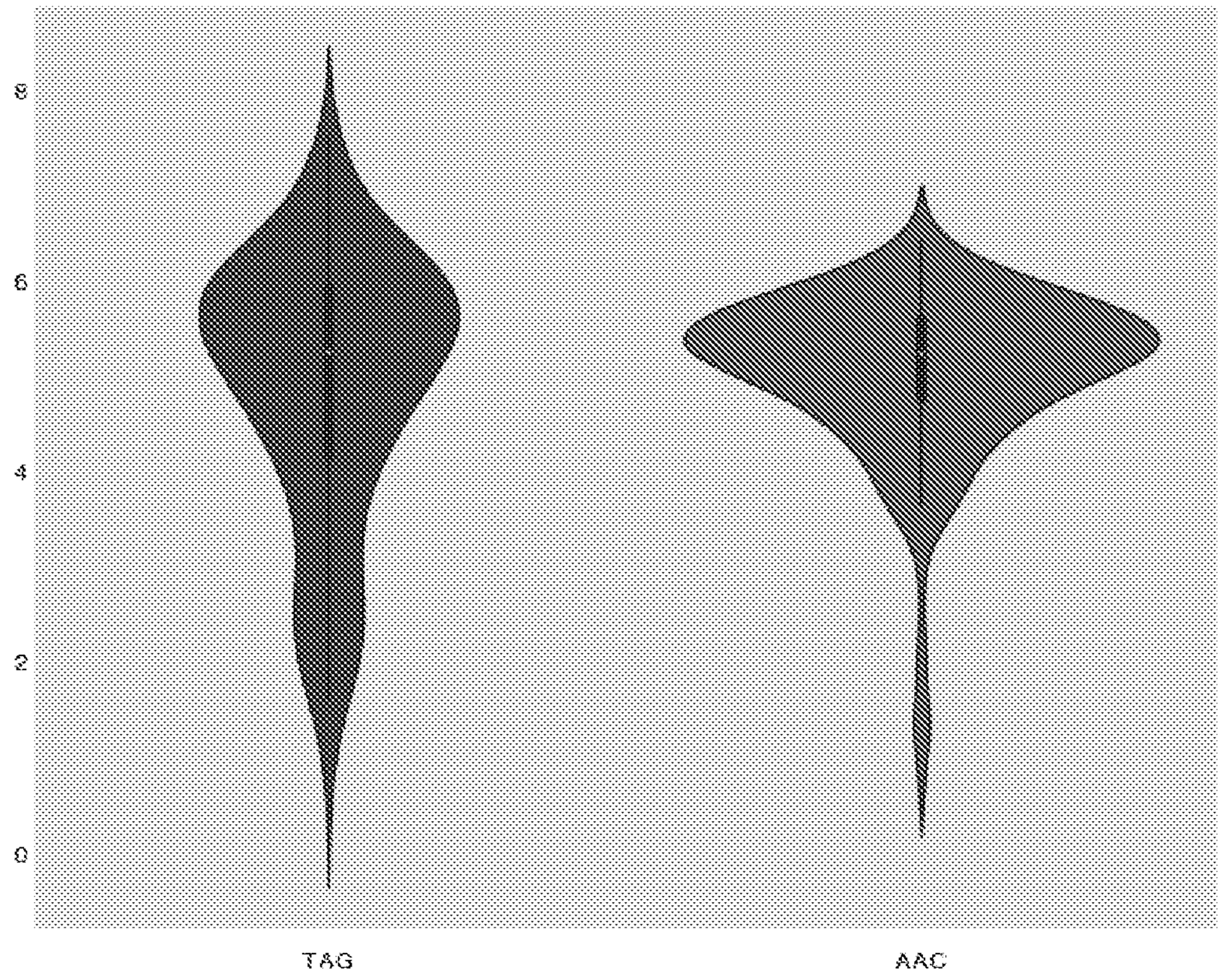


FIG. 33 (continued)

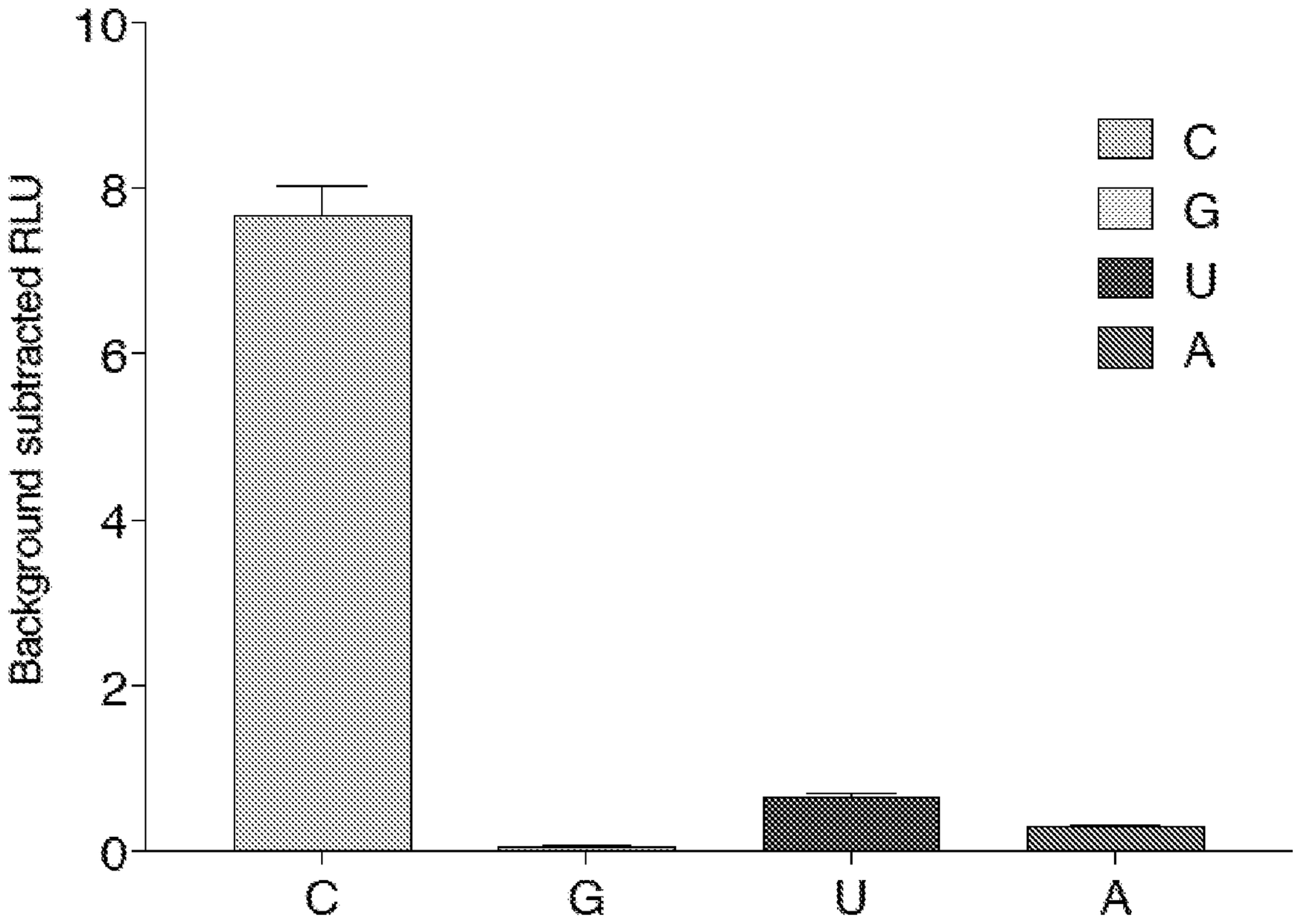

10
8
6
4
2
0
Background subtracted RLU
C
G
U
A
C
G
U
A
Base opposite target A

FIG. 34

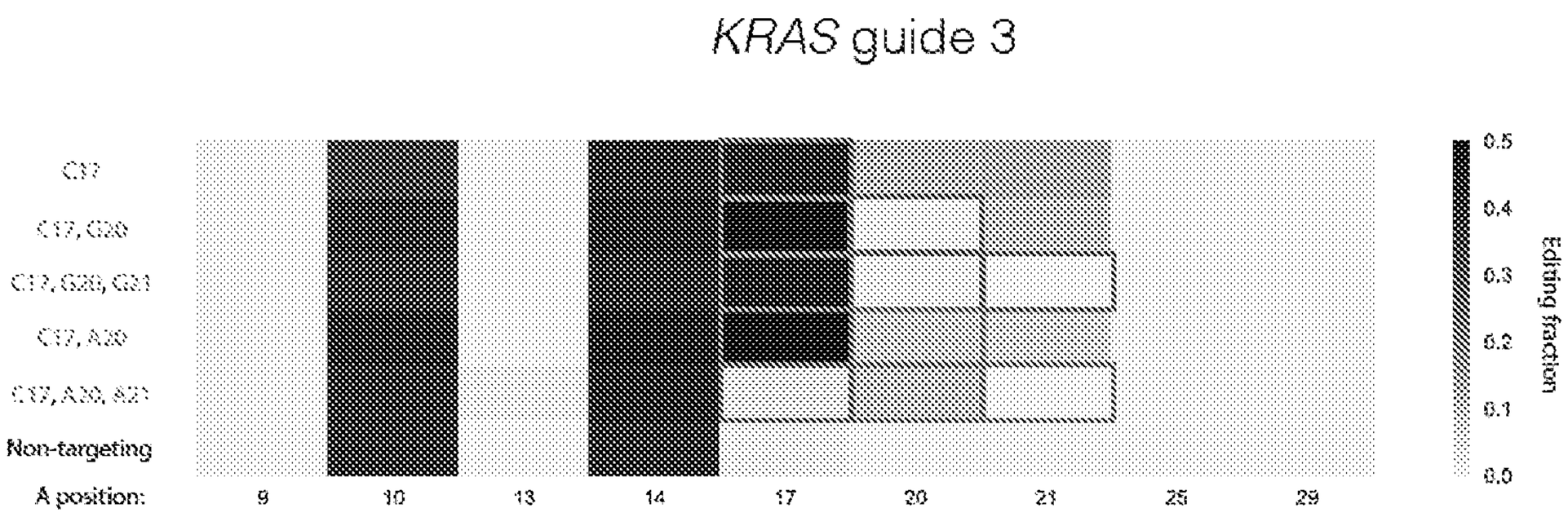
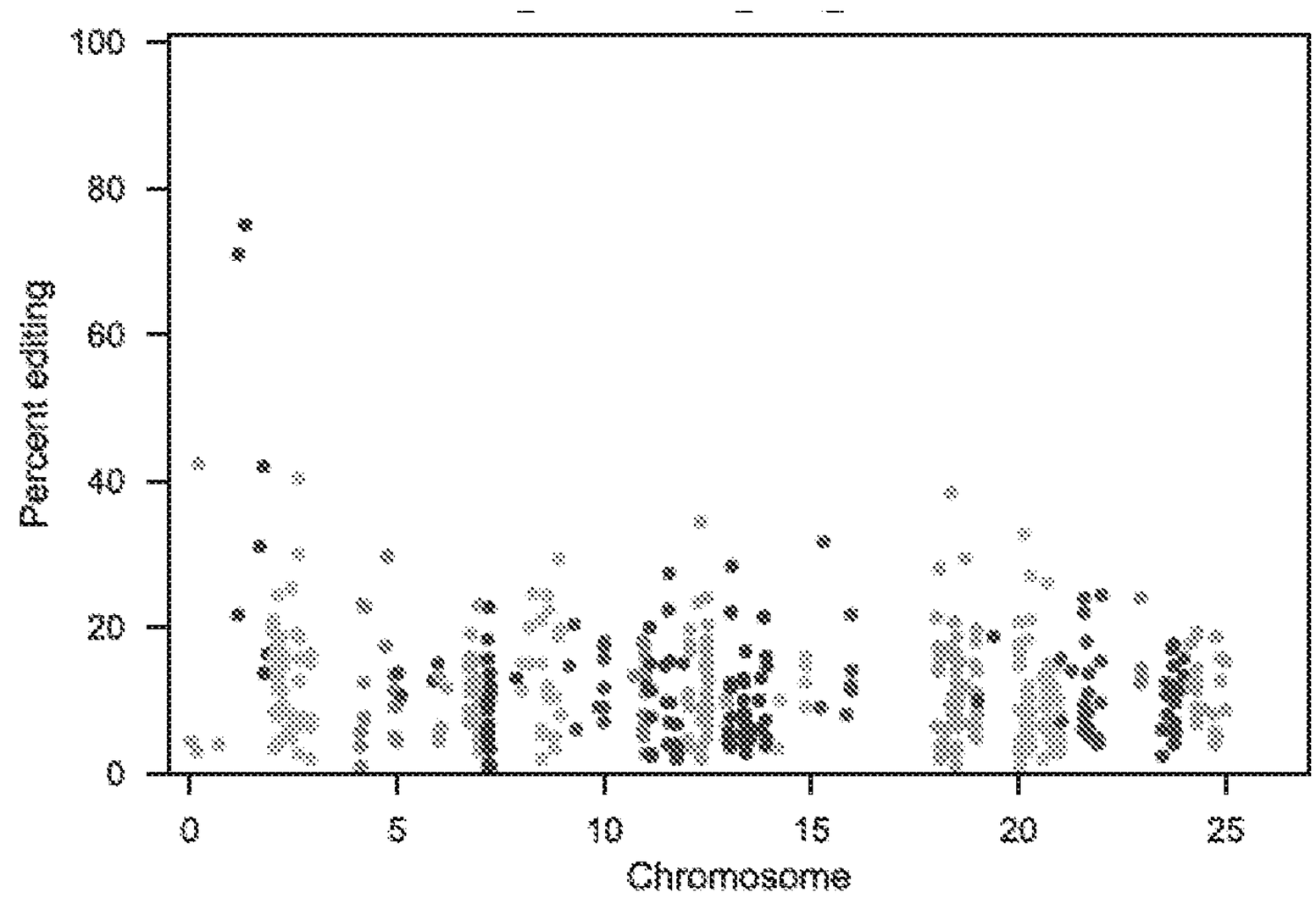

KRAS guide 3
C17
C17, G26
C17, G26, G31
C17, A26
C17, A26, A21
Non-targeting
A position:
9
10
13
14
17
20
21
25
29
0.5
0.4
0.3
0.2
0.1
0.0
Editing fraction

FIG. 35


A
100
80
60
40
20
0
Percent editing
0
5
10
15
20
25
Chromosome

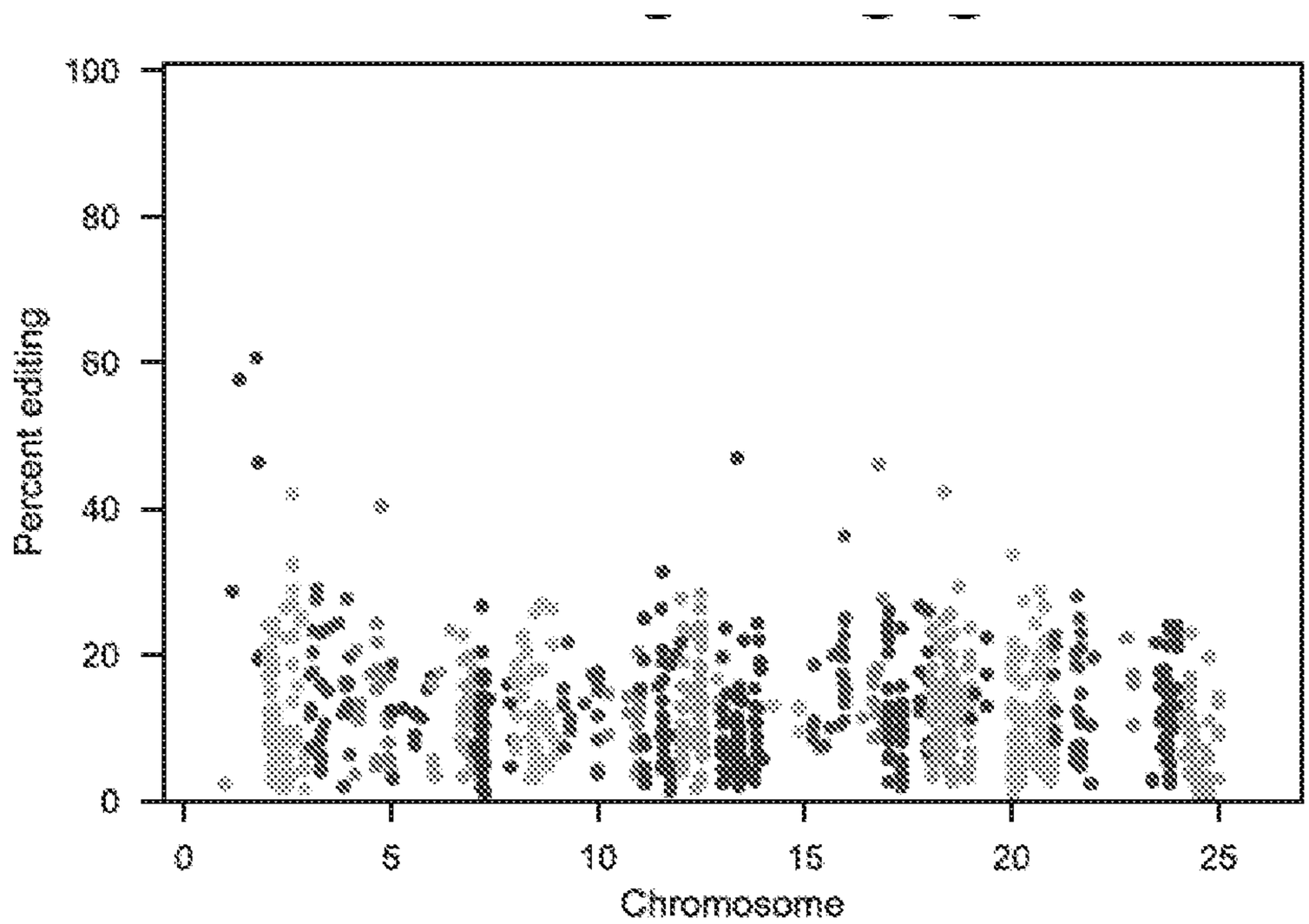
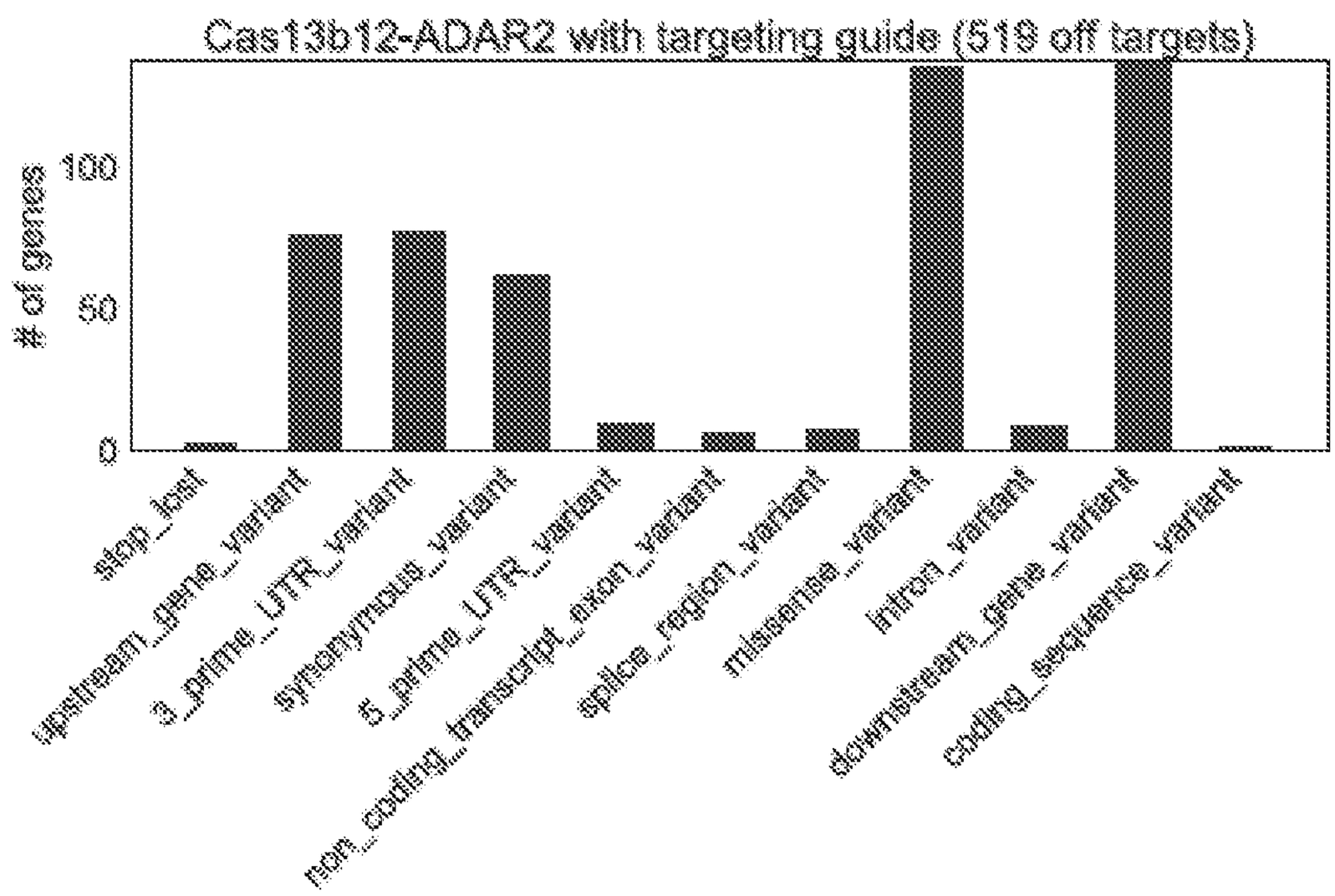

B
100
80
60
40
20
0
Percent editing
0
5
10
15
20
25
Chromosome

FIG. 36

A


Cas13b12-ADAR2 with targeting guide (519 off targets)
# of genes
100
50
0
essential_gene_variant
3_prime_UTR_variant
synonymous_variant
5_prime_UTR_variant
non_coding_transcript_exon_variant
splice_region_variant
missense_variant
intron_variant
downstream_gene_variant
coding_sequence_variant

B

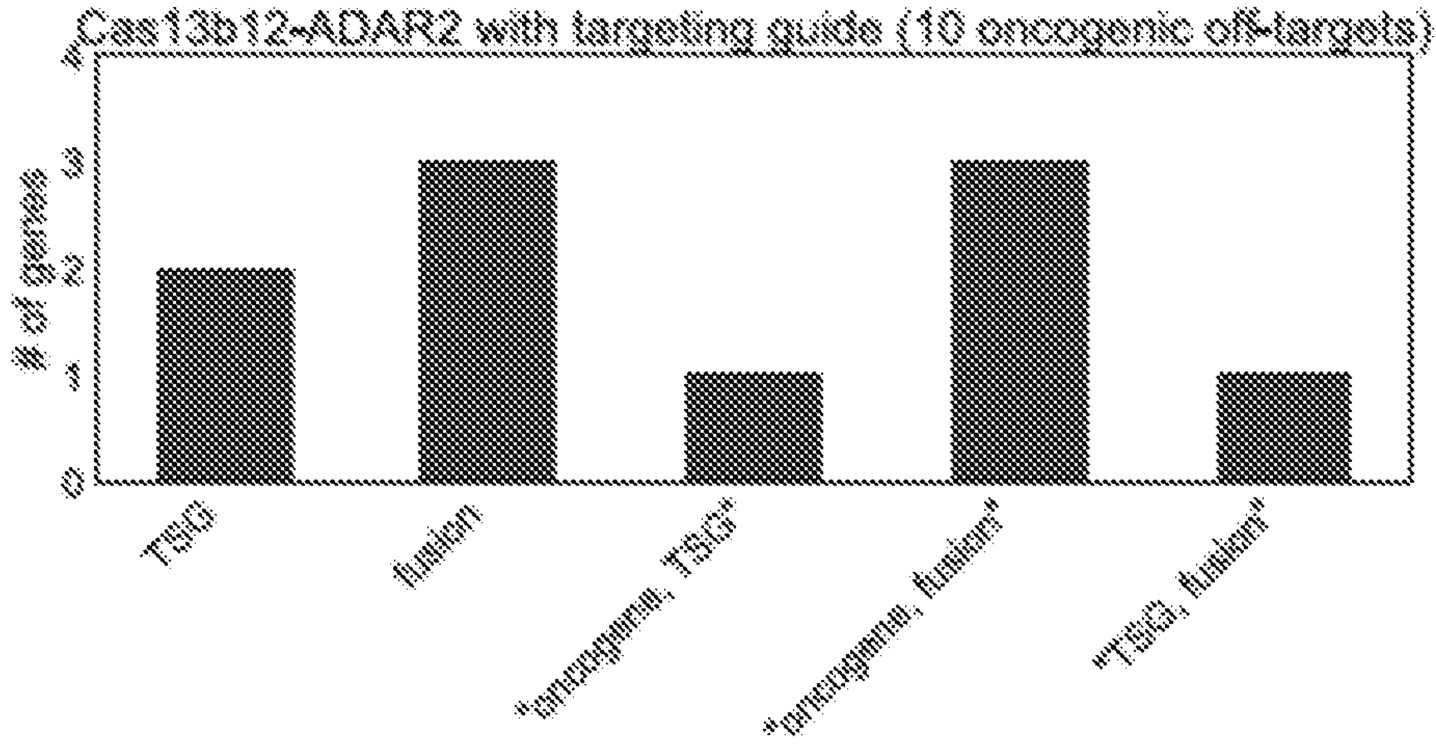
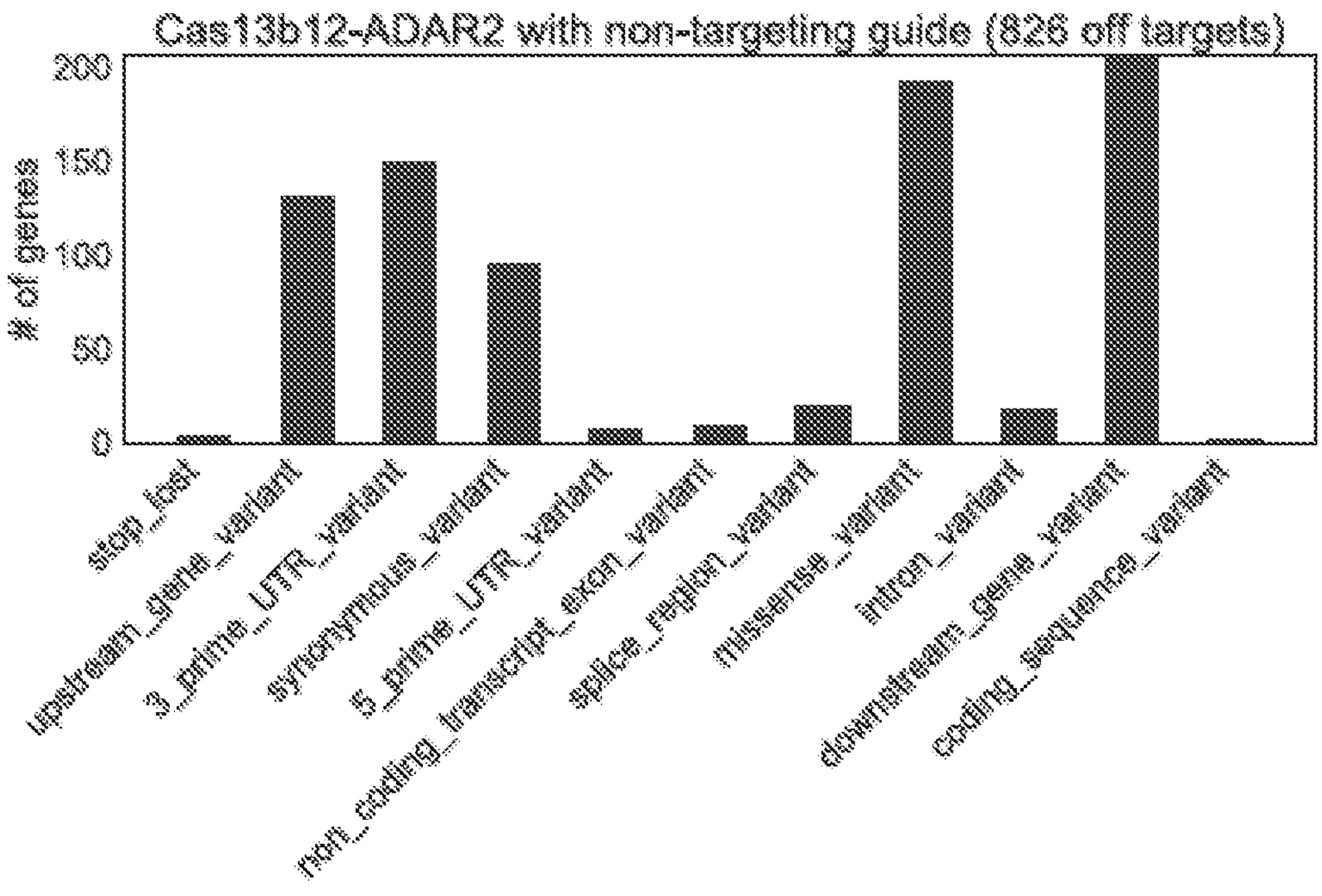

Cas13b12-ADAR2 with targeting guide (10 oncogenic off-targets)
# of guides
3
2
1
0
TSG
fusion
"oncogene, TSG"
"oncogene, fusion"
"TSG, fusion"

FIG. 37

C



D

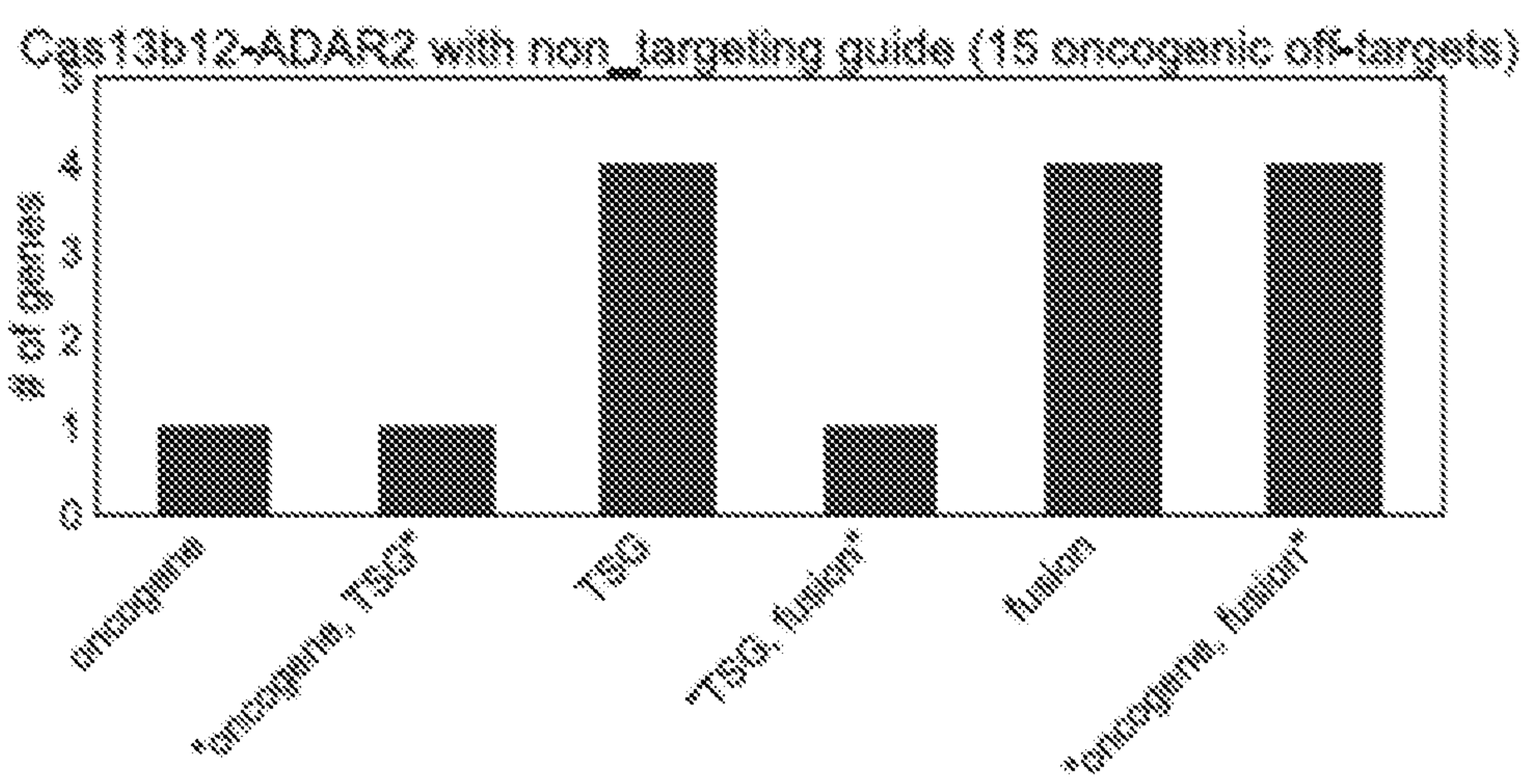


FIG. 37 (continued)

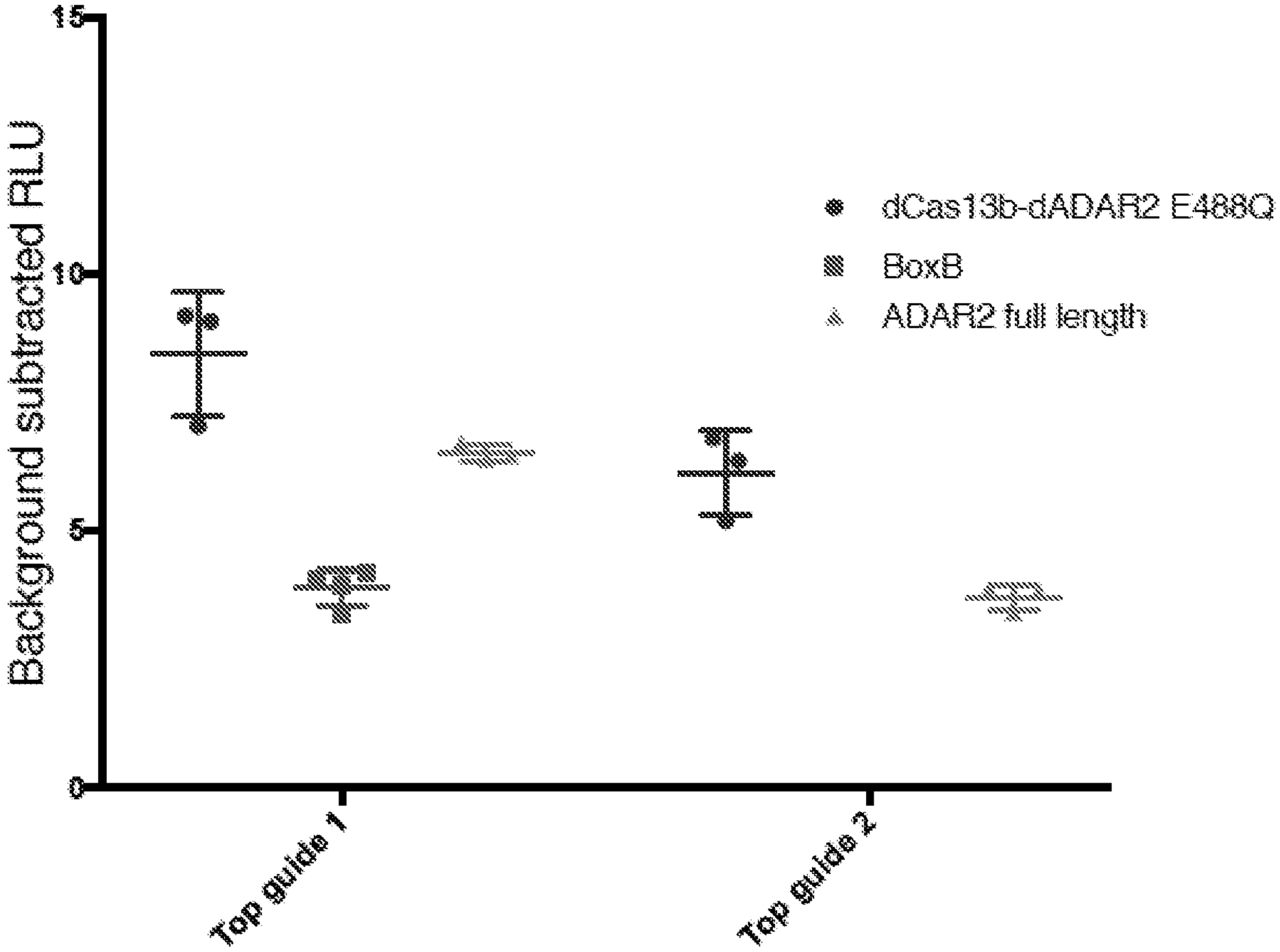
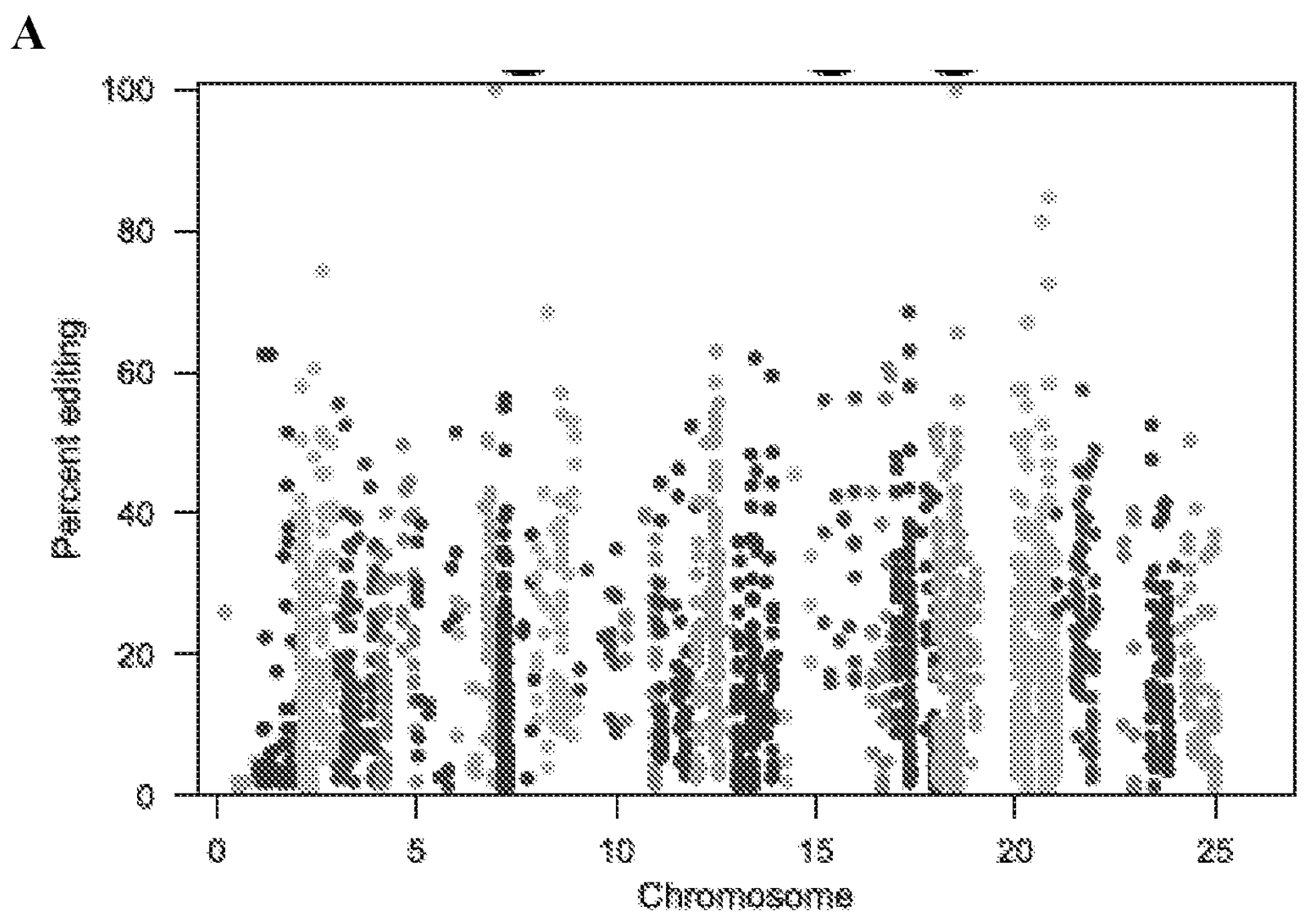
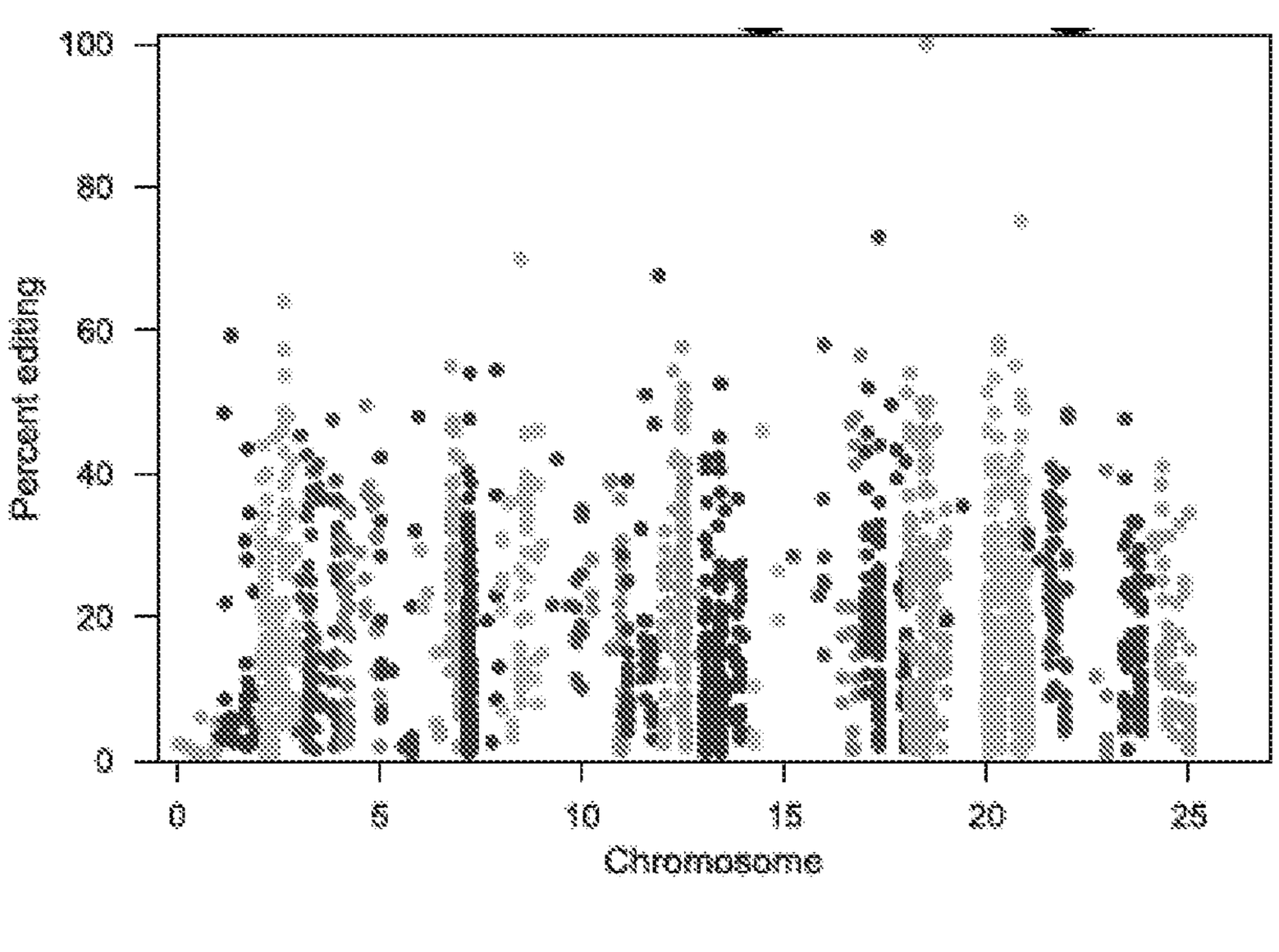

Background subtracted RLU
15
10
5
0
dCas13b-dADAR2 E488Q
BoxB
ADAR2 full length
Top guide 1
Top guide 2

FIG. 38

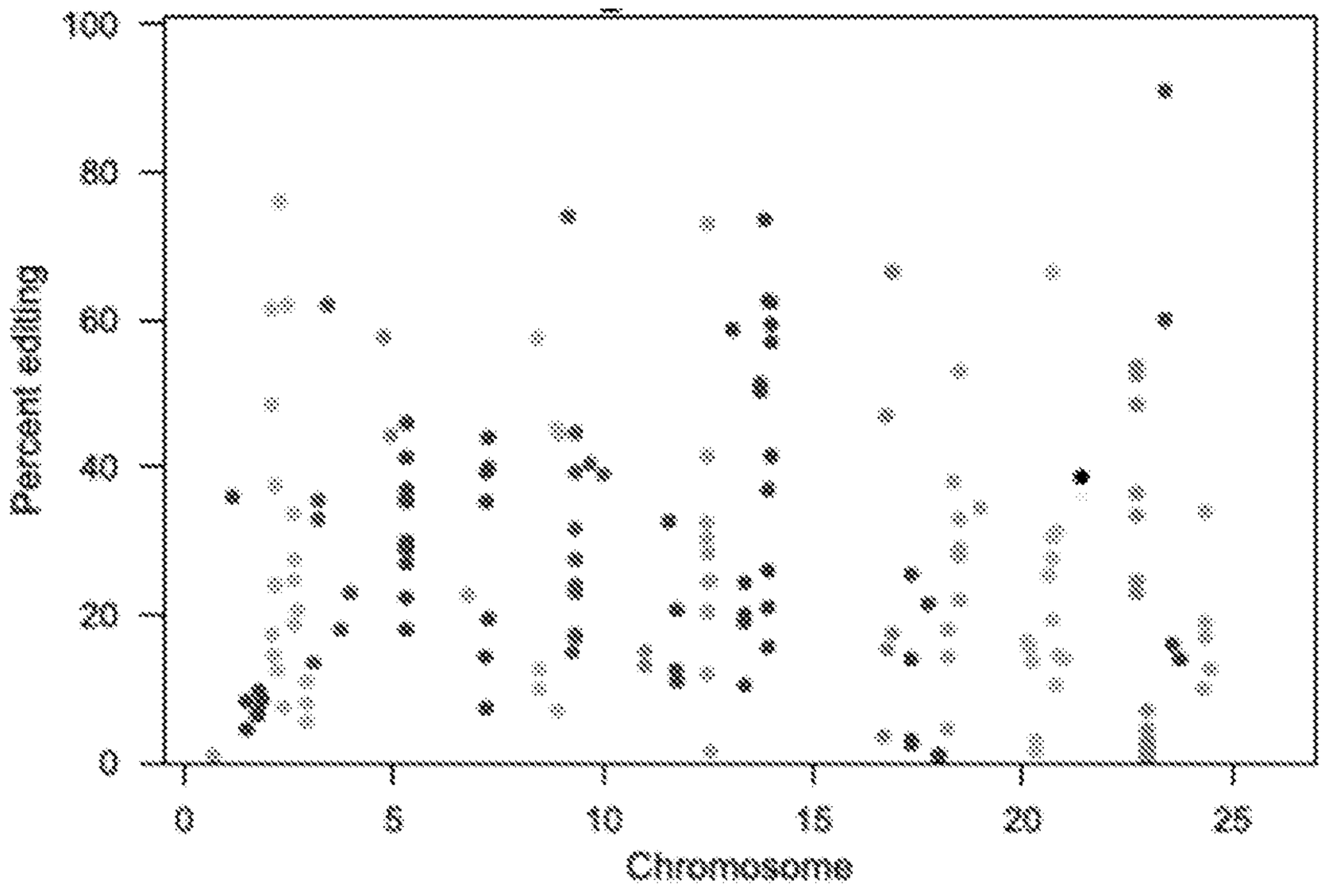

A
Percent editing
100
80
60
40
20
0
0
5
10
15
20
25
Chromosome


B
Percent editing
100
80
60
40
20
0
0
5
10
15
20
25
Chromosome

39

FIG.

C



D

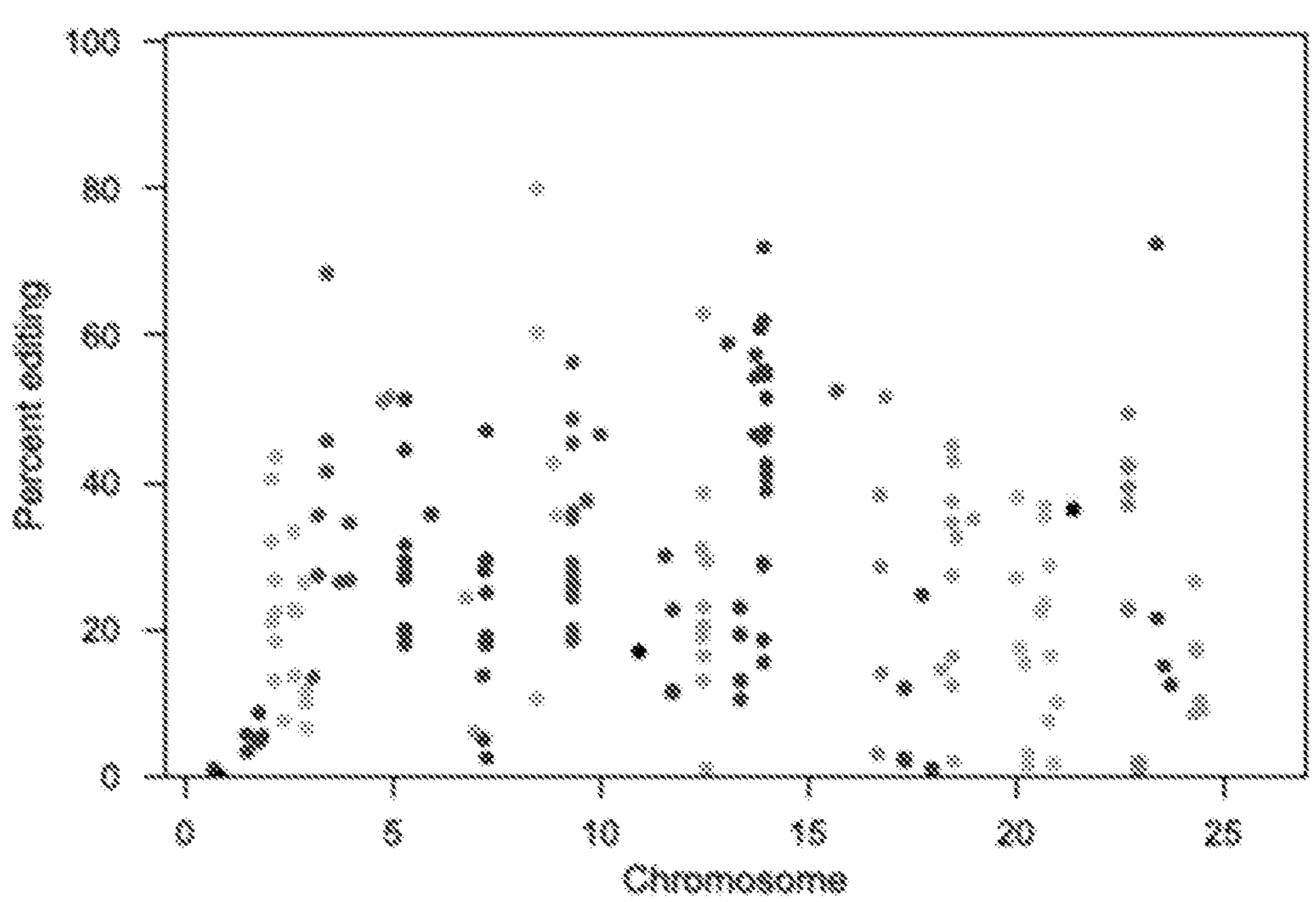


FIG. 39 (continued)

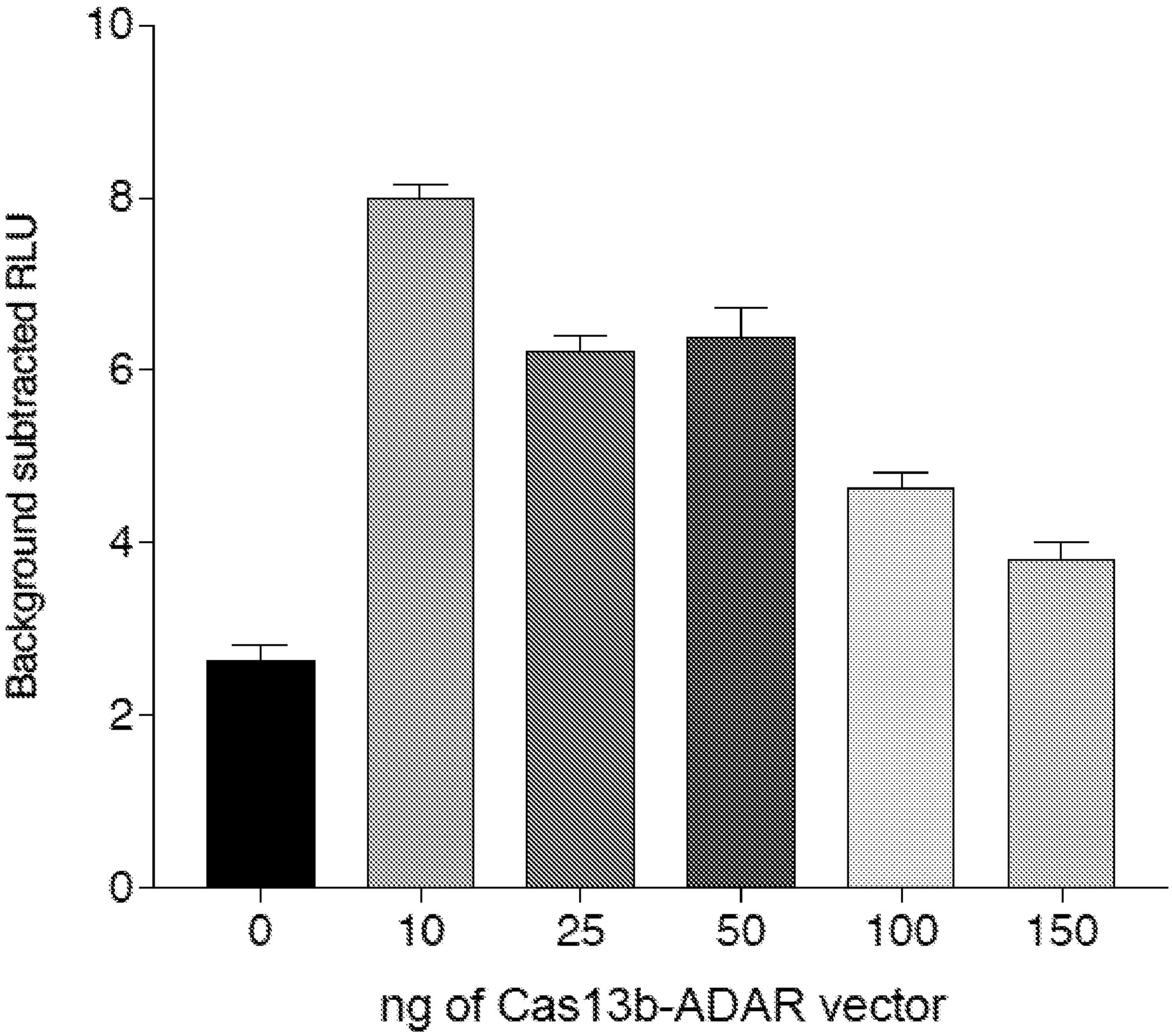
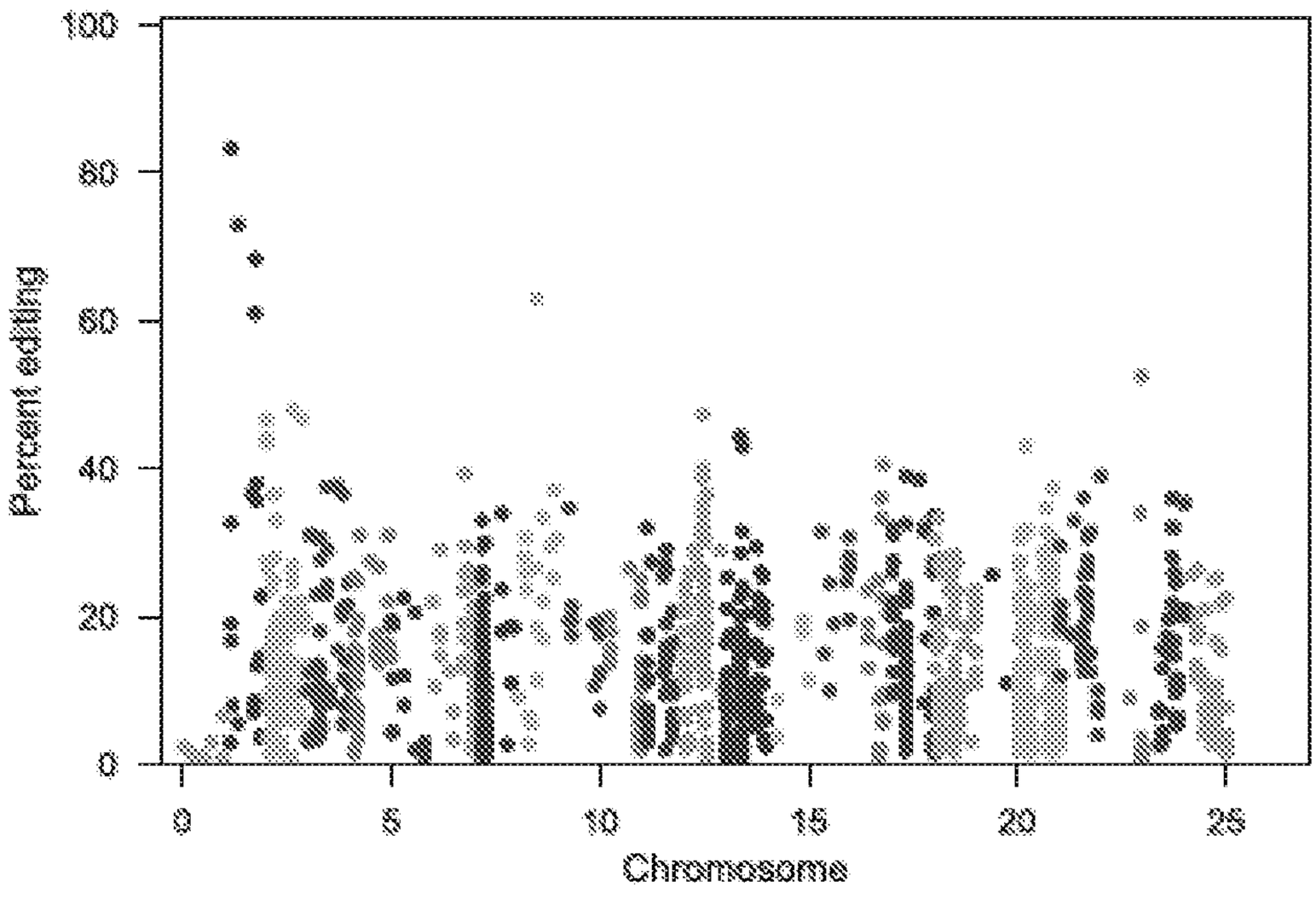

Background subtracted RLU
0
2
4
6
8
10
0
10
25
50
100
150
ng of Cas13b-ADAR vector

FIG. 40

A


100
80
60
40
20
0
Percent editing
0
5
10
15
20
25
Chromosome

B

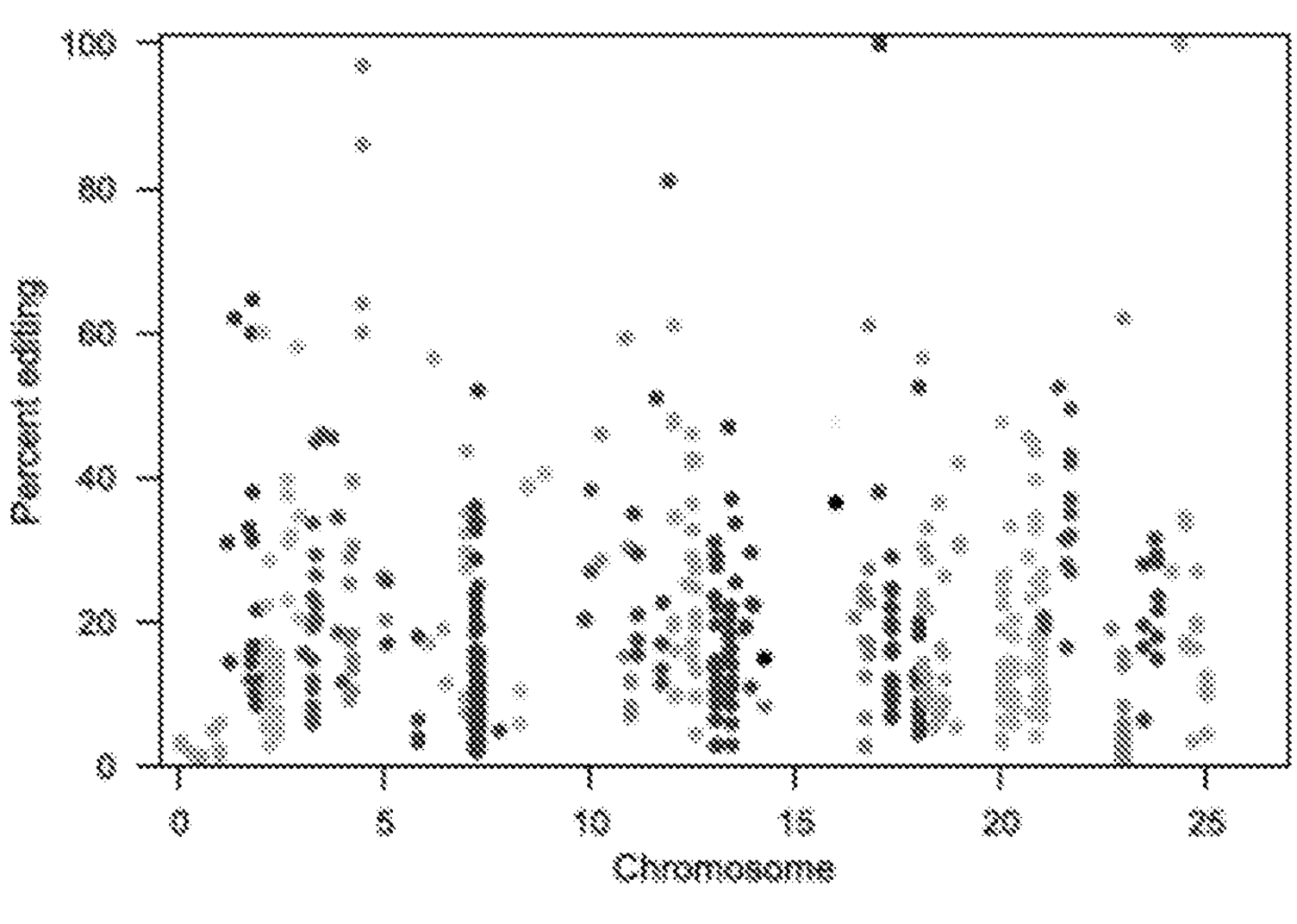
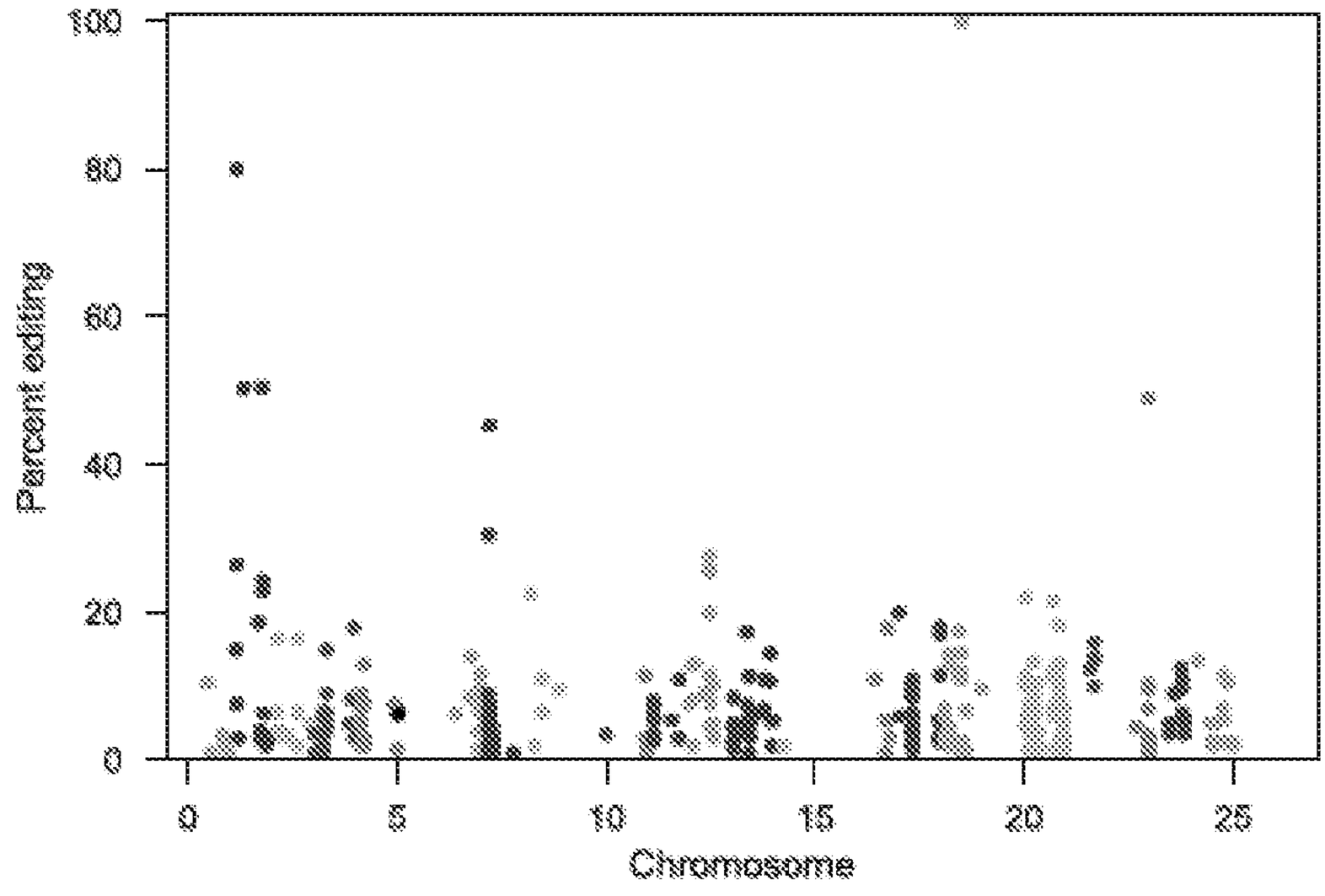

100
80
60
40
20
0
Percent editing
0
5
10
15
20
25
Chromosome

FIG. 41

C



D

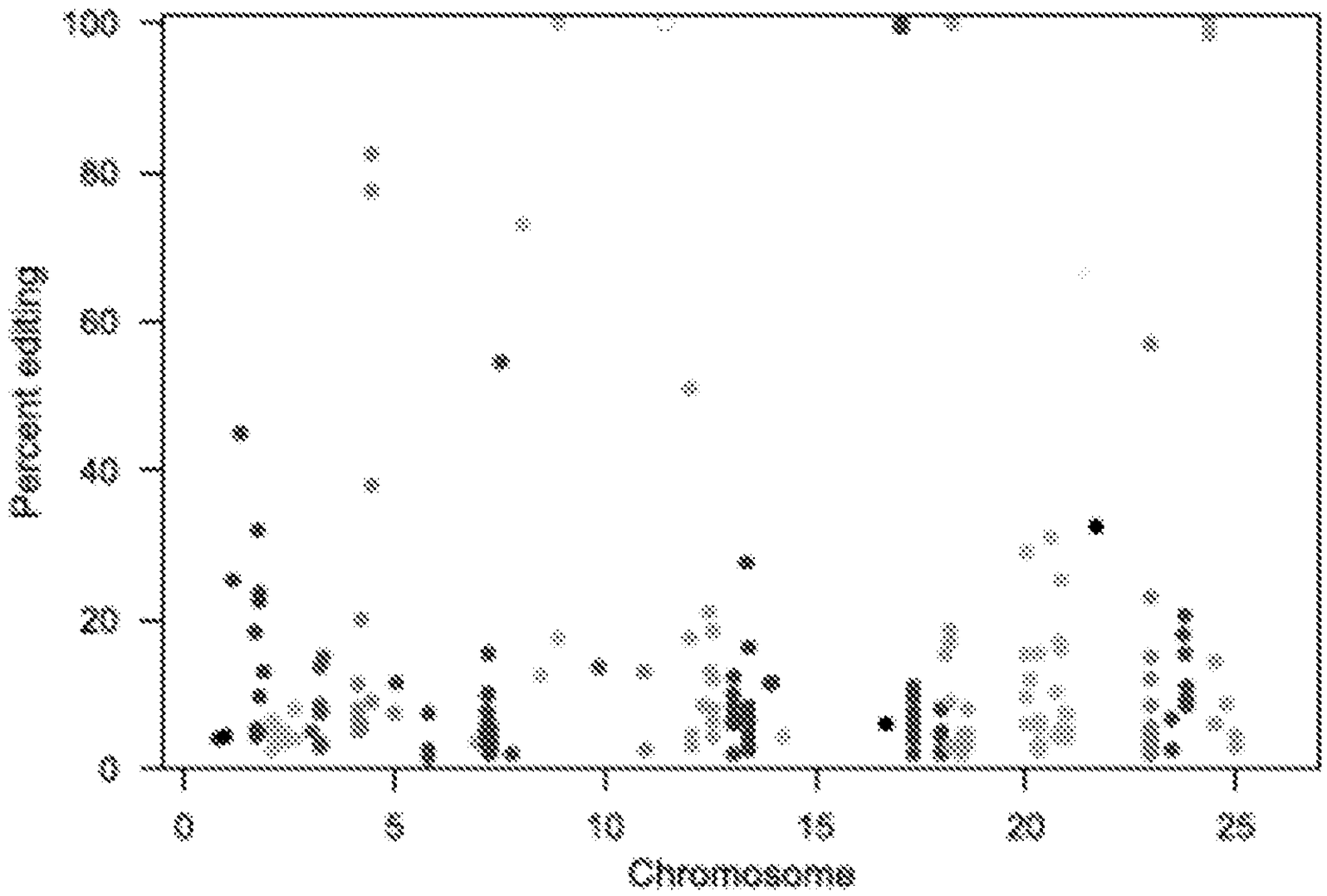
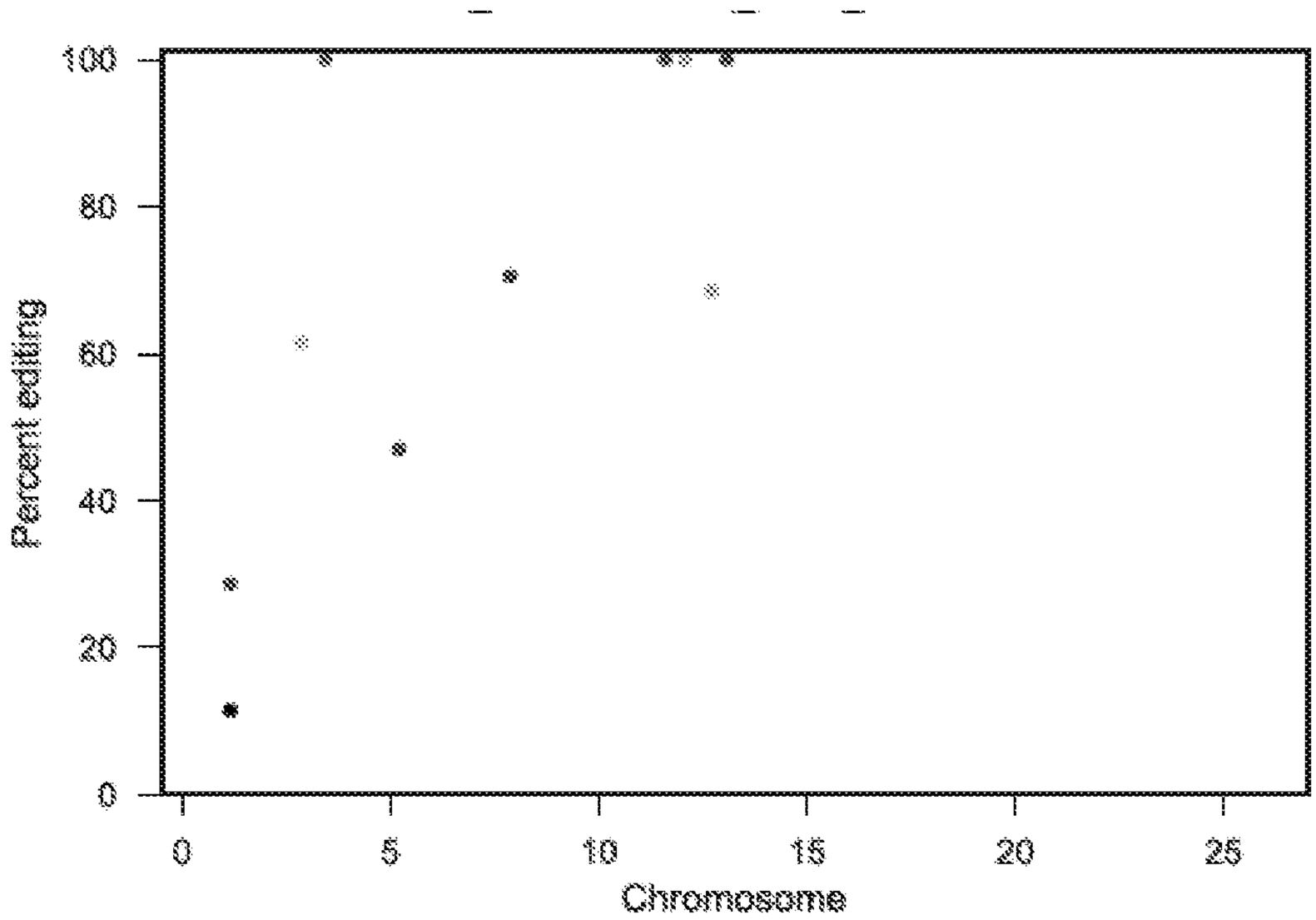


FIG. 41 (continued)

A


Percent editing
100
80
60
40
20
0
0
5
10
15
20
25
Chromosome

B

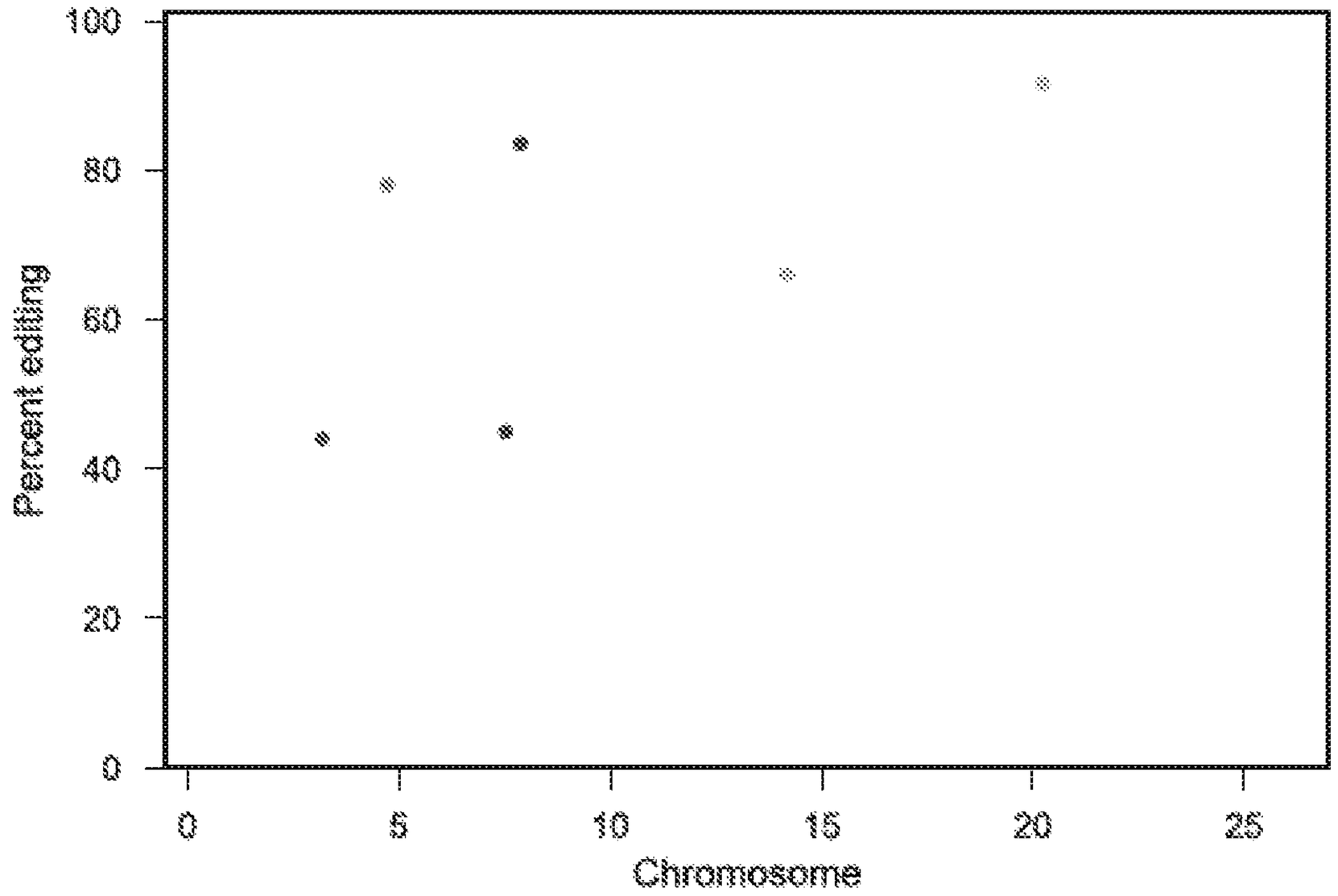
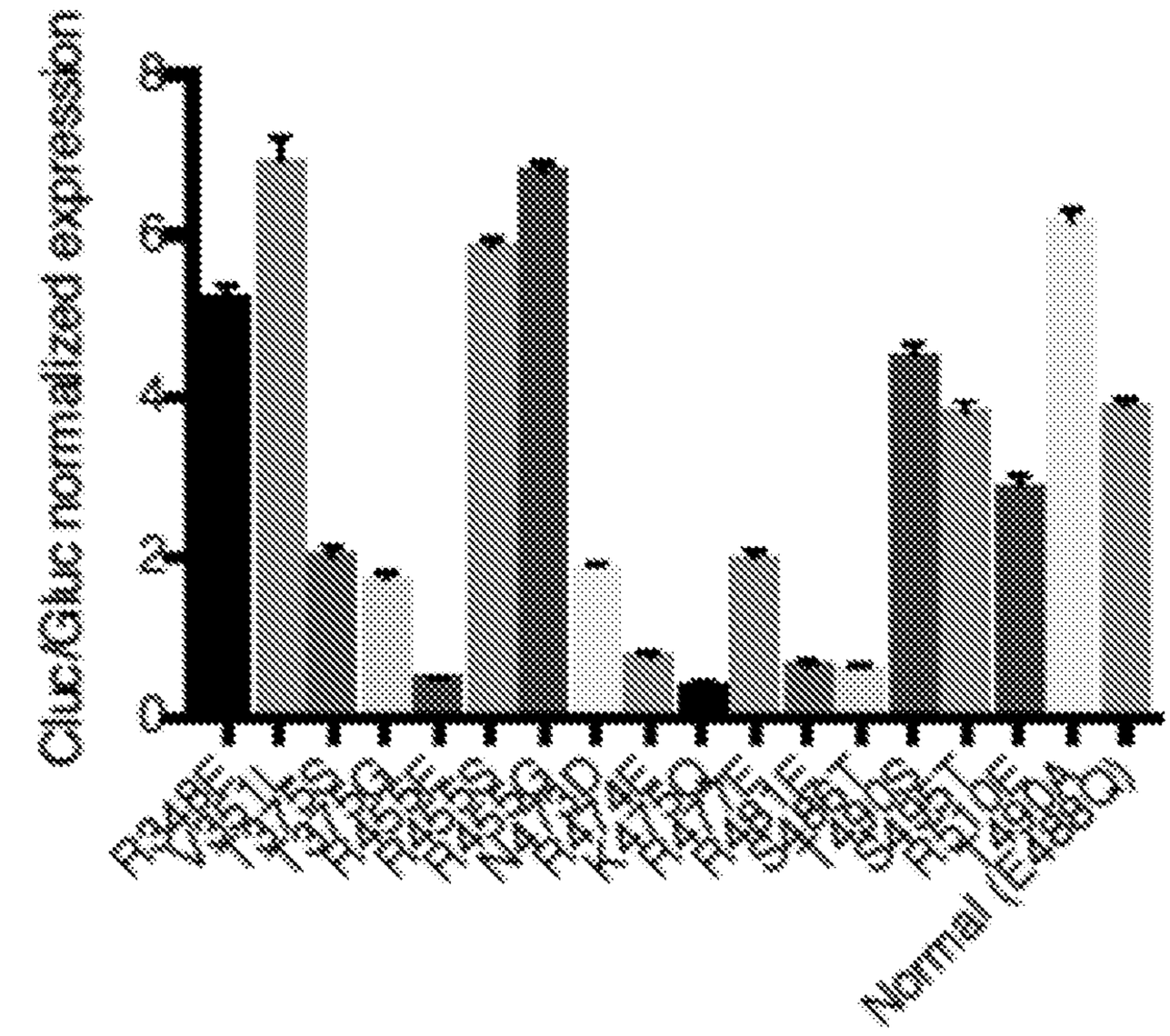

Percent editing
100
80
60
40
20
0
0
5
10
15
20
25
Chromosome

FIG. 42

A


Cluc/Gluc normalized expression
0
2
4
6
R-848
VBL1s6
T1s6
T1s6+R
R1s6
R1s6+R
NAM6
NAM6+R
K4T6
K4T6+R
R4B1E
R4B1E+R
CART
CART+R
CAST
CAST+R
R5T6
R5T6+R
4004
4004+R
Normal (EAE80)

B

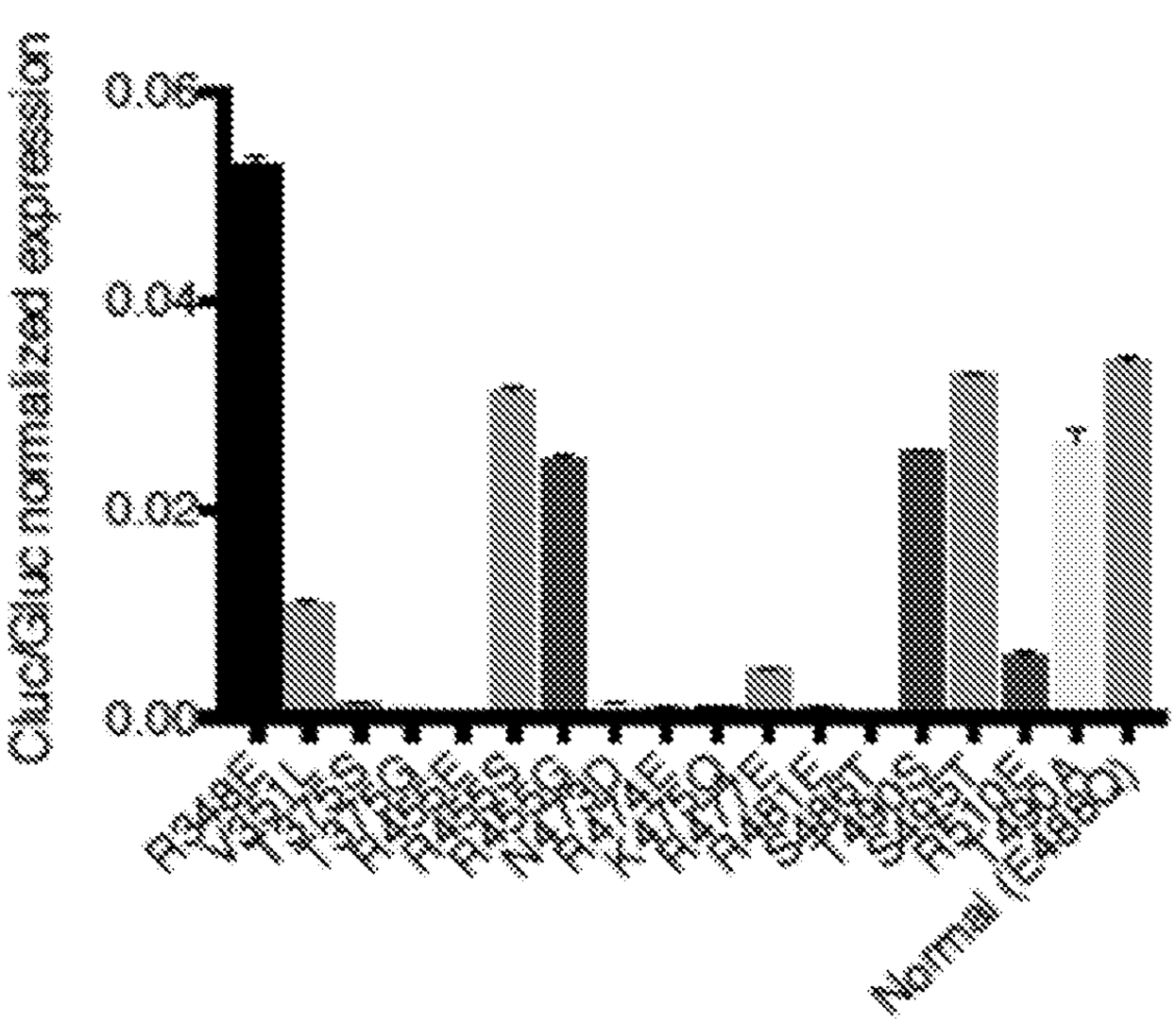
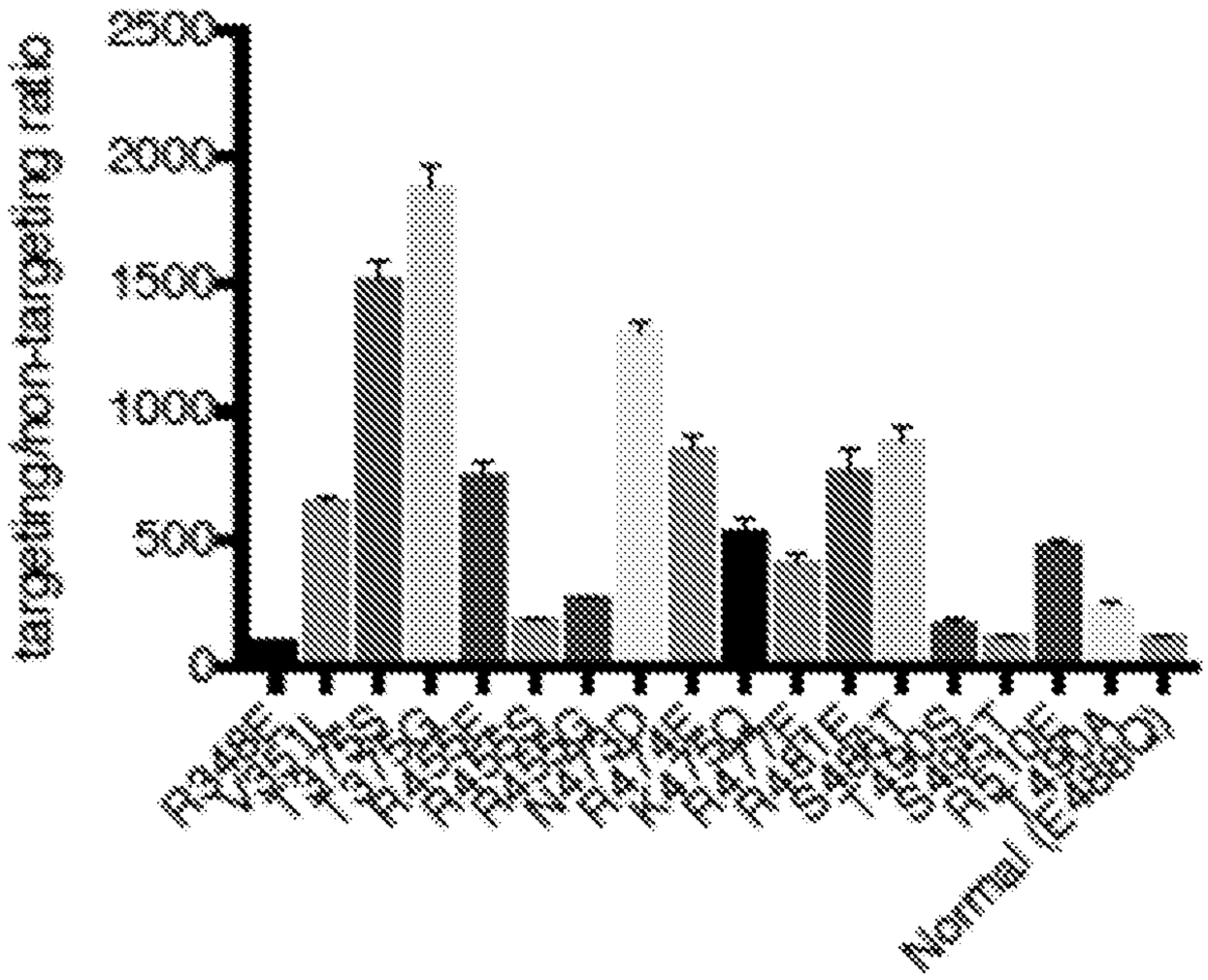

Cluc/Gluc normalized expression
0.00
0.02
0.04
0.06
R-848
VBL1s6
T1s6
T1s6+R
R1s6
R1s6+R
NAM6
NAM6+R
K4T6
K4T6+R
R4B1E
R4B1E+R
CART
CART+R
CAST
CAST+R
R5T6
R5T6+R
4004
4004+R
Normal (EAE80)

FIG. 43

C



D

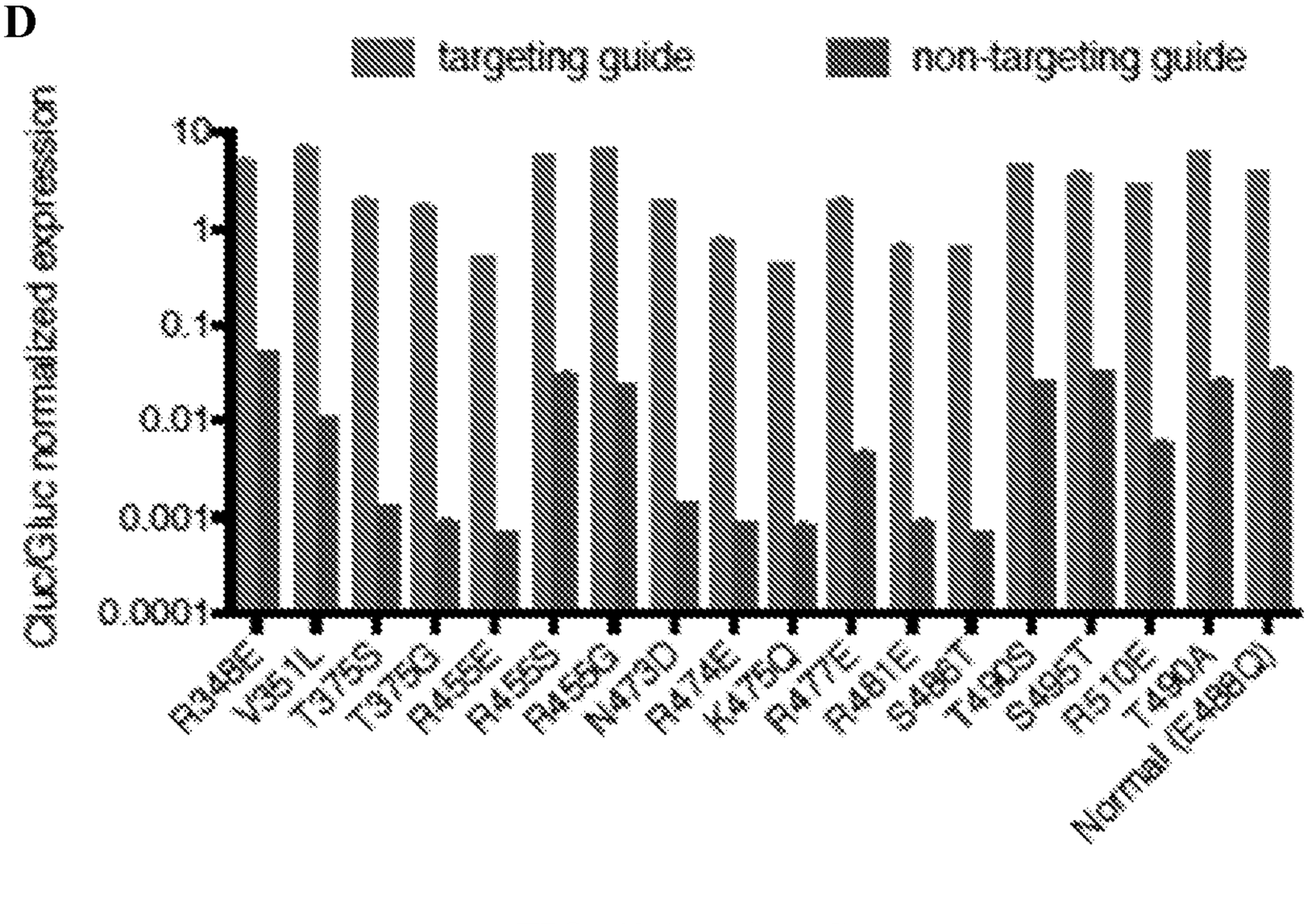


FIG. 43 (continued)

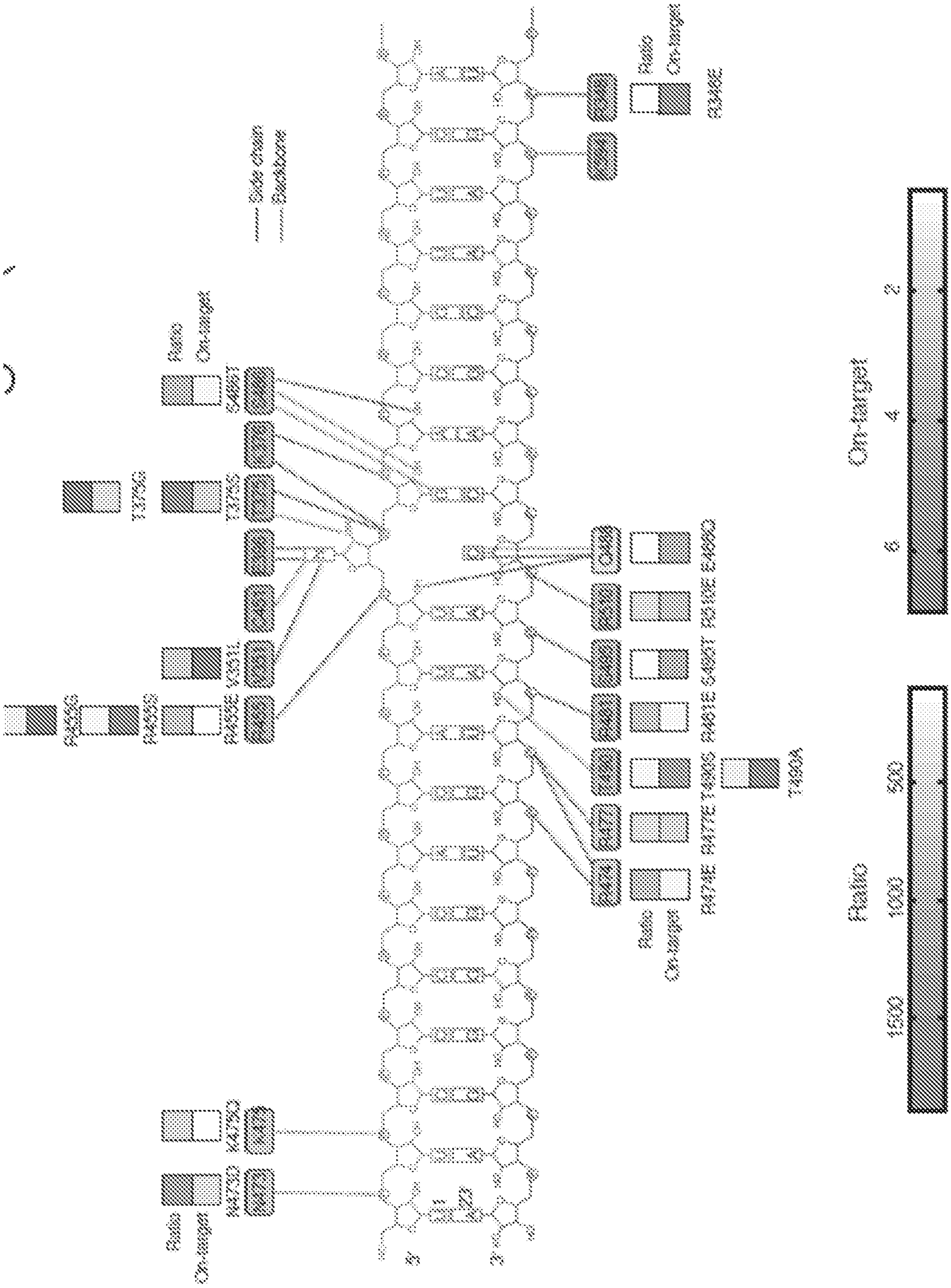

Side chain
Backbone
Ratio
On-target
Ratio
1500
1000
500
On-target
6
4
2

FIG. 44

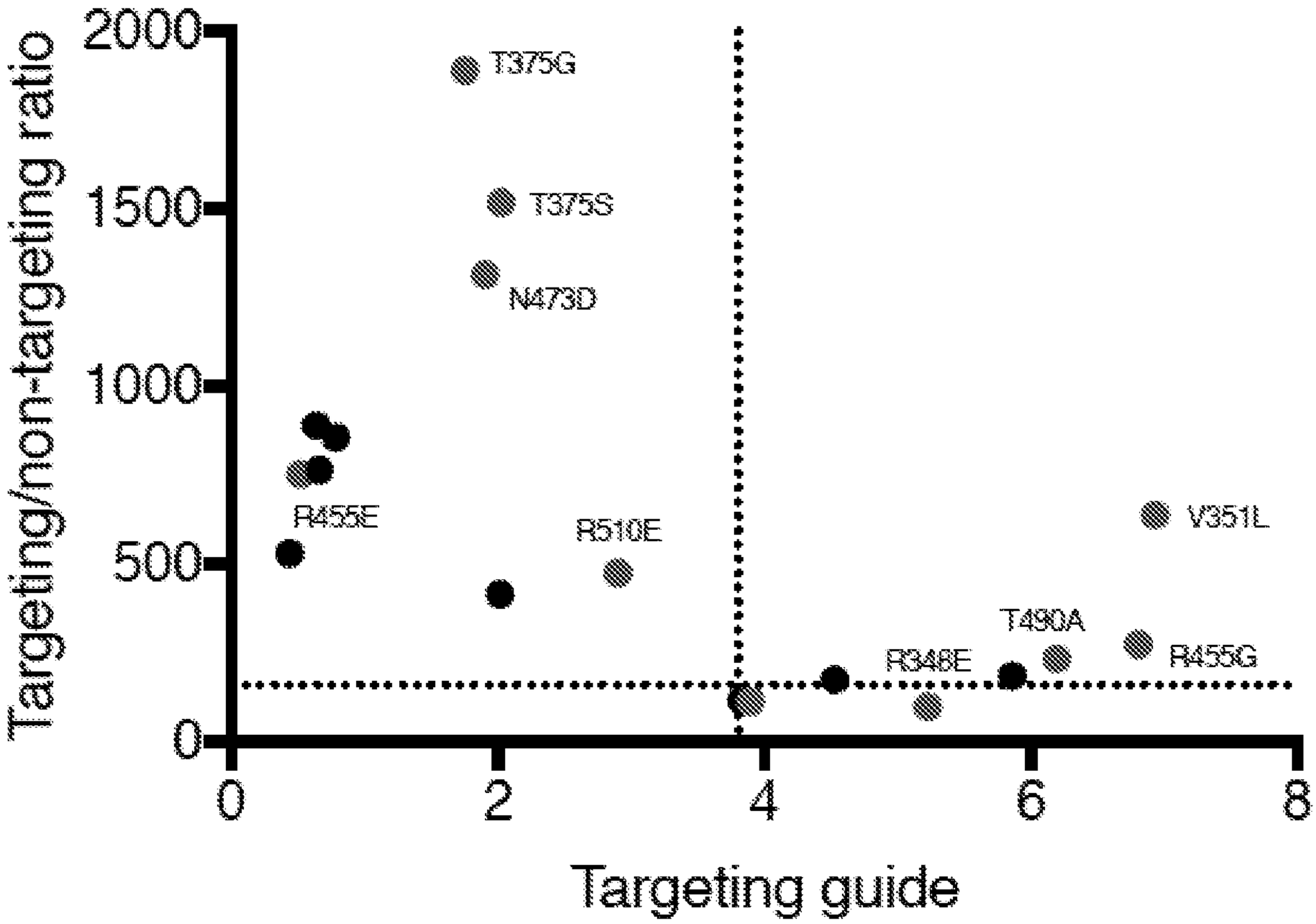

Targeting/non-targeting ratio
2000
1500
1000
500
0
0
2
4
6
8
Targeting guide
T375G
T375S
N473D
R455E
R510E
V351L
T490A
R346E
R455G

FIG. 45

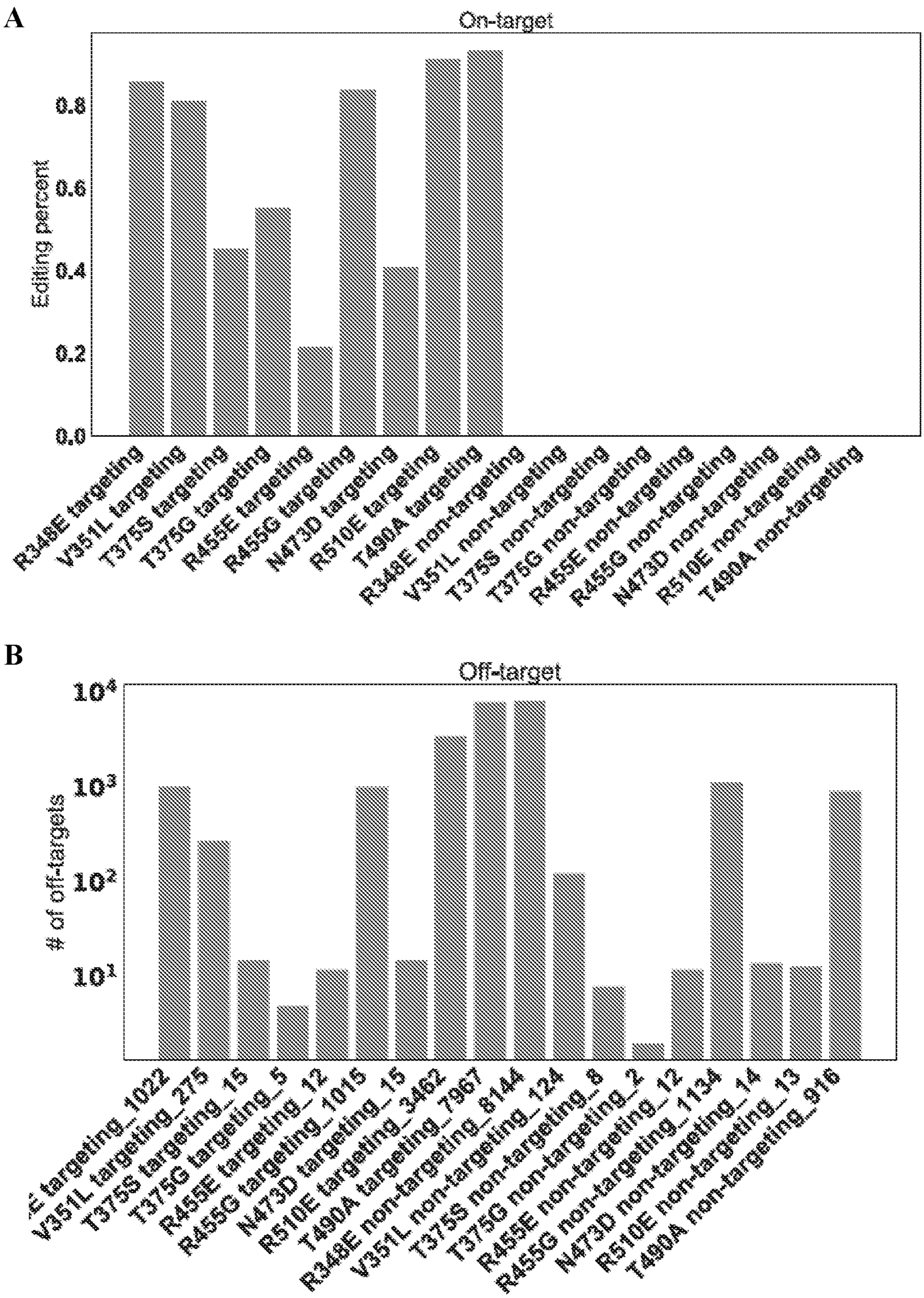

A
On-target
Editing percent
0.0
0.2
0.4
0.6
0.8
R348E targeting
V351L targeting
T375S targeting
T375G targeting
R455E targeting
R455G targeting
N473D targeting
R510E targeting
T490A targeting
R348E non-targeting
V351L non-targeting
T375S non-targeting
T375G non-targeting
R455E non-targeting
R455G non-targeting
N473D non-targeting
R510E non-targeting
T490A non-targeting
B
Off-target
# of off-targets
10^1
10^2
10^3
10^4
R348E targeting_1022
V351L targeting_275
T375S targeting_15
T375G targeting_5
R455E targeting_12
R455G targeting_1015
N473D targeting_15
R510E targeting_3462
T490A targeting_7967
R348E non-targeting_8144
V351L non-targeting_124
T375S non-targeting_8
T375G non-targeting_2
R455E non-targeting_12
R455G non-targeting_1134
N473D non-targeting_14
R510E non-targeting_13
T490A non-targeting_916

FIG. 46

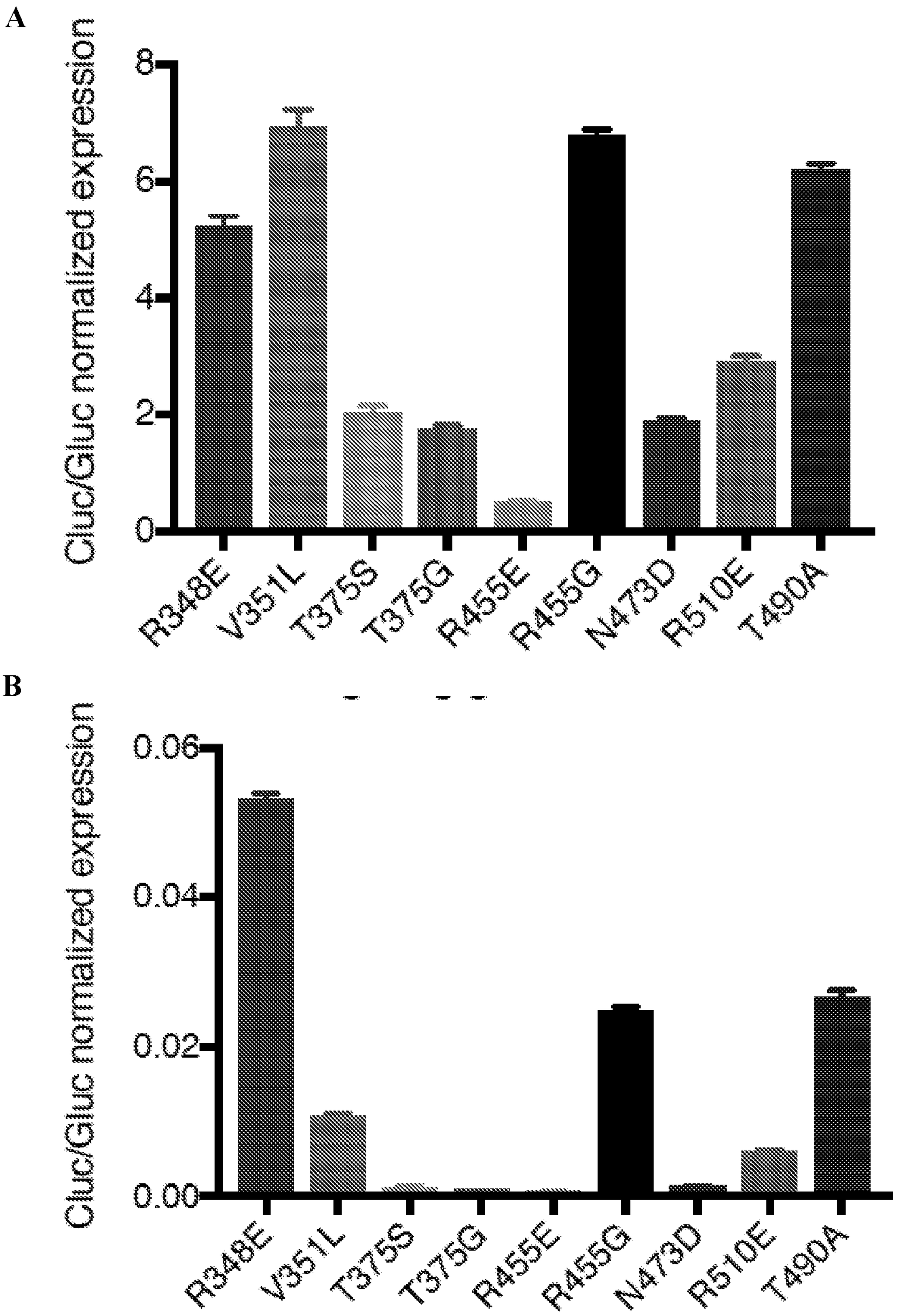
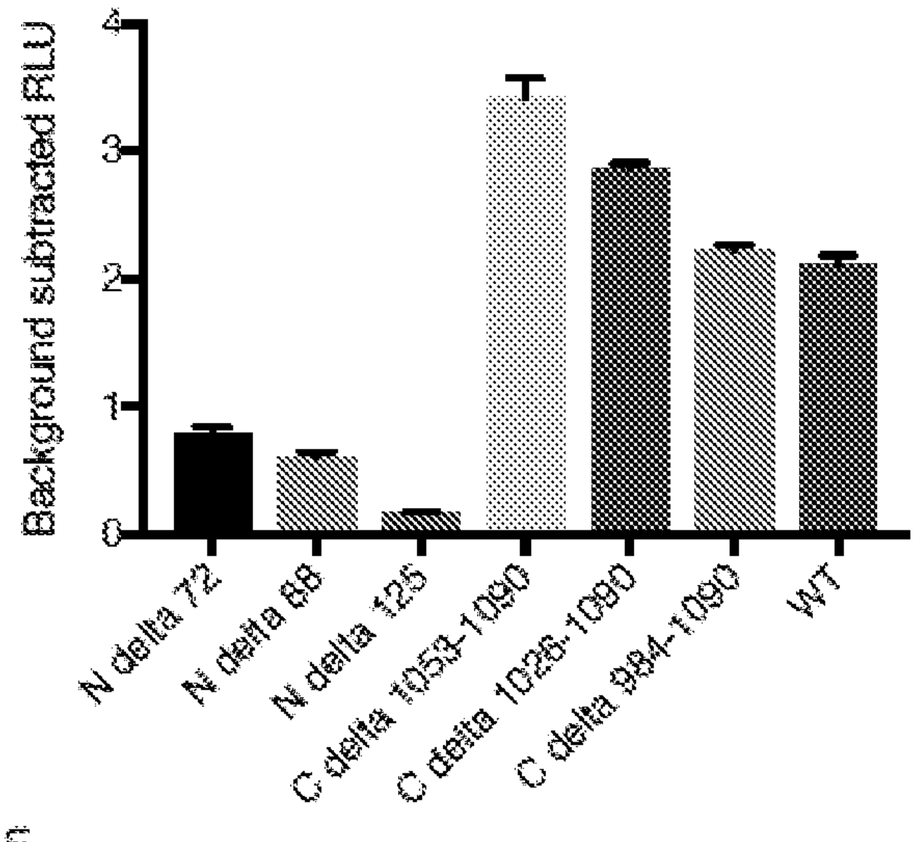
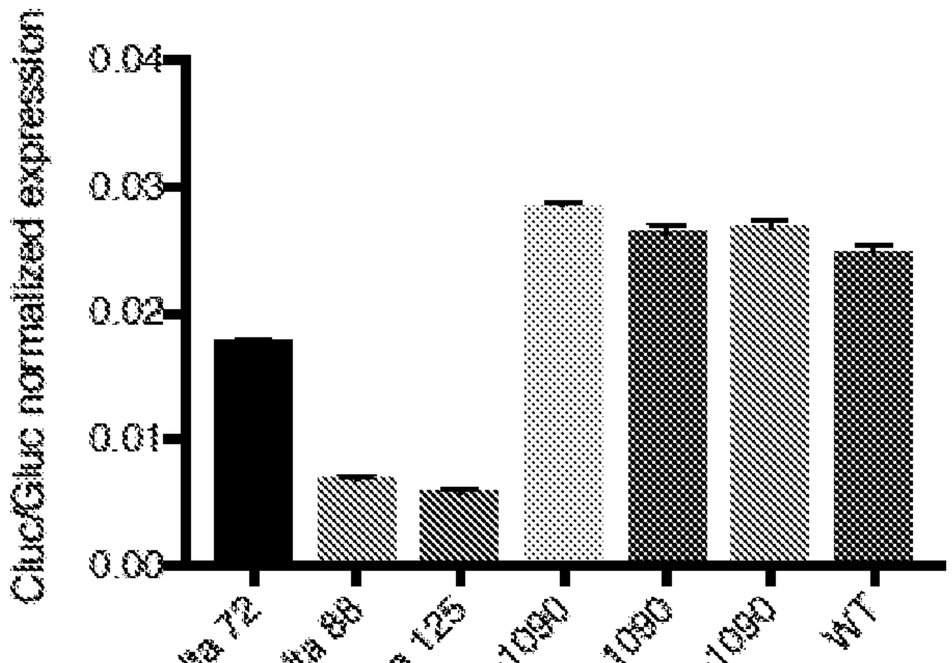

A
Cluc/Gluc normalized expression
0
2
4
6
8
R348E
V351L
T375S
T375G
R455E
R455G
N473D
R510E
T490A
B
Cluc/Gluc normalized expression
0.00
0.02
0.04
0.06
R348E
V351L
T375S
T375G
R455E
R455G
N473D
R510E
T490A

FIG. 47

A


Background subtracted RLU
0
1
2
3
4
N delta 72
N delta 88
N delta 125
C delta 1053-1090
C delta 1026-1090
C delta 984-1090
WT

B


Cluc/Gluc normalized expression
0.00
0.01
0.02
0.03
0.04
N delta 72
N delta 88
N delta 125
C delta 1053-1090
C delta 1026-1090
C delta 984-1090
WT

C

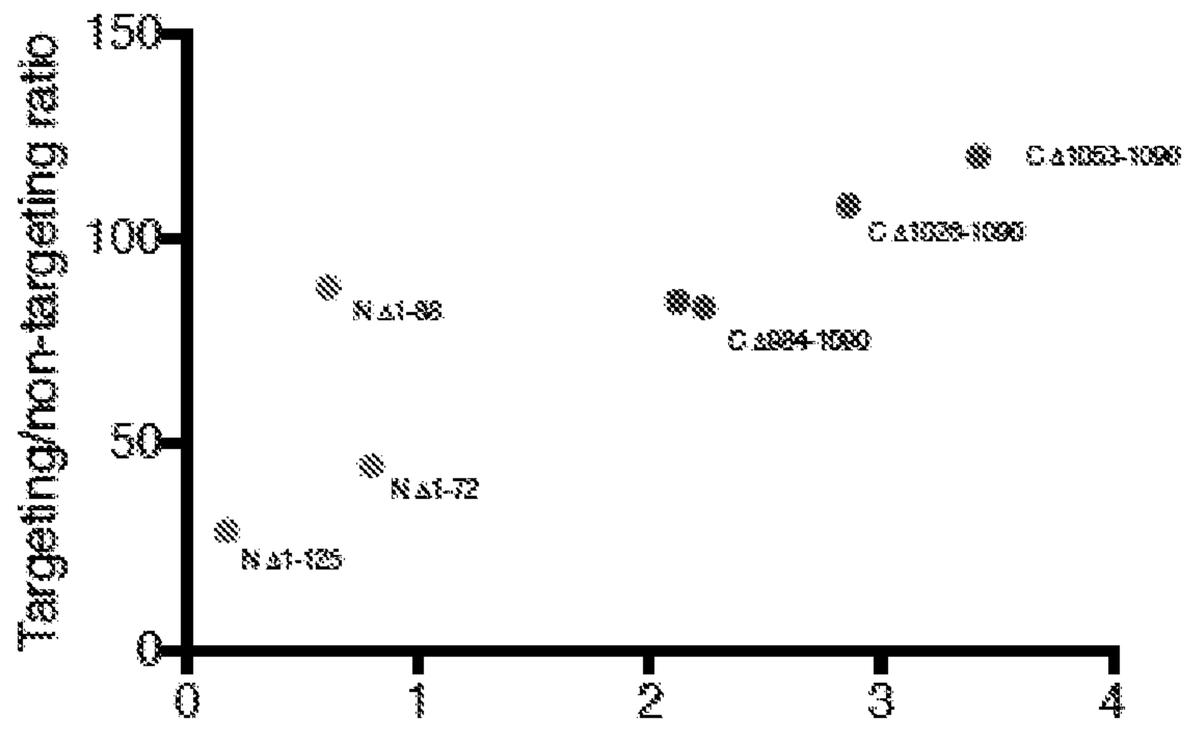

Targeting/non-targeting ratio
0
50
100
150
N Δ1-125
N Δ1-72
N Δ1-88
C Δ984-1090
C Δ1026-1090
C Δ1053-1090
0
1
2
3
4
Targeting guide

FIG. 48

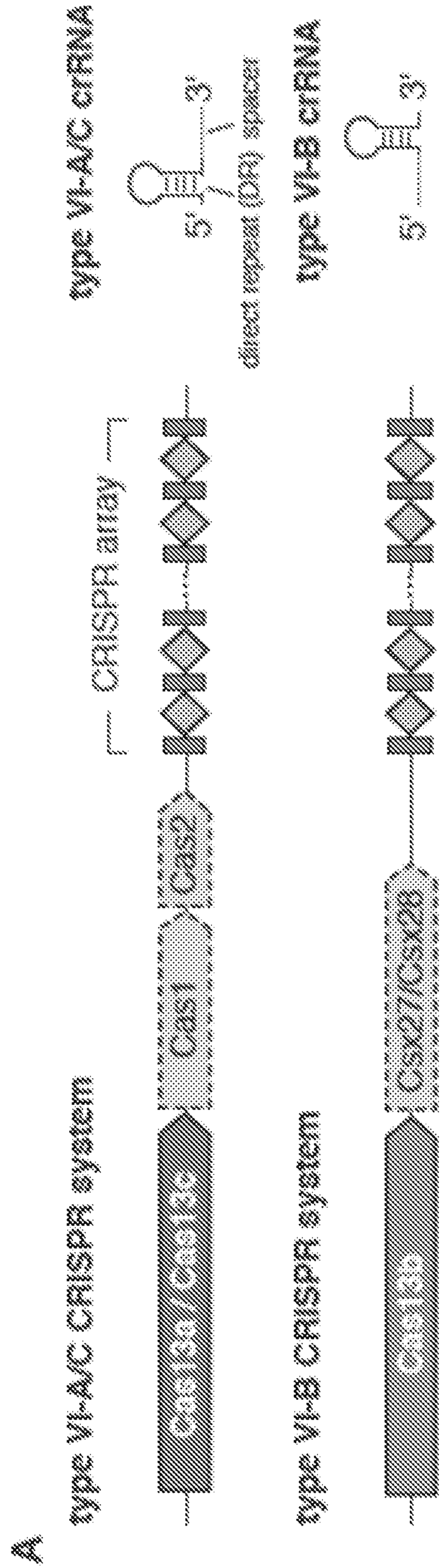
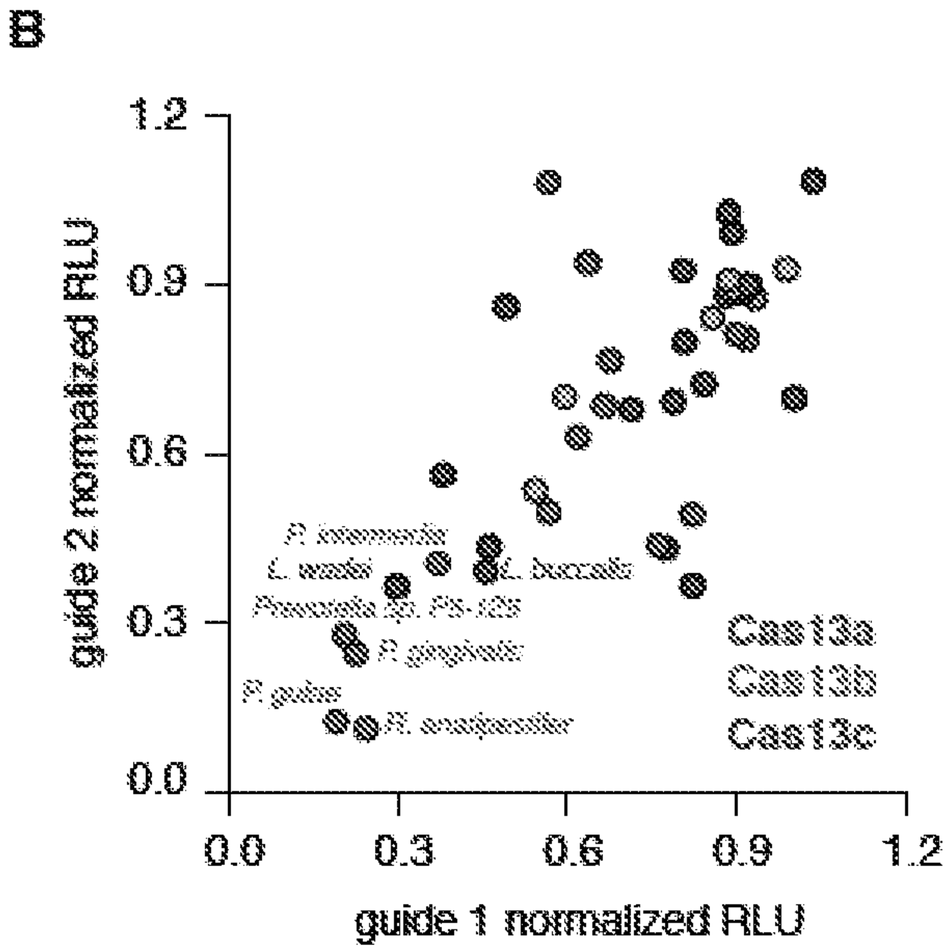
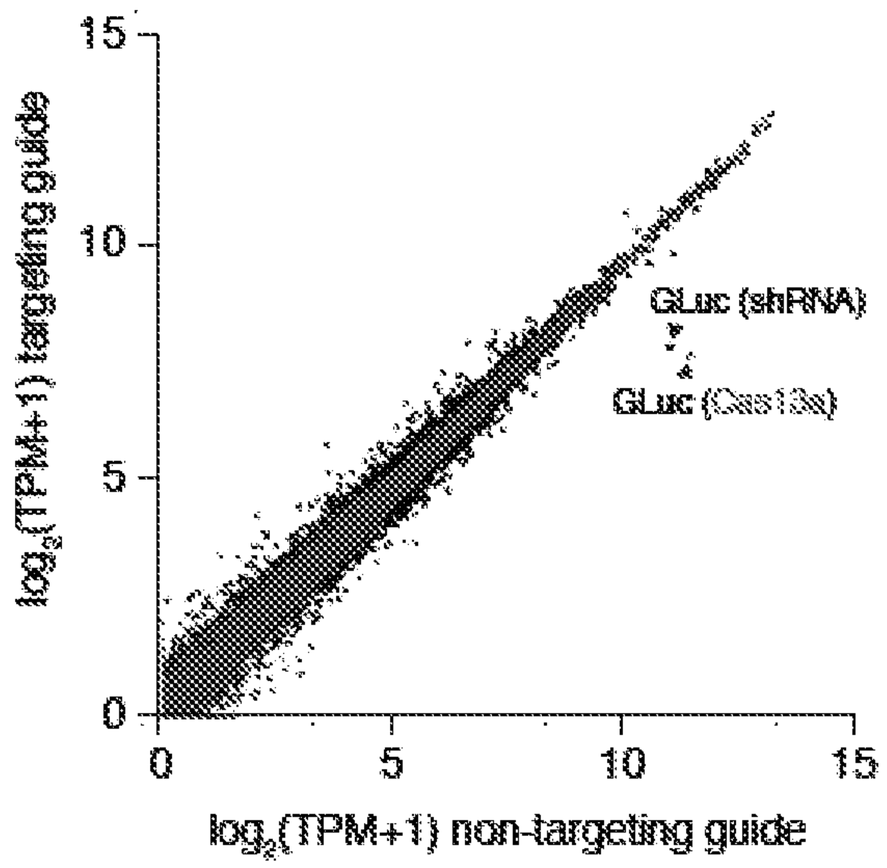
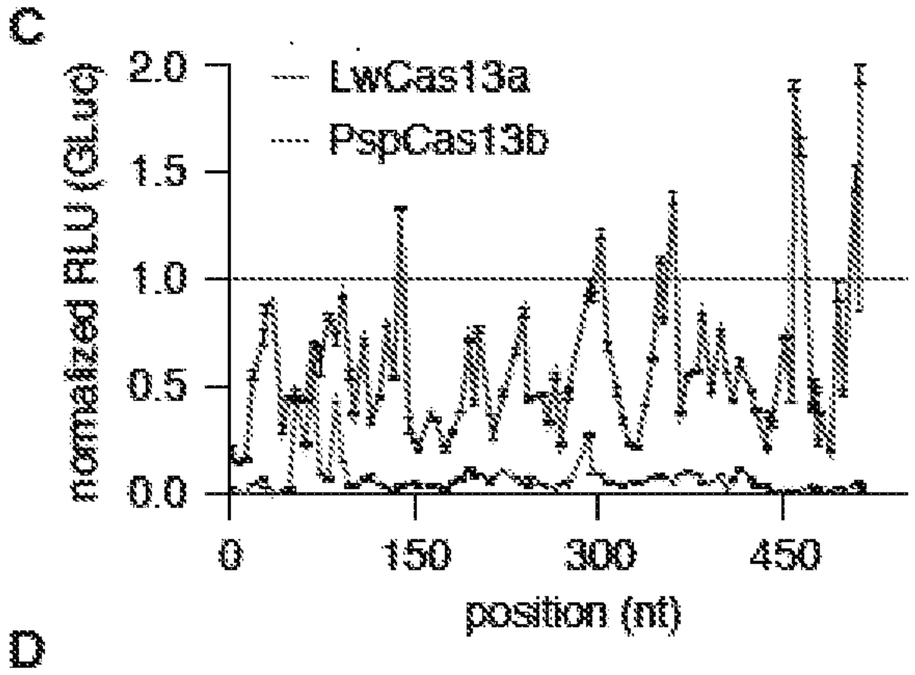
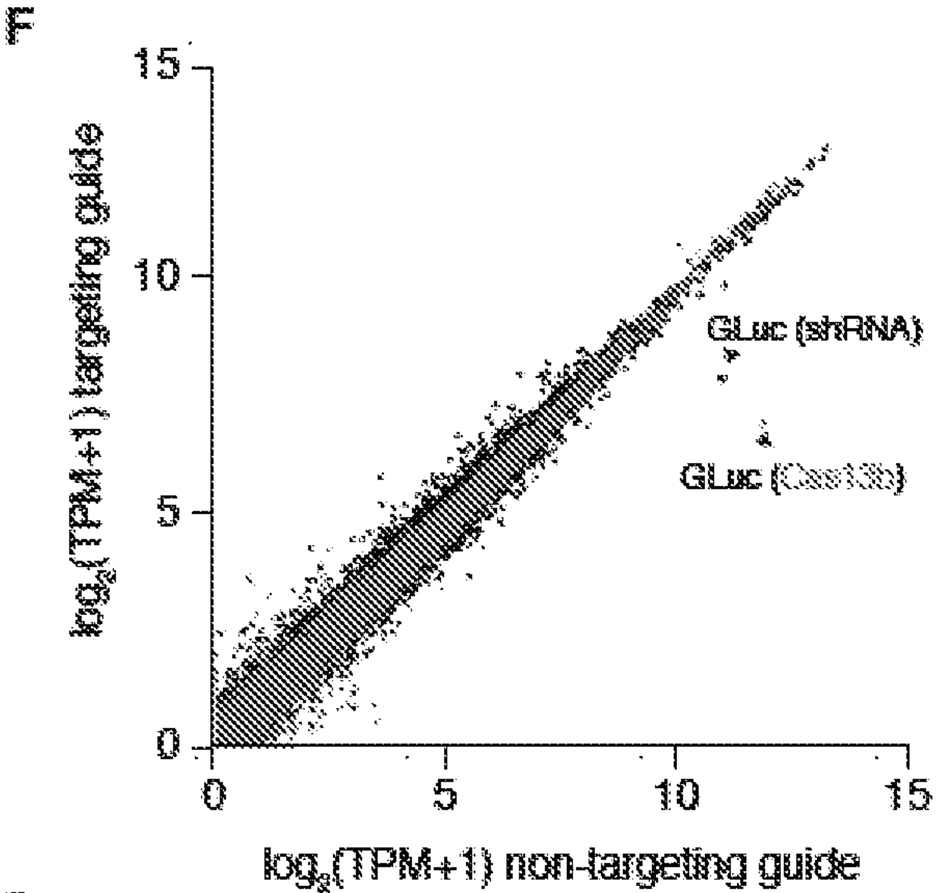

type VI-A/C crRNA
5'
3'
spacer
type VI-B crRNA
5'
3'
CRISPR array
Type VI-A/C CRISPR system
cas13a/c2c2c
Cas1
Cas2
Type VI-B CRISPR system
cas13b
Csx27/Csx28

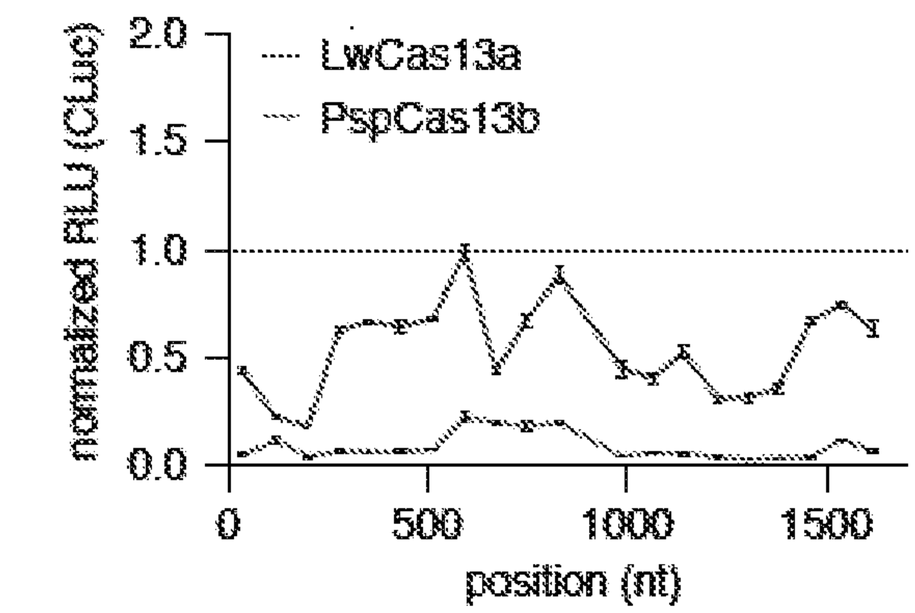
FIG. 49A











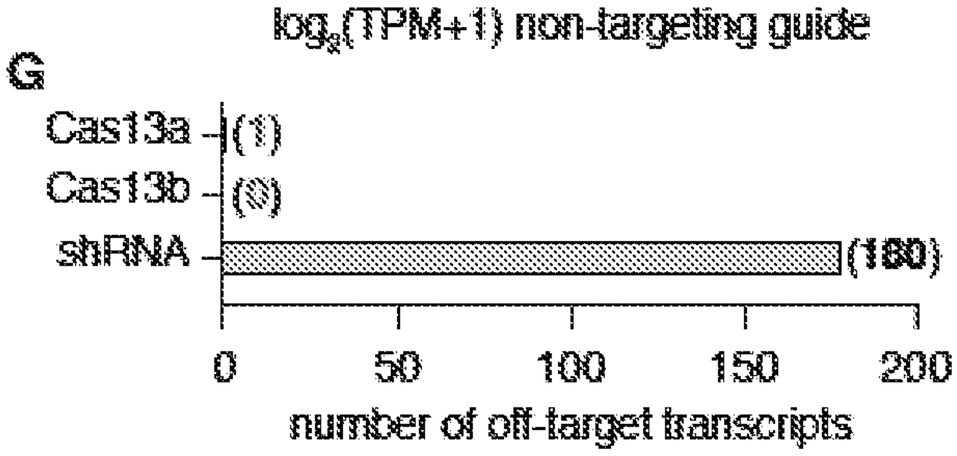


FIG. 49 (continued)

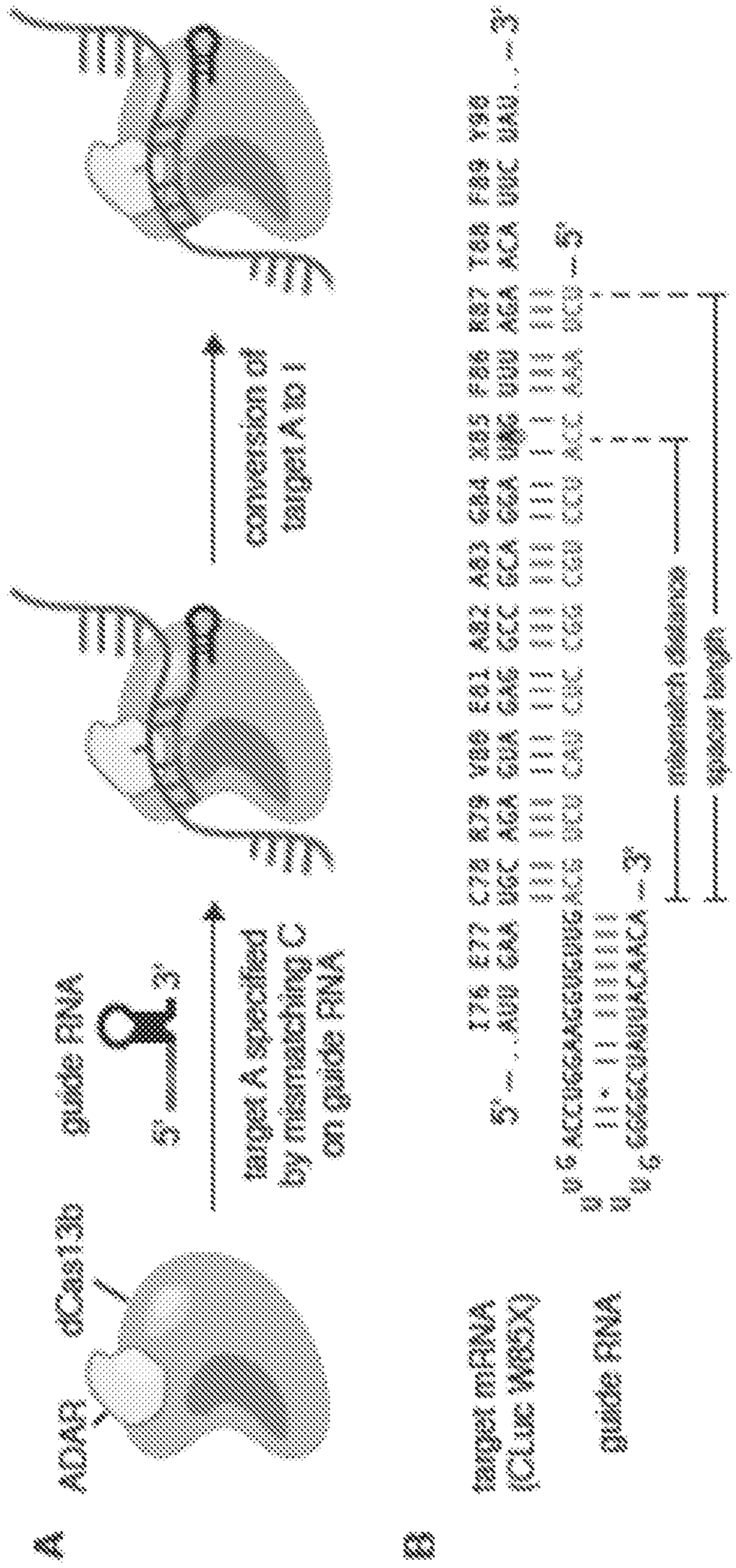

5'
3'
conversion of target A to 1
3'
5'
target A
guide RNA
5'
3'
target mRNA (CLUC UNSEQ)
guide RNA
% ACTIVE
% CONVERSION
guide RNA
target mRNA
5'
3'
target A
guide RNA
5'
3'
target A converted by trimming on guide RNA
guide RNA
target mRNA

FIG. 50

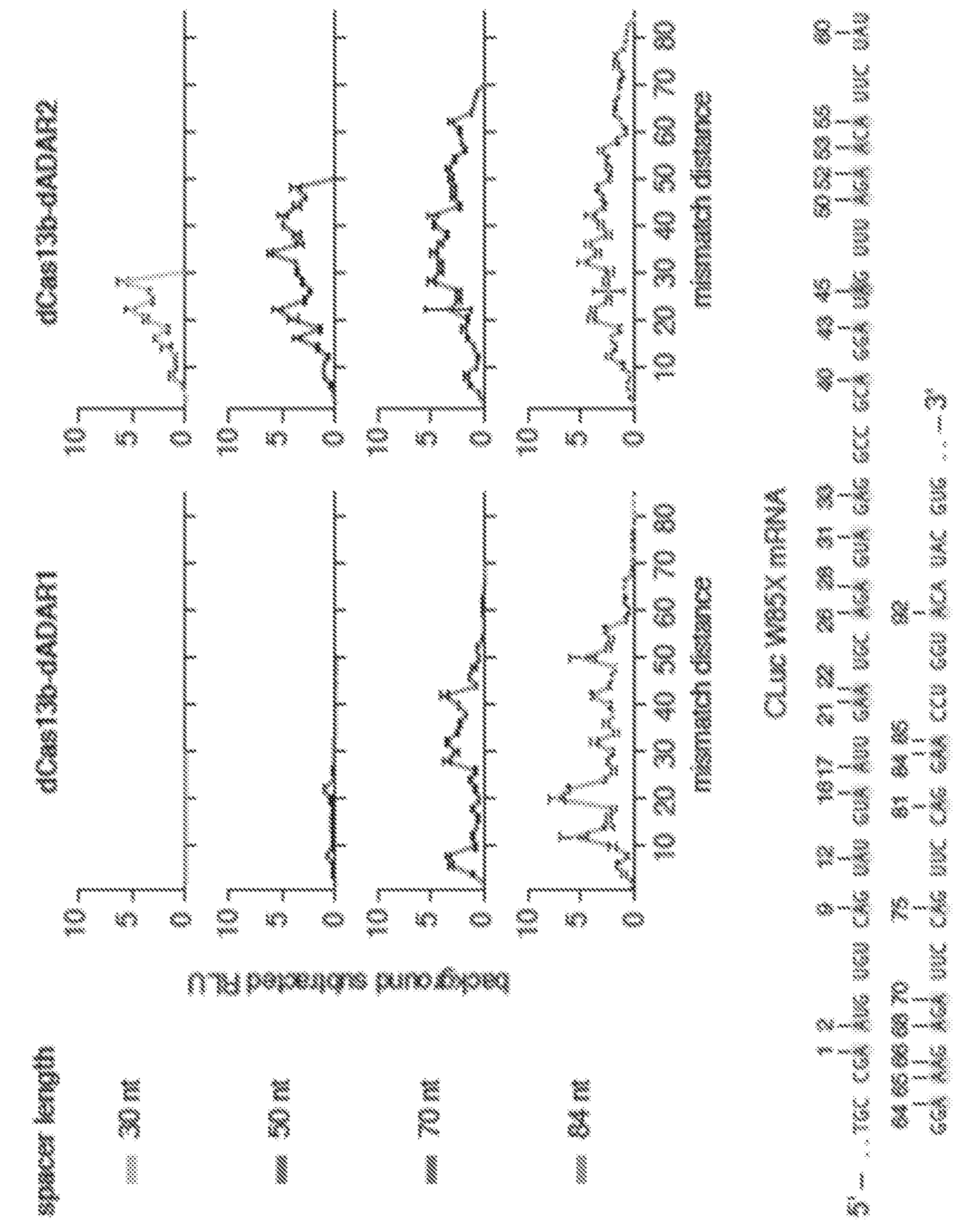

C
spacer length
30 nt
50 nt
70 nt
84 nt
dCas13b-dADAR1
dCas13b-dADAR2
background subtracted RLU
mismatch distance
mismatch distance
10
5
0
10 20 30 40 50 60 70 80
D
CLuc W85X mRNA
5'
3'

FIG. 50

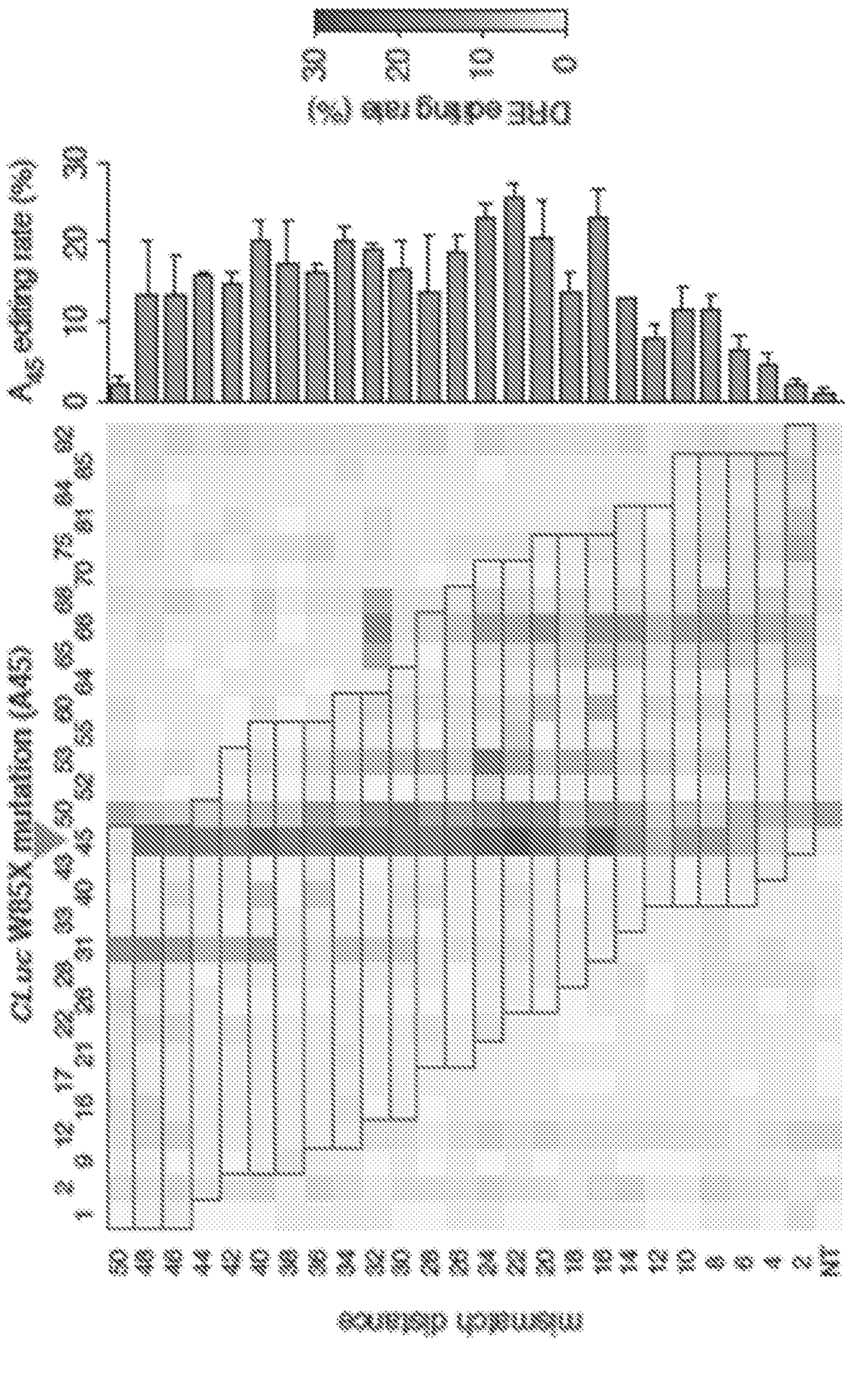

DNE editing rate (%)
0
20
40
60
80
100
Allele editing rate (%)
0
10
20
Cluc W85X mutation (A45)
1
8
15
22
29
36
43
50
57
64
71
78
85
92
mismatch distance
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1

FIG. 5D

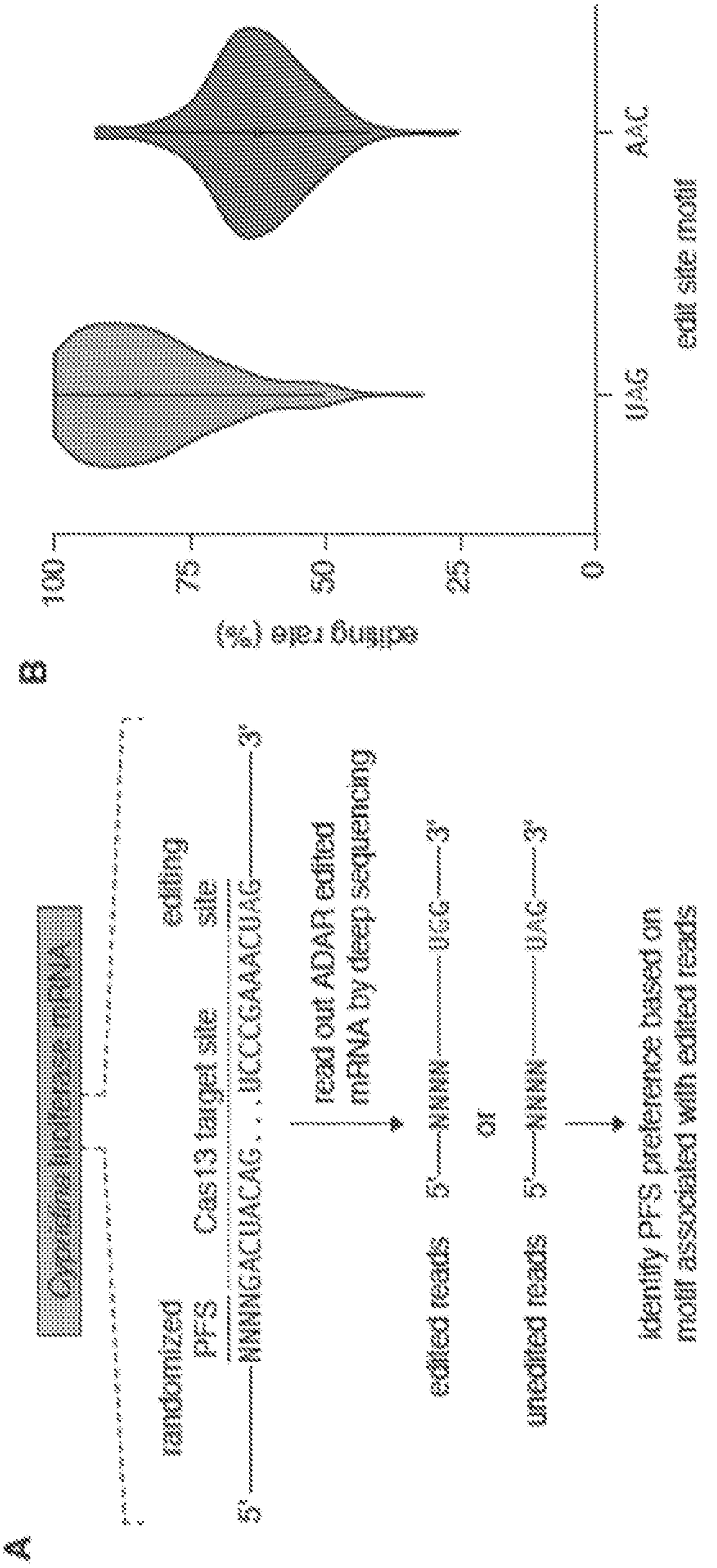
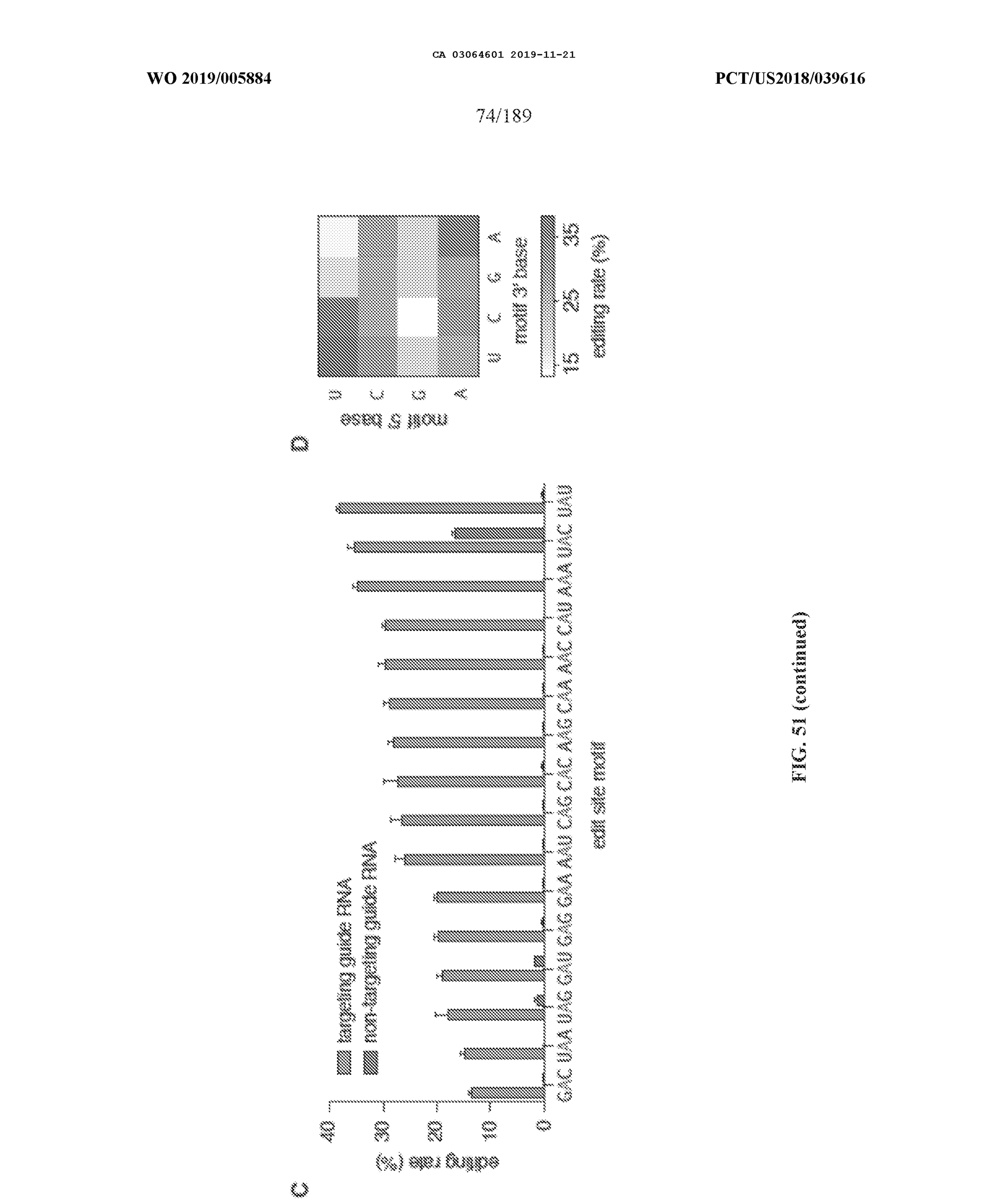

editing rate (%)
100
75
50
25
0
UAC
site motif
B
unedited reads
5'—NNNN——UAC——3'
or
edited reads
5'—NNNN——UGC——3'
read out ADAR edited
mRNA by deep sequencing
unbiased
PPS
editing
site
Cas13 target site
NNNNGACUACAG . . . UCCCGAAGCUAG
Cas13 mRNA

FIG. 5I


FIG. 51 (continued)
editing rate (%)
15
25
35
edit site motif
GAC UAA UAG GAU GAG GAA AAU CAG CAC AAG
GAC AAC CAU AAA UAC UAG
motif 3' base
A
G
C
U
motif 5' base
A
G
C
U
targeting guide RNA
non-targeting guide RNA
editing rate (%)
0
10
20
30
40

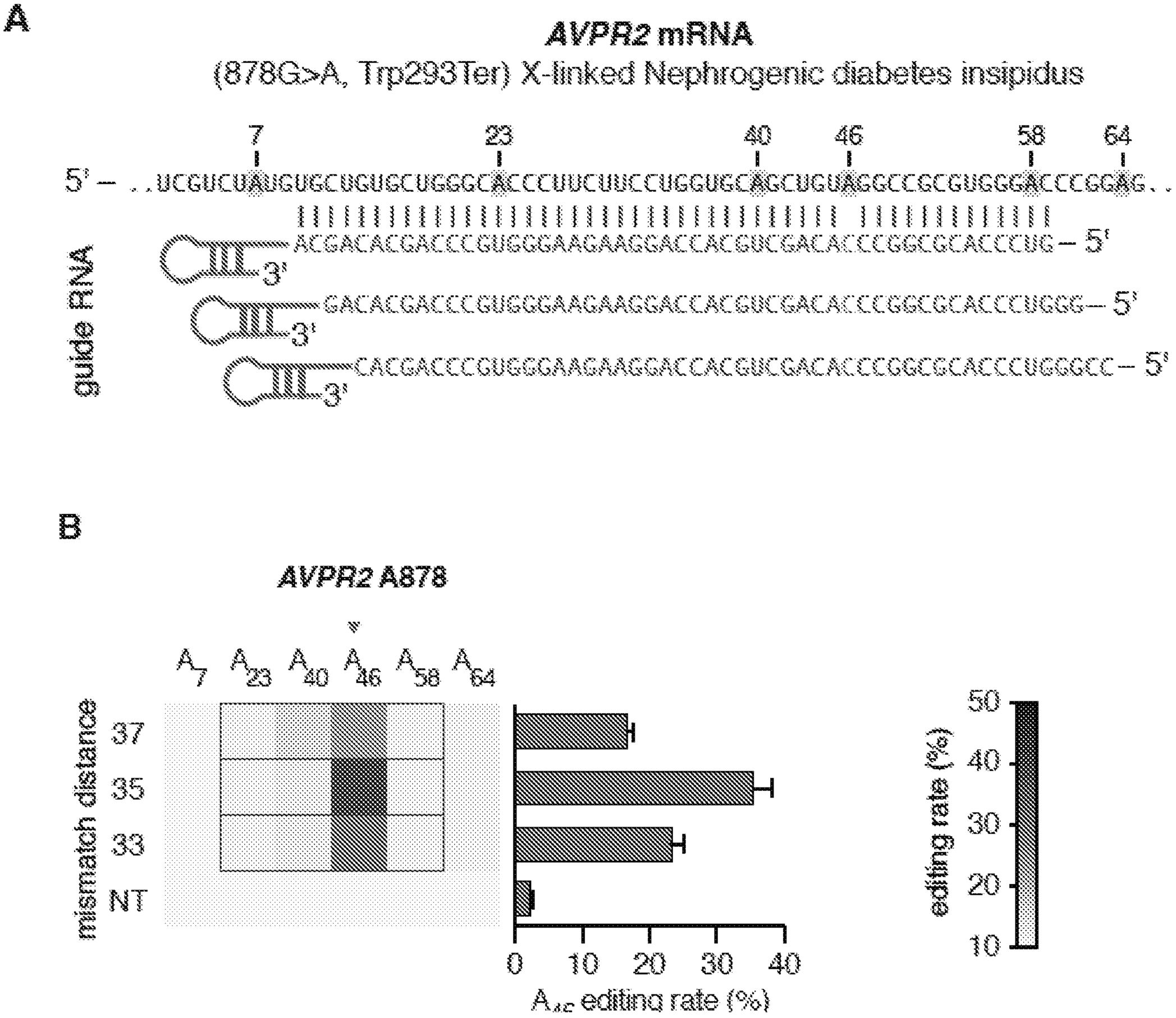
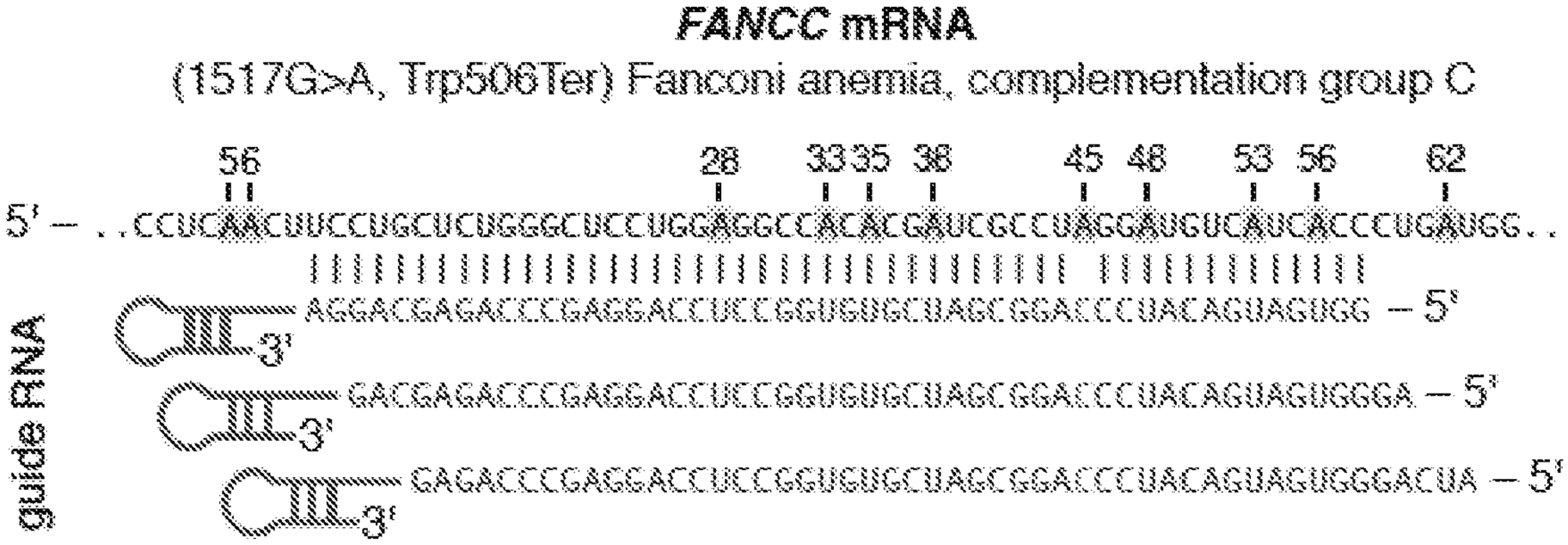

A
AVPR2 mRNA
(878G>A, Trp293Ter) X-linked Nephrogenic diabetes insipidus
7
23
40
46
58
64
5' – . . UCGUCUAUGUGCUGUGCUGGGCACCCUUCUUCCUGGUGCAGCUGUAGGCCGCGUGGGACCCGGAG . . – 3'
guide RNA
ACGACACGACCCGUGGGAAGAAGGACCACGUCGACACCCGGCGCACCCUG – 5'
3'
GACACGACCCGUGGGAAGAAGGACCACGUCGACACCCGGCGCACCCUGGG – 5'
3'
CACGACCCGUGGGAAGAAGGACCACGUCGACACCCGGCGCACCCUGGGCC – 5'
3'
B
AVPR2 A878
A7 A23 A40 A46 A58 A64
mismatch distance
37
35
33
NT
0 10 20 30 40
A46 editing rate (%)
editing rate (%)
50
40
30
20
10

FIG. 52

C



D

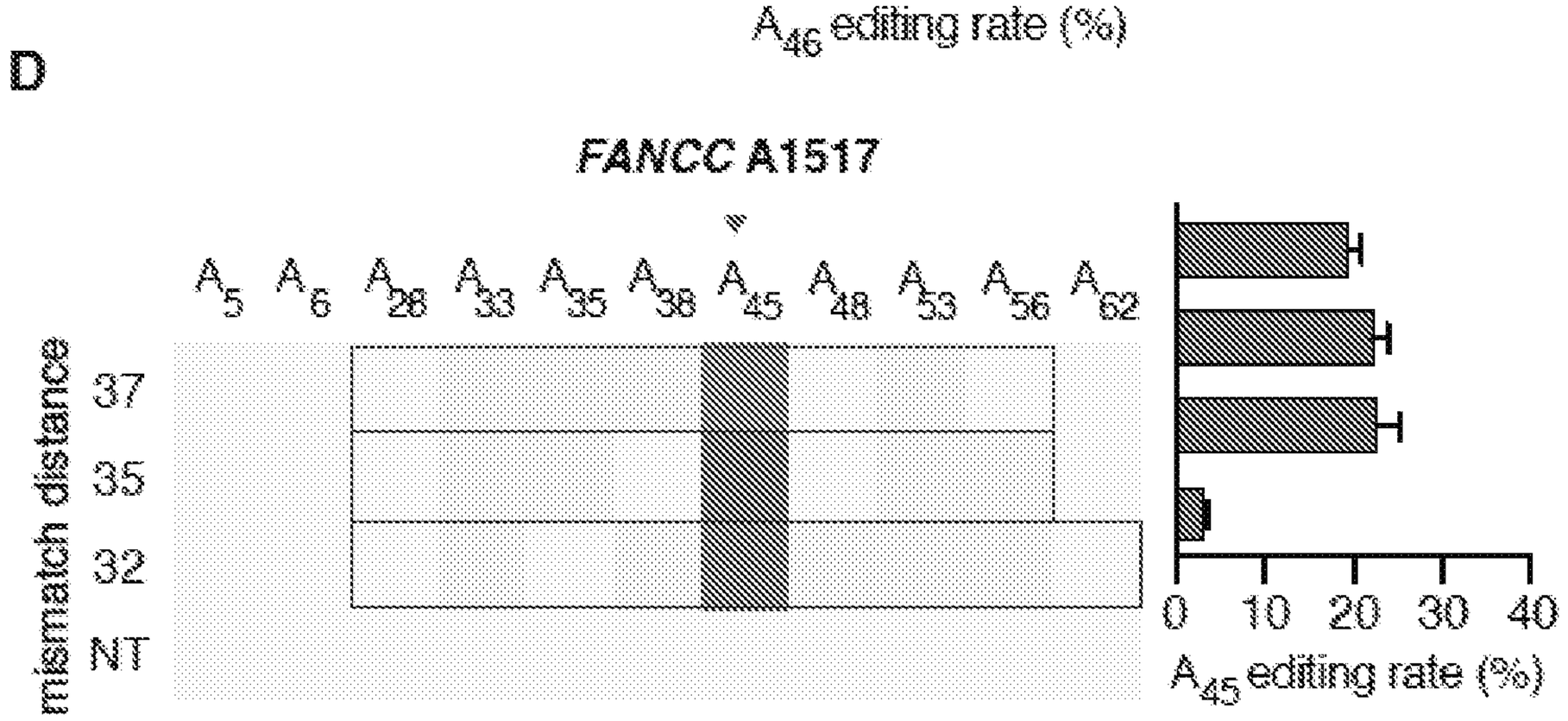
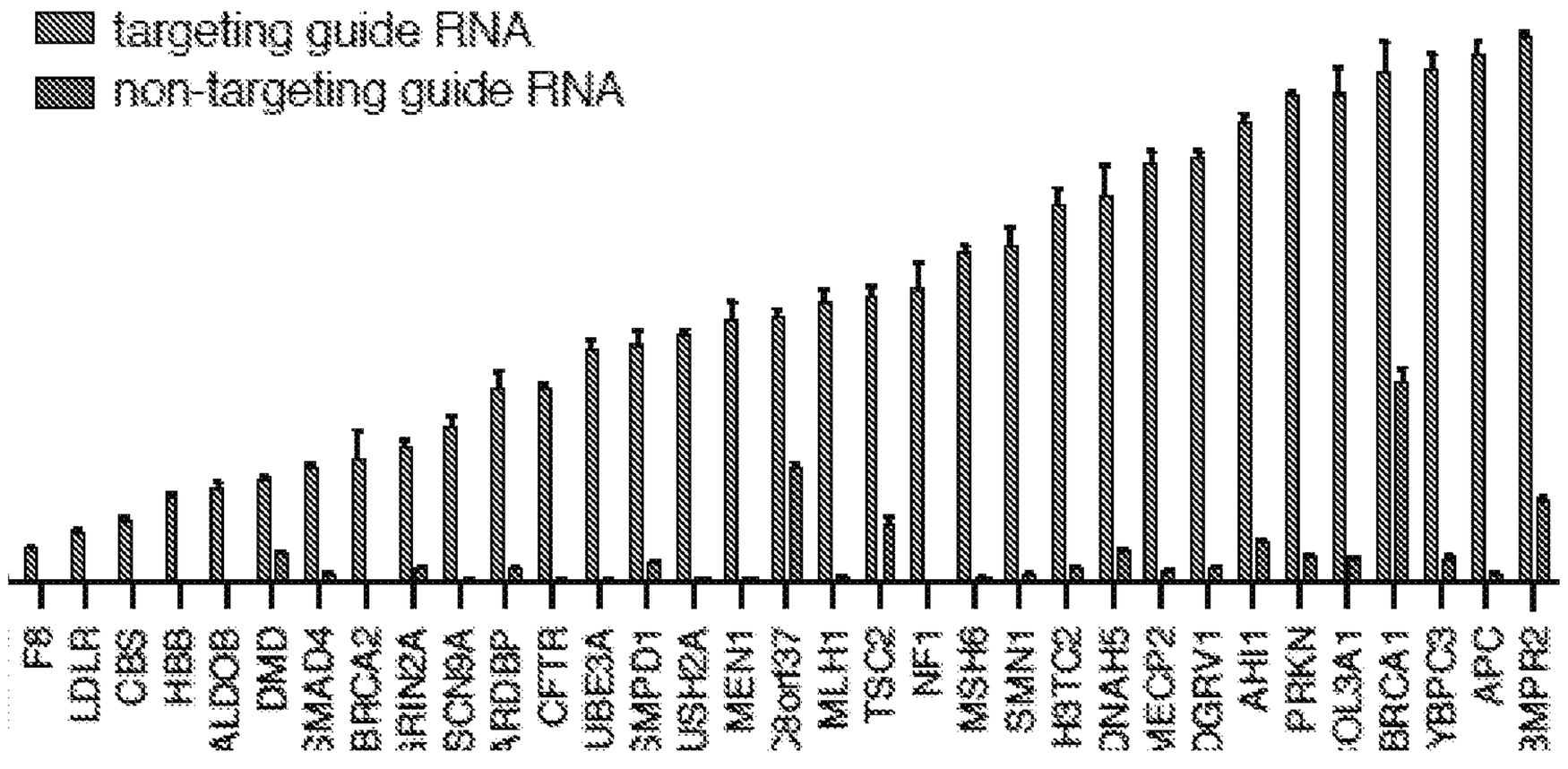
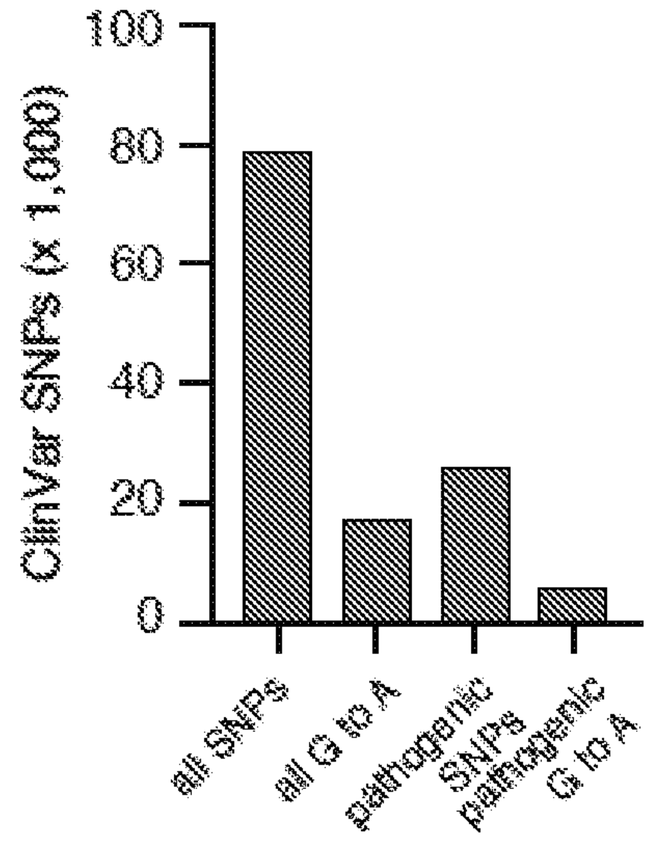


FIG. 52 (continued)

E



F



G

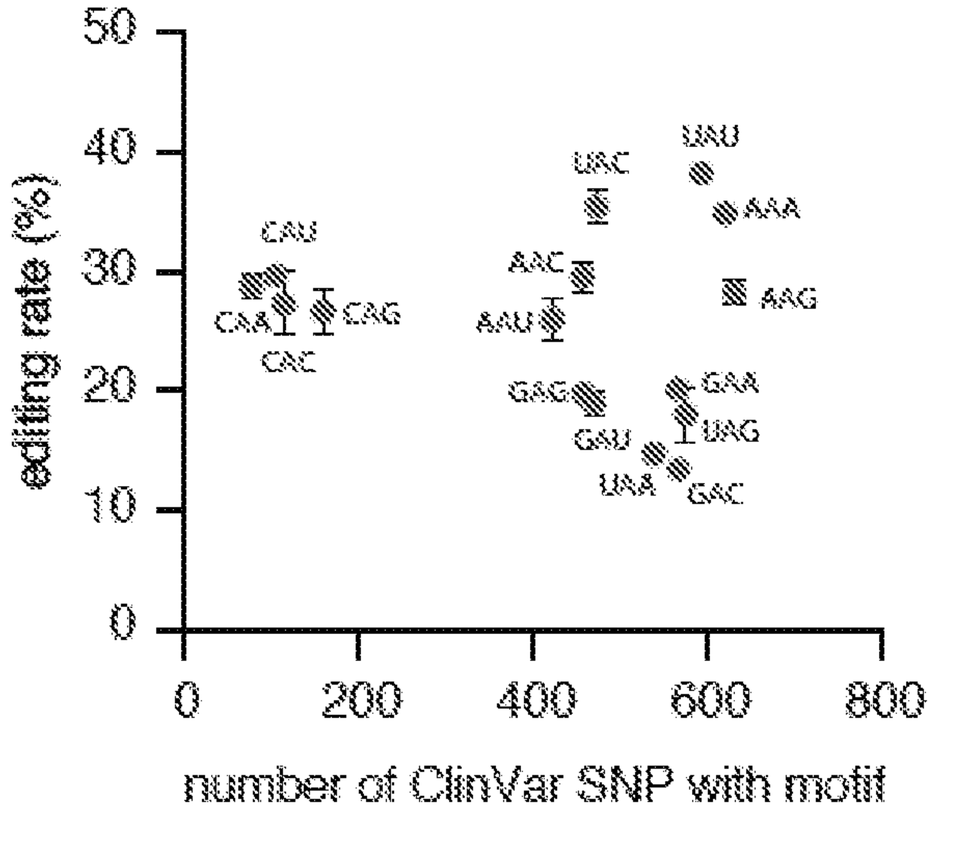


FIG. 52 (continued)

CA 03064601 2019-11-21



FIG. 53

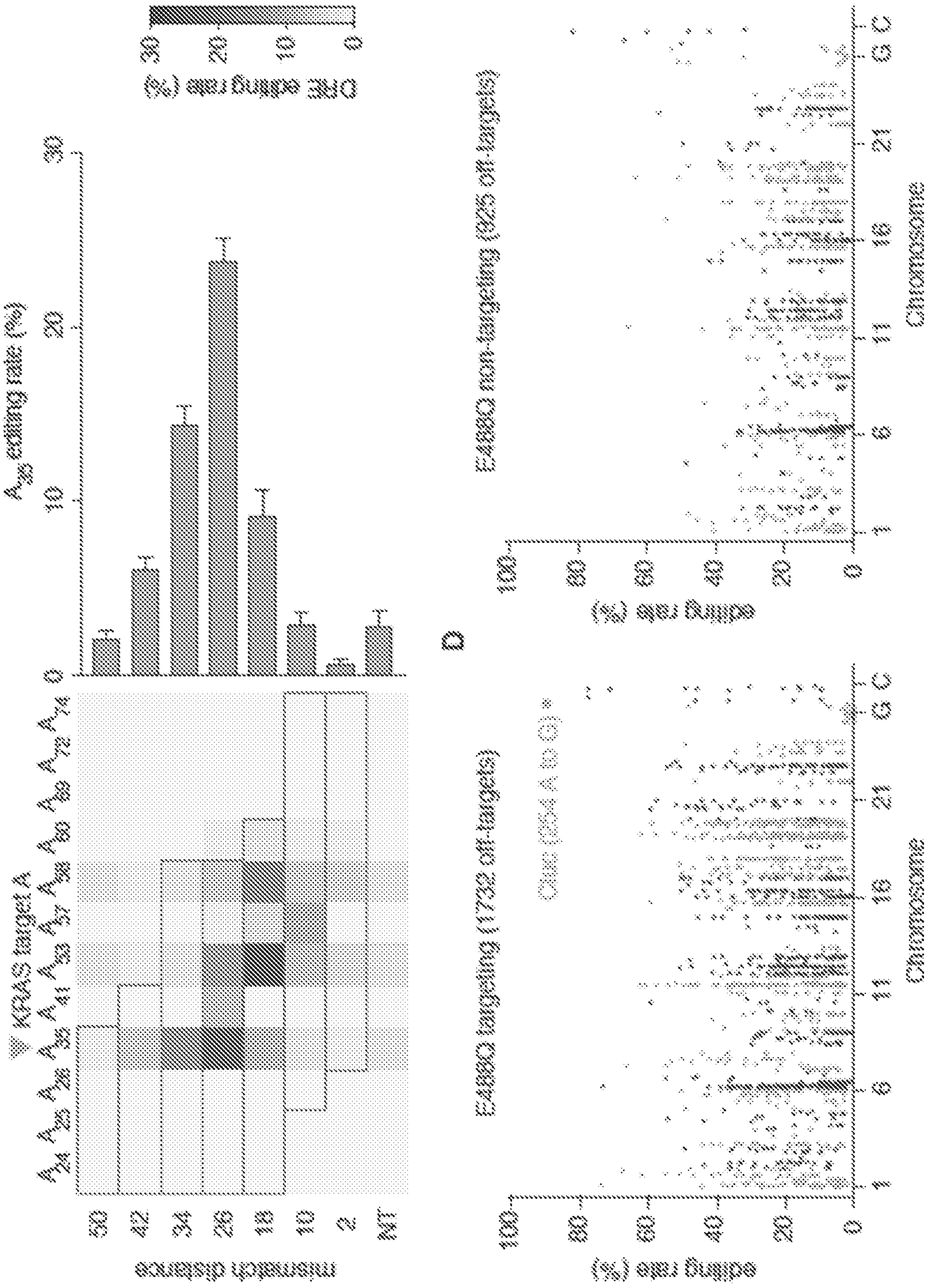


FIG. 53 (continued)

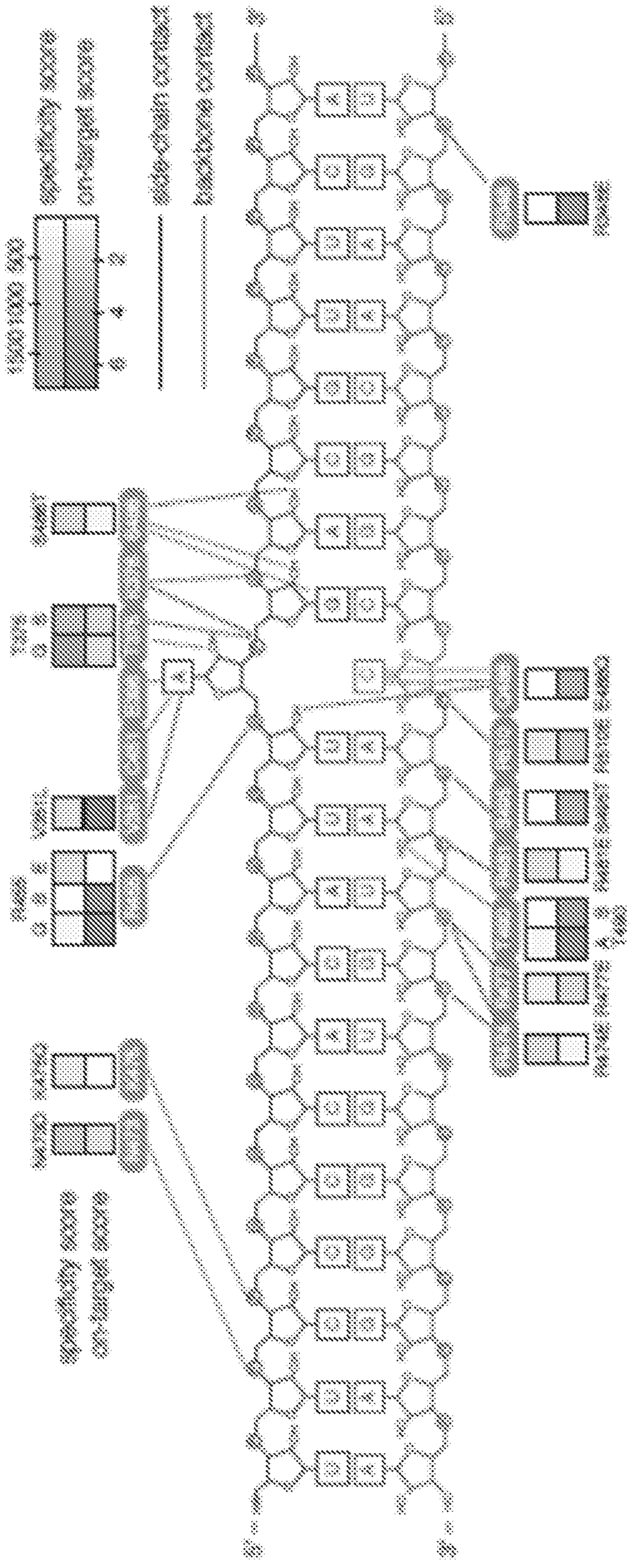


FIG. 54

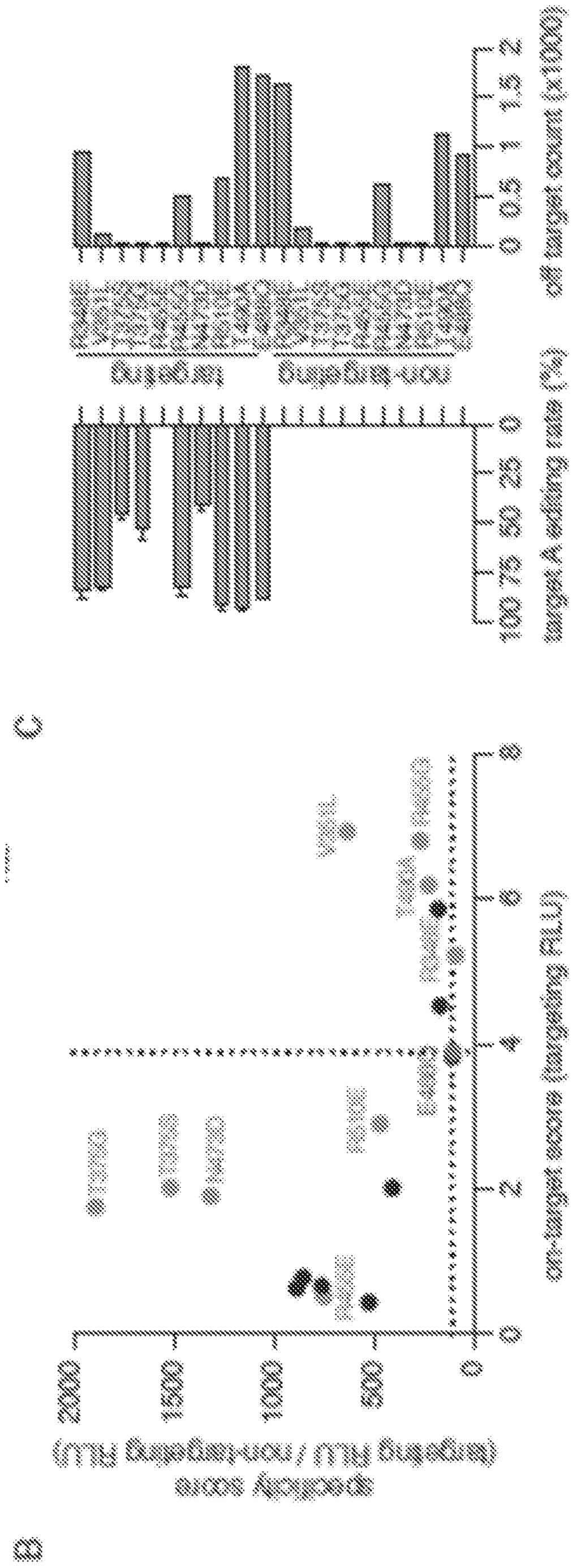


FIG. 54 (continued)

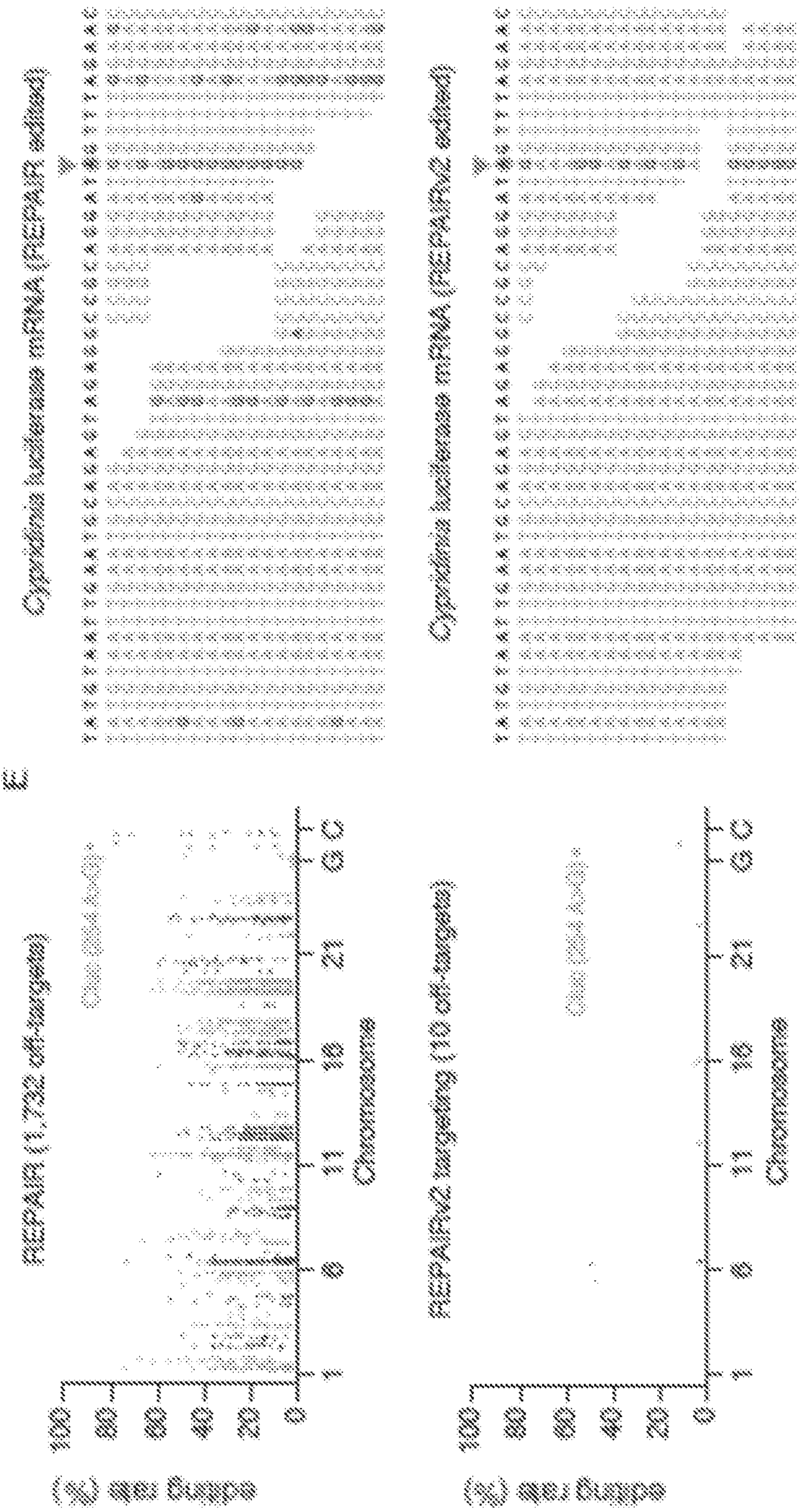


FIG. 54 (continued)

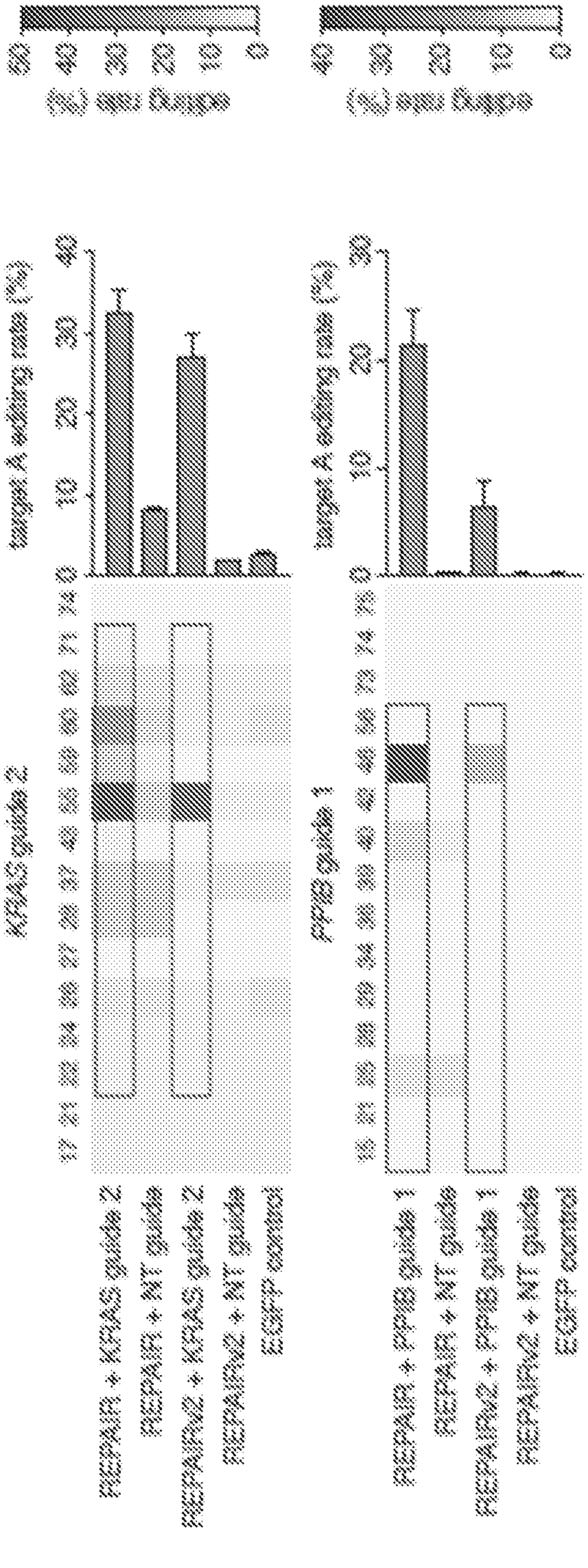


FIG. 54 (continued)

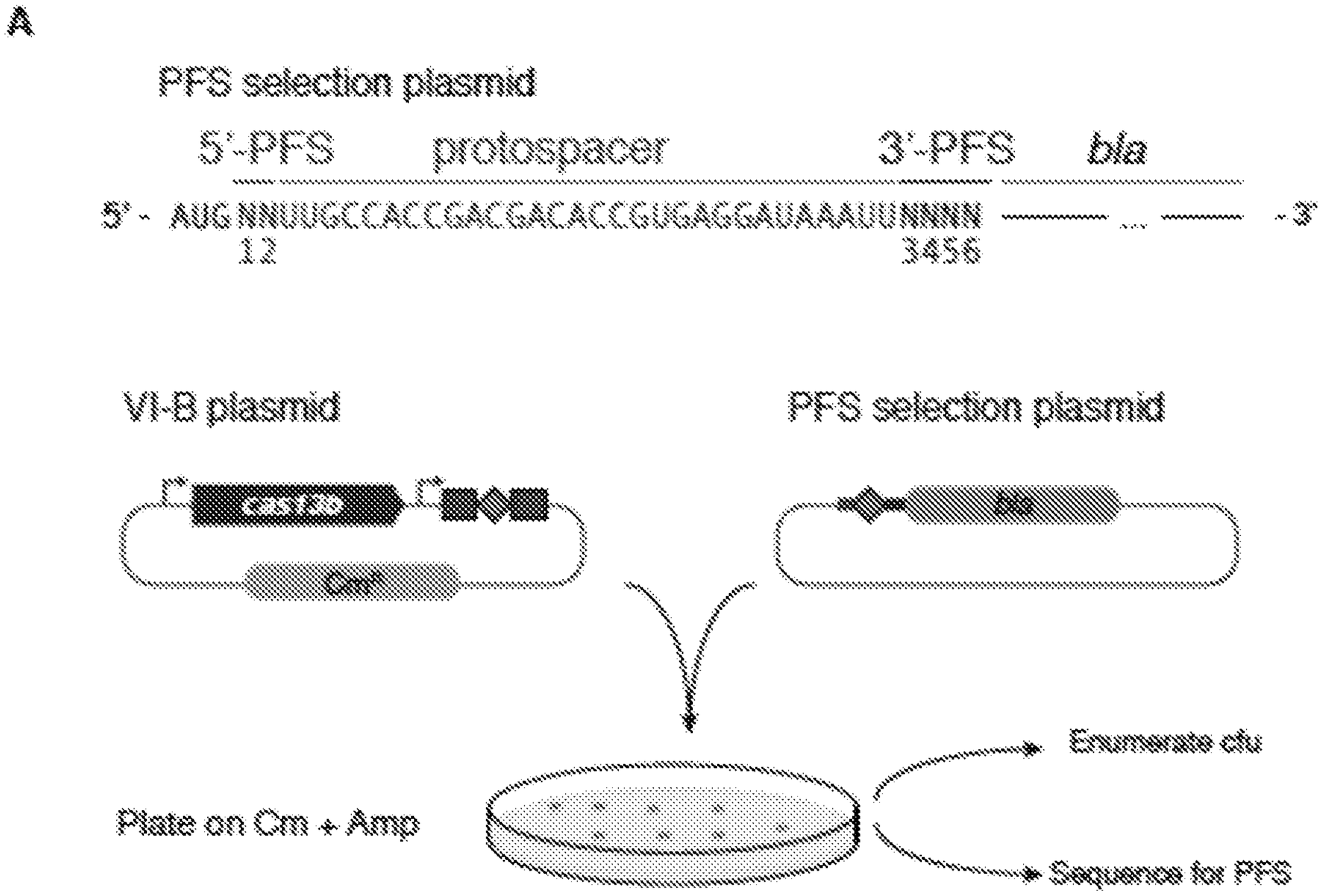

A
PFS selection plasmid
5'-PFS
protospacer
3'-PFS
bla
5'- AUG NNUUGCCACCGACGACACCGUGAGGAUAAAUUNNNN -3'
12
3456
VI-B plasmid
PFS selection plasmid
Cm
Amp
Plate on Cm + Amp
Enumerate cfu
Sequence for PFS

FIG. 55

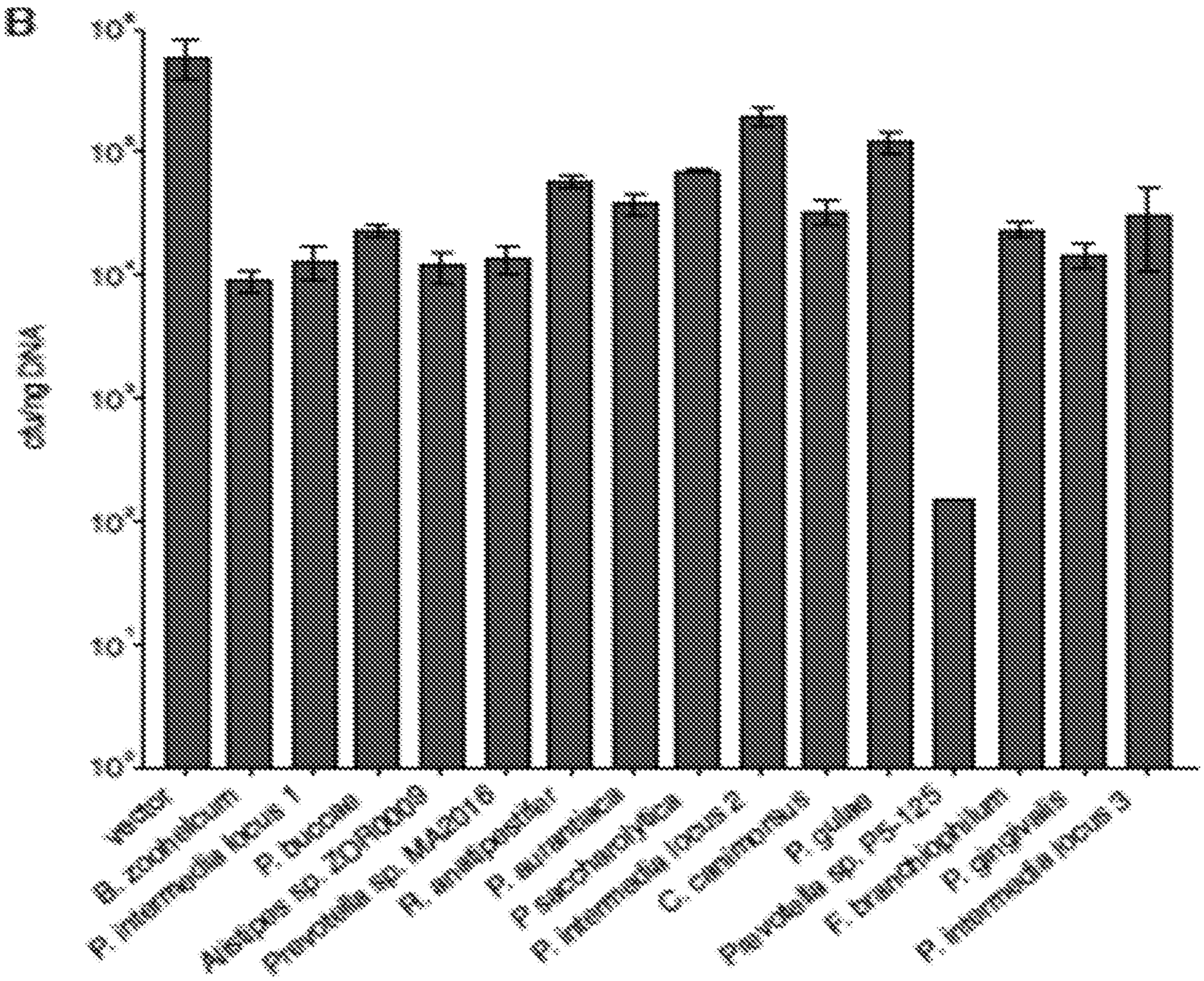


FIG. 55 (continued)

C

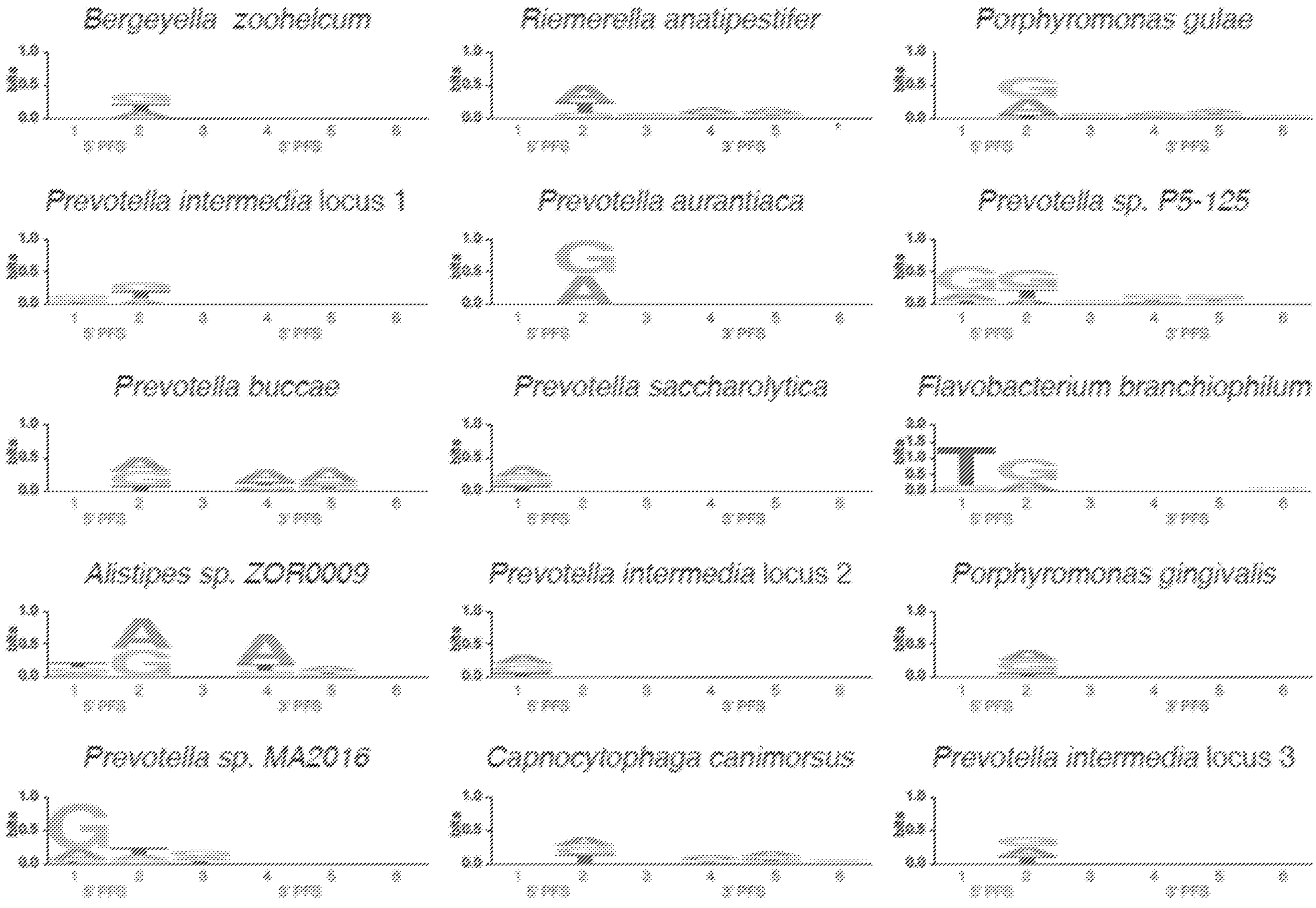
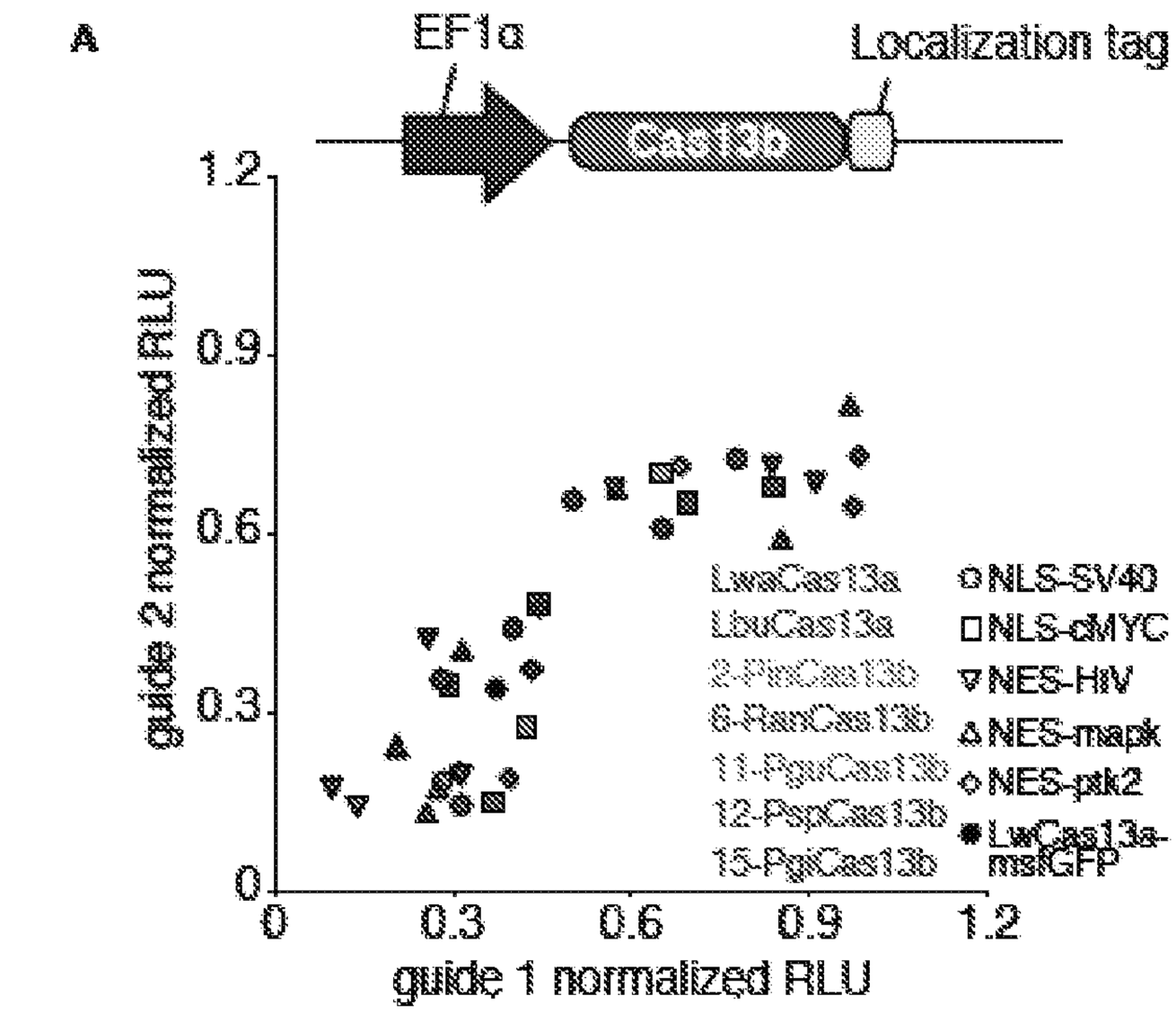

Bergeyella zoohelcum
Riemerella anatipestifer
Porphyromonas gulae
Prevotella intermedia locus 1
Prevotella aurantiaca
Prevotella sp. P5-125
Prevotella buccae
Prevotella saccharolytica
Flavobacterium branchiophilum
Alistipes sp. ZOR0009
Prevotella intermedia locus 2
Porphyromonas gingivalis
Prevotella sp. MA2016
Capnocytophaga canimorsus
Prevotella intermedia locus 3

FIG. 55 cont.


A
EF1a
Localization tag
Cas13b
guide 2 normalized RLU
guide 1 normalized RLU
0
0.3
0.6
0.9
1.2
LwaCas13a
LbuCas13a
2-PinCas13b
6-RanCas13b
11-PguCas13b
12-PspCas13b
15-PguCas13b
NLS-SV40
NLS-cMYC
NES-HIV
NES-mapk
NES-ptk2
LwaCas13a-msfGFP

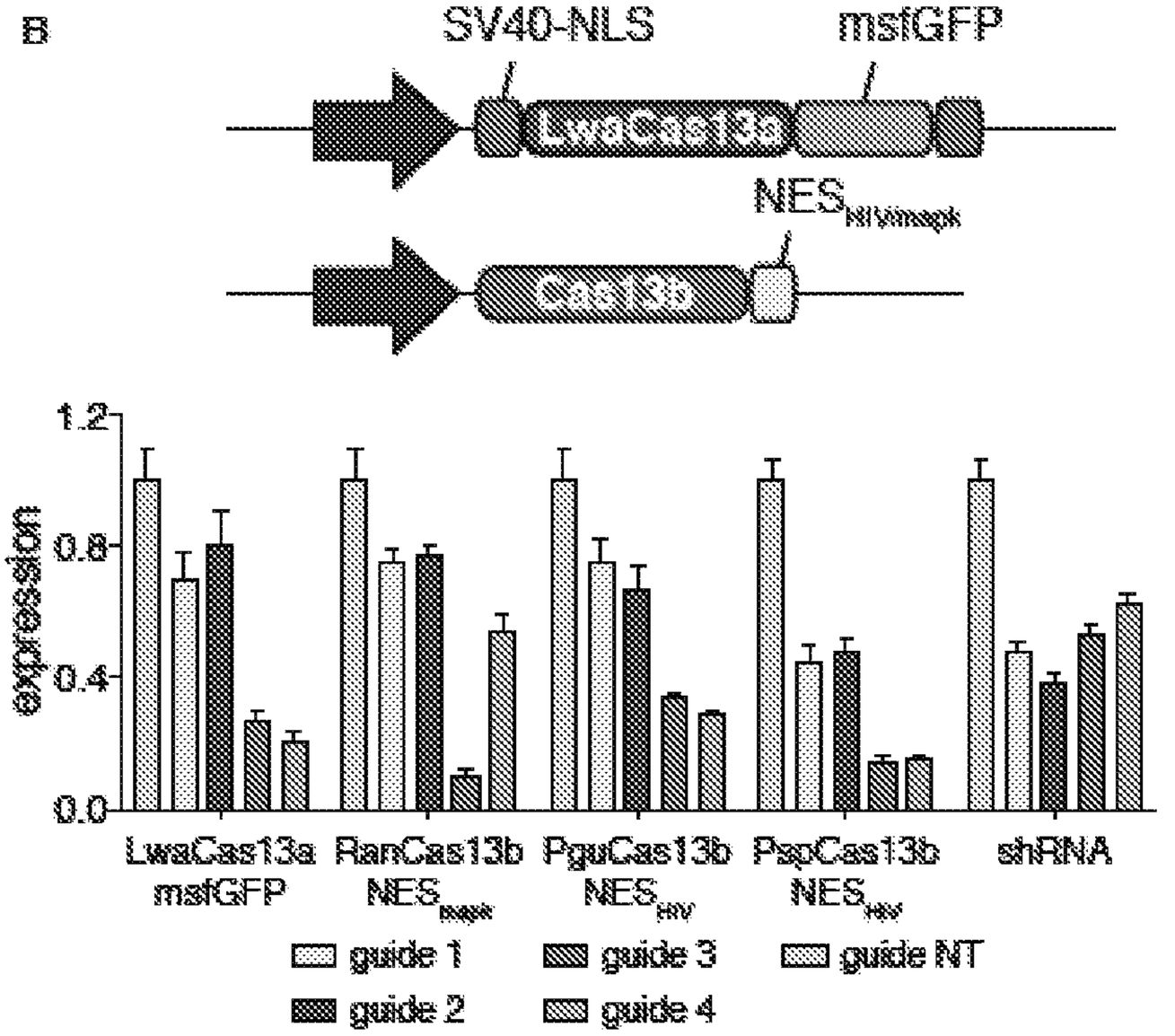

B
SV40-NLS
msfGFP
LwaCas13a
NES
Cas13b
expression
0.0
0.4
0.8
1.2
LwaCas13a
RanCas13b
PguCas13b
PspCas13b
shRNA
msfGFP
guide 1
guide 2
guide 3
guide 4
guide NT

FIG. 56

C

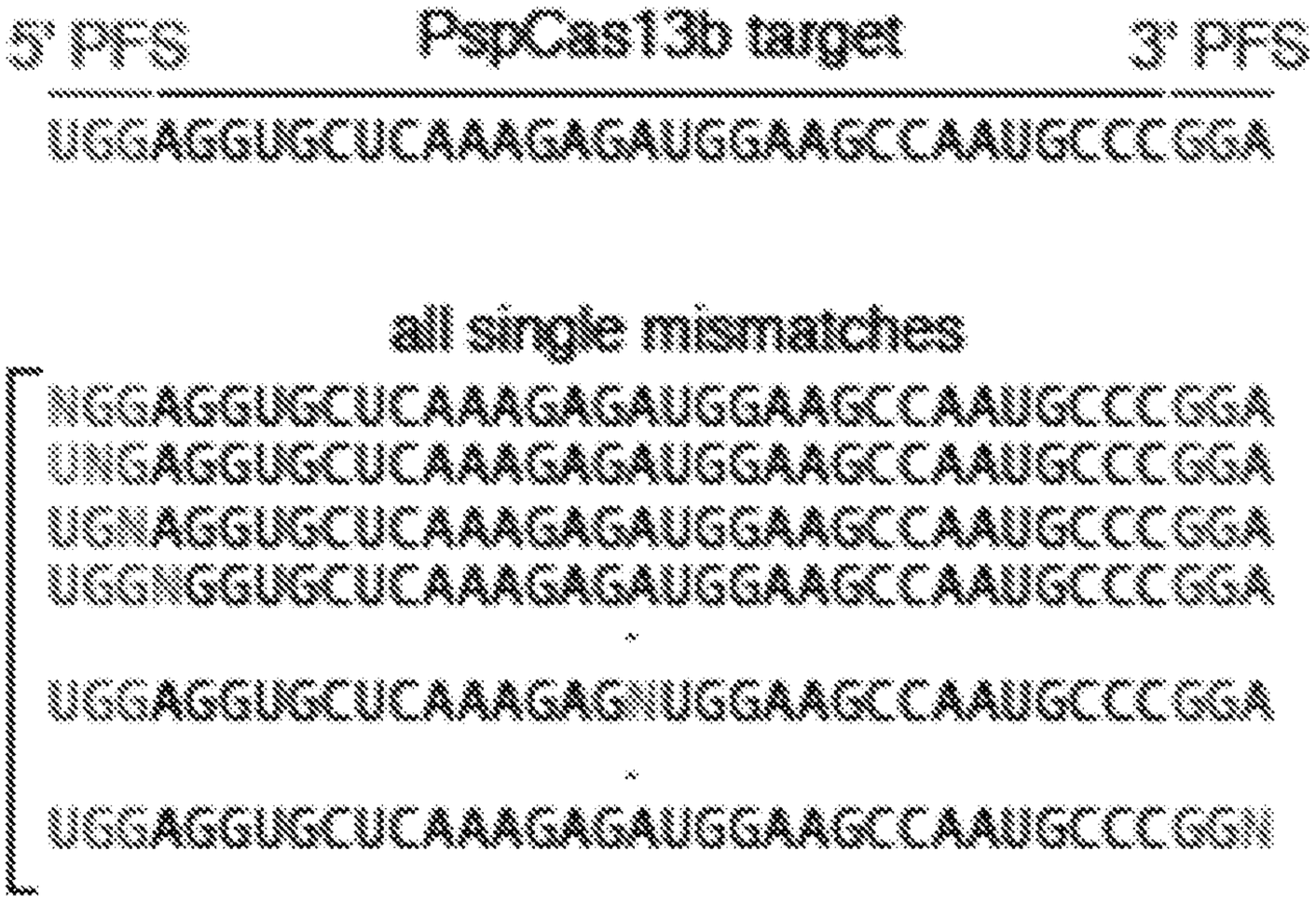


FIG. 56 (continued)

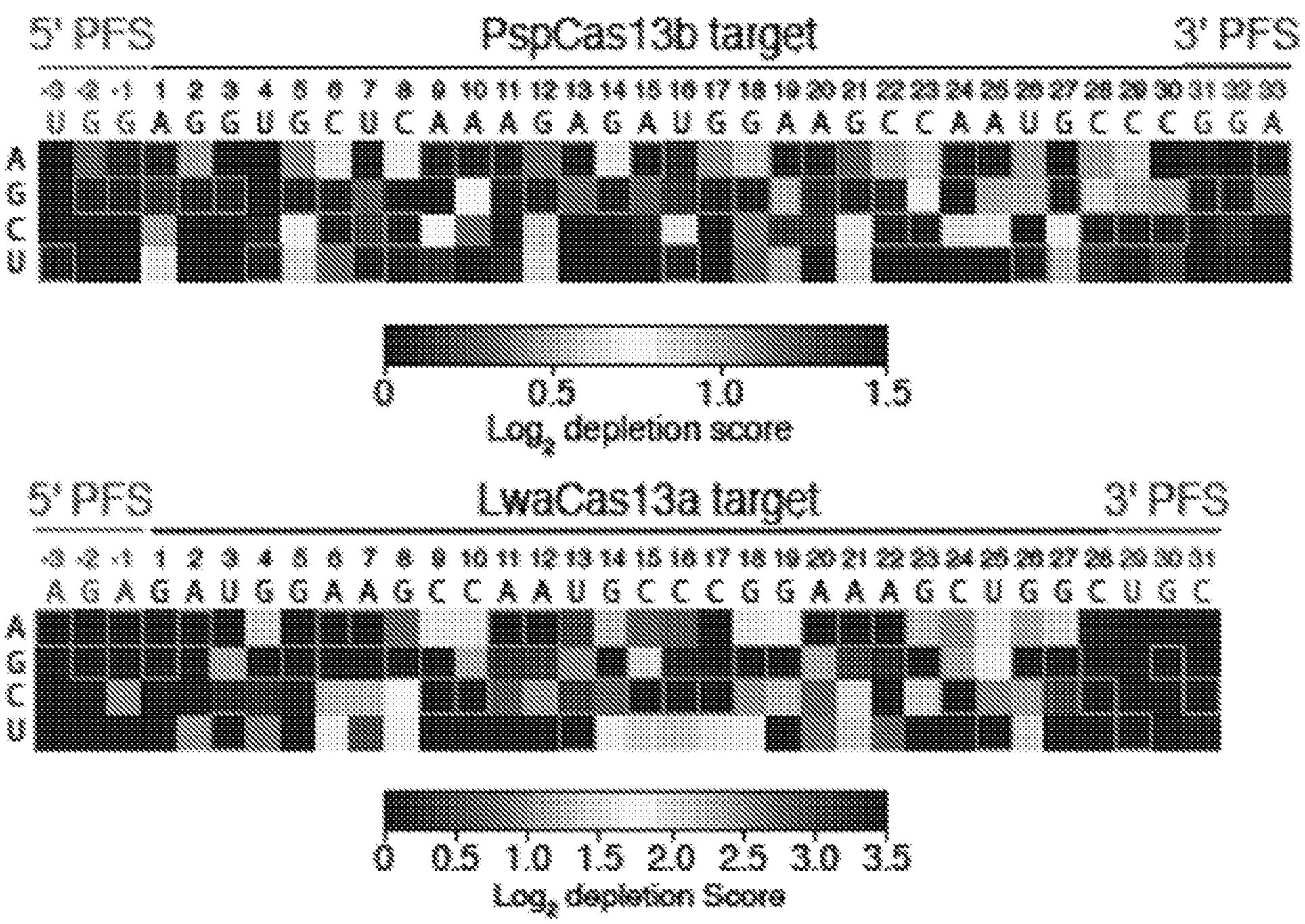
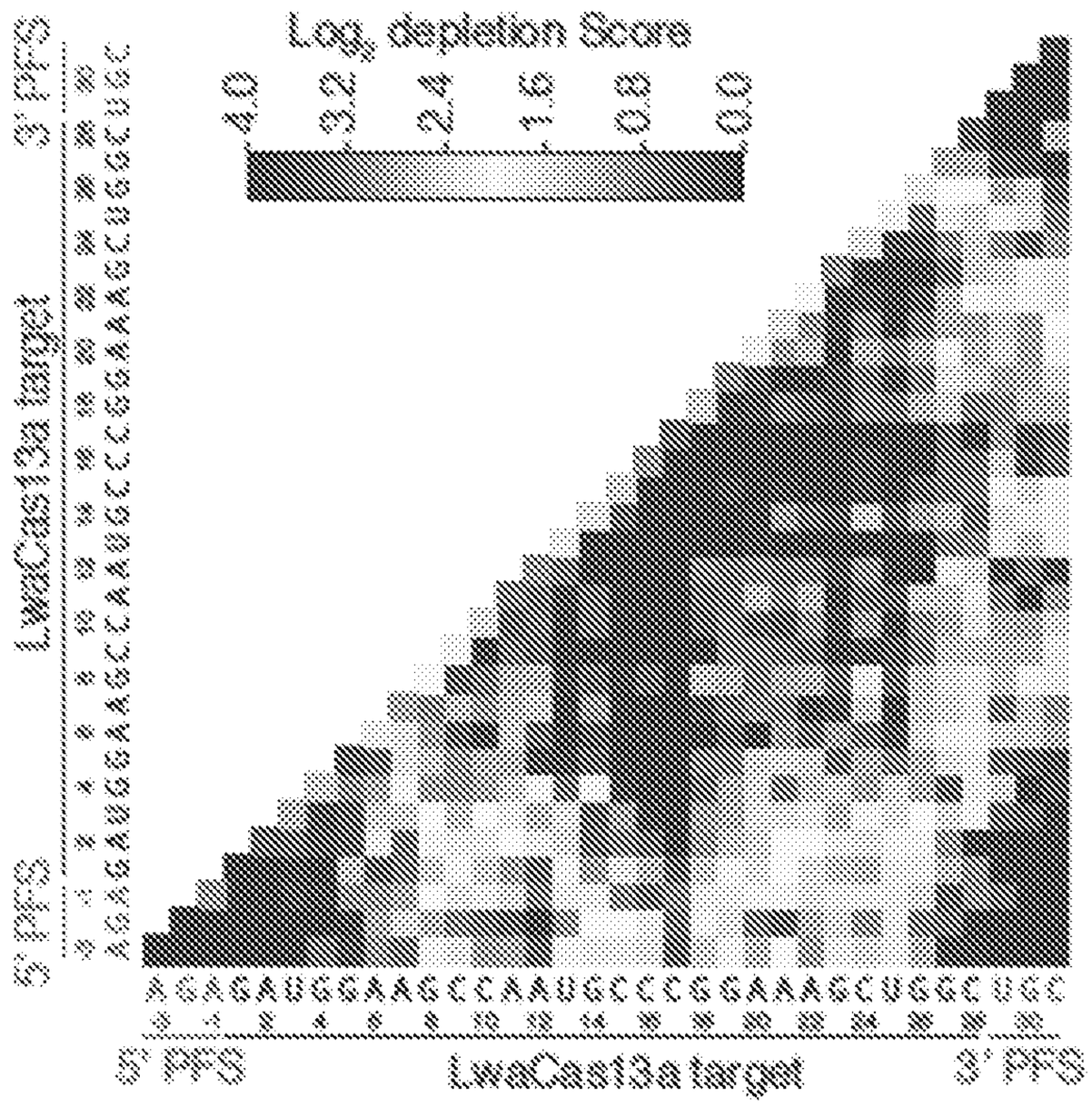


FIG. 56 (continued)

E


Log2 depletion Score
4.0
3.2
2.4
1.6
0.8
0.0
5' PFS
LwaCas13a target
3' PFS
5' PFS
LwaCas13a target
3' PFS
A G A G A U G G A A G C C A A U G C C C C G G A A A G C U G G C U G C
-8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10

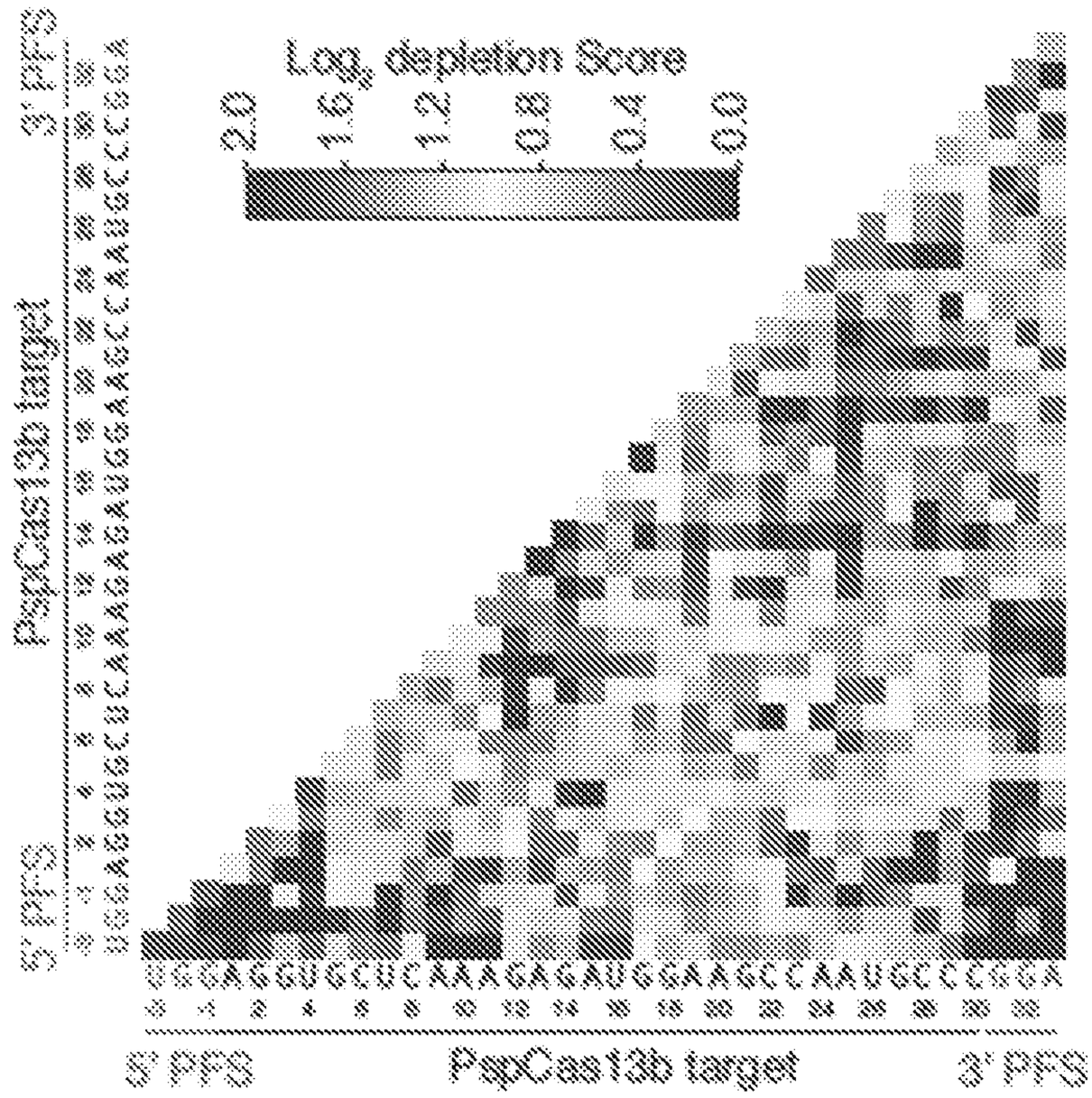

Log2 depletion Score
2.0
1.6
1.2
0.8
0.4
0.0
5' PFS
PspCas13b target
3' PFS
5' PFS
PspCas13b target
3' PFS
U G C G A G G U G C C U C A A A G A G A U G G A A G C C A A U G C C C C G G A
-8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10

FIG. 56 cont.

A

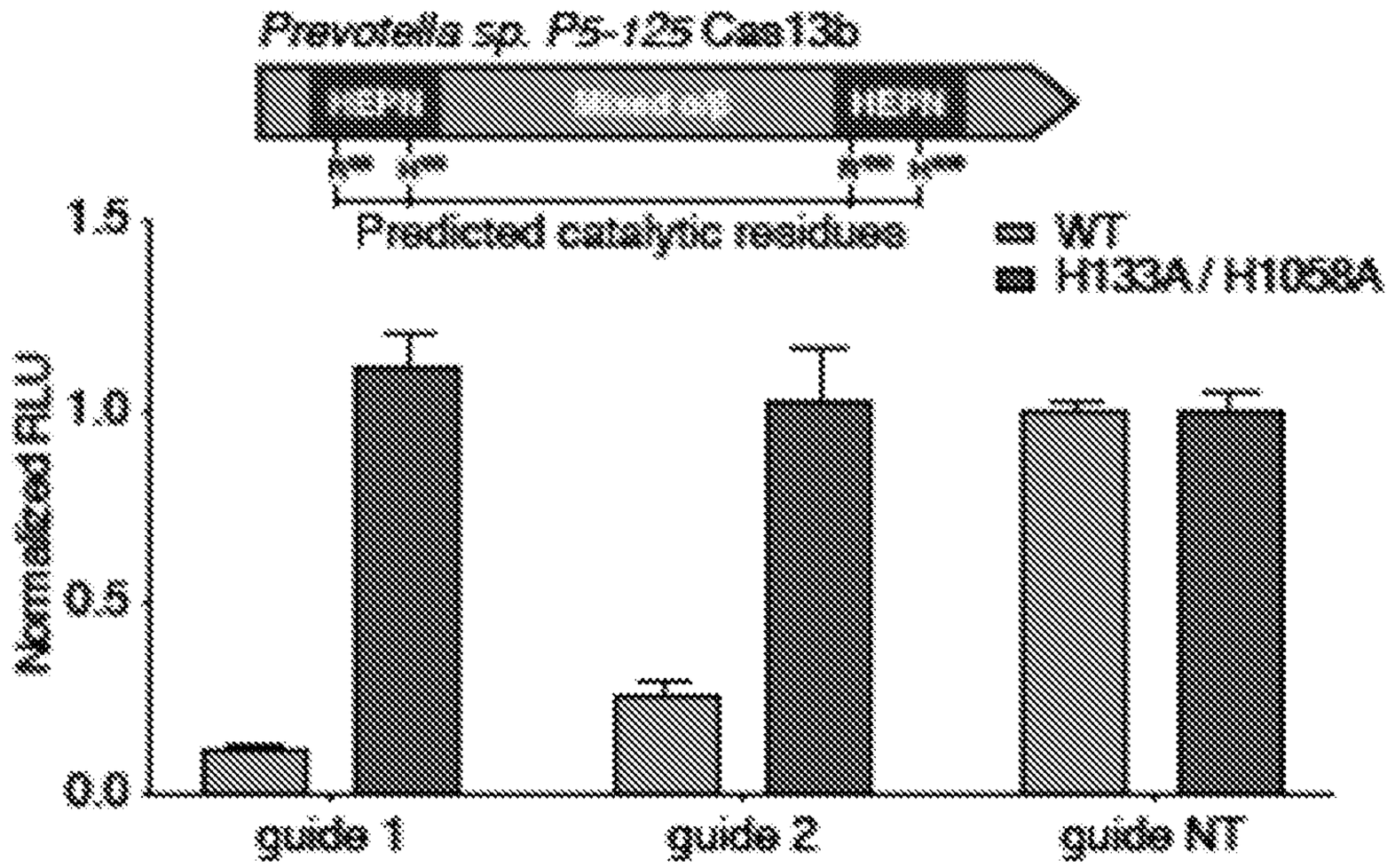

Prevotella sp. P5-125 Cas13b
NES
NES
Predicted catalytic residues
WT
H133A / H1058A
Normalized RLU
1.5
1.0
0.5
0.0
guide 1
guide 2
guide NT

FIG. 57

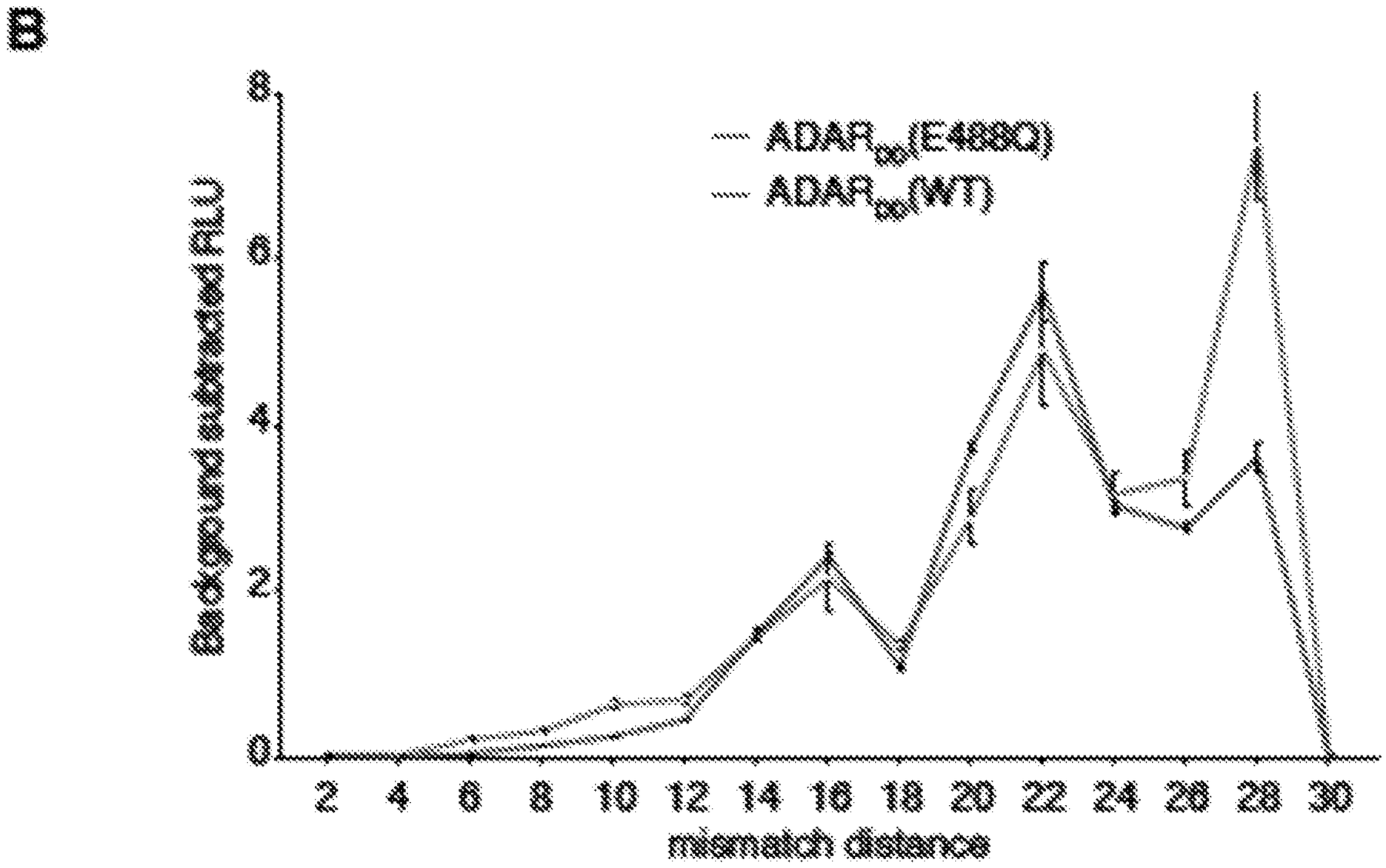


FIG. 57 (continued)

C

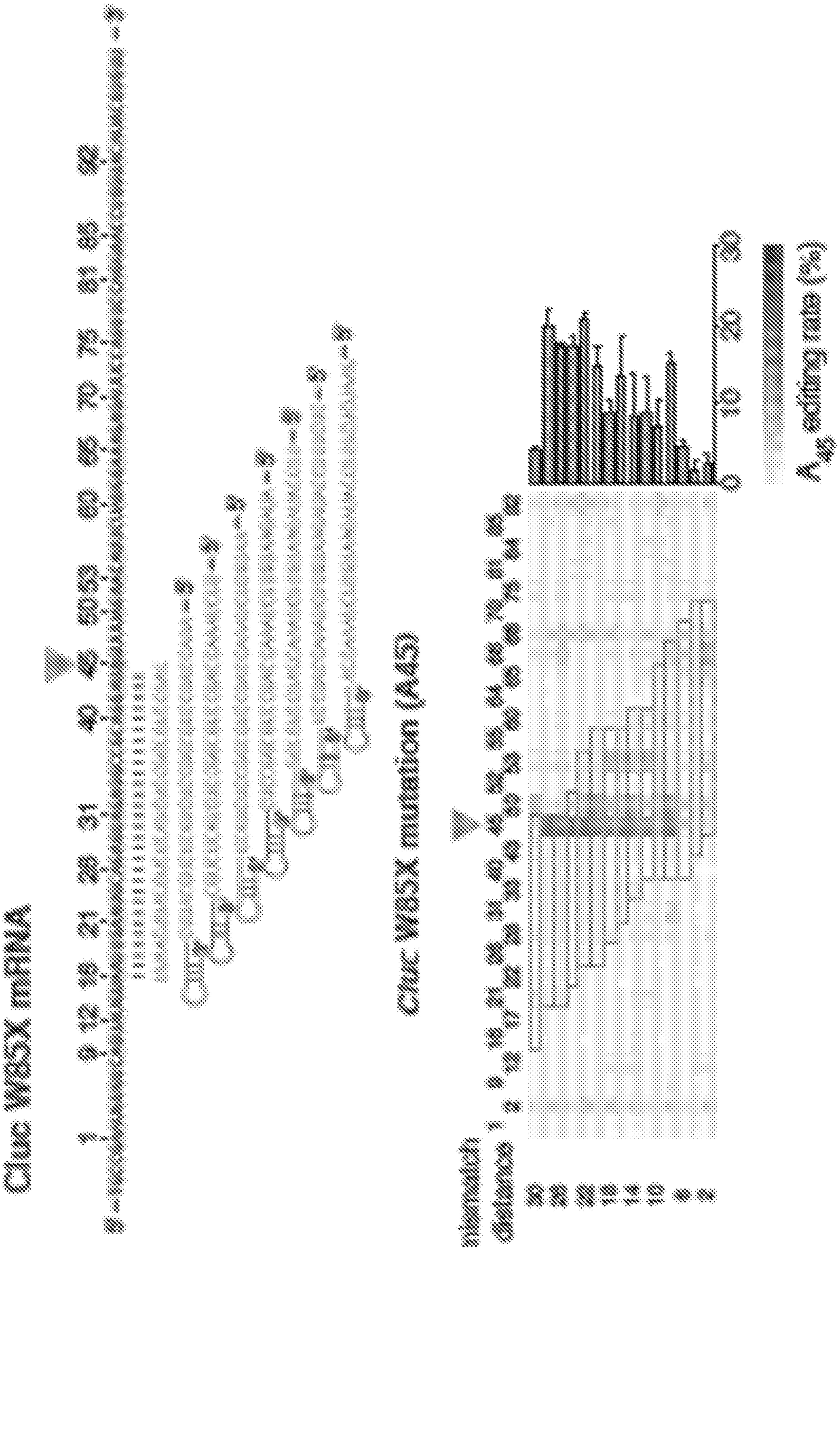


FIG. 57 (continued)

D

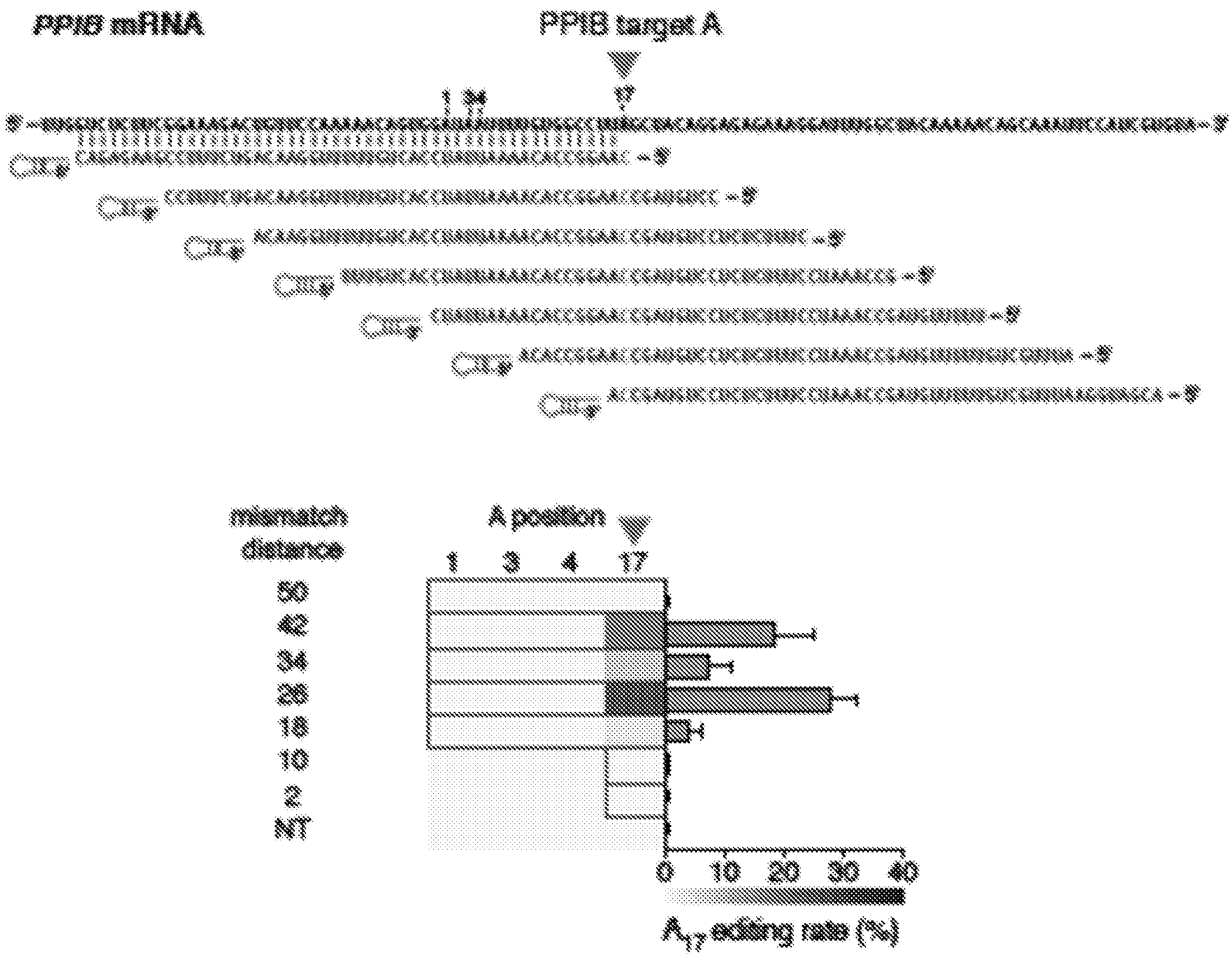


FIG. 57 (continued)

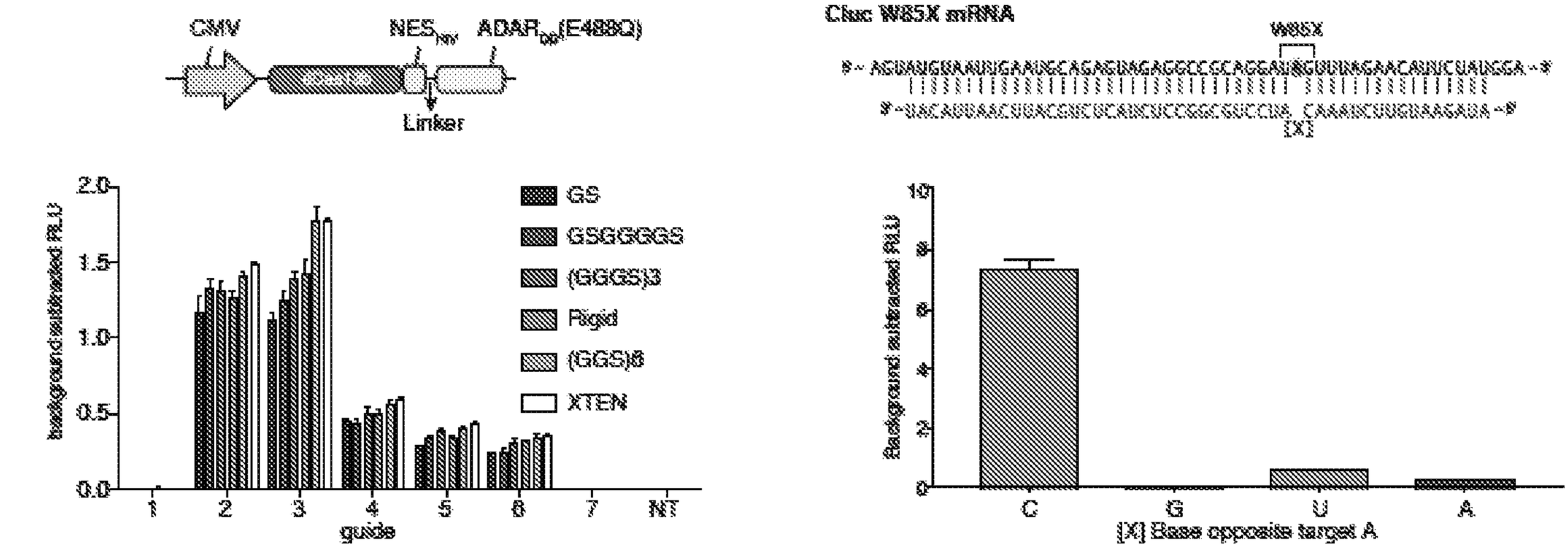


FIG. 57 (continued)

A

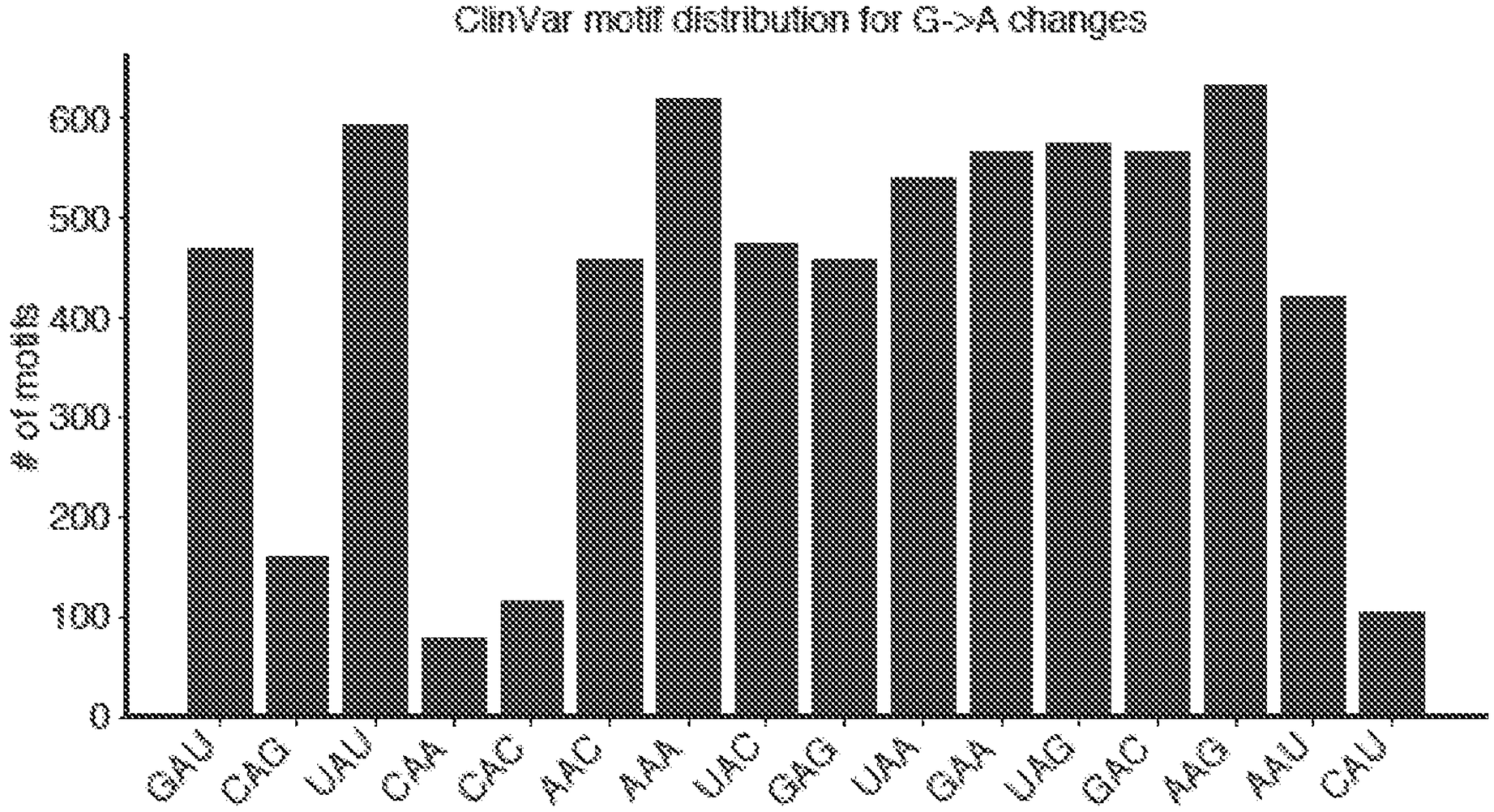

ClinVar motif distribution for G->A changes
# of motifs
0
100
200
300
400
500
600
GAU
CAG
UAU
CAA
CAC
AAC
AAA
UAC
GAG
UAA
GAA
UAG
GAC
AAG
AAU
CAU

FIG. 58

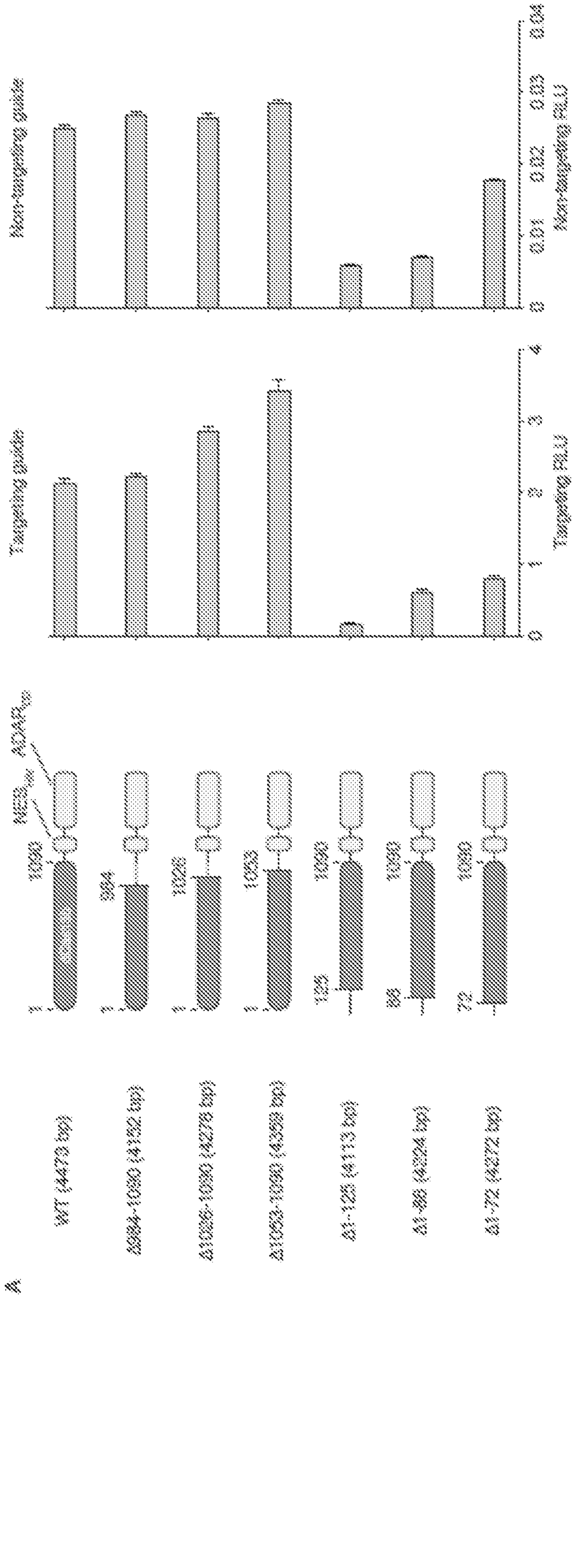

Non-targeting guide
Targeting guide
Indel Quantity (%)
0
0.01
0.02
0.03
0.04
Indel Quantity (%)
0
1
2
3
4
WT (4429 bp)
Δ1004-1030 (4152 bp)
Δ1026-1030 (4279 bp)
Δ1033-1060 (4359 bp)
Δ1-1125 (4113 bp)
Δ1-800 (4324 bp)
Δ1-772 (4272 bp)
1030
1004
1026
1033
1060
1125
800
772
NES
ADAR

FIG. 59

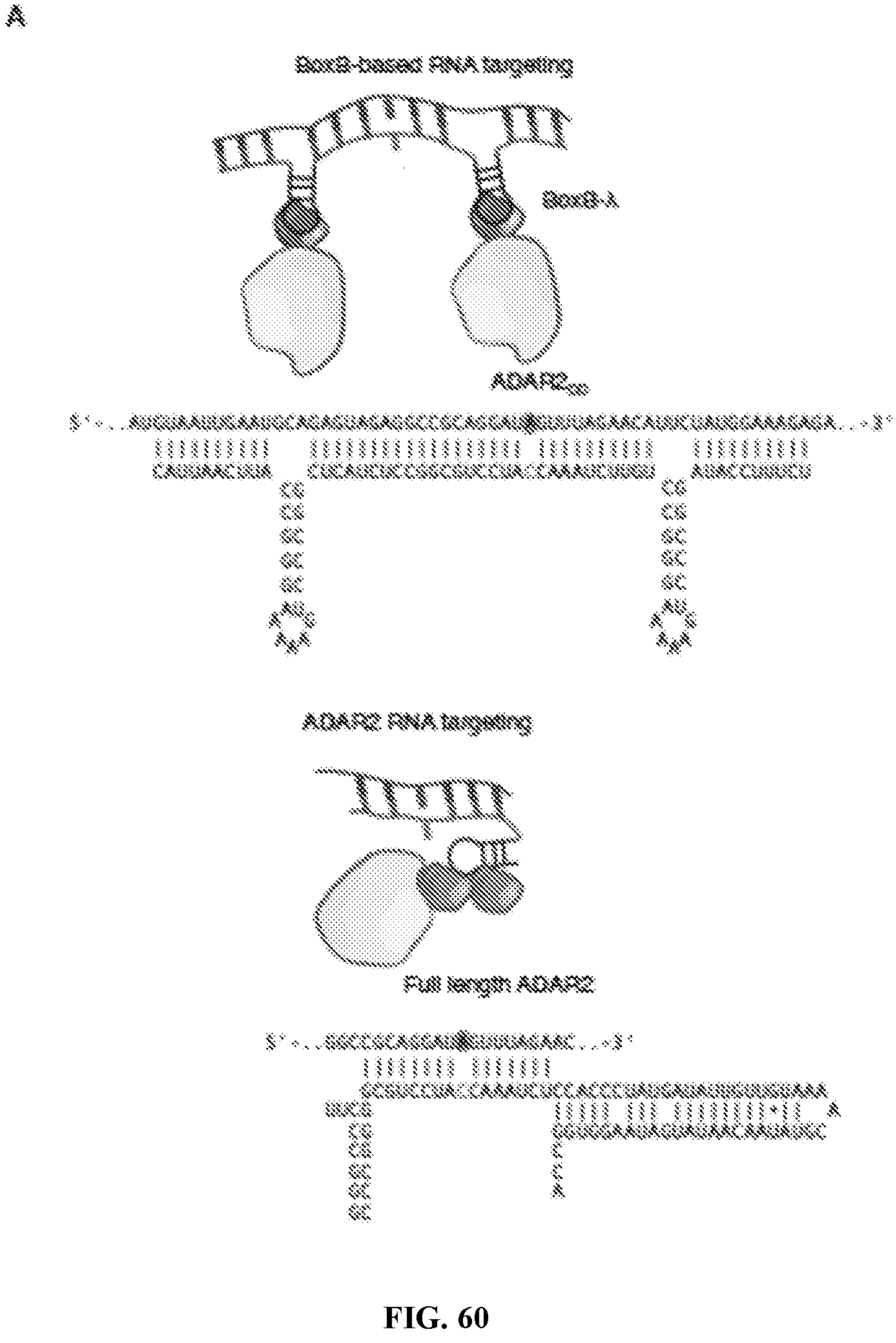

A
Bcnib-based RNA targeting
Bcnib-A
ADAR2dd
ADAR2 RNA targeting
Full length ADAR2

FIG. 60

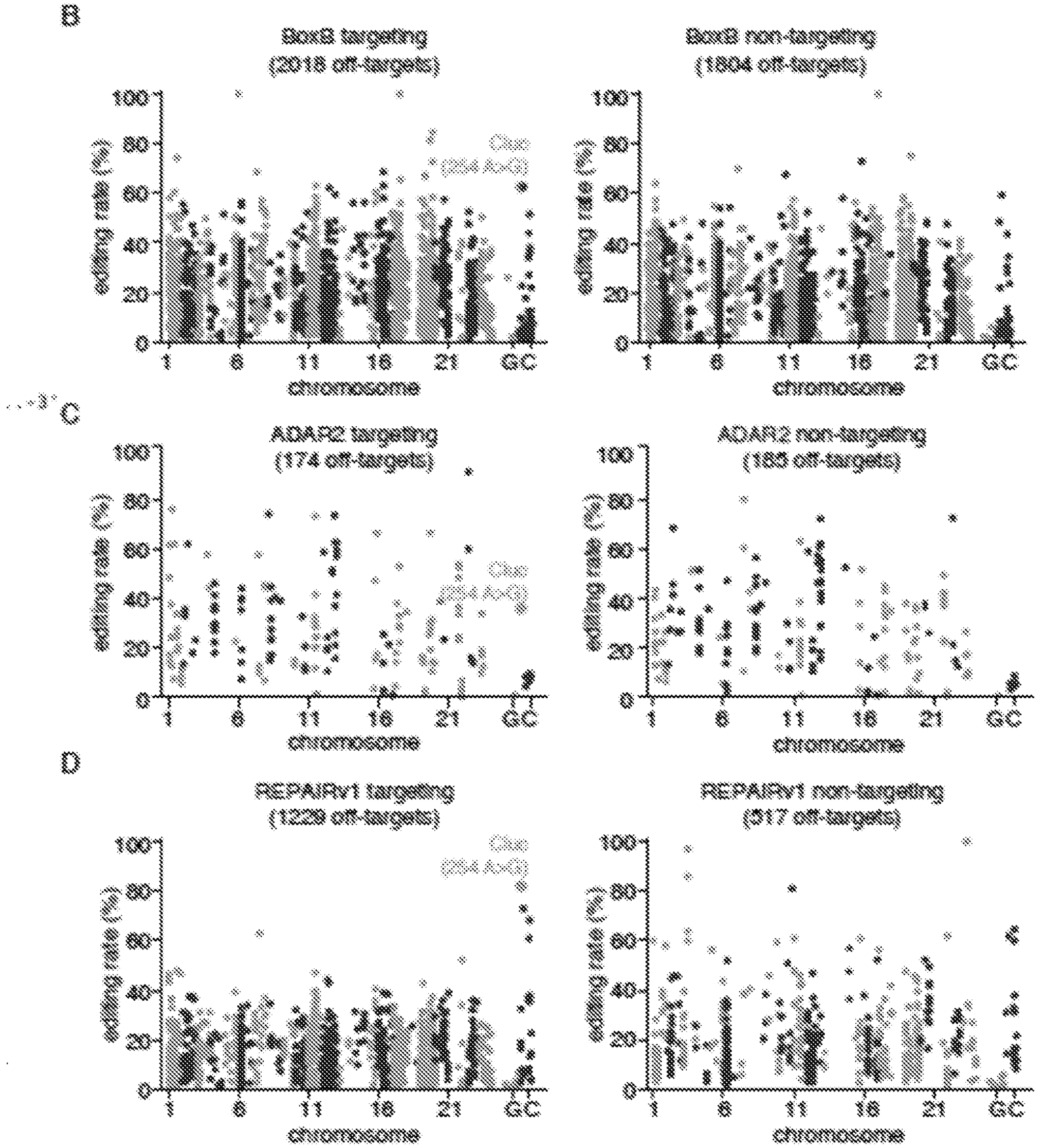


FIG. 60 (continued)

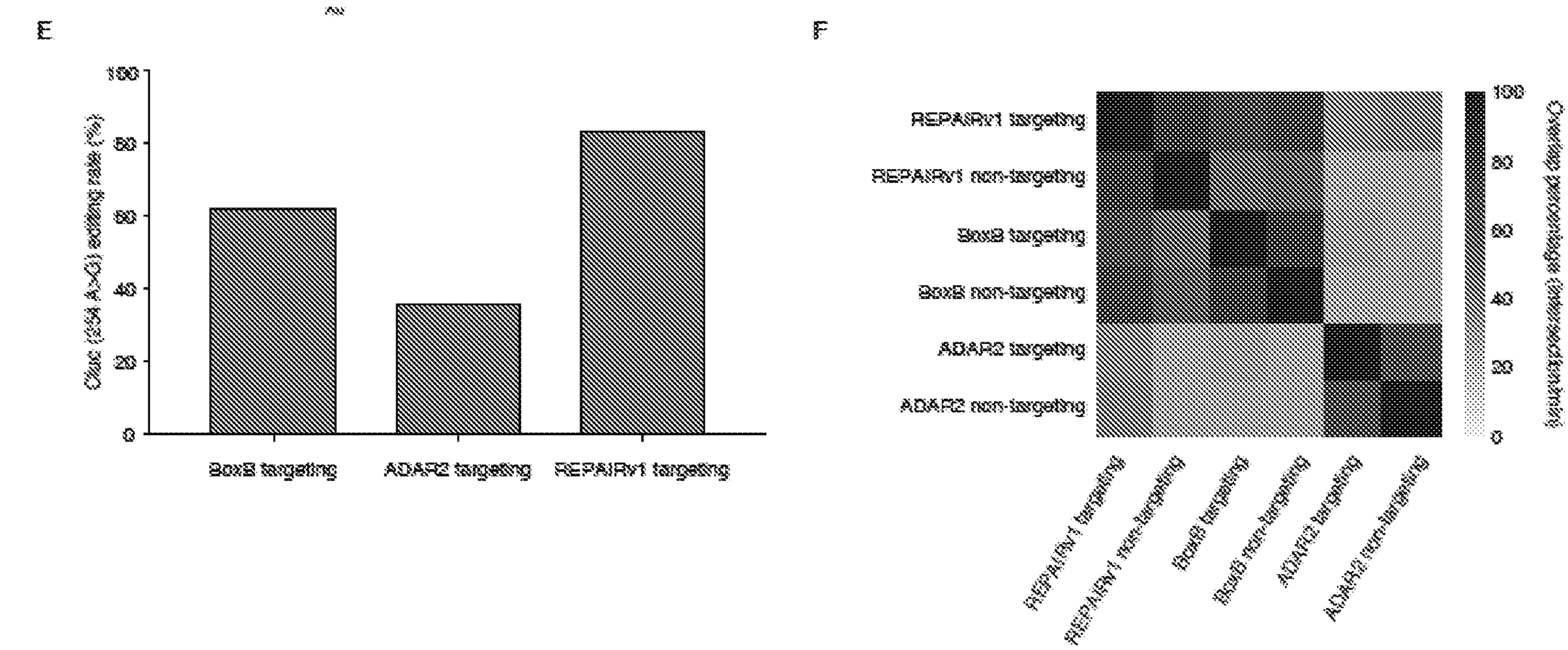


FIG. 60 (continued)

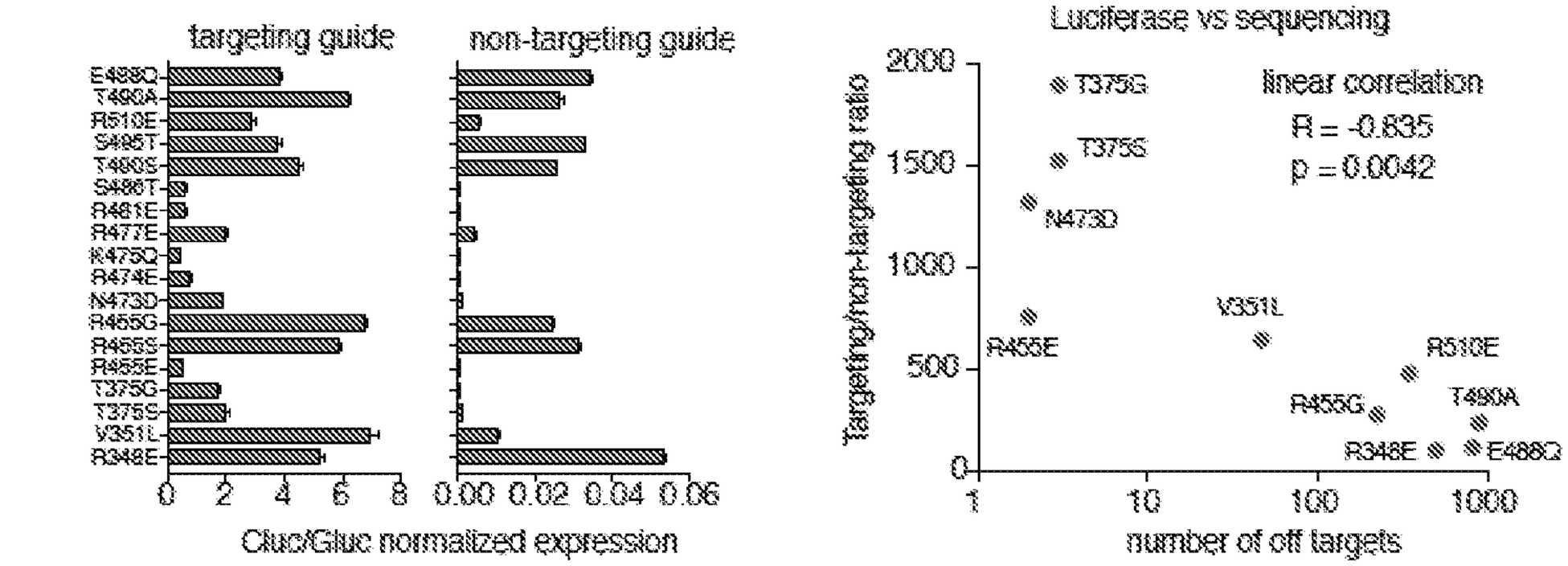

A
targeting guide
non-targeting guide
E488Q
T490A
R510E
S485T
T490S
S486T
R461E
R477E
N475D
R474E
N473D
R455G
R455S
R455E
T375G
T375S
V351L
R348E
0 2 4 6 8
0.00 0.02 0.04 0.06
Cluc/Gluc normalized expression
B
Luciferase vs sequencing
linear correlation
R = -0.835
p = 0.0042
Targeting/non-targeting ratio
2000
1500
1000
500
0
1 10 100 1000
number of off targets
C
number of off targets
1
10
100
1000
10000
0 20 40 60 80 100
Cluc (254 A>G) editing rate (%)

FIG. 61

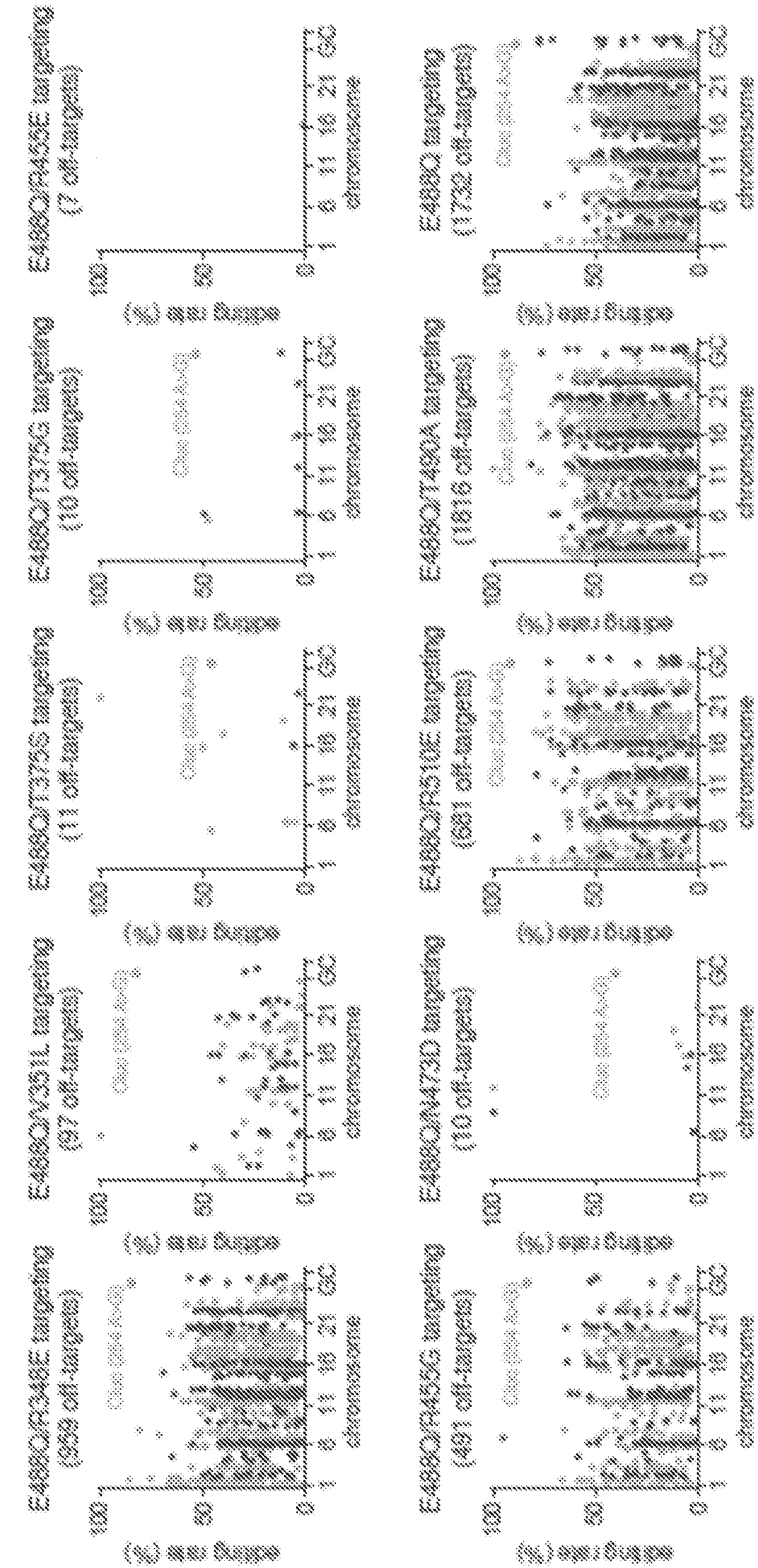

CA9/CAIX expression
(7 cell lines)
CA9/CAIX expression
(10 cell lines)
CA9/CAIX expression
(11 cell lines)
CA9/CAIX expression
(16/16 cell lines)
CA9/CAIX expression
(6/11 cell lines)
CA9/CAIX expression
(8/8 cell lines)
CA9/CAIX expression
(3/7 cell lines)
CA9/CAIX expression
(4/9 cell lines)
(%) specific lysis
Effector:Target ratio

FIG. 62

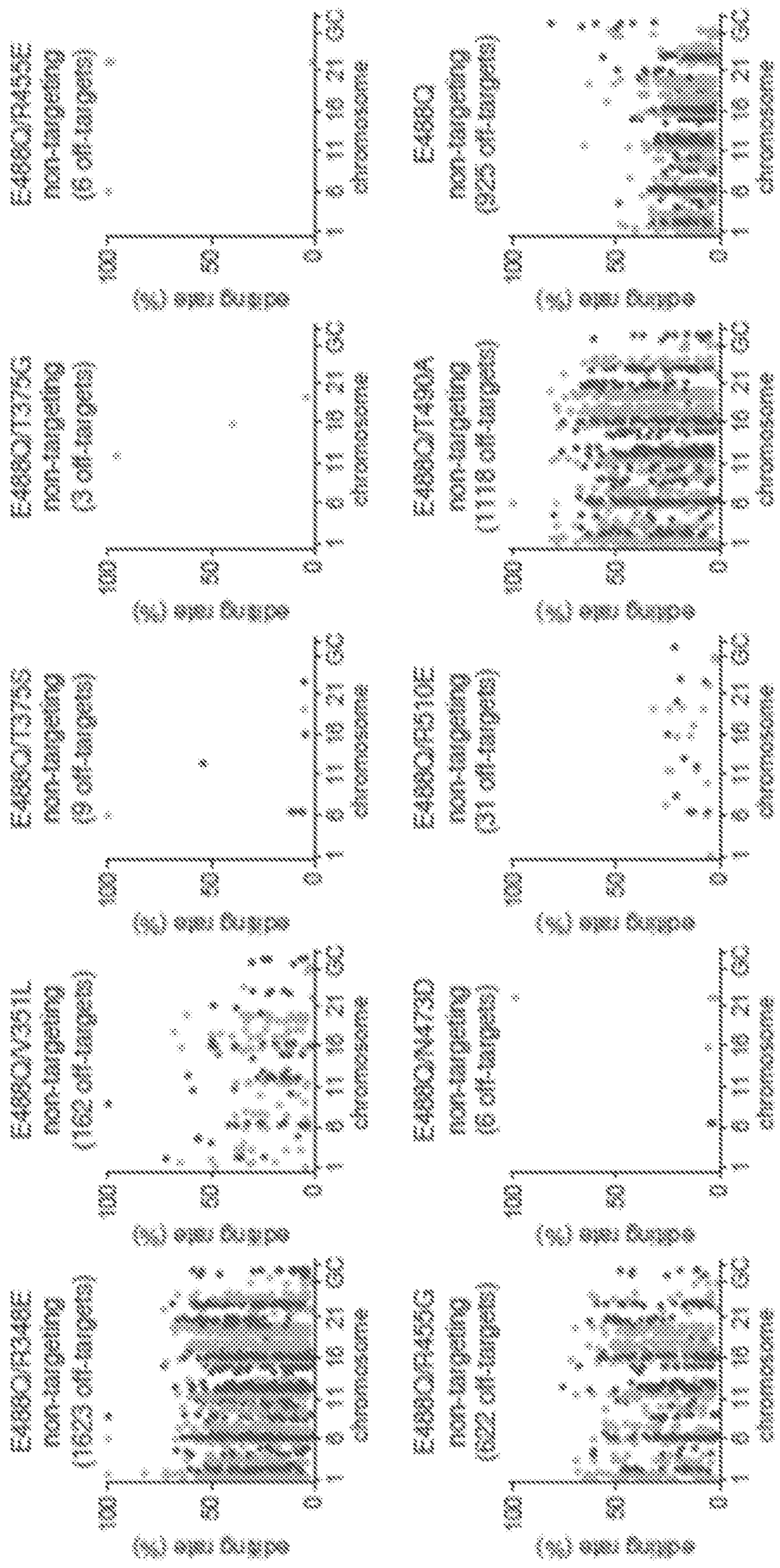

Cancer/Stroma non-tumor (6 cm-tumor)
Cancer/Stroma non-tumor (2 cm-tumor)
Cancer/Stroma non-tumor (11 cm-tumor)
Cancer/Stroma non-tumor (31 cm-tumor)
Cancer/Stroma non-tumor (16 cm-tumor)
Cancer/Stroma non-tumor (22 cm-tumor)
(%) ssa Bugs
chromosome

FIG. 62 cont.

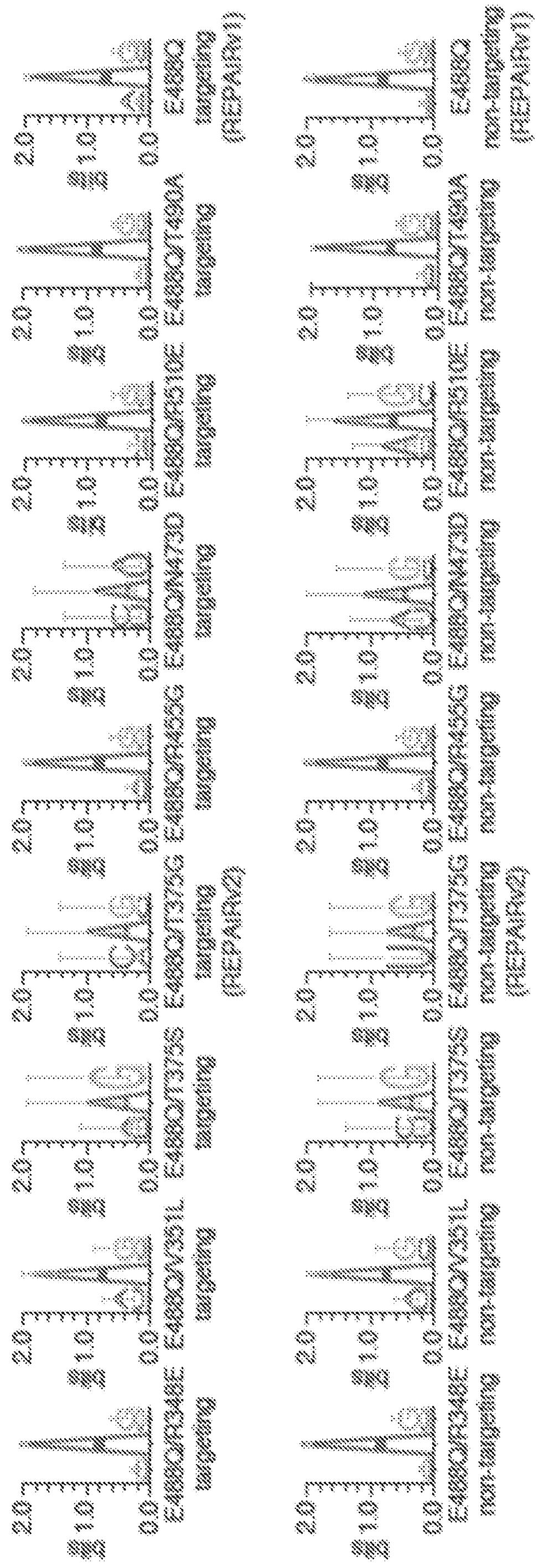
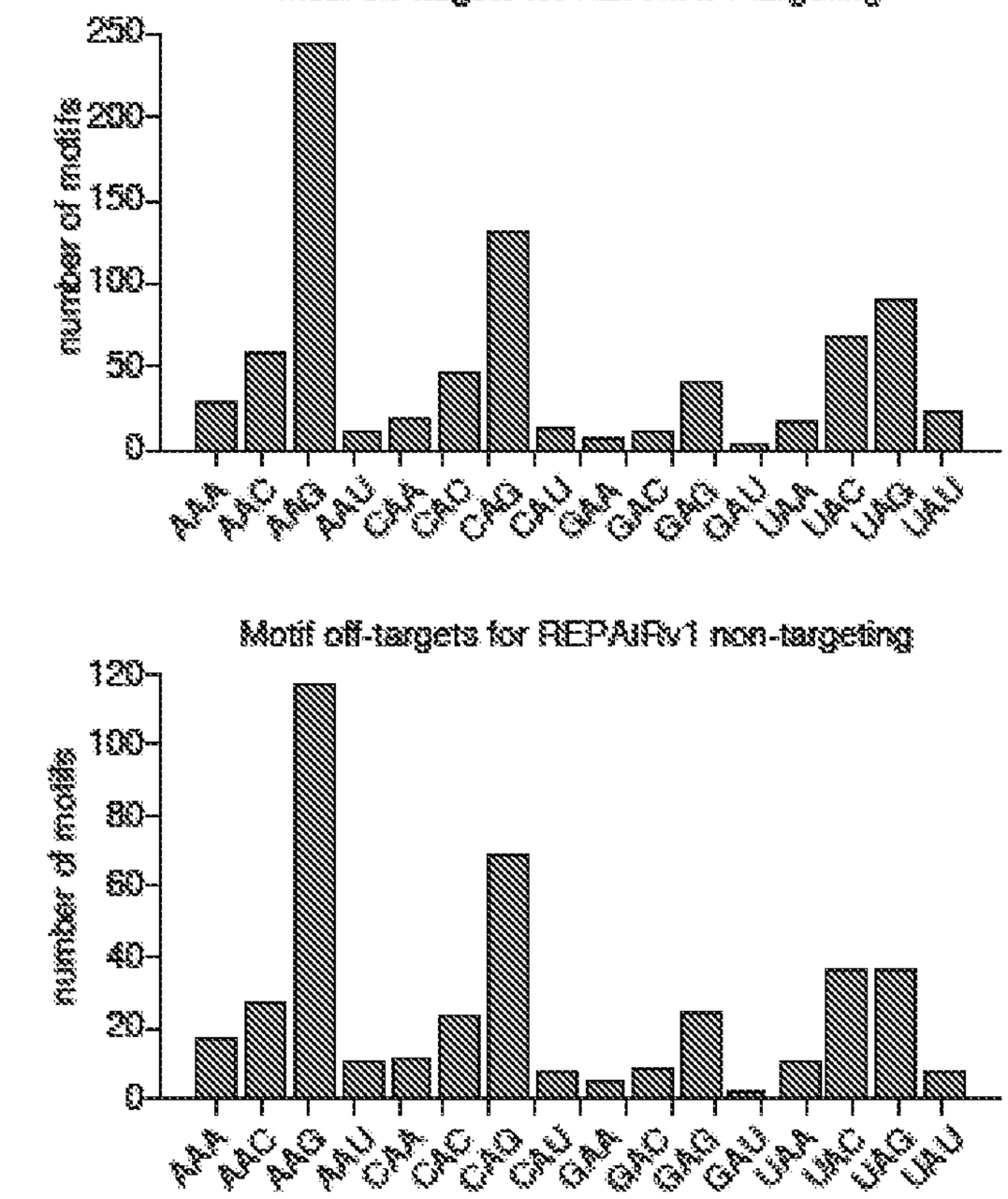

Ratio
2.0
1.5
1.0
0.5
0.0
CAMK2N1
CAMK2N2
CAMK2G
CAMK2D
CAMK2A
non-targeting
targeting
(KEPA1/KEPA2)

FIG. 63



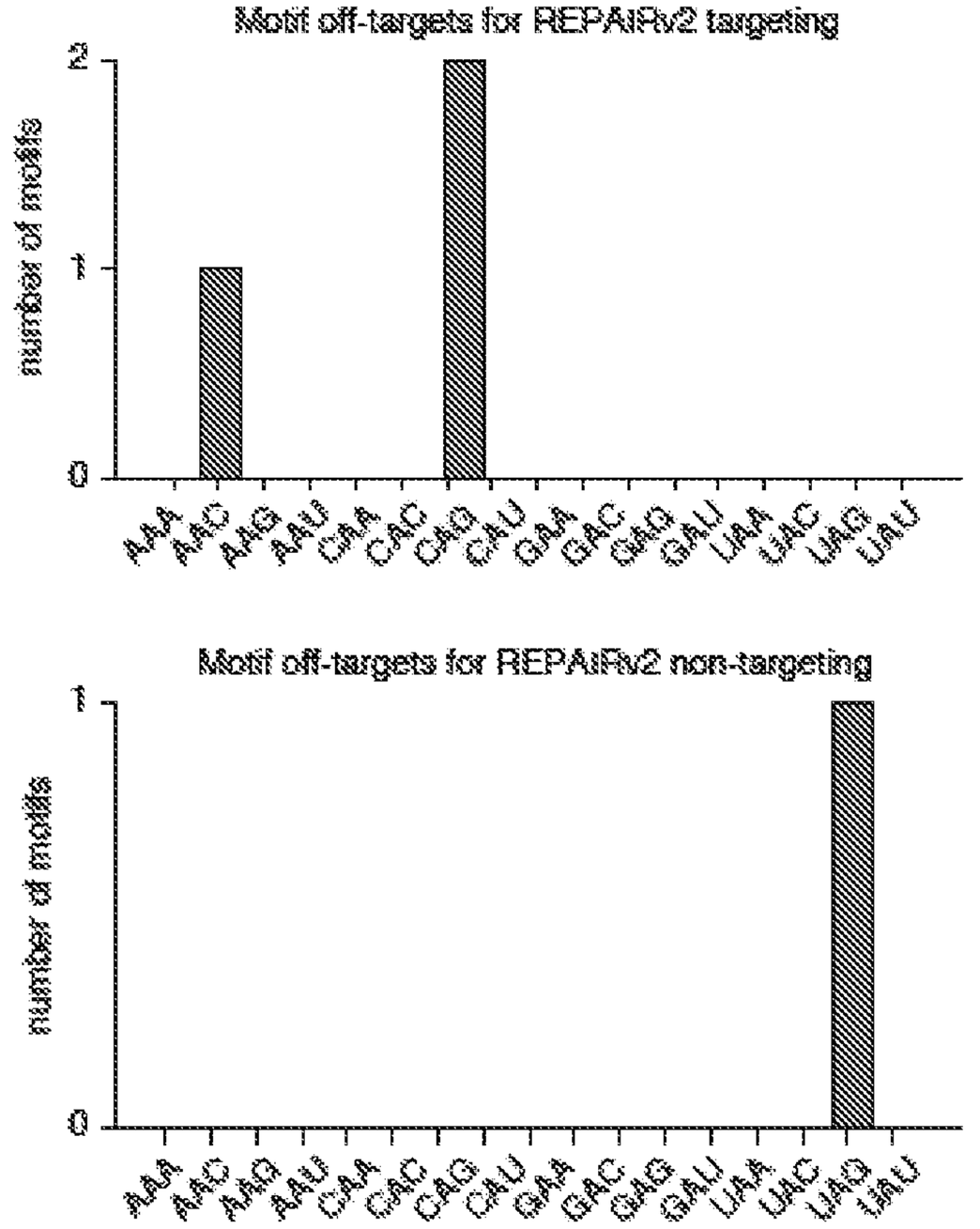


FIG. 63 (continued)

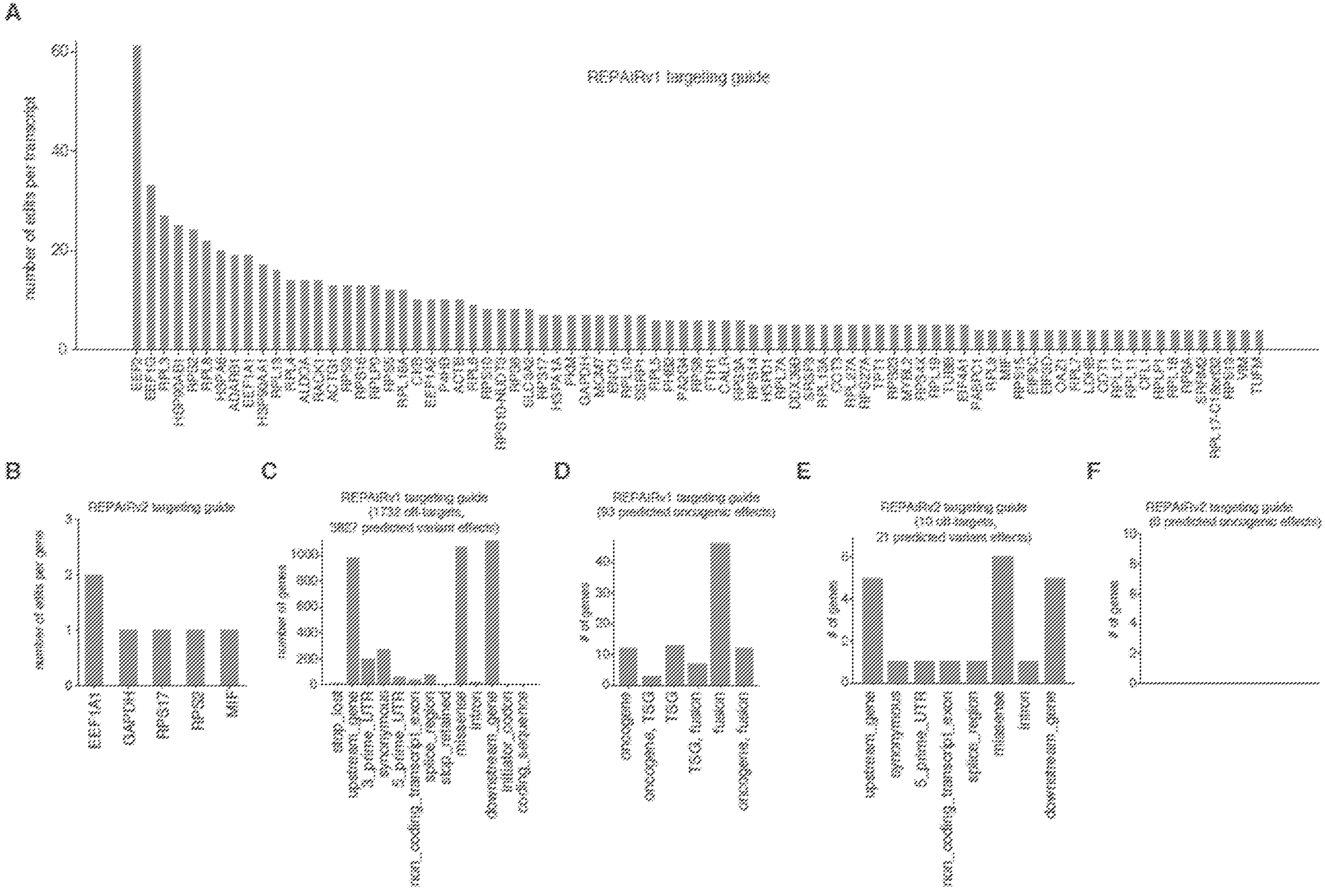

A
PEPAIRv1 targeting guide
number of edits per sequence
B
PEPAIRv2 targeting guide
C
PEPAIRv1 targeting guide
(1732 off-targets,
3867 predicted variant effects)
D
PEPAIRv1 targeting guide
(93 predicted oncogenic effects)
E
PEPAIRv2 targeting guide
(10 off-targets,
21 predicted variant effects)
F
PEPAIRv2 targeting guide
(0 predicted oncogenic effects)

FIG. 64

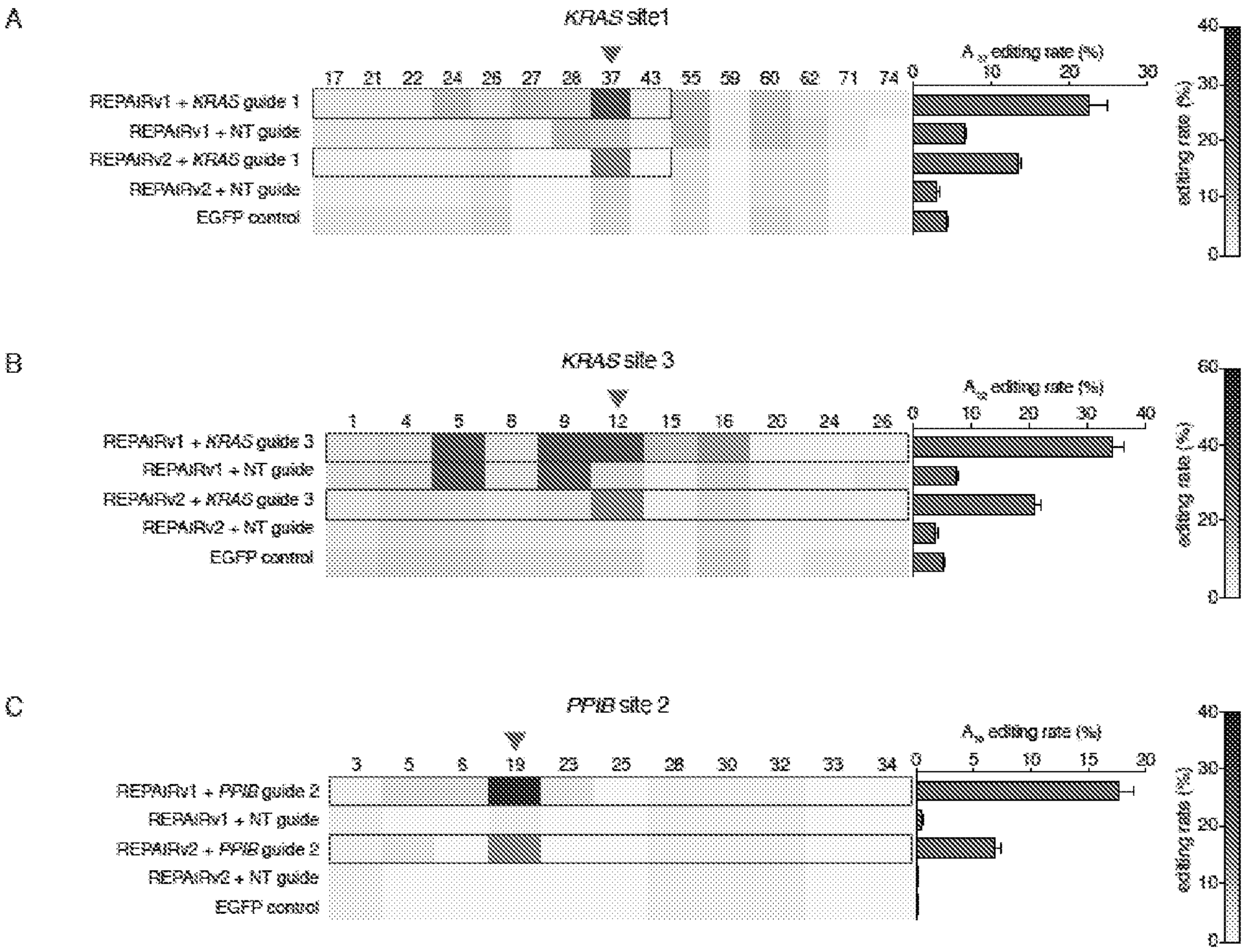

A
KRAS site1
REPAIRv1 + KRAS guide 1
REPAIRv1 + NT guide
REPAIRv2 + KRAS guide 1
REPAIRv2 + NT guide
EGFP control
17 21 22 24 26 27 28 37 43 55 59 60 62 71 74
A to G editing rate (%)
editing rate (%)
B
KRAS site 3
REPAIRv1 + KRAS guide 3
REPAIRv1 + NT guide
REPAIRv2 + KRAS guide 3
REPAIRv2 + NT guide
EGFP control
1 4 5 8 9 12 15 16 20 24 26
A to G editing rate (%)
editing rate (%)
C
PPIB site 2
REPAIRv1 + PPIB guide 2
REPAIRv1 + NT guide
REPAIRv2 + PPIB guide 2
REPAIRv2 + NT guide
EGFP control
3 5 6 19 23 25 28 30 32 33 34
A to G editing rate (%)
editing rate (%)

FIG. 65

A

Initial Codon		Edited Codon
Tyr	→	Cys
stop	→	Trp
His	→	Arg
Gln	→	Arg
Asn	→	Ser / Asp
Lys	→	Arg / Glu
Ser	→	Gly
Arg	→	Gly
Thr	→	Ala
Ile	→	Val
Met	→	Val

B

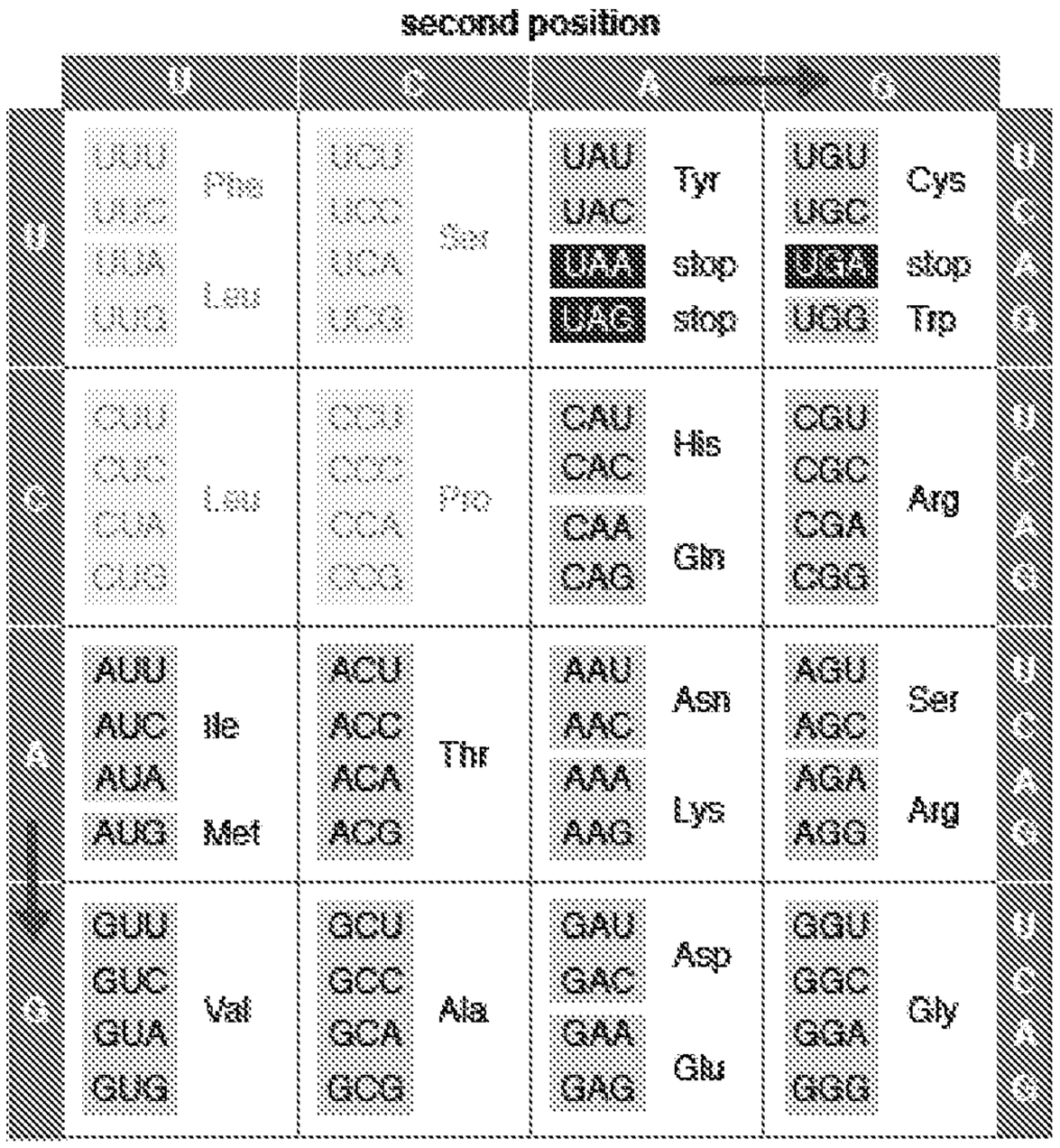


FIG. 66

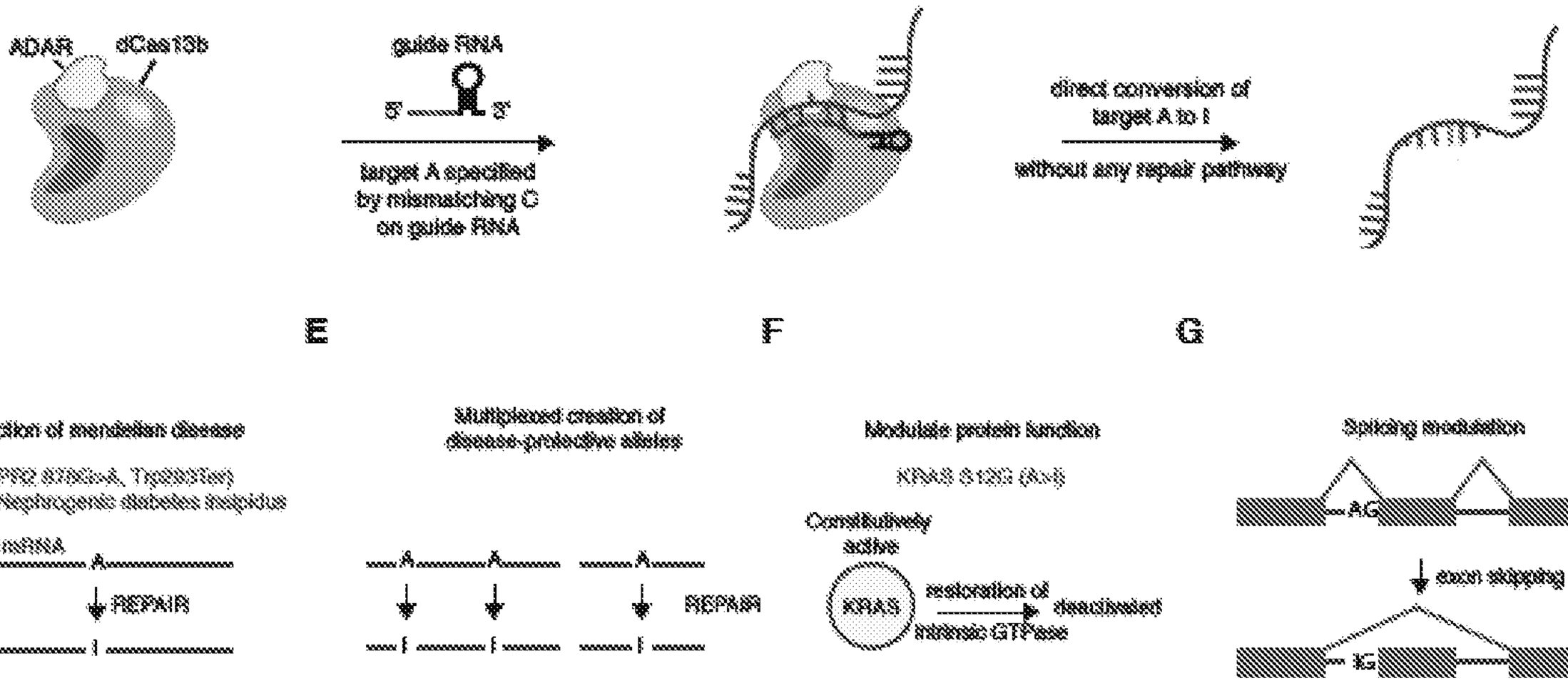


FIG. 66
(continued)

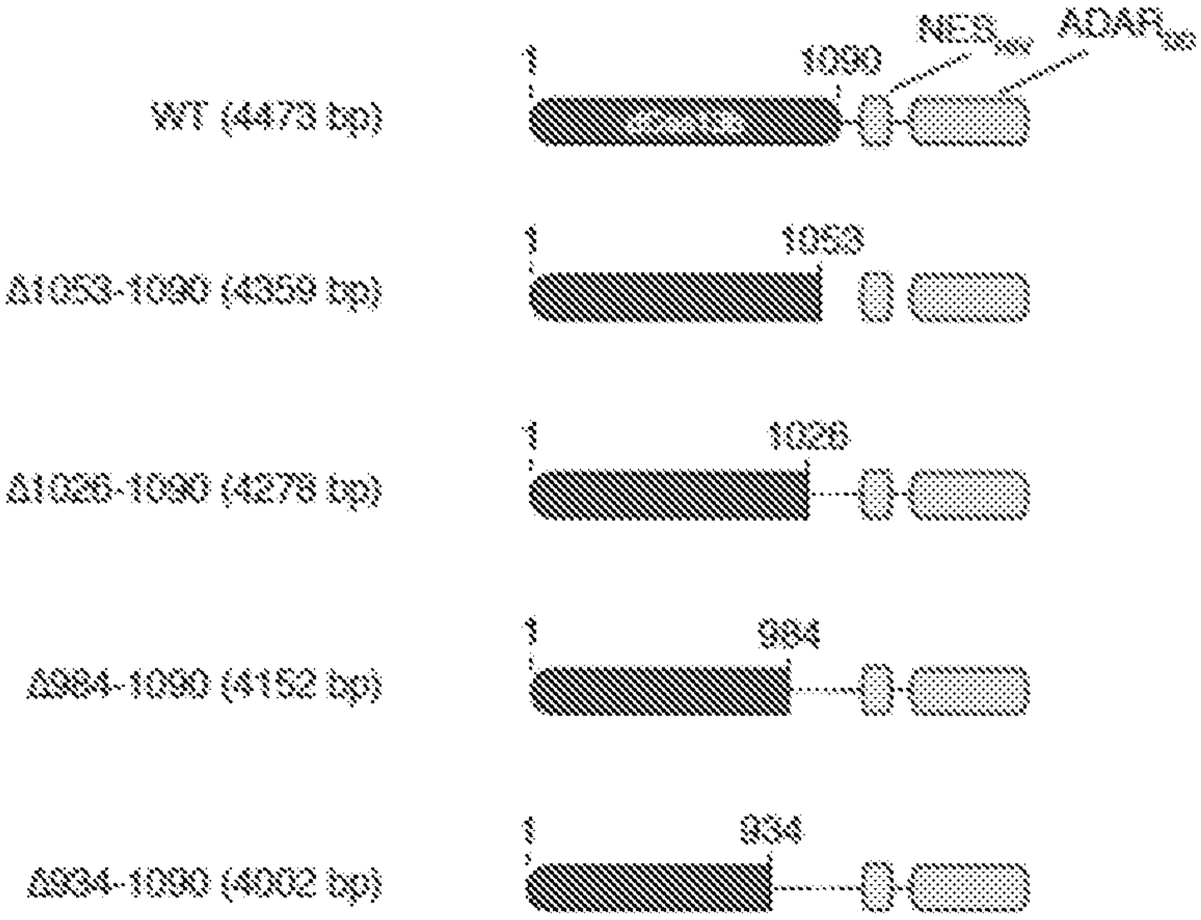
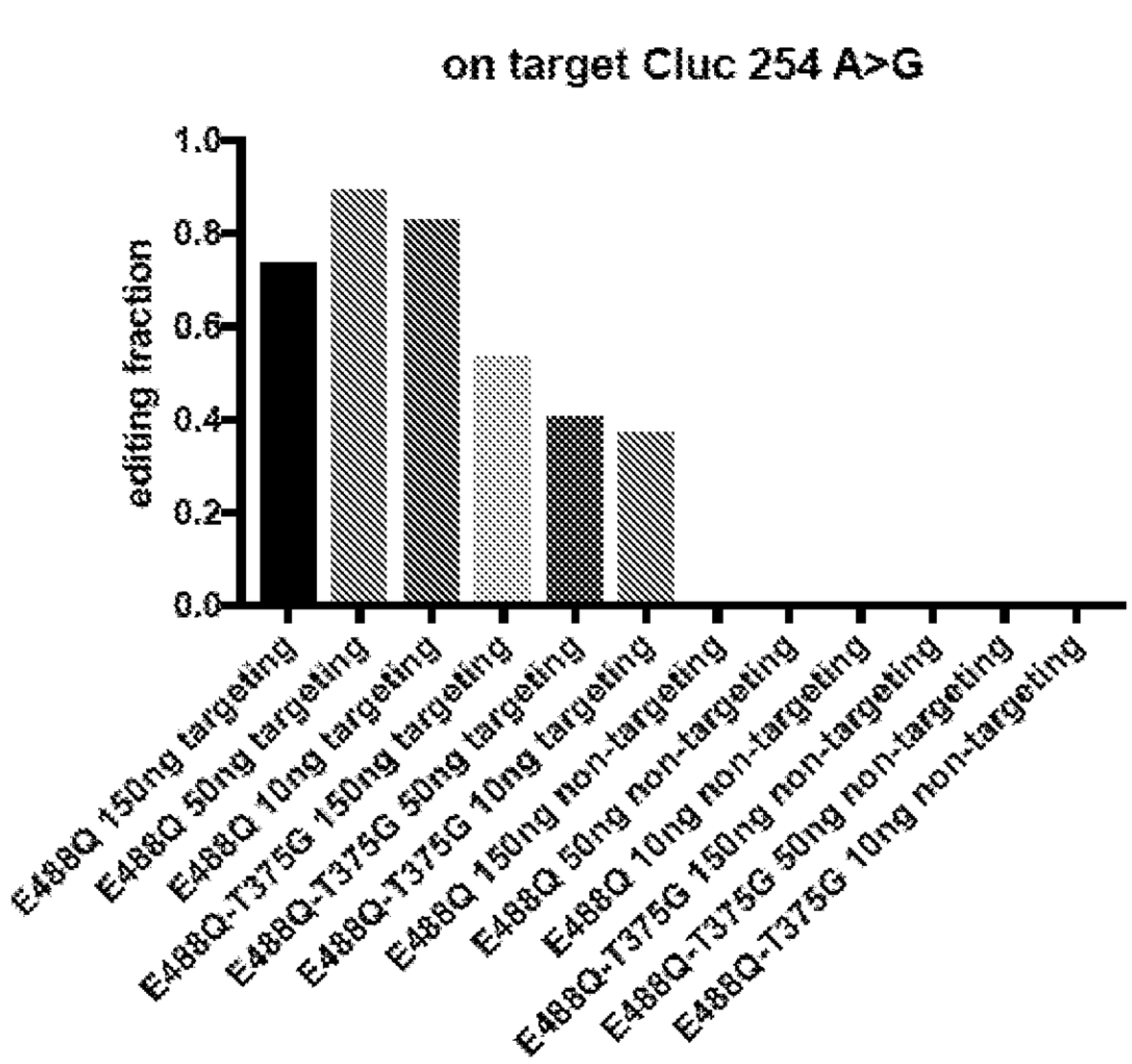

WT (4473 bp)
1090
NES
ADAR
Δ1053-1090 (4359 bp)
1053
Δ1026-1090 (4278 bp)
1026
Δ984-1090 (4152 bp)
984
Δ934-1090 (4002 bp)
934
Δ884-1090 (3852 bp)
884
Δ834-1090 (3702 bp)
834
Δ784-1090 (3552 bp)
784
Δ734-1090 (3402 bp)
734

FIG. 67


on target Cluc 254 A>G
editing fraction
1.0
0.8
0.6
0.4
0.2
0.0
E488Q 150ng targeting
E488Q 50ng targeting
E488Q 10ng targeting
E488Q-T375G 150ng targeting
E488Q-T375G 50ng targeting
E488Q-T375G 10ng targeting
E488Q 150ng non-targeting
E488Q 50ng non-targeting
E488Q 10ng non-targeting
E488Q-T375G 150ng non-targeting
E488Q-T375G 50ng non-targeting
E488Q-T375G 10ng non-targeting

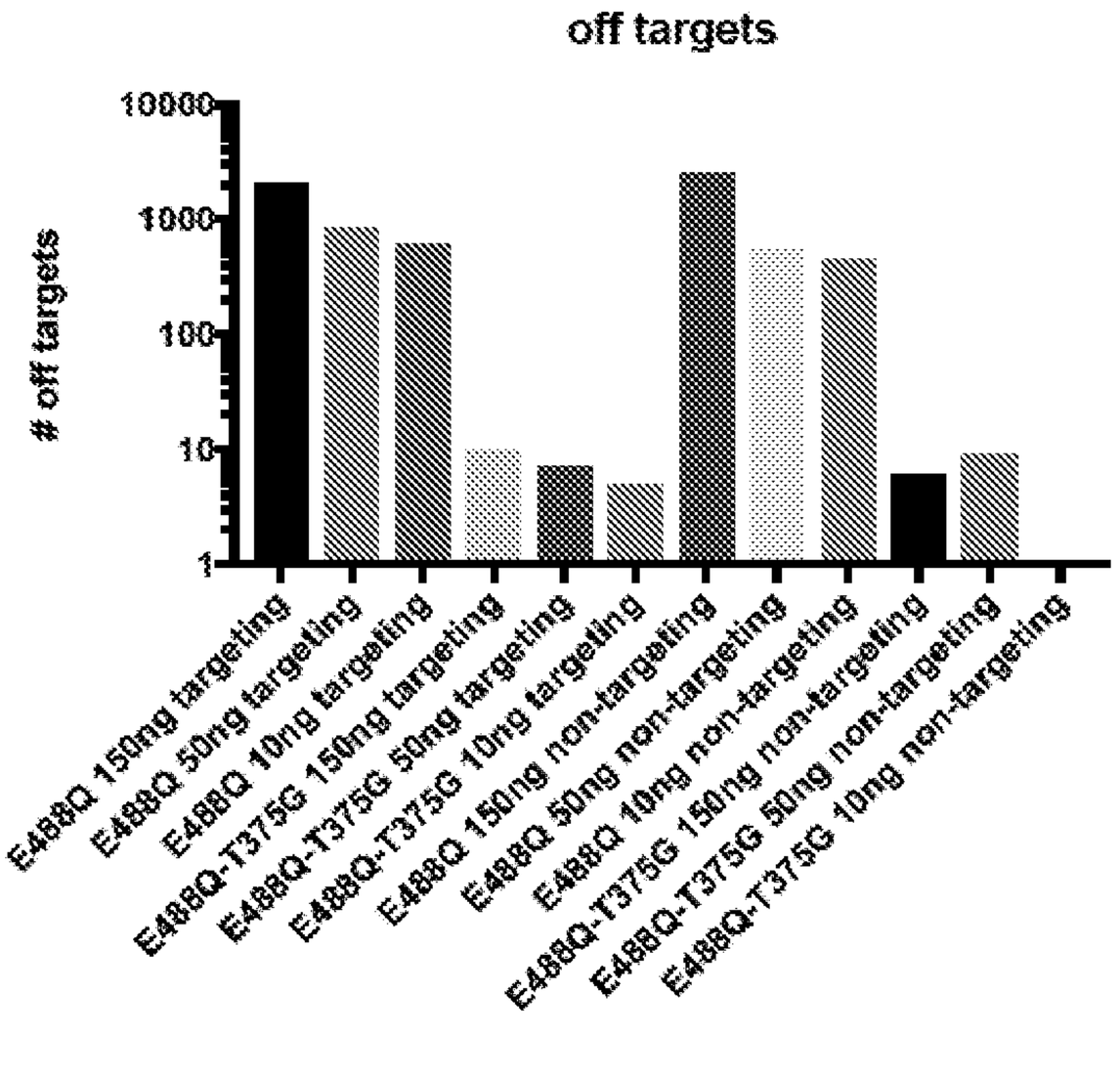

off targets
off targets
10000
1000
100
10
1
E488Q 150ng targeting
E488Q 50ng targeting
E488Q 10ng targeting
E488Q-T375G 150ng targeting
E488Q-T375G 50ng targeting
E488Q-T375G 10ng targeting
E488Q 150ng non-targeting
E488Q 50ng non-targeting
E488Q 10ng non-targeting
E488Q-T375G 150ng non-targeting
E488Q-T375G 50ng non-targeting
E488Q-T375G 10ng non-targeting

FIG. 68

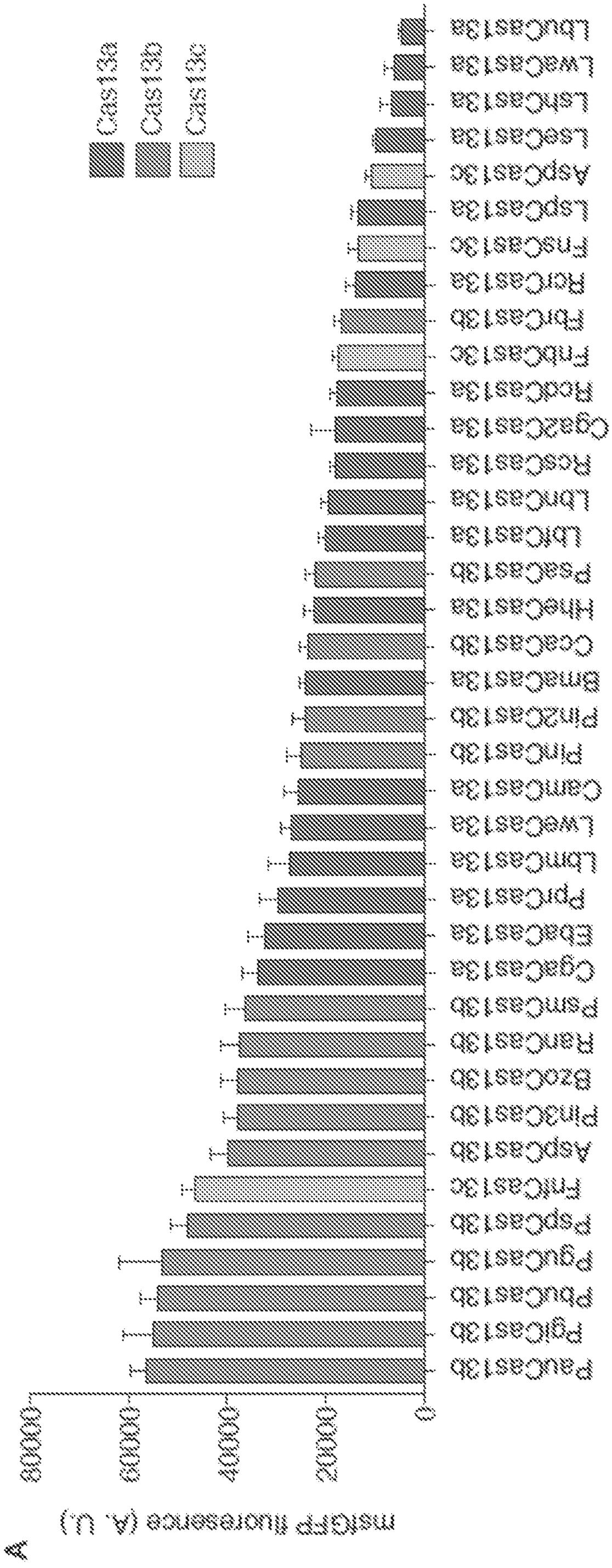

Cas13a
Cas13b
Cas13c
LbuCas13a
LwaCas13a
LshCas13a
LseCas13a
AspCas13c
LspCas13a
FnsCas13c
RcrCas13a
FbrCas13b
FnbCas13c
RcdCas13a
Cga2Cas13a
RcsCas13a
LbnCas13a
LbiCas13a
PsaCas13b
HheCas13a
CcaCas13b
BmaCas13a
Pin2Cas13b
EbaCas13a
PprCas13a
LbmCas13a
LweCas13a
CamCas13a
PinCas13b
FnlCas13c
AspCas13b
Pin3Cas13b
BzoCas13b
RanCas13b
PsmCas13b
CgaCas13a
EbaCas13a
PprCas13a
LbmCas13a
LweCas13a
CamCas13a
PinCas13b
FnlCas13c
PspCas13b
PguCas13b
PbuCas13b
PgiCas13b
PauCas13b
0
20000
40000
60000
80000
msfGFP fluorescence (A. u.)

FIG. 69

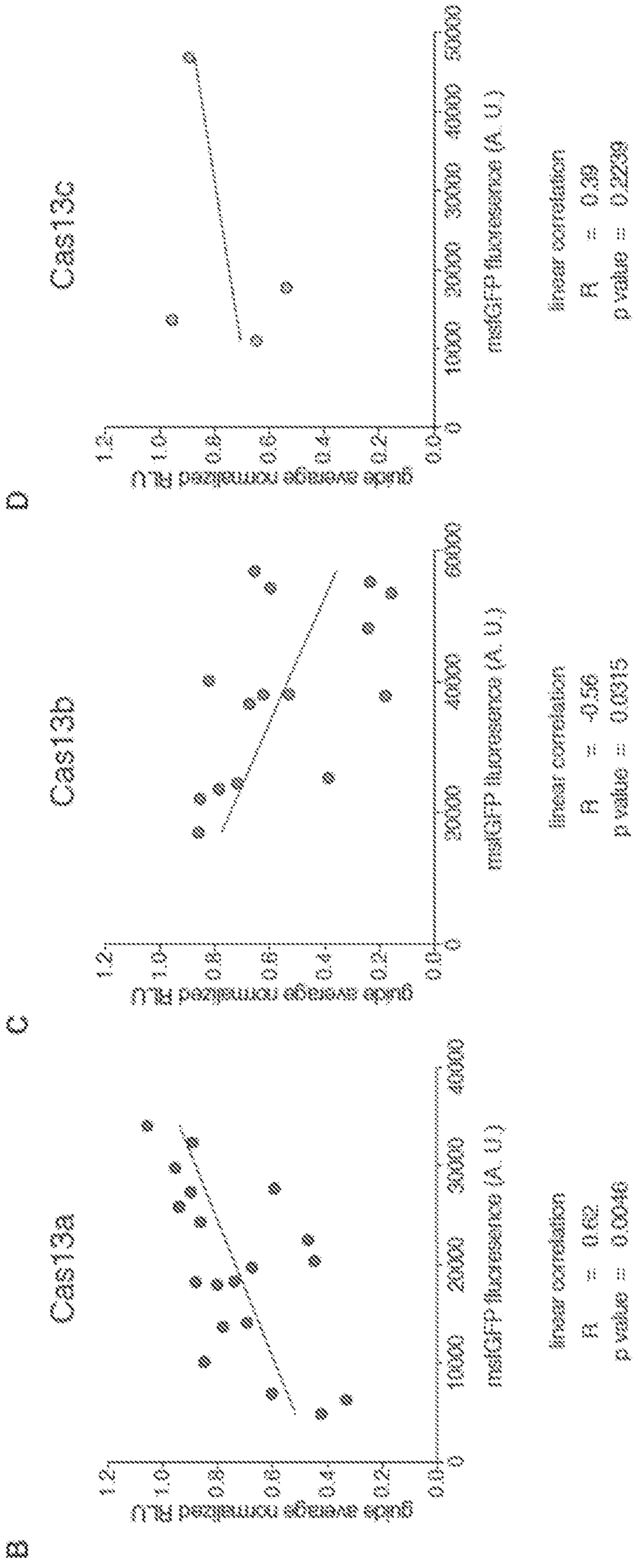


FIG. 69 (continued)

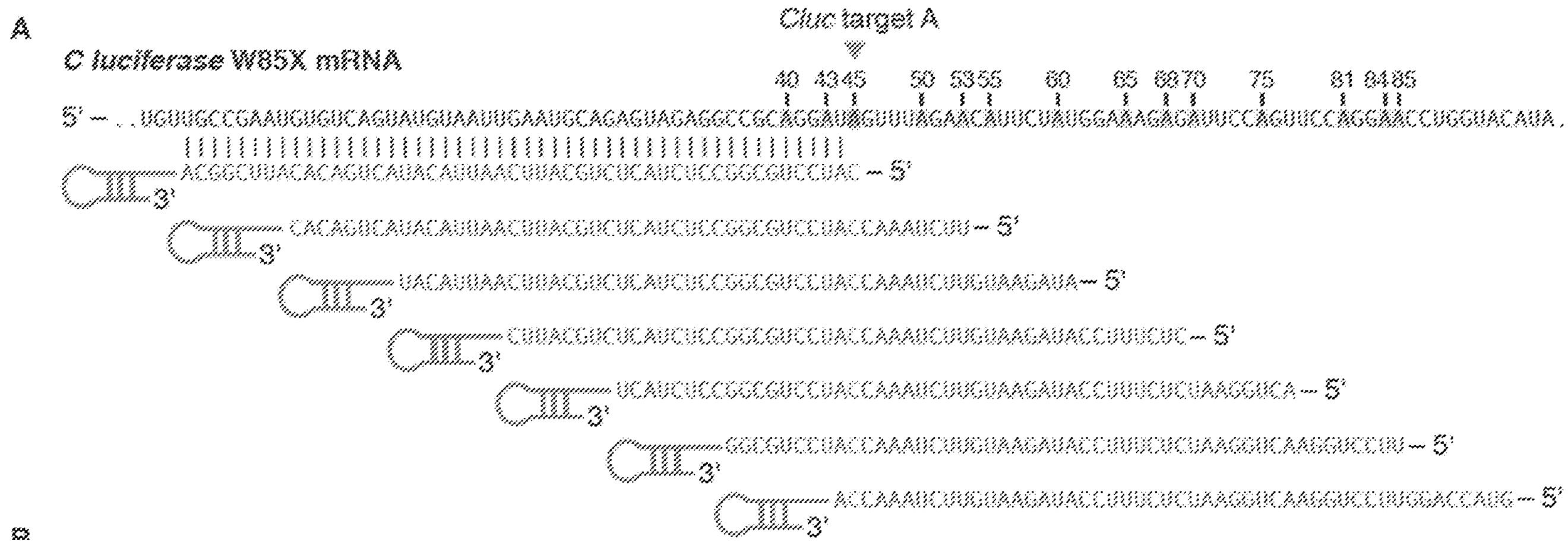

A
C luciferase W85X mRNA
Cluc target A
40 43 45 50 53 55 60 65 68 70 75 81 84 85
5' – . . UGUUGCCGAAUGUGUCAGUAUGUAAUUGAAUGCAGAGUAGAGGCCGCAGGAUAGUUUAGAACAUUCUAUGGAAAGAGAUUCCAGUUCCAGGAACCUGGUACAUA . . – 3'
ACGGCUUACACAGUCAUACAUUAACUUACGUCUCAUCUCCGGCGUCCUAC – 5'
CACAGUCAUACAUUAACUUACGUCUCAUCUCCGGCGUCCUACCAAAUCUU – 5'
UACAUUAACUUACGUCUCAUCUCCGGCGUCCUACCAAAUCUAGUAAGAUA – 5'
CUUACGUCUCAUCUCCGGCGUCCUACCAAAUCUAGUAAGAUACCUUUCUC – 5'
UCAUCUCCGGCGUCCUACCAAAUCUAGUAAGAUACCUUUCUCUAAGGUCA – 5'
GGCGUCCUACCAAAUCUAGUAAGAUACCUUUCUCUAAGGUCAAGGUCCUU – 5'
ACCAAAUCUAGUAAGAUACCUUUCUCUAAGGUCAAGGUCCUAGGACCAUG – 5'

FIG. 70

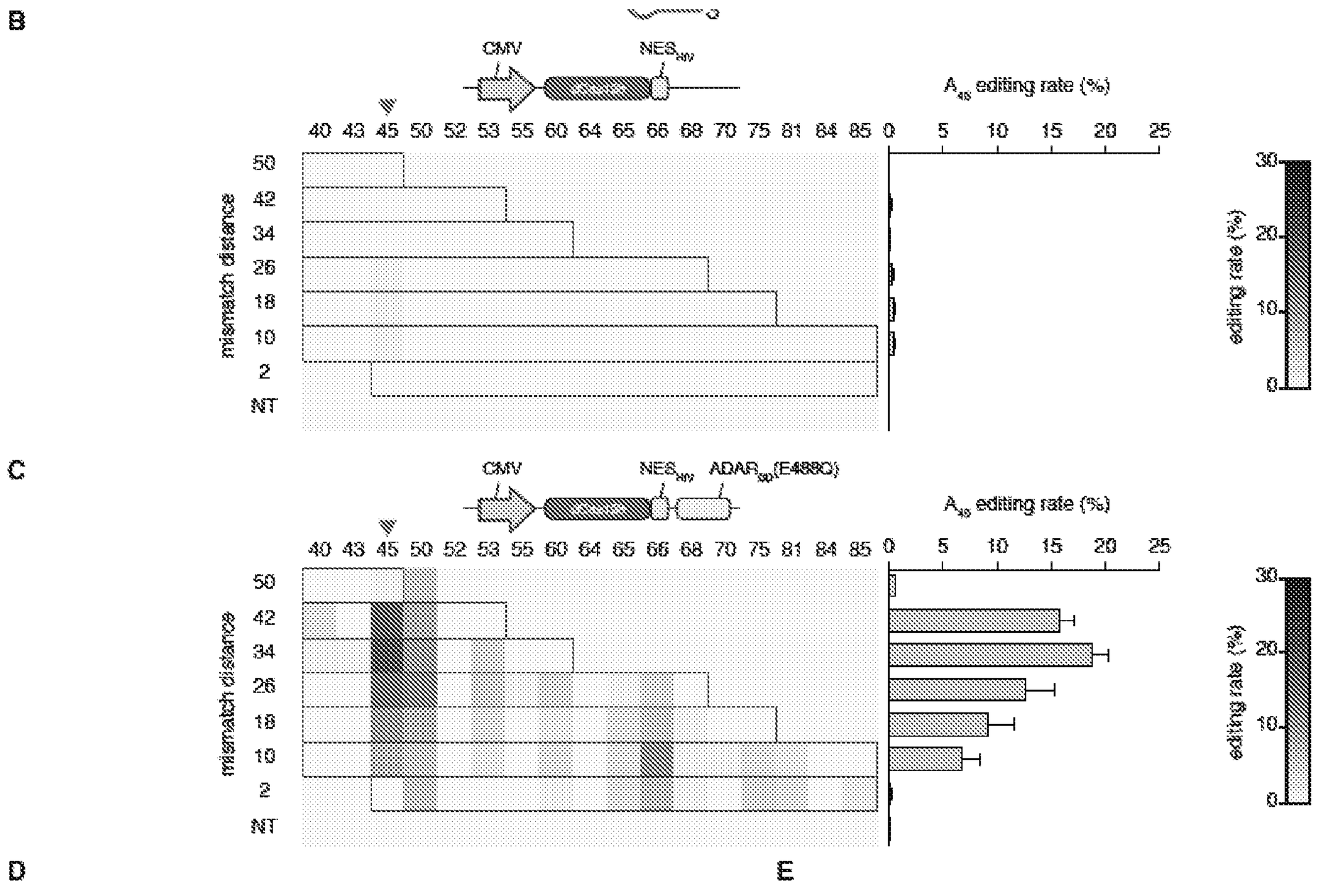


FIG. 70 (continued)

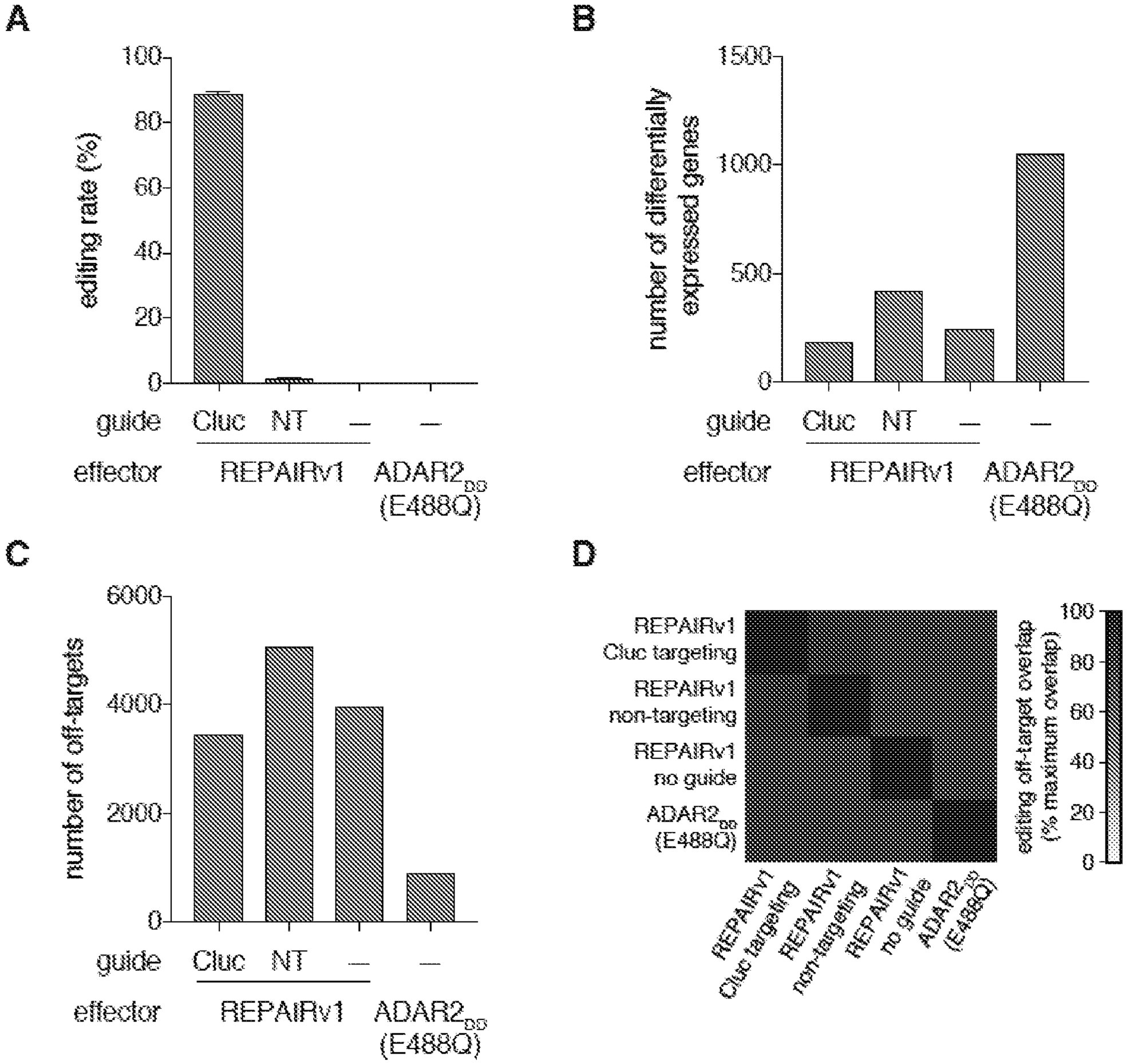
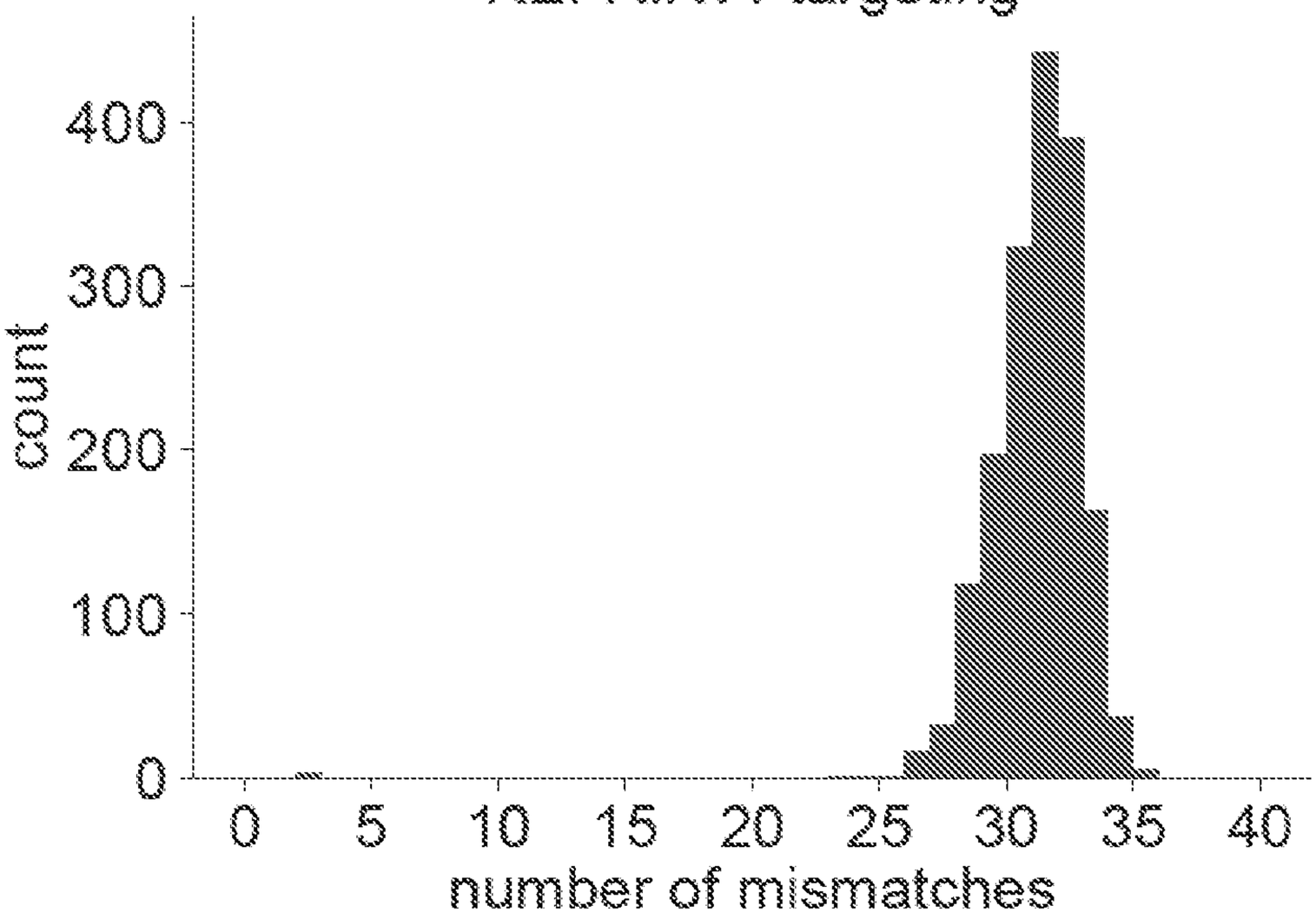

A
editing rate (%)
100
80
60
40
20
0
guide
Cluc
NT
—
—
effector
REPAIRv1
ADAR2dd (E488Q)
B
number of differentially expressed genes
1500
1000
500
0
guide
Cluc
NT
—
—
effector
REPAIRv1
ADAR2dd (E488Q)
C
number of off-targets
6000
4000
2000
0
guide
Cluc
NT
—
—
effector
REPAIRv1
ADAR2dd (E488Q)
D
REPAIRv1 Cluc targeting
REPAIRv1 non-targeting
REPAIRv1 no guide
ADAR2dd (E488Q)
editing off-target overlap (% maximum overlap)
100
80
60
40
20
0

FIG. 71

A


REPAIRv1 targeting
count
0
100
200
300
400
0 5 10 15 20 25 30 35 40
number of mismatches

B

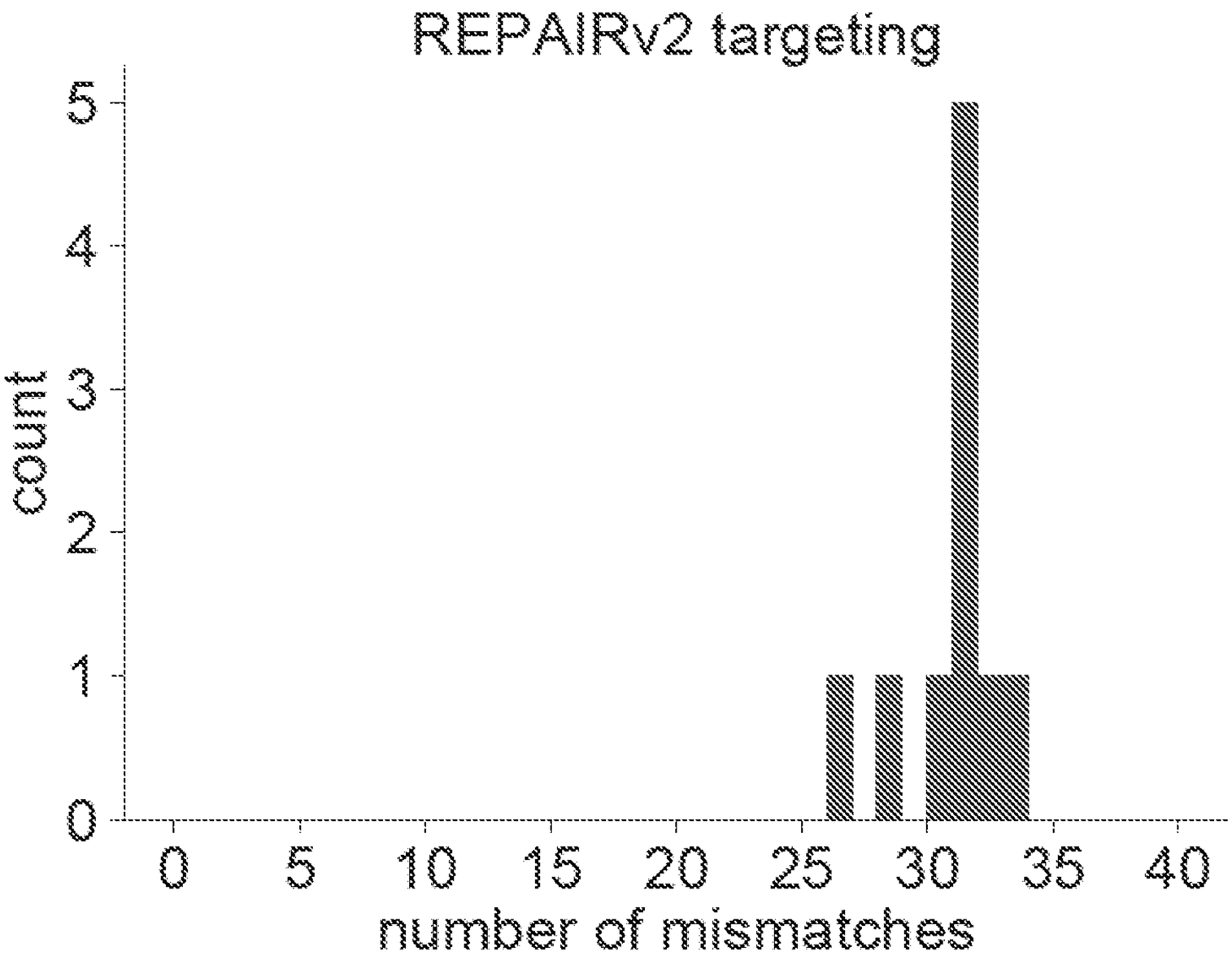
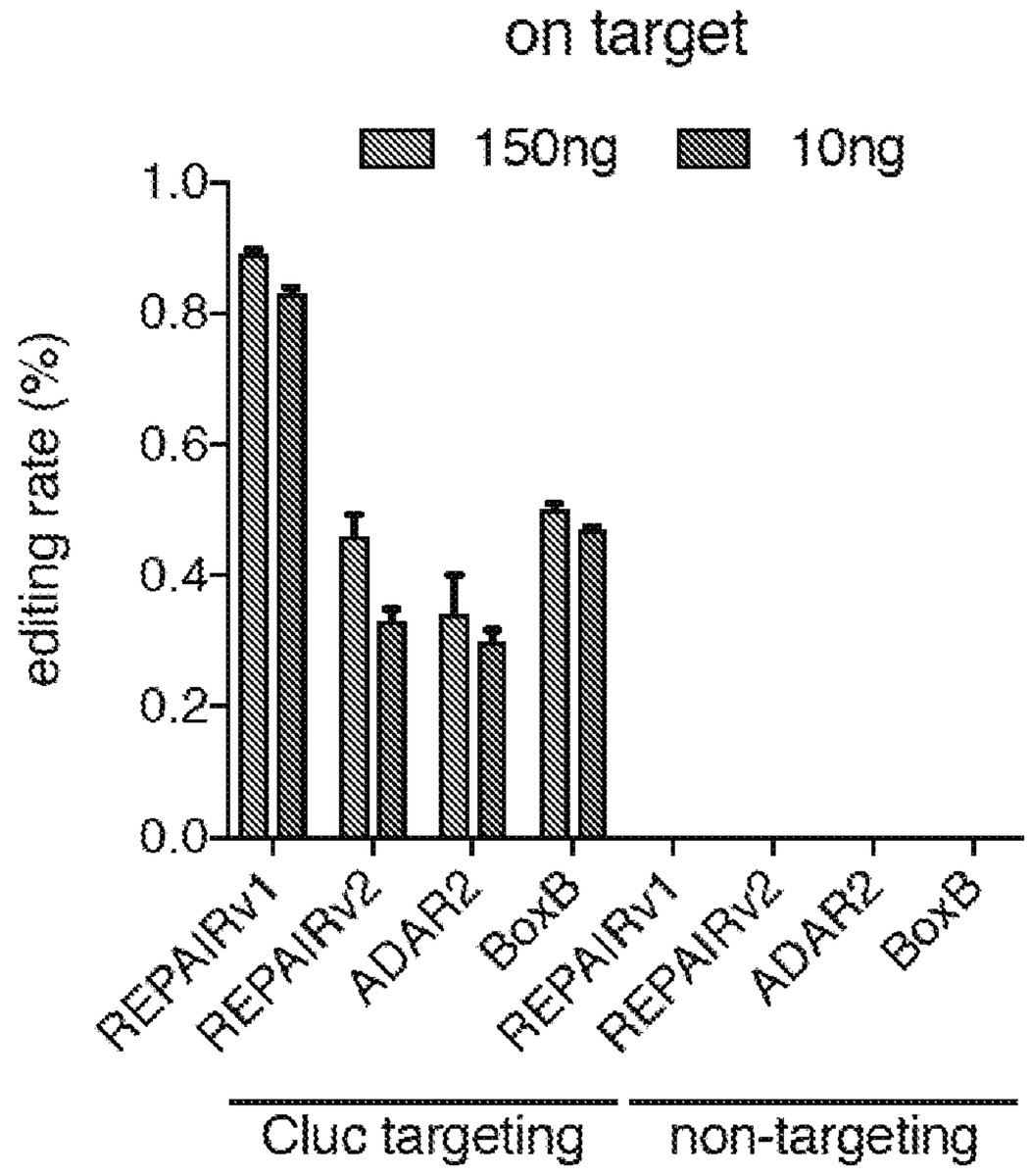

REPAIRv2 targeting
count
0
1
2
3
4
5
0 5 10 15 20 25 30 35 40
number of mismatches

FIG. 72

A


on target
150ng
10ng
editing rate (%)
1.0
0.8
0.6
0.4
0.2
0.0
REPAIRv1
REPAIRv2
ADAR2
BoxB
REPAIRv1
REPAIRv2
ADAR2
BoxB
Cluc targeting
non-targeting

B

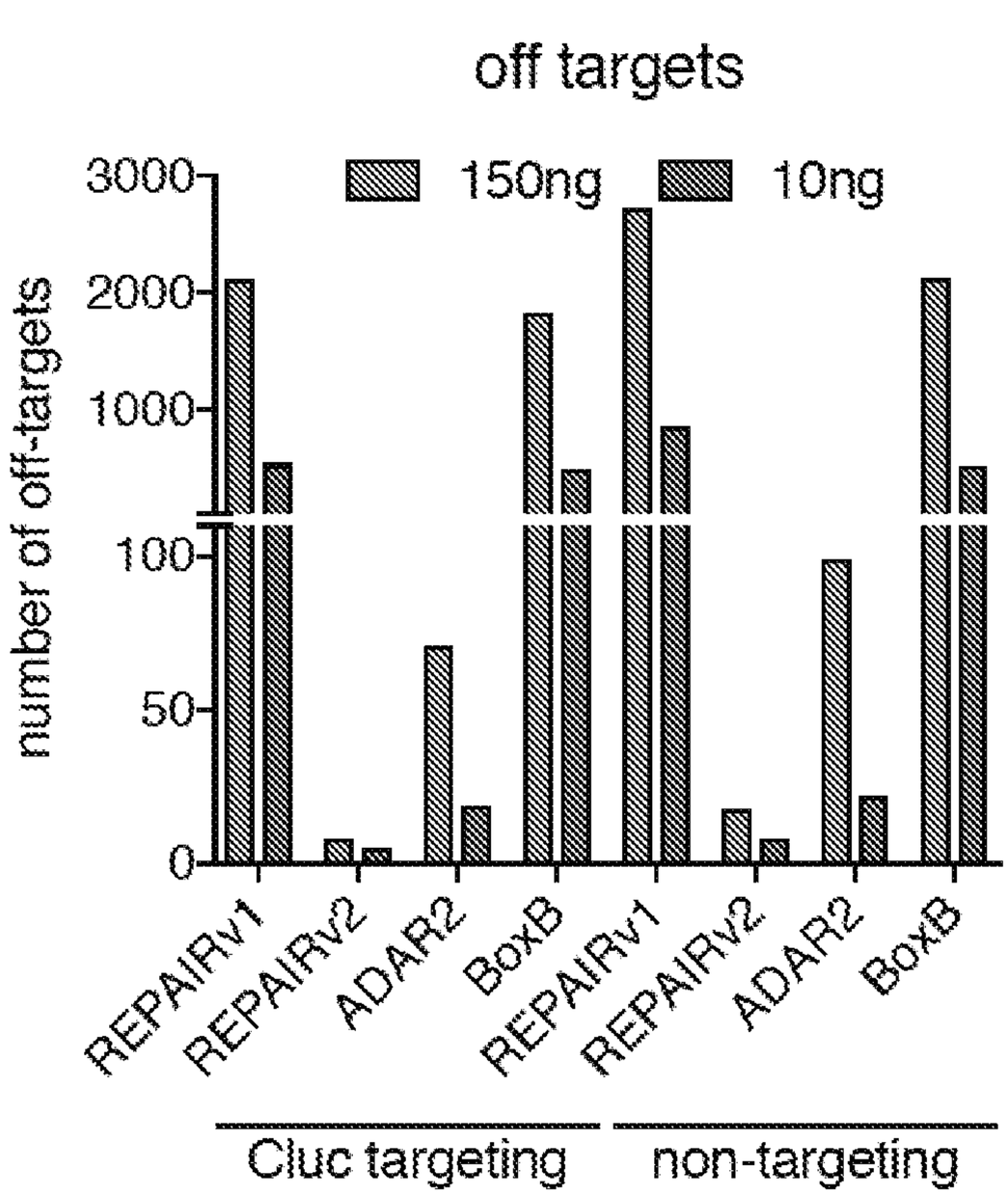

off targets
150ng
10ng
number of off-targets
3000
2000
1000
100
50
0
REPAIRv1
REPAIRv2
ADAR2
BoxB
REPAIRv1
REPAIRv2
ADAR2
BoxB
Cluc targeting
non-targeting

FIG. 73

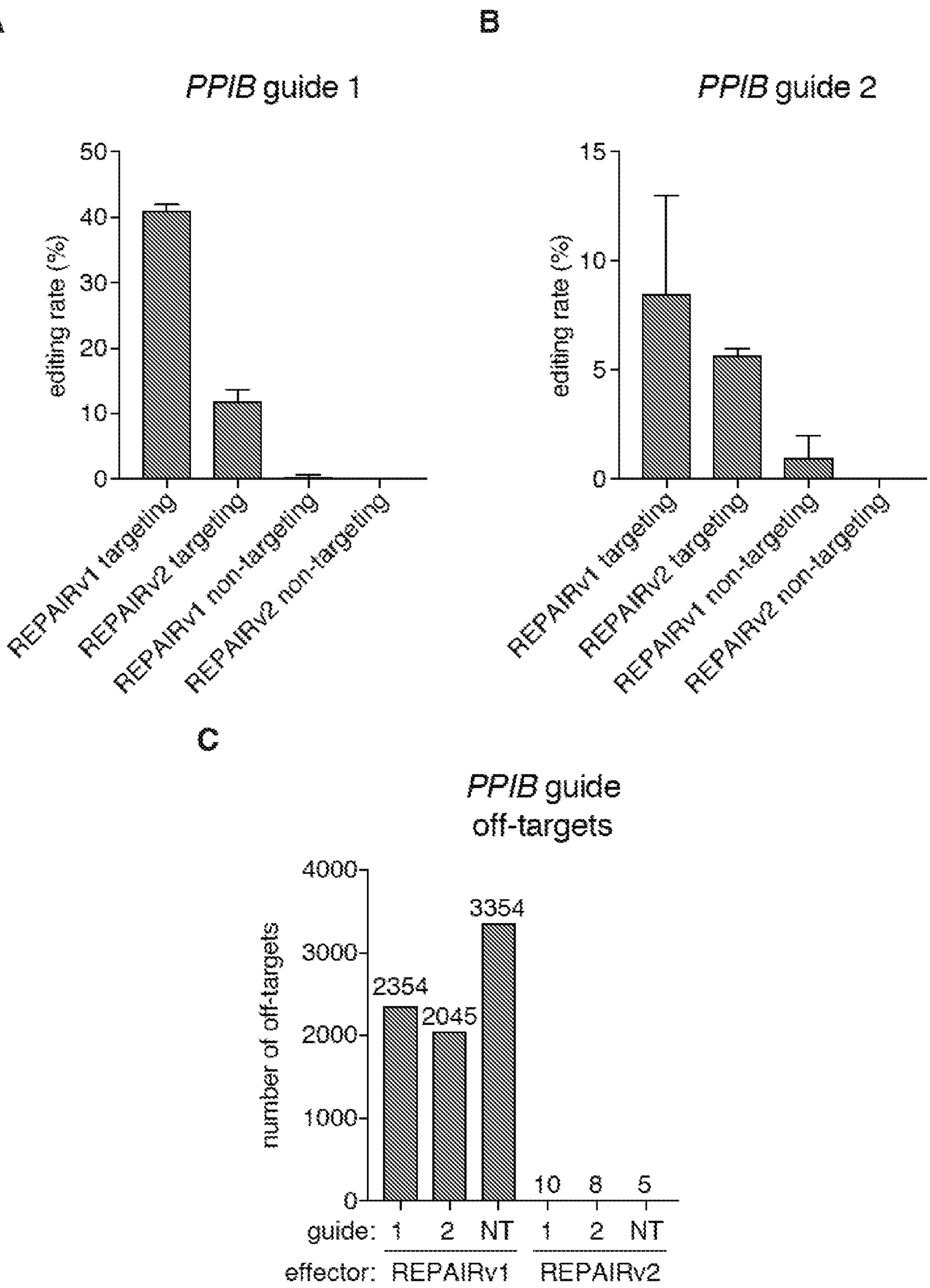

A
PPIB guide 1
editing rate (%)
50
40
30
20
10
0
REPAIRv1 targeting
REPAIRv2 targeting
REPAIRv1 non-targeting
REPAIRv2 non-targeting
B
PPIB guide 2
editing rate (%)
15
10
5
0
REPAIRv1 targeting
REPAIRv2 targeting
REPAIRv1 non-targeting
REPAIRv2 non-targeting
C
PPIB guide
off-targets
number of off-targets
4000
3000
2000
1000
0
2354
2045
3354
10
8
5
guide: 1 2 NT 1 2 NT
effector: REPAIRv1 REPAIRv2

FIG. 74

D

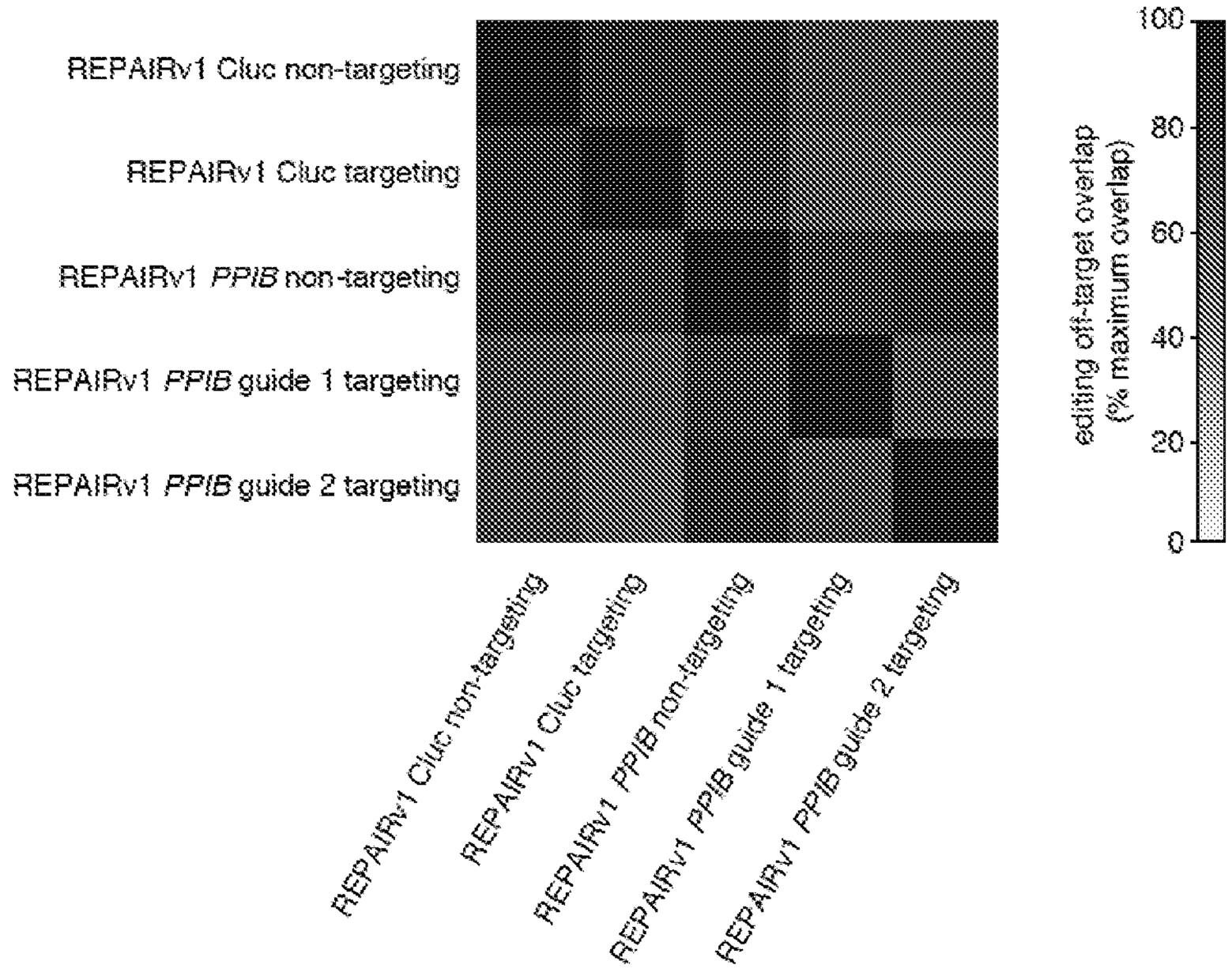


FIG. 74 (continued)

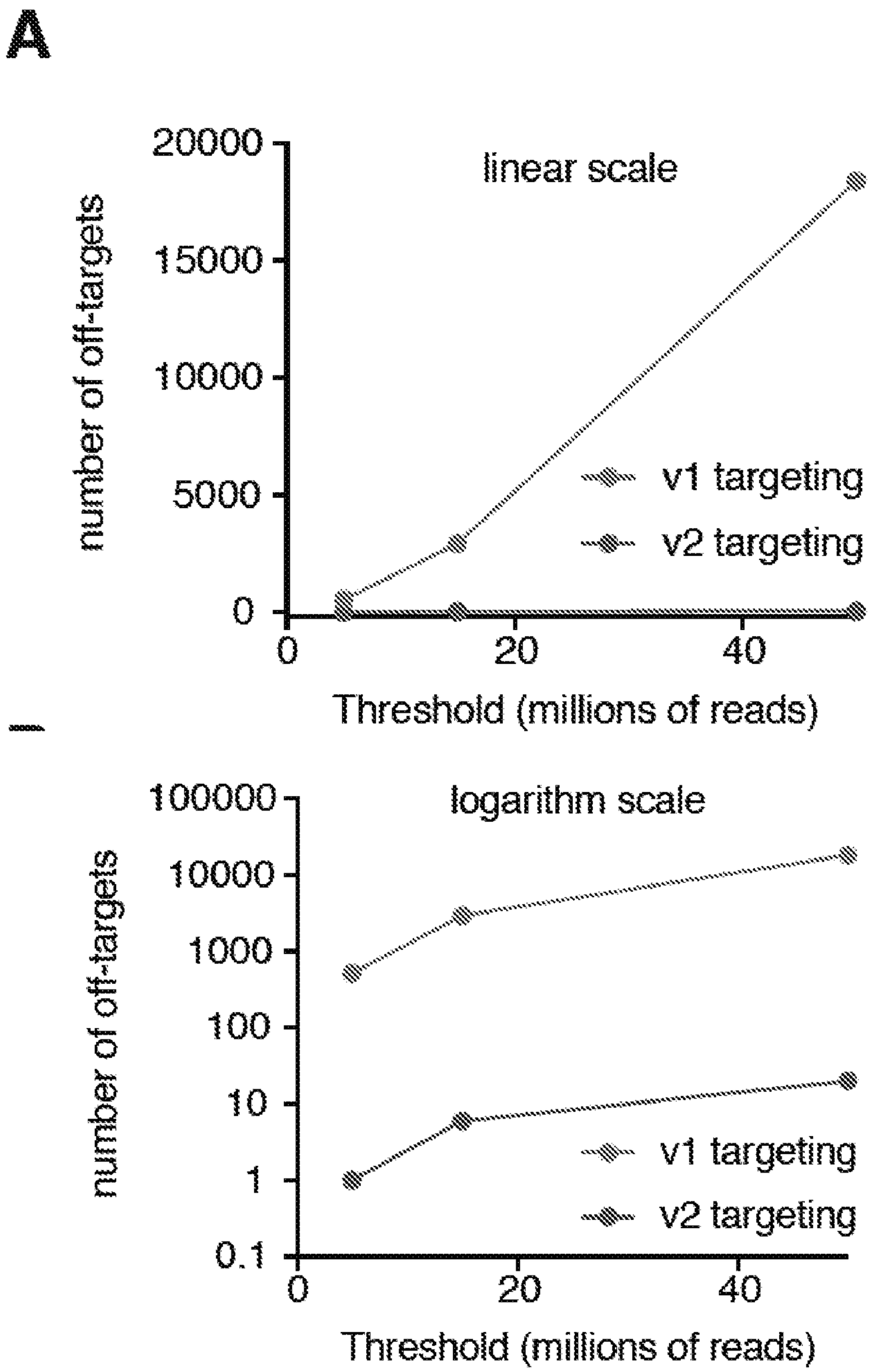

A
linear scale
number of off-targets
20000
15000
10000
5000
0
0
20
40
Threshold (millions of reads)
v1 targeting
v2 targeting
logarithm scale
number of off-targets
100000
10000
1000
100
10
1
0.1
0
20
40
Threshold (millions of reads)
v1 targeting
v2 targeting

FIG. 75

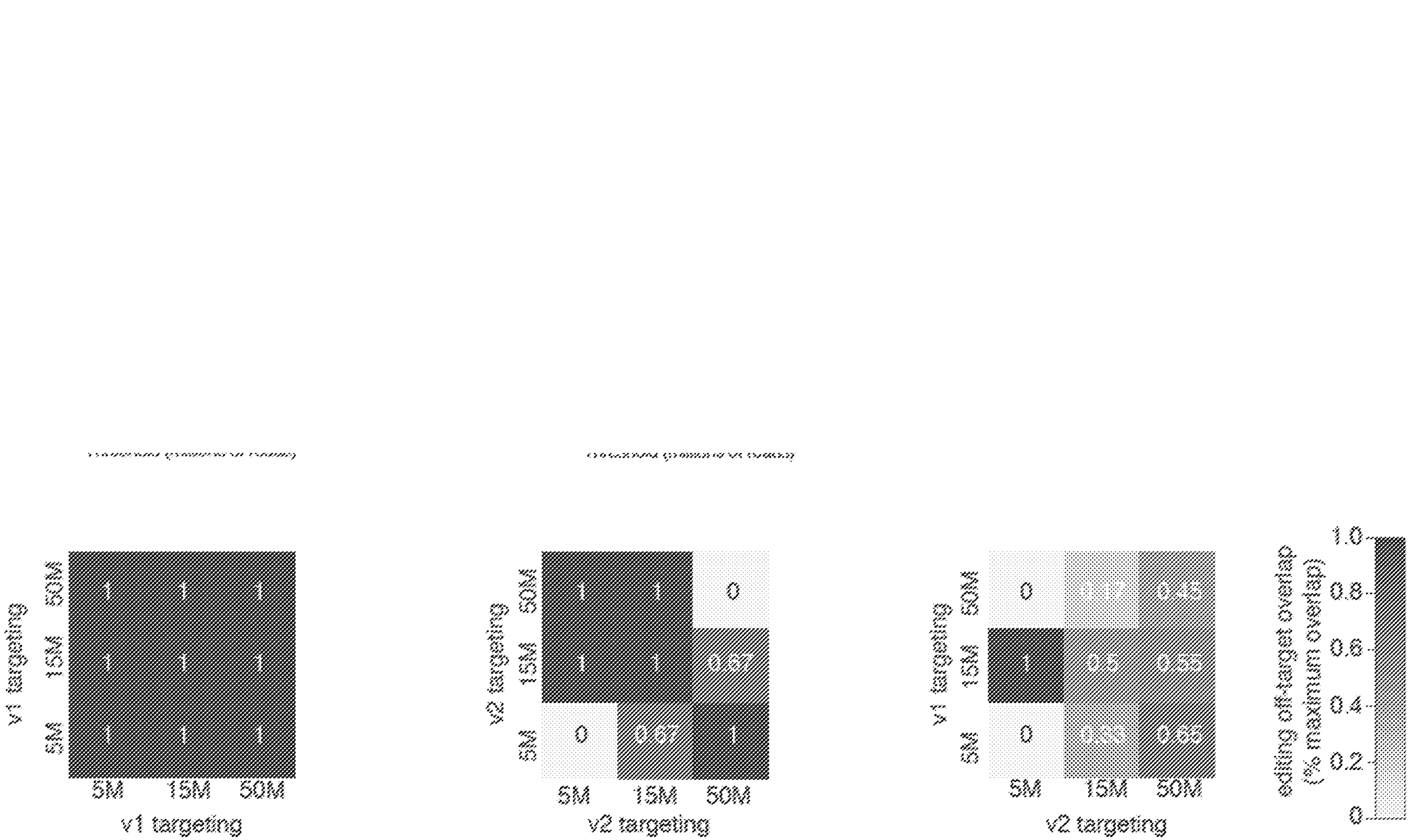


FIG. 75 (continued)

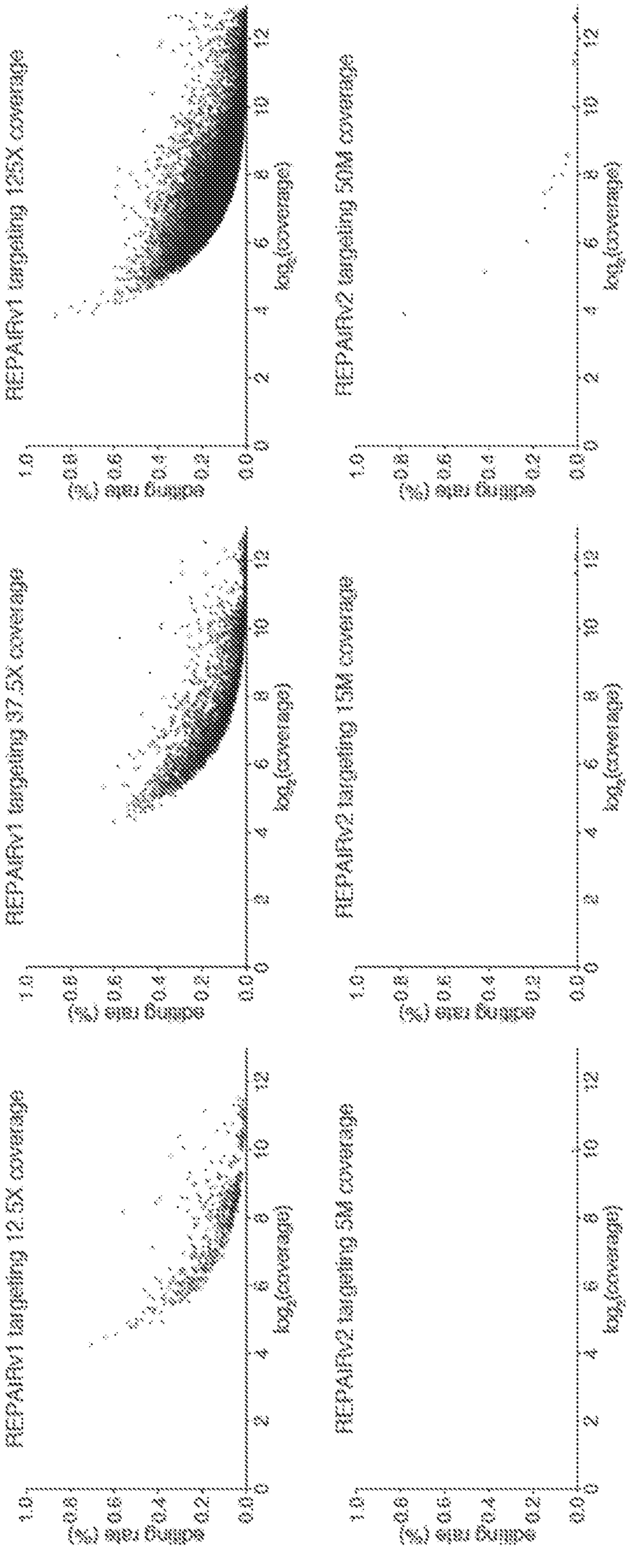


FIG. 75 (continued)

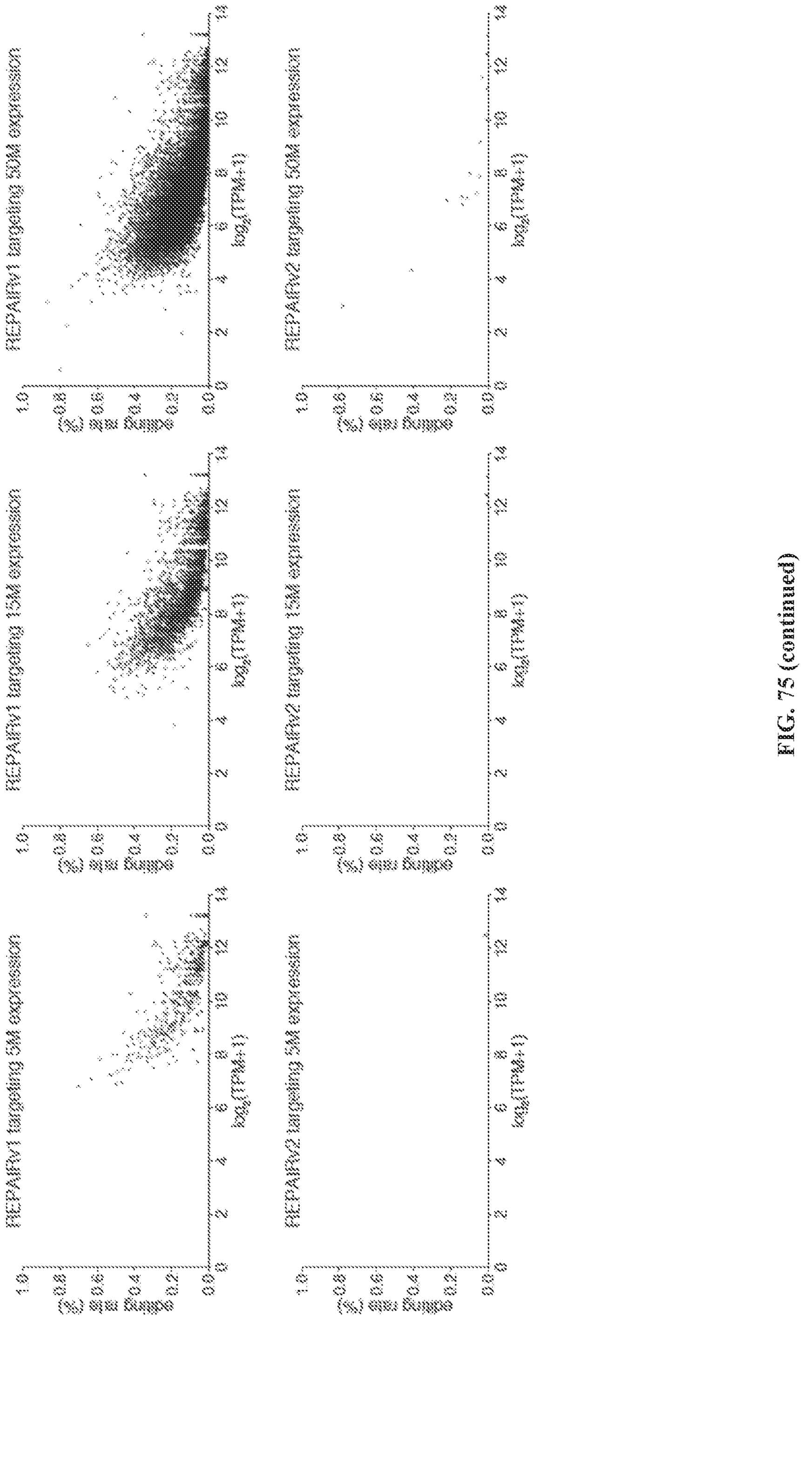


FIG. 75 (continued)

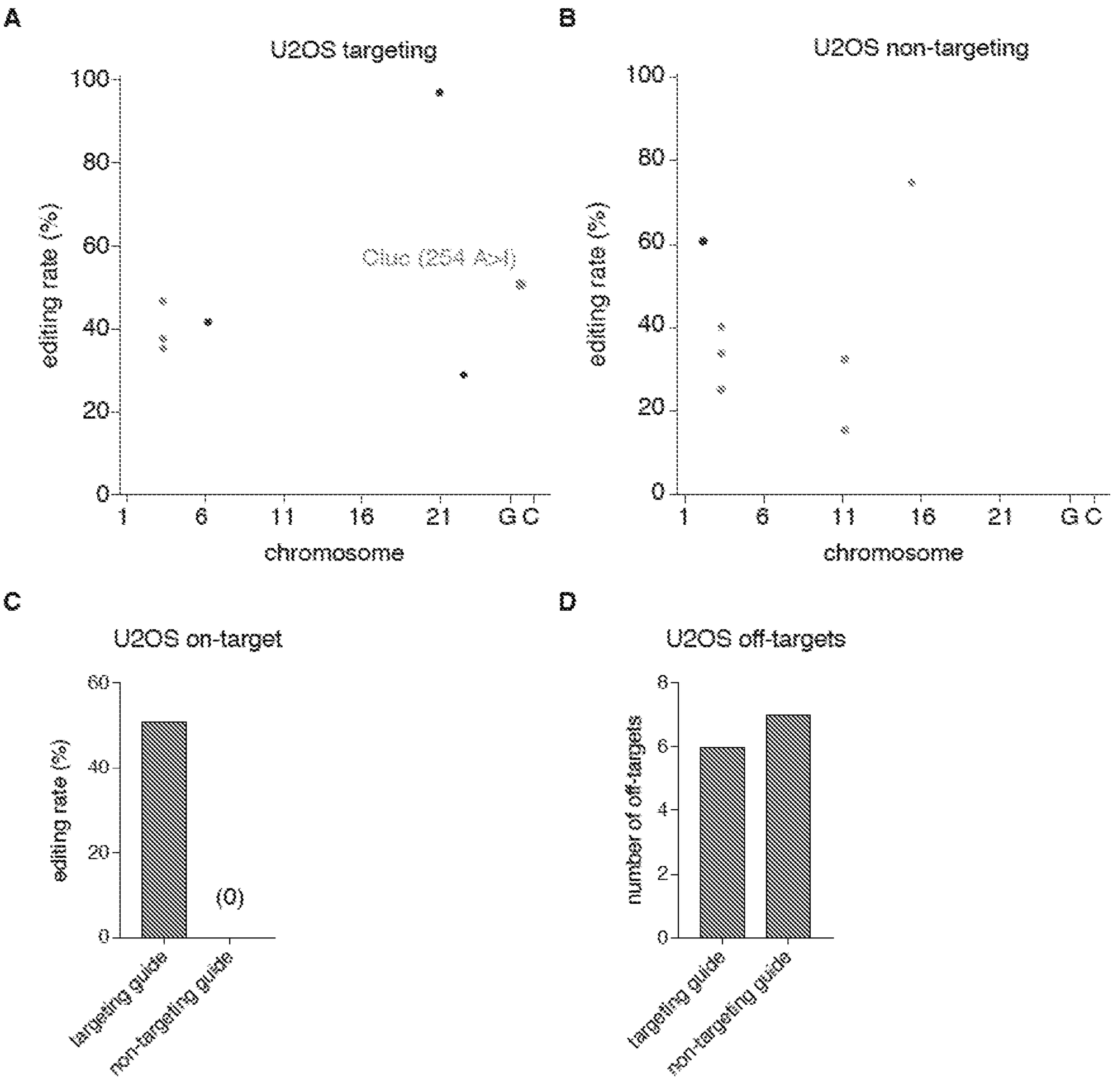

A
U2OS targeting
editing rate (%)
100
80
60
40
20
0
Cluc (254 A>I)
1
6
11
16
21
G
C
chromosome
B
U2OS non-targeting
editing rate (%)
chromosome
C
U2OS on-target
editing rate (%)
(0)
targeting guide
non-targeting guide
D
U2OS off-targets
number of off-targets
8
6
4
2
0
targeting guide
non-targeting guide

FIG. 76

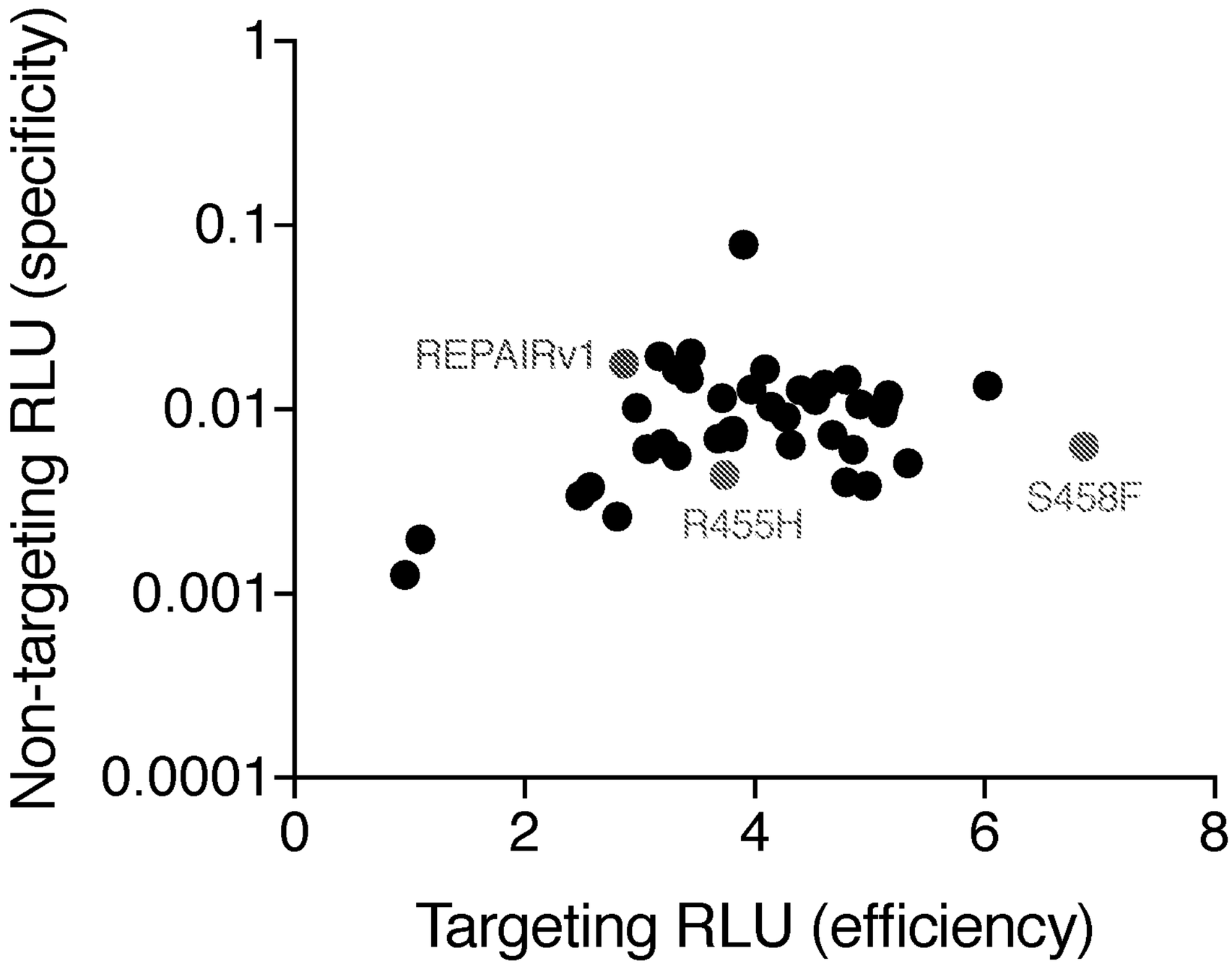

Non-targeting RLU (specificity)
1
0.1
0.01
0.001
0.0001
0
2
4
6
8
Targeting RLU (efficiency)
REPAIRv1
R455H
S458F

FIG. 77

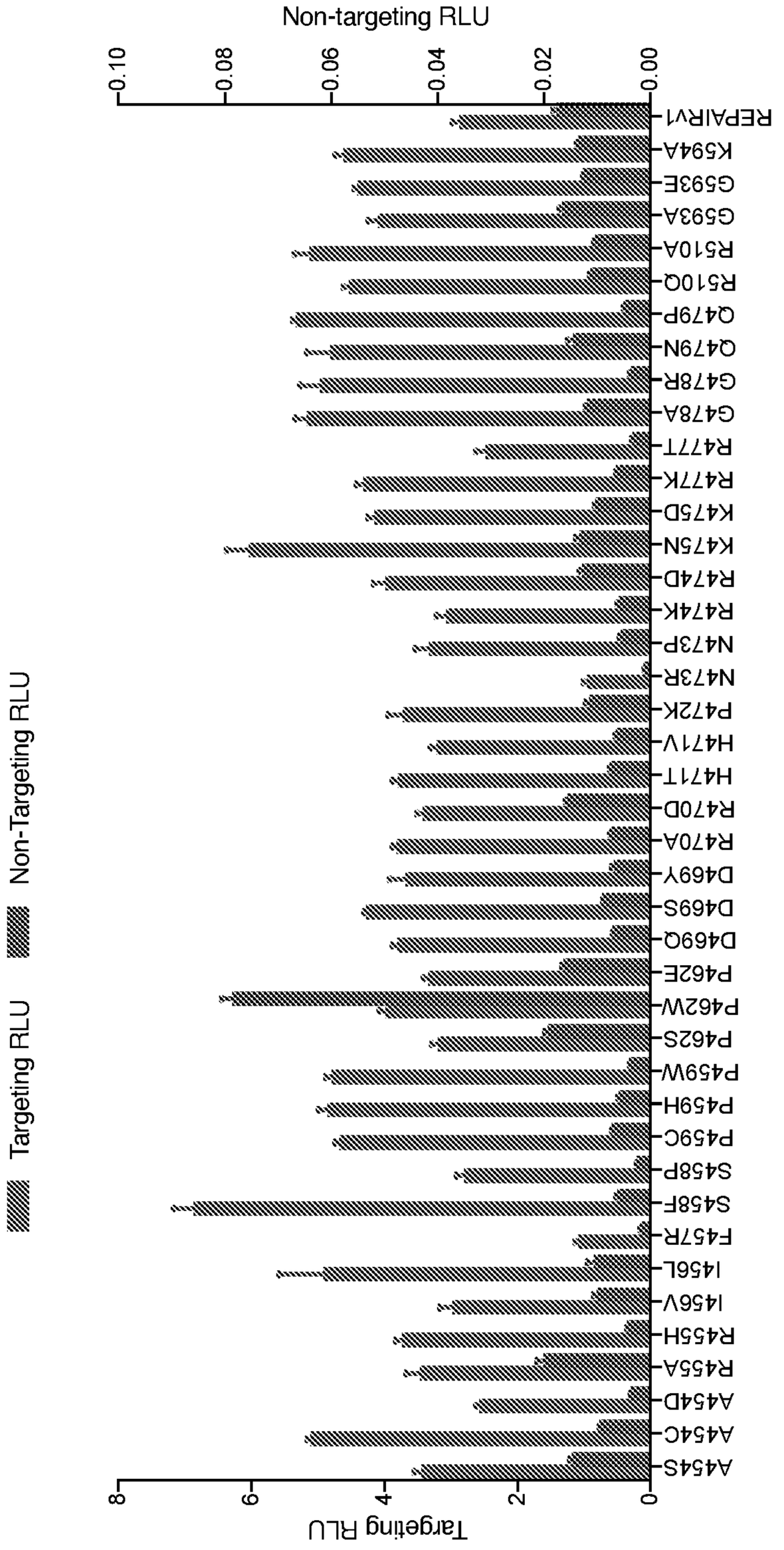

Non-targeting RLU
0.10
0.08
0.06
0.04
0.02
0.00
REPAIRv1
K594A
G593E
G593A
R510A
R510Q
Q479P
Q479N
G478R
G478A
R477T
R477K
K475D
K475N
R474D
R474K
N473P
N473R
P472K
H471V
H471T
R470D
R470A
D469Y
D469S
D469Q
P462E
P462W
P462S
P459W
P459H
P459C
S458P
S458F
F457R
I456L
I456V
R455H
R455A
A454D
A454C
A454S
Targeting RLU
Non-Targeting RLU
Targeting RLU
ADAR Mutant

FIG. 78

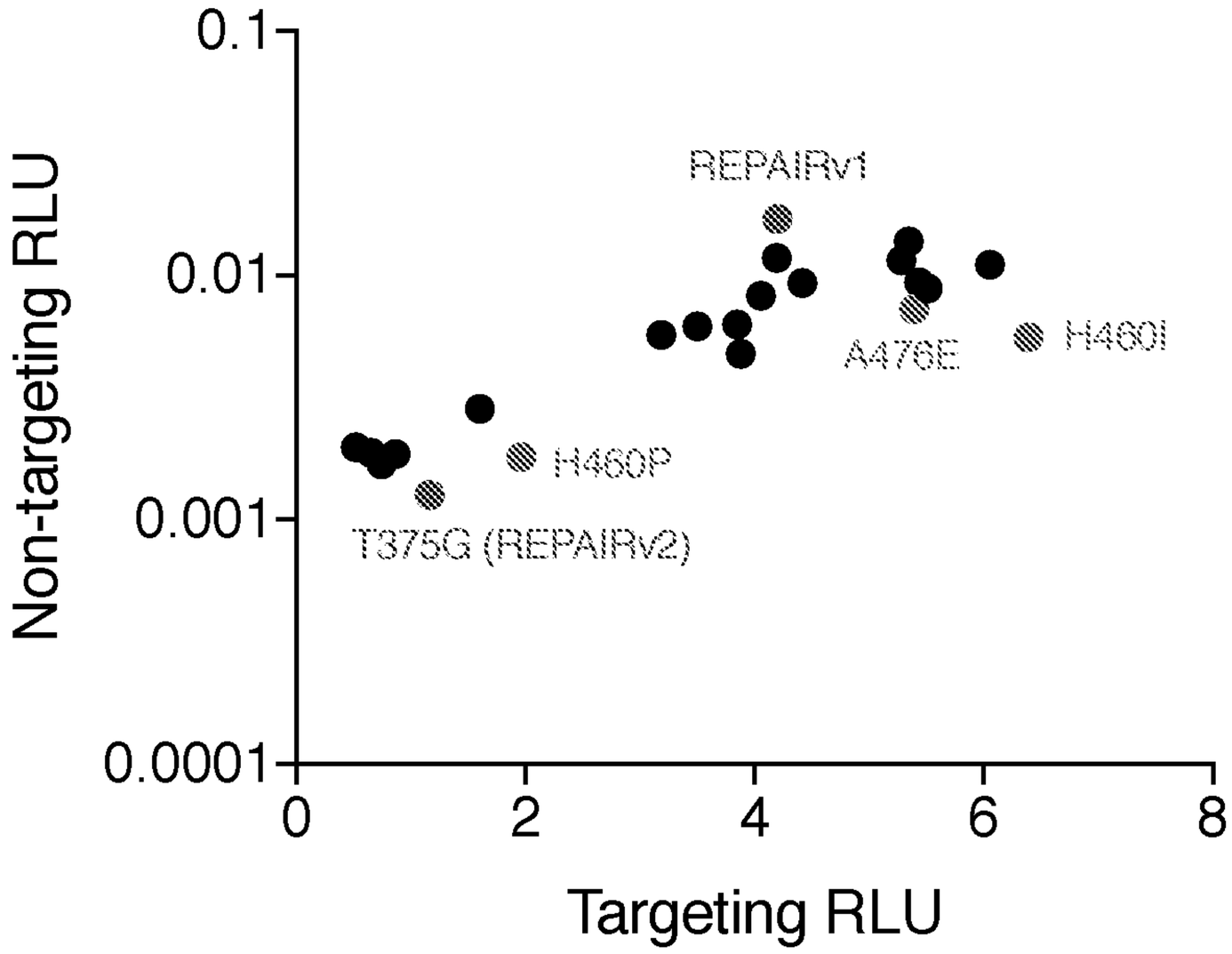

0.1
0.01
0.001
0.0001
Non-targeting RLU
0
2
4
6
8
Targeting RLU
REPAIRv1
A476E
H460I
H460P
T375G (REPAIRv2)

FIG. 79

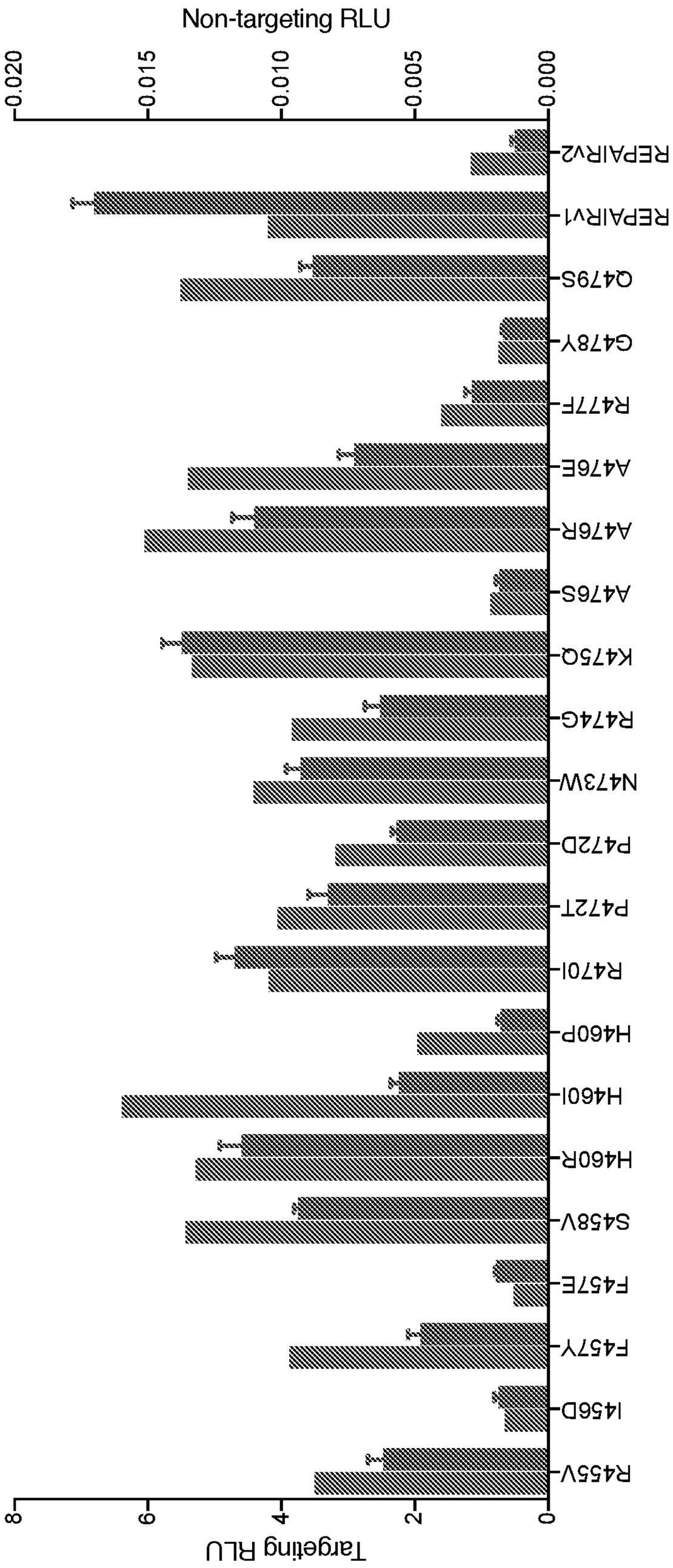

Non-targeting RLU
0.020
0.015
0.010
0.005
0.000
REPAIRv2
REPAIRv1
Q479S
G478Y
R477F
A476E
A476R
A476S
K475Q
R474G
N473W
P472D
P472T
R470I
H460P
H460I
H460R
S458V
F457E
F457Y
I456D
R455V
Non-Targeting RLU
Targeting RLU
ADAR Mutant
Targeting RLU
0
2
4
6
8

FIG. 80

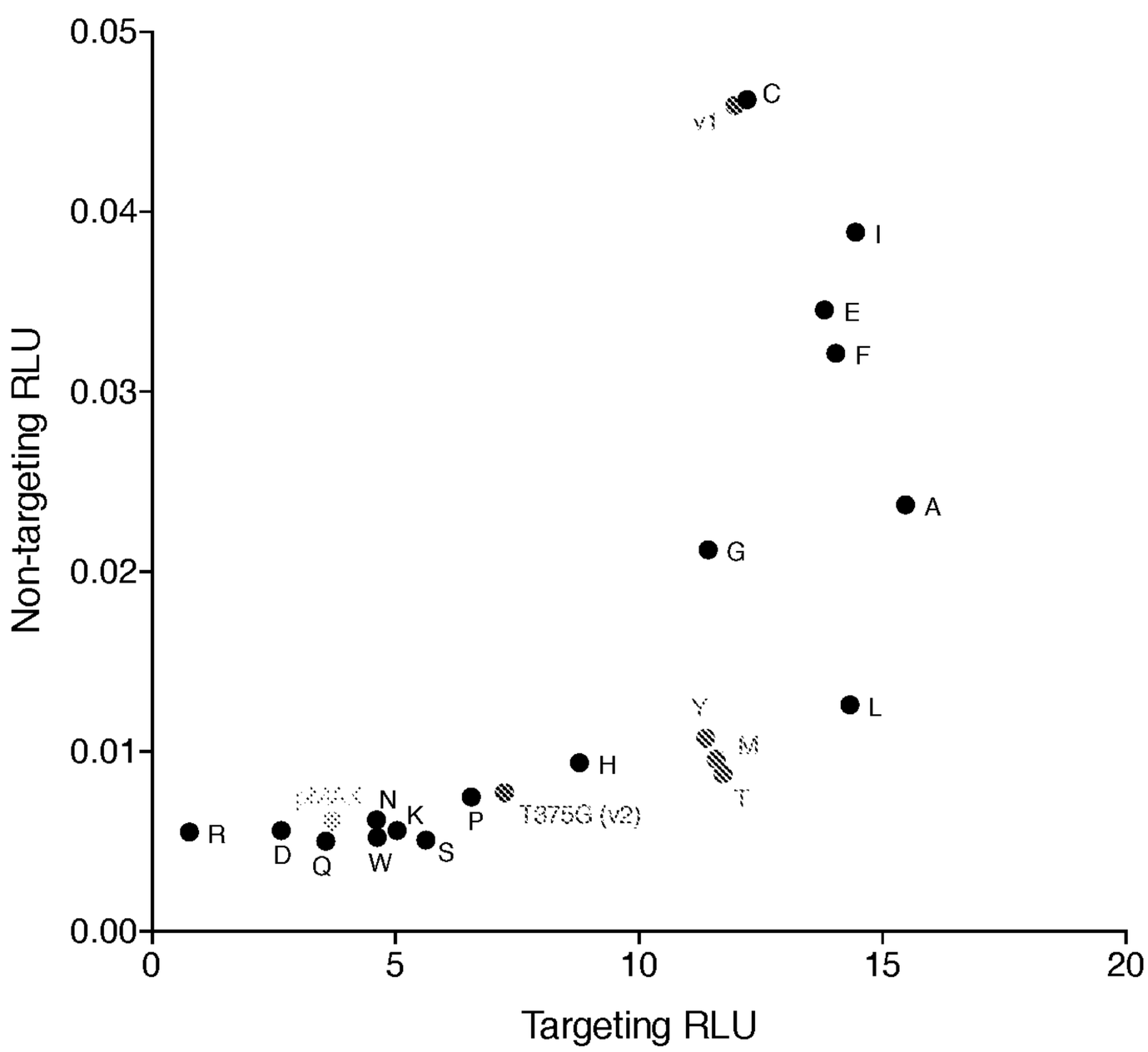

V351 mutagenesis
Non-targeting RLU
Targeting RLU
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
C
I
E
F
A
G
L
Y
M
T
H
N
K
P
T375G (v2)
R
D
Q
W
S

FIG. 81

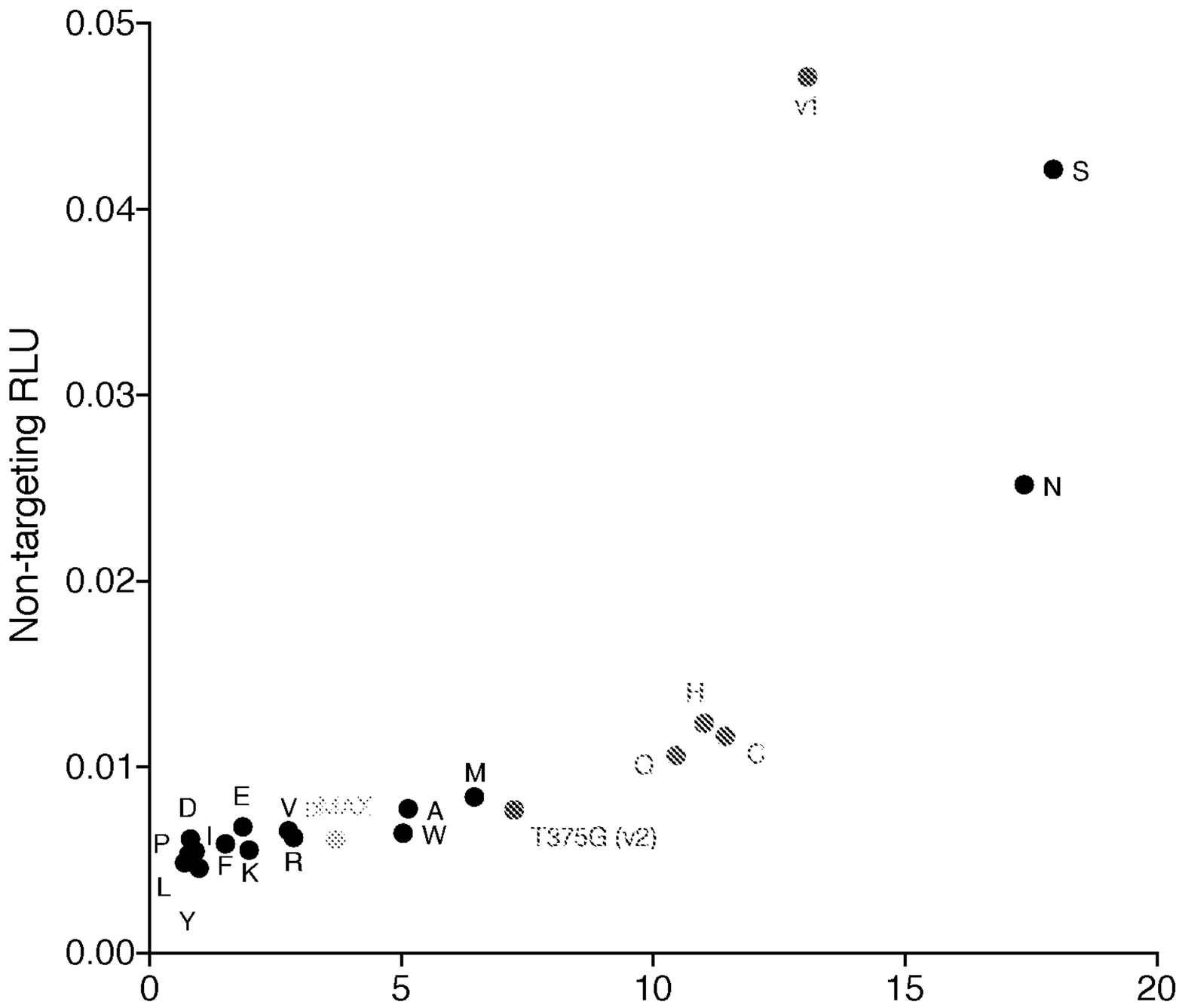

T375 mutagenesis
Non-targeting RLU
Targeting RLU
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
V1
S
N
H
C
M
A
W
T375G (v2)
D
E
V
P
I
F
K
R
L
Y

FIG. 82

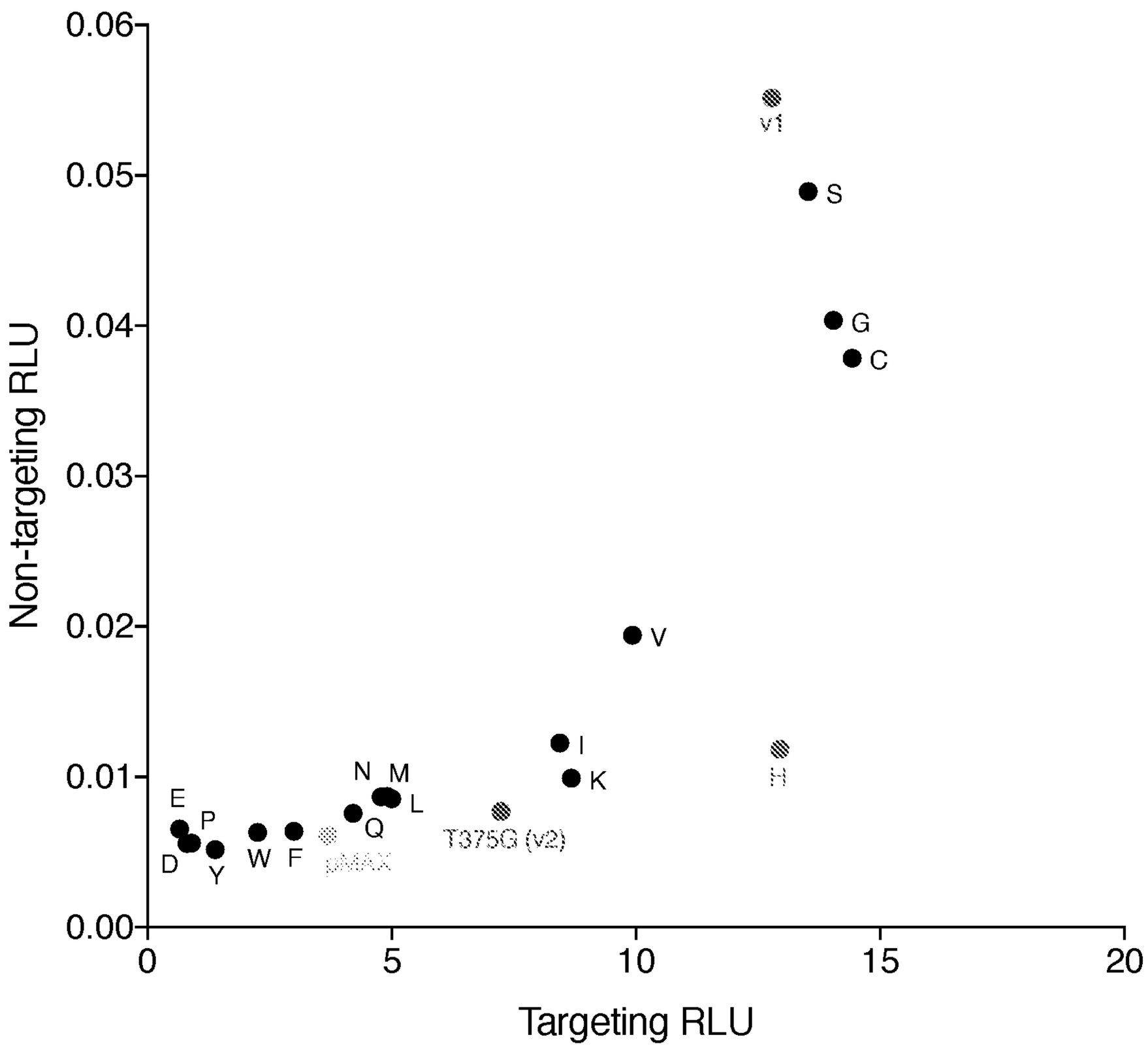

R455 mutagenesis
Non-targeting RLU
Targeting RLU
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
v1
S
G
C
V
I
K
H
N
M
L
Q
E
P
D
Y
W
F
pMAX
T375G (v2)

FIG. 83

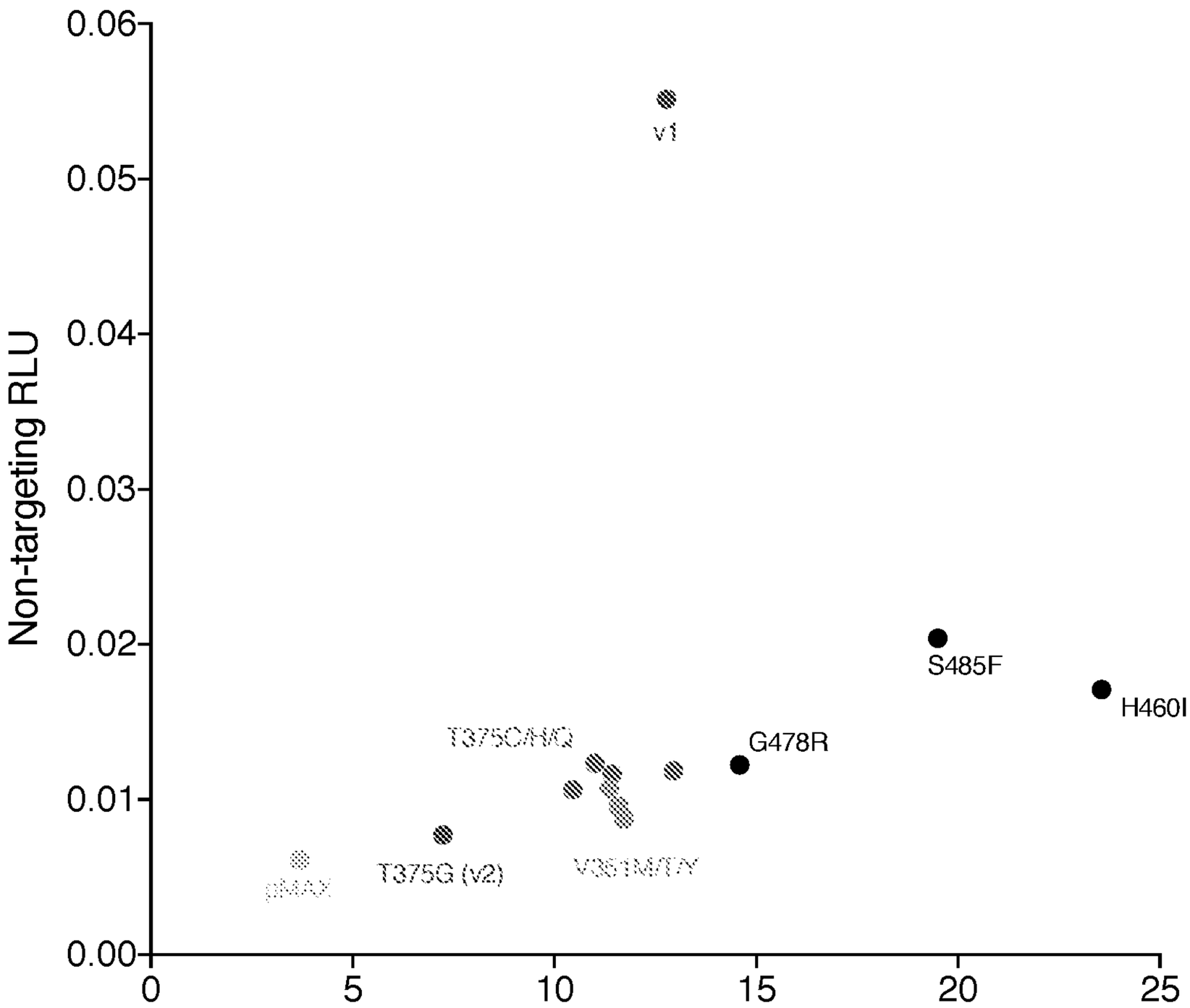

Selected mutants
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Non-targeting RLU
0
5
10
15
20
25
Targeting RLU
v1
S485F
H460I
T375C/H/Q
G478R
T375G (v2)
V361M/T/Y
pMAX

FIG. 84

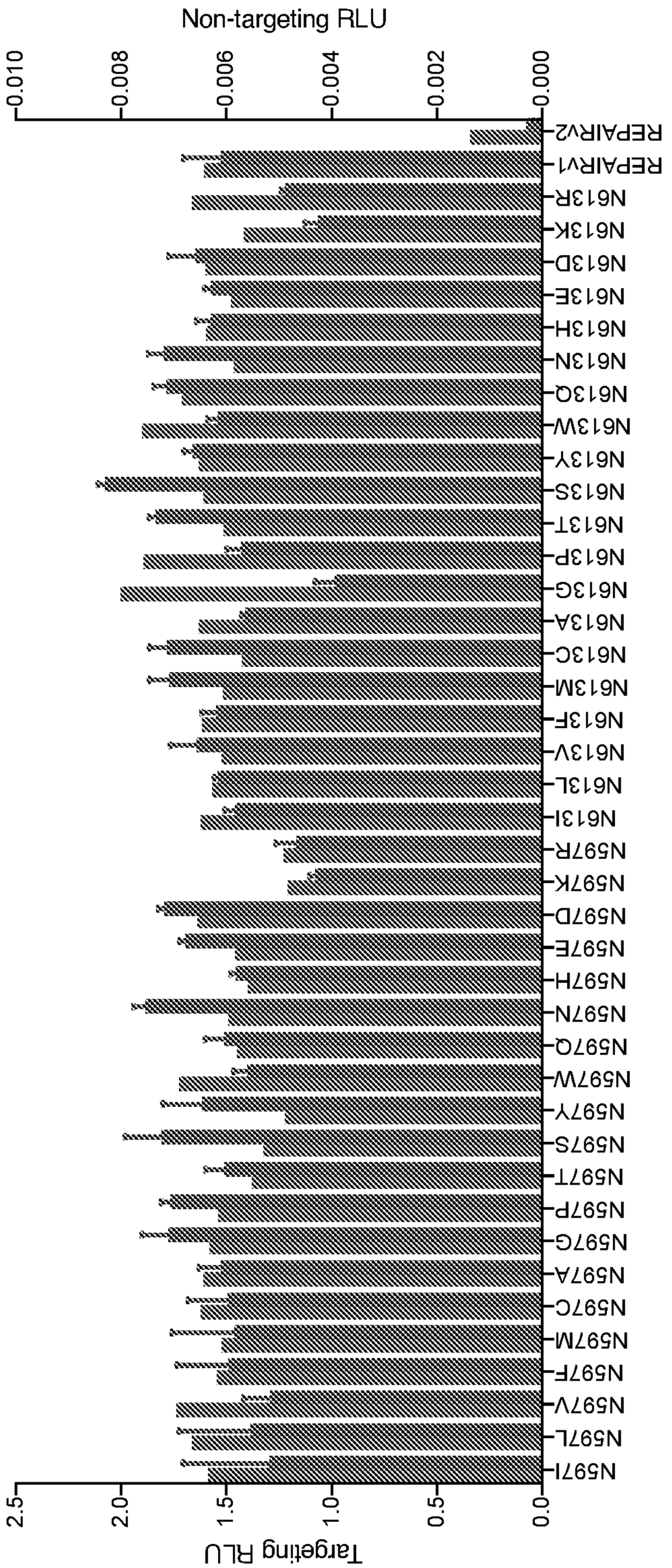

Non-targeting RLU
0.010
0.008
0.006
0.004
0.002
0.000
REPAIRv2
REPAIRv1
N613R
N613K
N613D
N613E
N613H
N613N
N613Q
N613W
N613Y
N613S
N613T
N613P
N613G
N613A
N613C
N613M
N613F
N613V
N613L
N613I
N597R
N597K
N597D
N597E
N597H
N597N
N597Q
N597W
N597Y
N597S
N597T
N597P
N597G
N597A
N597C
N597M
N597F
N597V
N597L
N597I
2.5
2.0
1.5
1.0
0.5
0.0
Targeting RLU
Non-targeting RLU
Targeting RLU
ADAR Mutant

FIG. 85

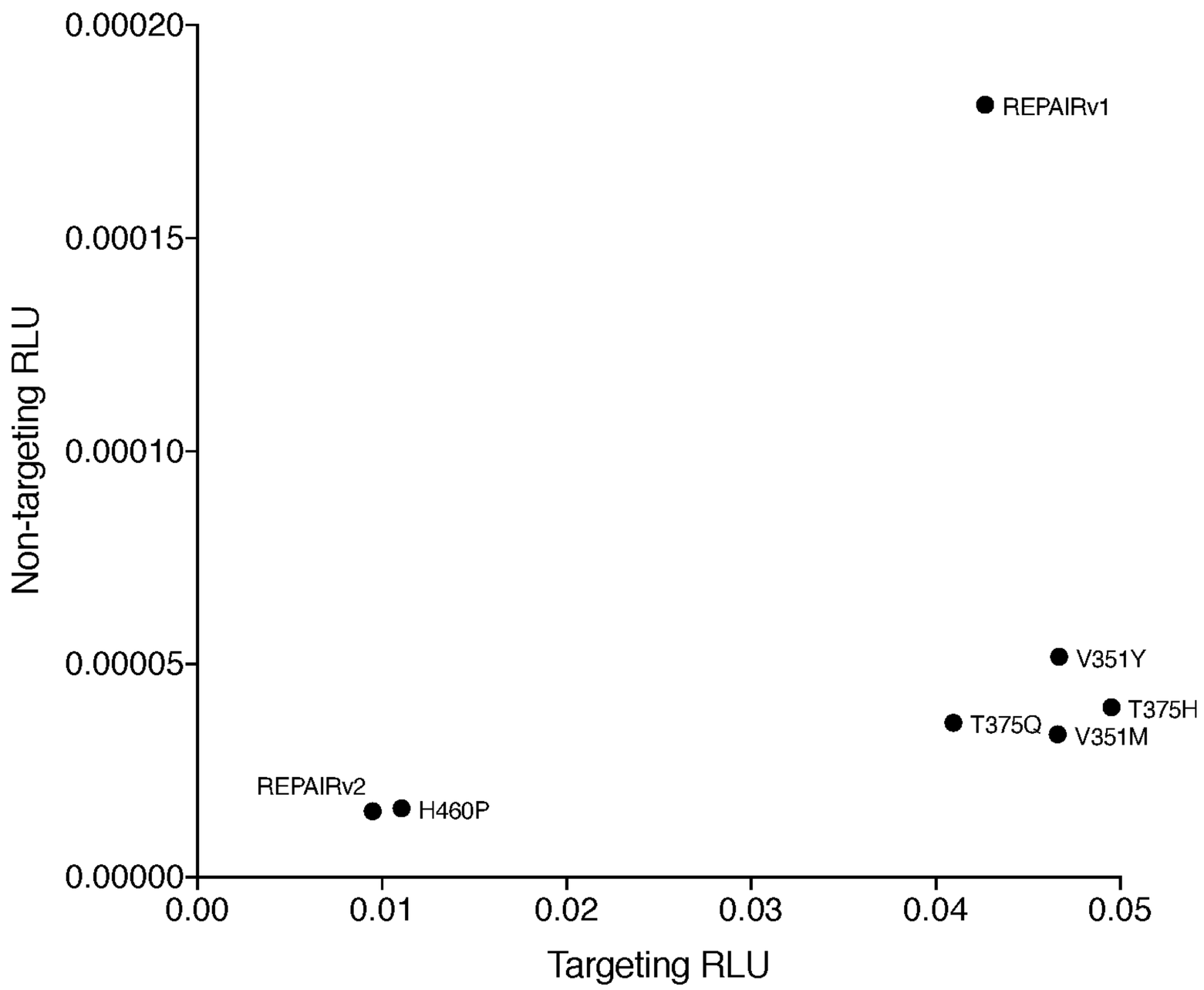

0.00020
0.00015
0.00010
0.00005
0.00000
Non-targeting RLU
0.00
0.01
0.02
0.03
0.04
0.05
Targeting RLU
REPAIRv1
V351Y
T375H
T375Q
V351M
REPAIRv2
H460P

FIG. 86

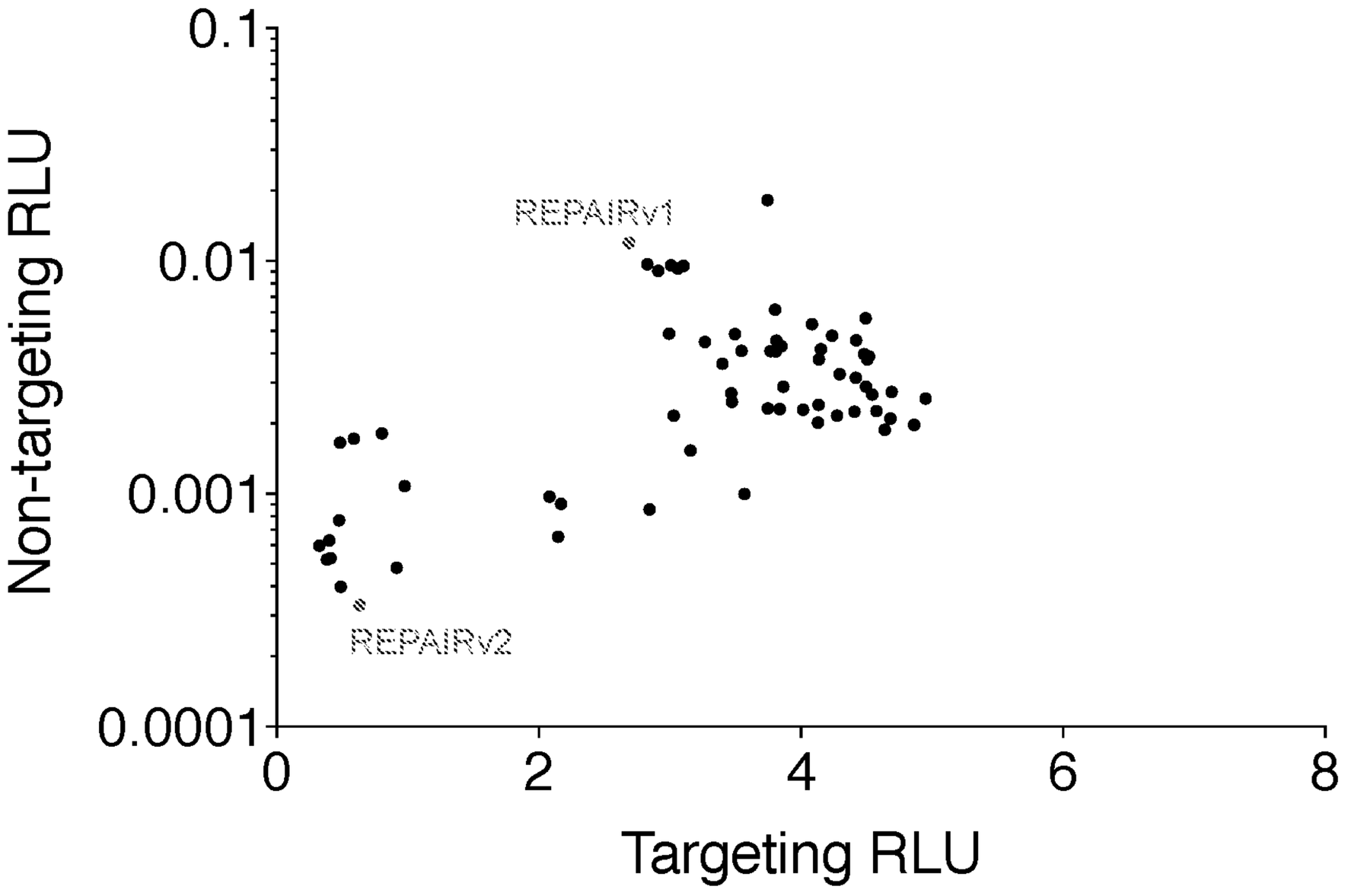
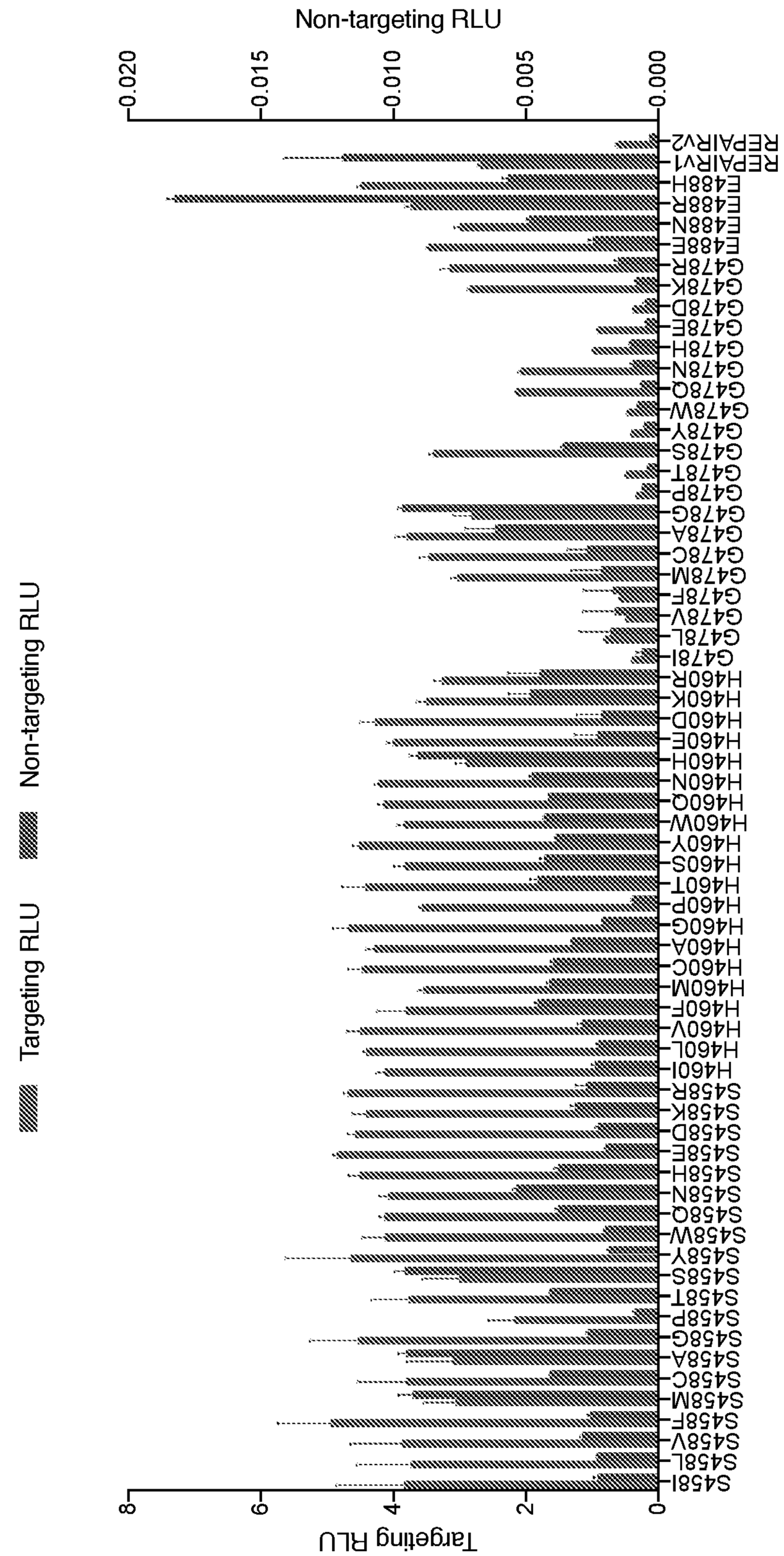

0.1
0.01
0.001
0.0001
Non-targeting RLU
0
2
4
6
8
Targeting RLU
REPAIRv1
REPAIRv2

FIG. 87



ADAR Mutant

FIG. 88

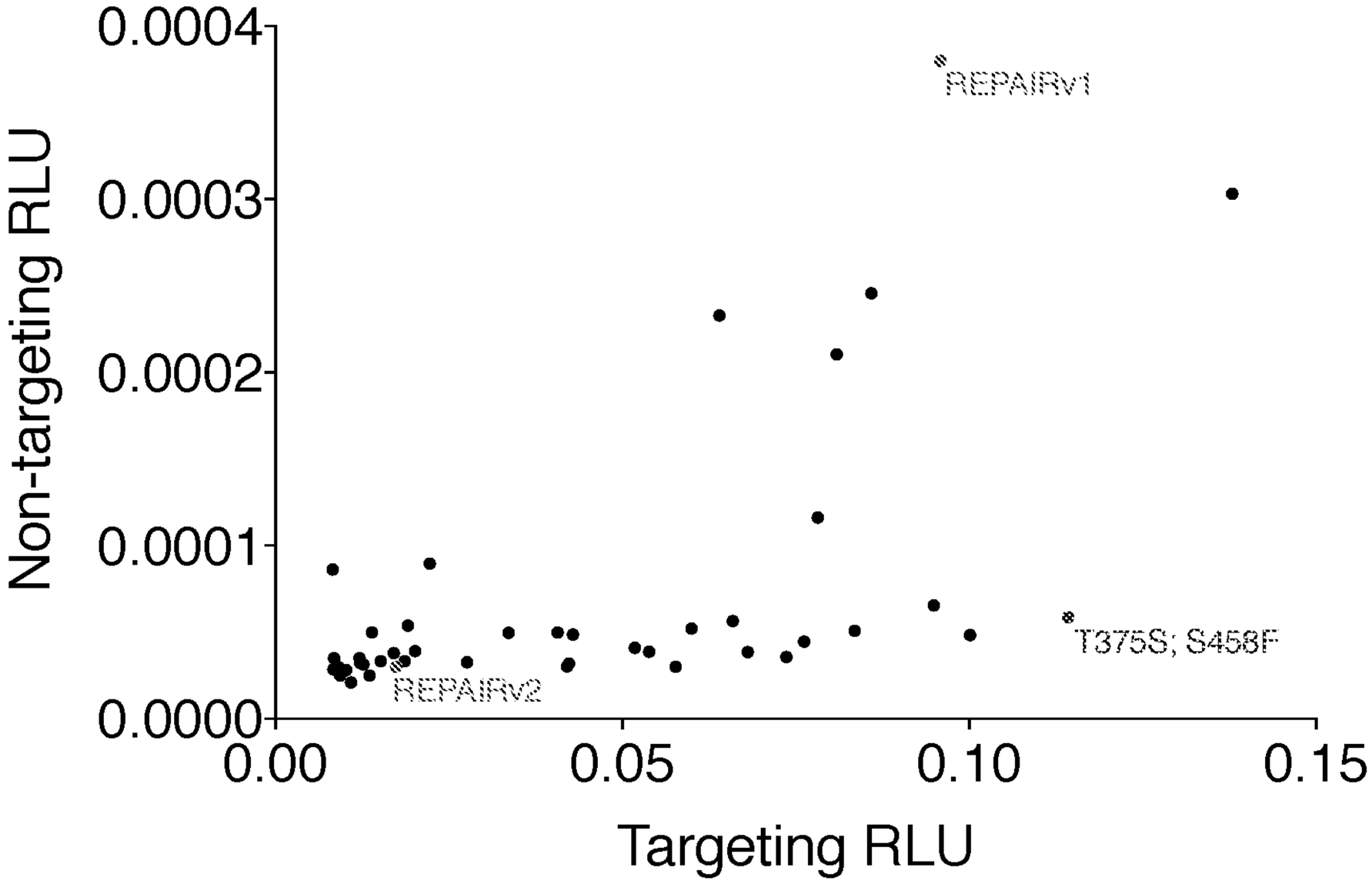

Targeting vs. Avg. Non-Targeting
Non-targeting RLU
Targeting RLU
0.0000
0.0001
0.0002
0.0003
0.0004
0.00
0.05
0.10
0.15
REPAIRv1
REPAIRv2
T375S; S458F

FIG. 89

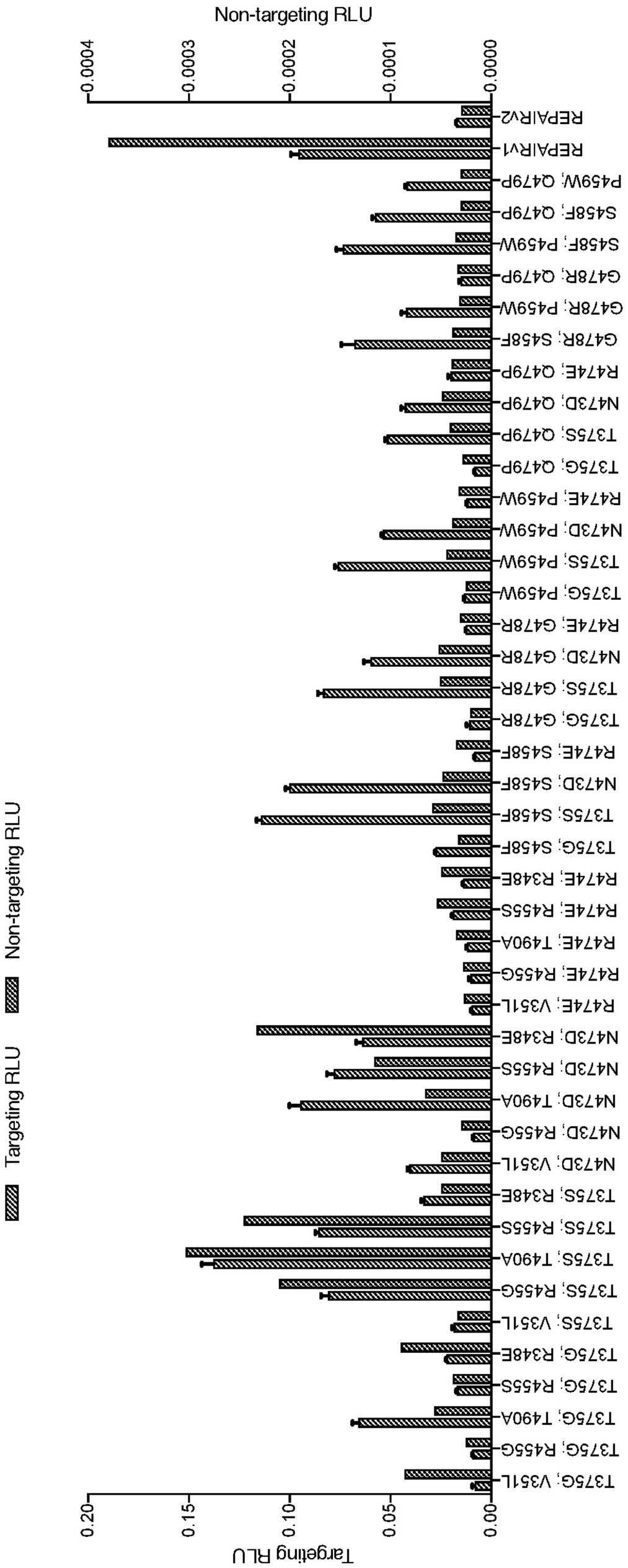

Non-targeting RLU
0.0004
0.0003
0.0002
0.0001
0.0000
REPAIRv2
REPAIRv1
P459W; Q479P
S458F; Q479P
S458F; P459W
G478R; Q479P
G478R; P459W
G478R; S458F
R474E; Q479P
N473D; Q479P
T375S; Q479P
T375G; Q479P
R474E; P459W
N473D; P459W
T375S; P459W
T375G; P459W
R474E; G478R
N473D; G478R
T375S; G478R
T375G; G478R
R474E; S458F
N473D; S458F
T375S; S458F
T375G; S458F
R474E; R348E
R474E; R455S
R474E; T490A
R474E; R455G
R474E; V351L
N473D; R348E
N473D; R455S
N473D; T490A
N473D; R455G
N473D; V351L
T375S; R348E
T375S; R455S
T375S; T490A
T375S; R455G
T375S; V351L
T375G; R348E
T375G; R455S
T375G; T490A
T375G; R455G
T375G; V351L
Targeting RLU
0.20
0.15
0.10
0.05
0.00
Targeting RLU
Non-targeting RLU
ADAR Mutant

FIG. 90

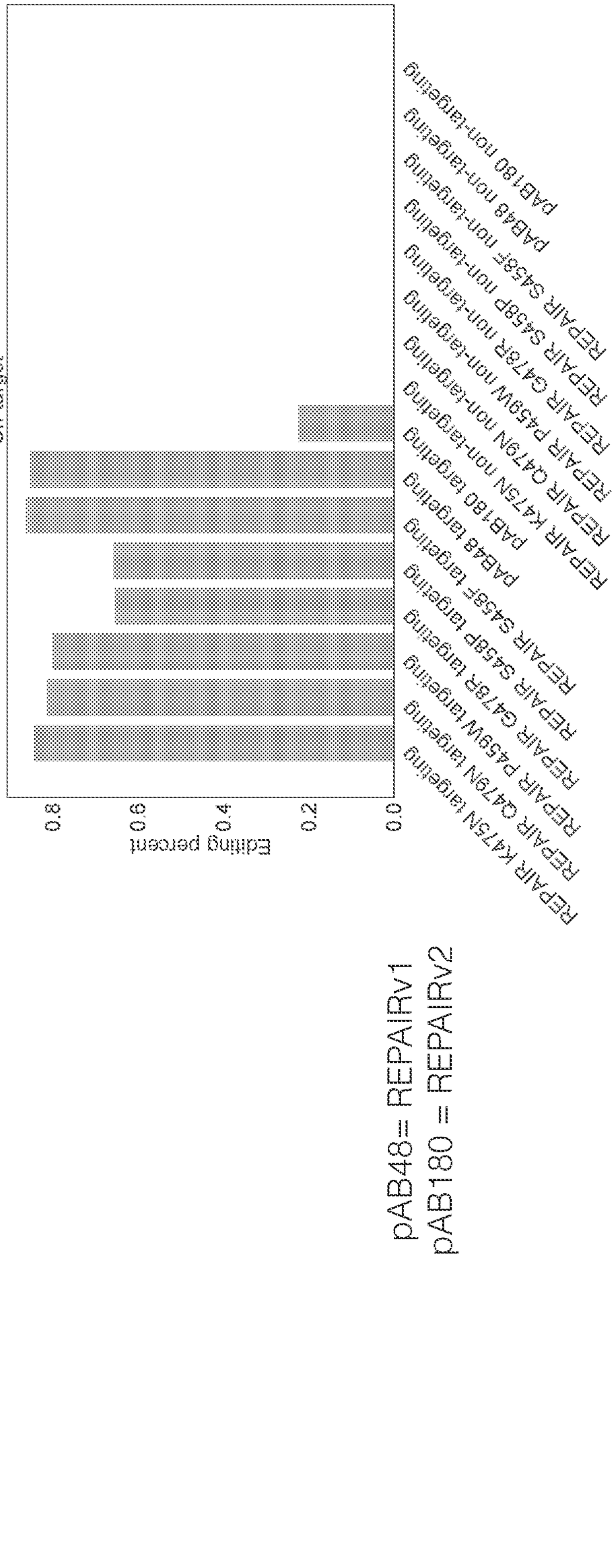

On-target
Editing percent
0.0
0.2
0.4
0.6
0.8
REPAIR K475N targeting
REPAIR C479N targeting
REPAIR P459W targeting
REPAIR G478R targeting
REPAIR S458P targeting
REPAIR S458F targeting
pAB48 targeting
pAB180 targeting
REPAIR K475N non-targeting
REPAIR C479N non-targeting
REPAIR P459W non-targeting
REPAIR G478R non-targeting
REPAIR S458P non-targeting
REPAIR S458F non-targeting
pAB48 non-targeting
pAB180 non-targeting
pAB48= REPAIRv1
pAB180 = REPAIRv2

FIG. 91

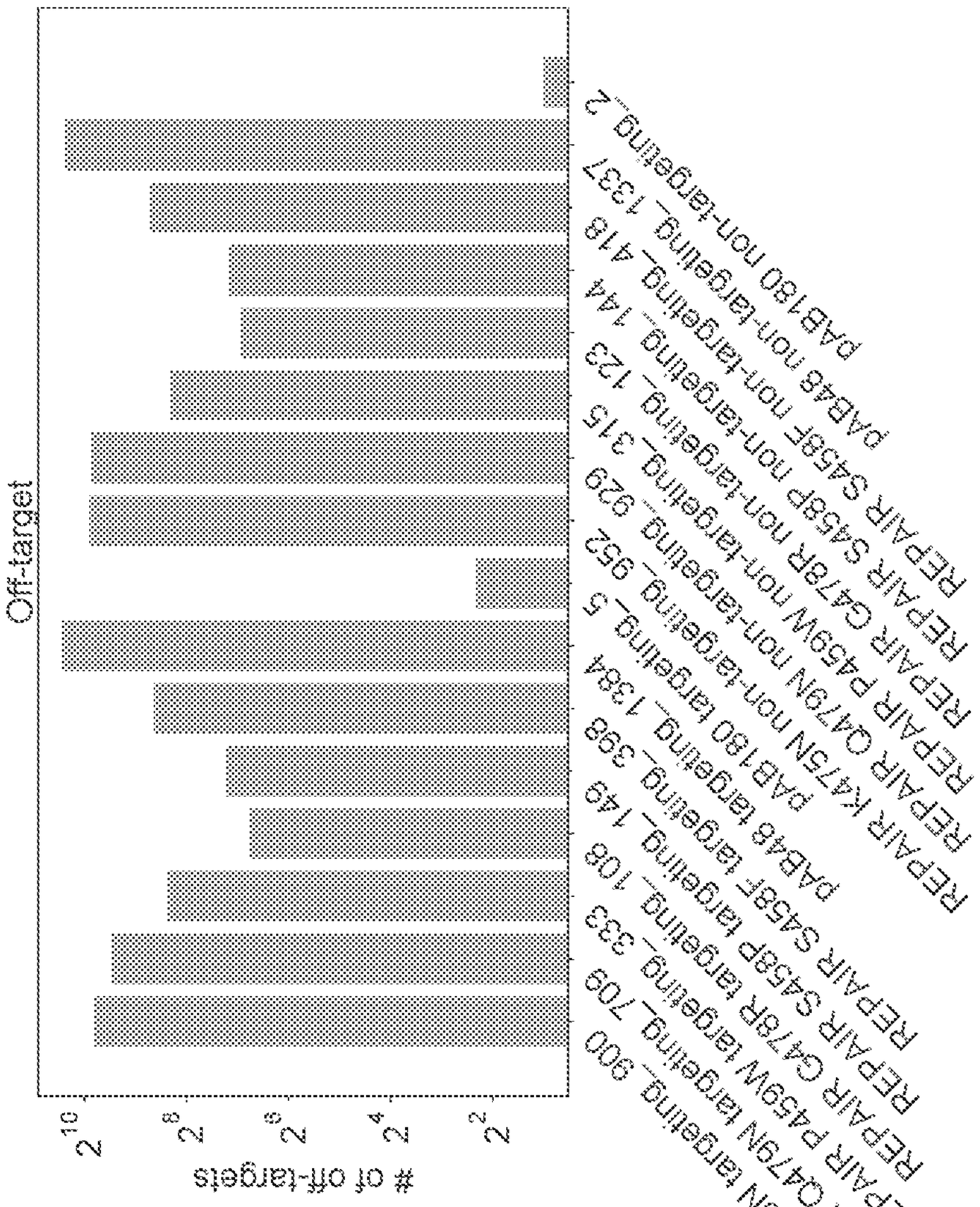

Off-target
# of off-targets
2^{10}
2^{8}
2^{6}
2^{4}
2^{2}
REPAIR K475N targeting_ 900
REPAIR Q479N targeting_ 709
REPAIR P459W targeting_ 333
REPAIR G478R targeting_ 123
REPAIR S458P targeting_ 108
REPAIR pAB48 targeting_ 149
REPAIR pAB180 targeting_ 398
REPAIR K475N non-targeting_ 952
REPAIR Q479N non-targeting_ 929
REPAIR P459W non-targeting_ 315
REPAIR G478R non-targeting_ 144
REPAIR S458P non-targeting_ 418
REPAIR pAB48 non-targeting_ 1337
REPAIR pAB180 non-targeting_ 2

FIG. 92

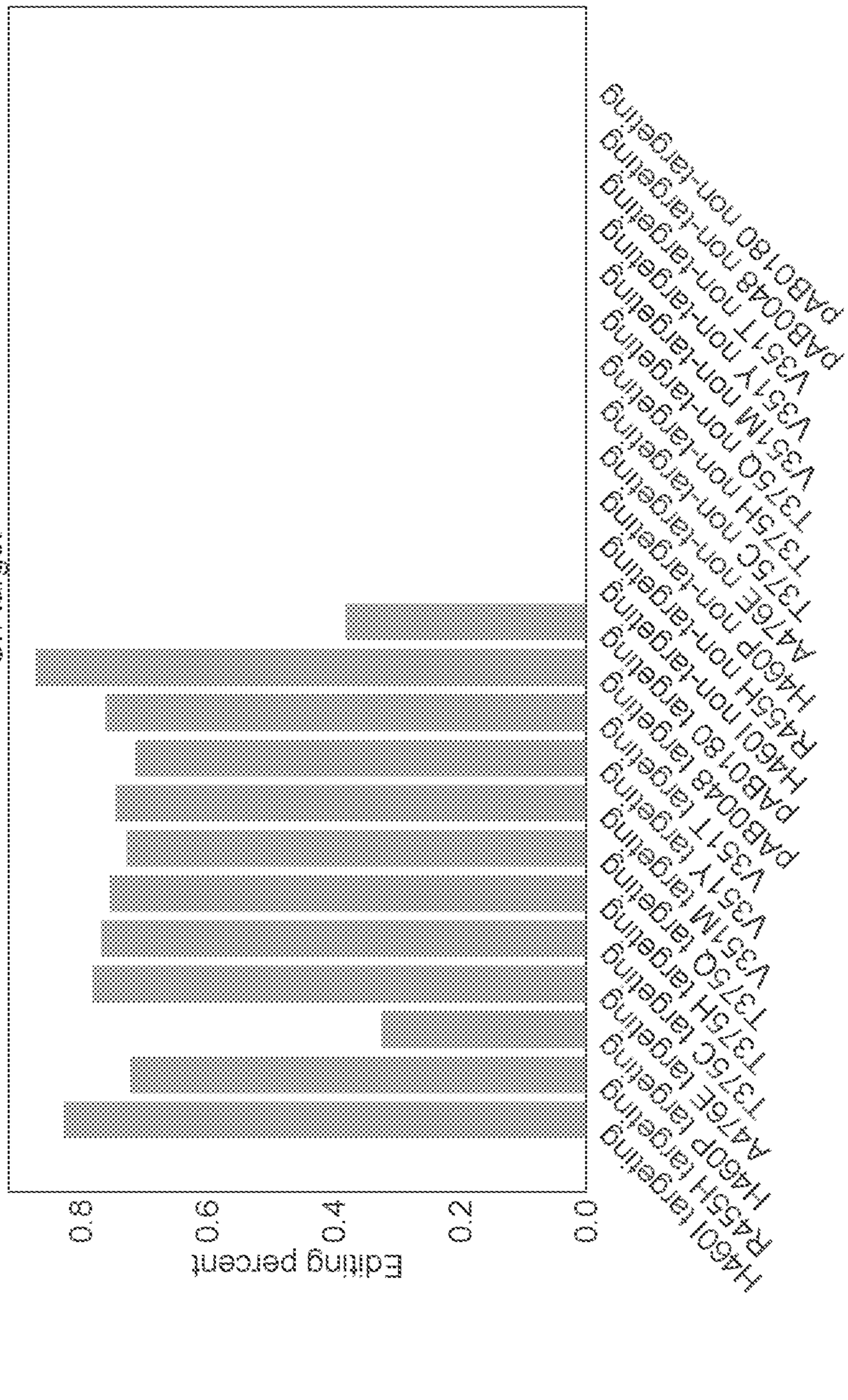

On-target
Editing percent
0.0
0.2
0.4
0.6
0.8
Bulgeless
Bulgeless non-targeting
pAB0180 non-targeting
pAB0048 non-targeting
V351Y non-targeting
V351M non-targeting
T375H non-targeting
T375C non-targeting
A476E non-targeting
A476P non-targeting
R455H non-targeting
H460I non-targeting
pAB0048 targeting
H460I targeting
R455H targeting
H460P targeting
A476E targeting
T375C targeting
T375H targeting
V351M targeting
V351Y targeting

FIG. 93

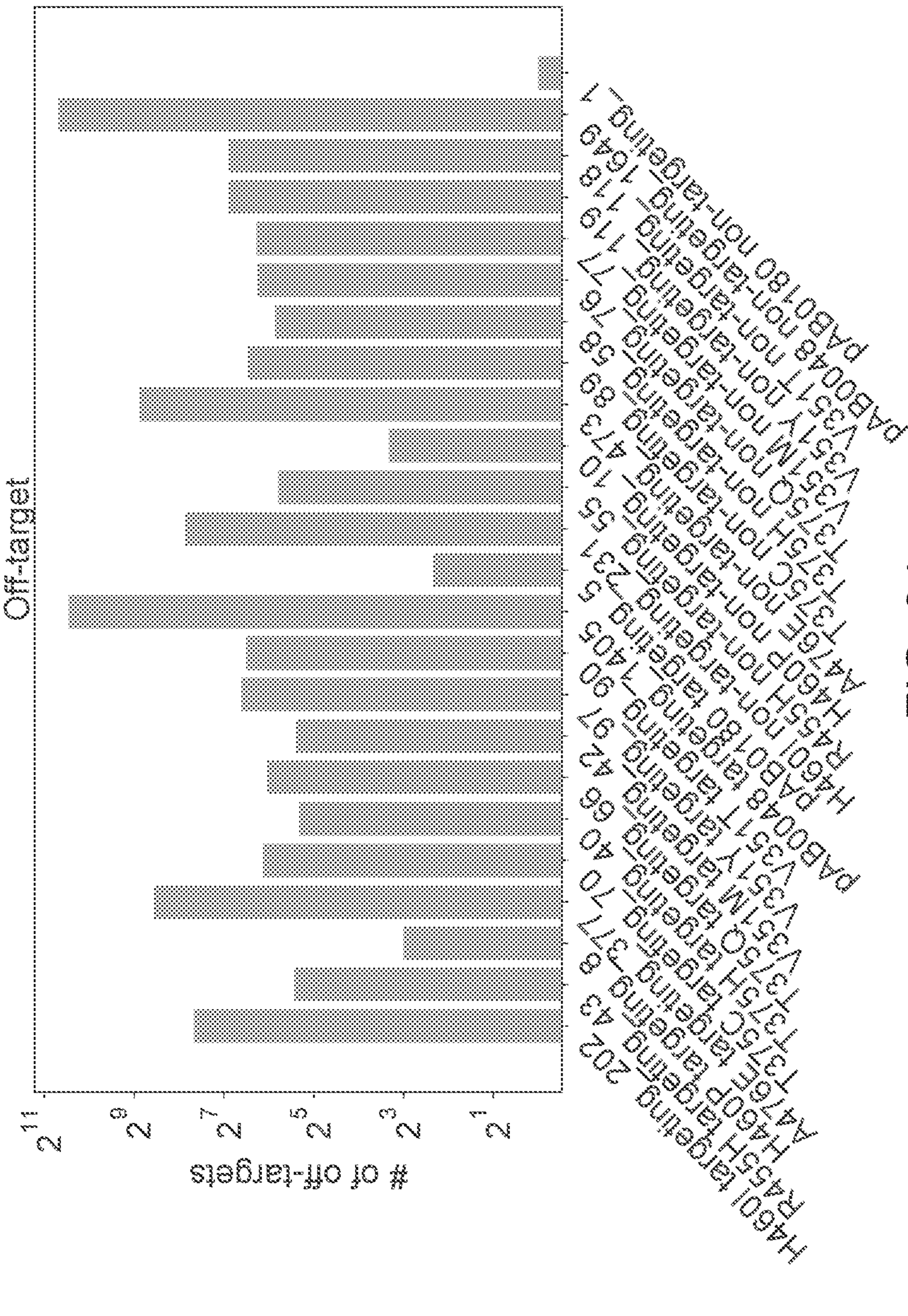

Off-target
# of off-targets
2^11
2^9
2^7
2^5
2^3
2^1
pAB00A8 non-targeting 1 49
pAB00A8 non-targeting 1 42
pAB00A8 non-targeting 1 69
pAB00A8 non-targeting 1 78
pAB00A8 non-targeting 1 79
pAB00A8 non-targeting 1 80
pAB00A8 non-targeting 1 97
pAB00A8 non-targeting 1 90
pAB00A8 non-targeting 1 105
pAB00A8 non-targeting 1 95
pAB00A8 non-targeting 1 98
pAB00A8 non-targeting 1 70
pAB00A8 non-targeting 1 37
pAB00A8 non-targeting 1 43
pAB00A8 non-targeting 1 54
pAB00A8 non-targeting 1 55
pAB00A8 non-targeting 1 23
pAB00A8 non-targeting 1 40
pAB00A8 non-targeting 1 49
pAB00A8 non-targeting 1 66
pAB00A8 non-targeting 1 42
pAB00A8 non-targeting 1 37
pAB00A8 non-targeting 1 70
pAB00A8 non-targeting 1 43
H460 targeting Barge1 202
R453H A460m Barge1 T37S/S53N/V35I
A549 T37S/S53H Barge1 T37S/S53N/V35I
H460 targeting Barge1 202
R453H A460m Barge1 T37S/S53N/V35I
A549 T37S/S53H Barge1 T37S/S53N/V35I

FIG. 94

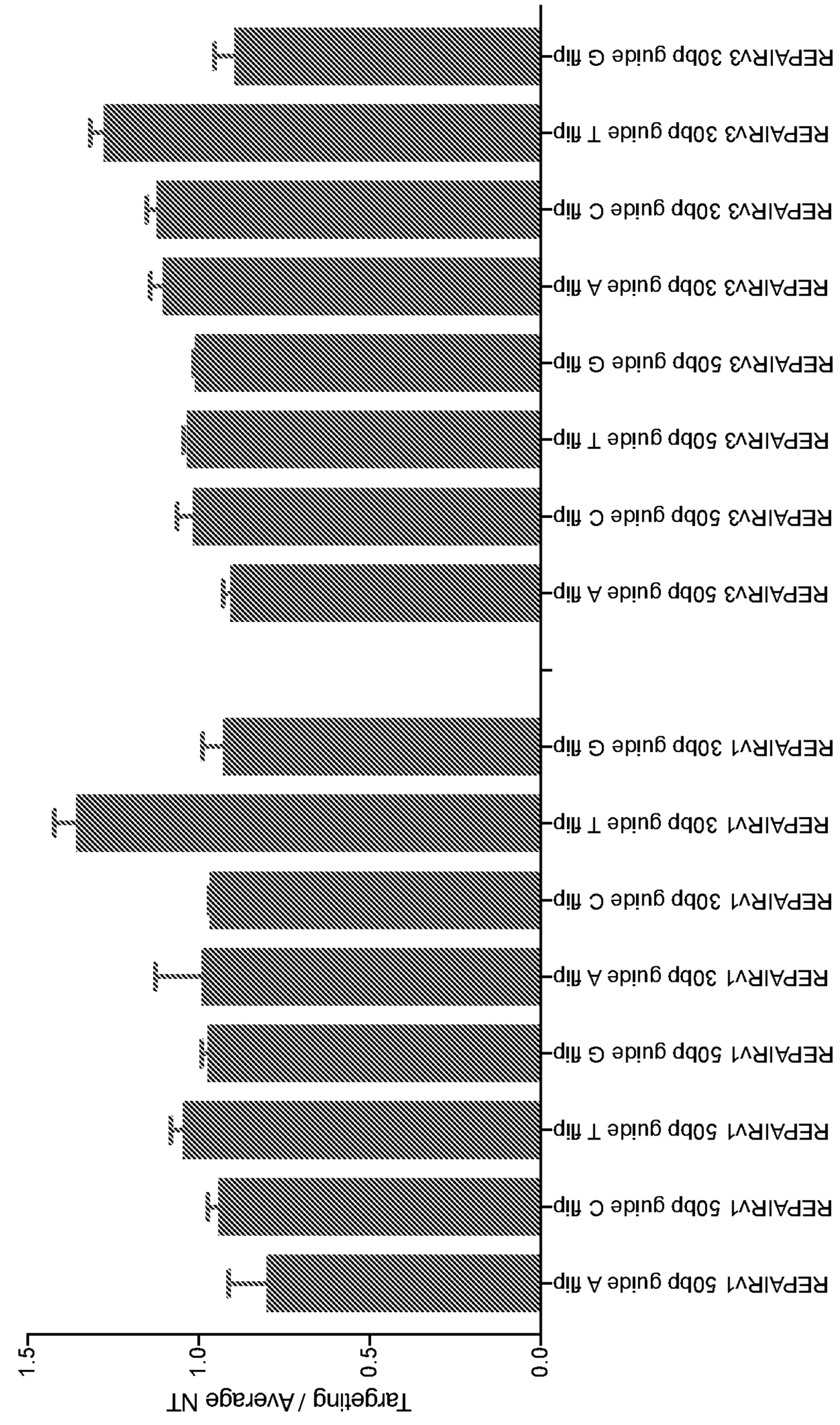

Targeting / Average NT
0.0
0.5
1.0
1.5
REPAIRv1 50bp guide A flip
REPAIRv1 50bp guide C flip
REPAIRv1 50bp guide T flip
REPAIRv1 50bp guide G flip
REPAIRv1 30bp guide A flip
REPAIRv1 30bp guide C flip
REPAIRv1 30bp guide T flip
REPAIRv1 30bp guide G flip
REPAIRv3 50bp guide A flip
REPAIRv3 50bp guide C flip
REPAIRv3 50bp guide T flip
REPAIRv3 50bp guide G flip
REPAIRv3 30bp guide A flip
REPAIRv3 30bp guide C flip
REPAIRv3 30bp guide T flip
REPAIRv3 30bp guide G flip
Construct-Guide Pairing

FIG. 95

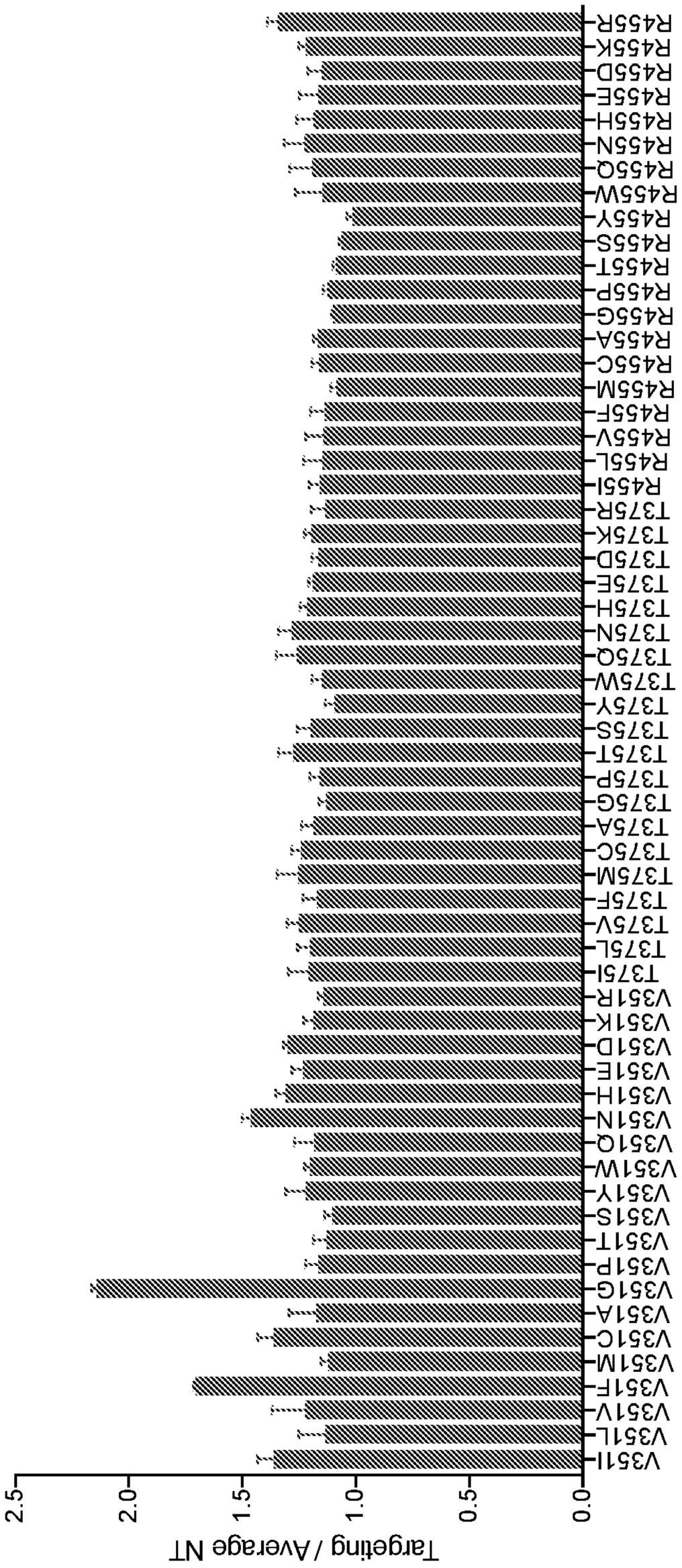

R455R
R455K
R455D
R455E
R455H
R455N
R455Q
R455W
R455Y
R455S
R455T
R455P
R455G
R455A
R455C
R455M
R455F
R455V
R455L
R455I
T375R
T375K
T375D
T375E
T375H
T375N
T375Q
T375W
T375Y
T375S
T375T
T375P
T375G
T375A
T375C
T375M
T375F
T375V
T375L
T375I
V351R
V351K
V351D
V351E
V351H
V351N
V351Q
V351W
V351Y
V351S
V351T
V351P
V351G
V351A
V351C
V351M
V351F
V351V
V351L
V351I
ADAR mutant
2.5
2.0
1.5
1.0
0.5
0.0
Targeting / Average NT

FIG. 96

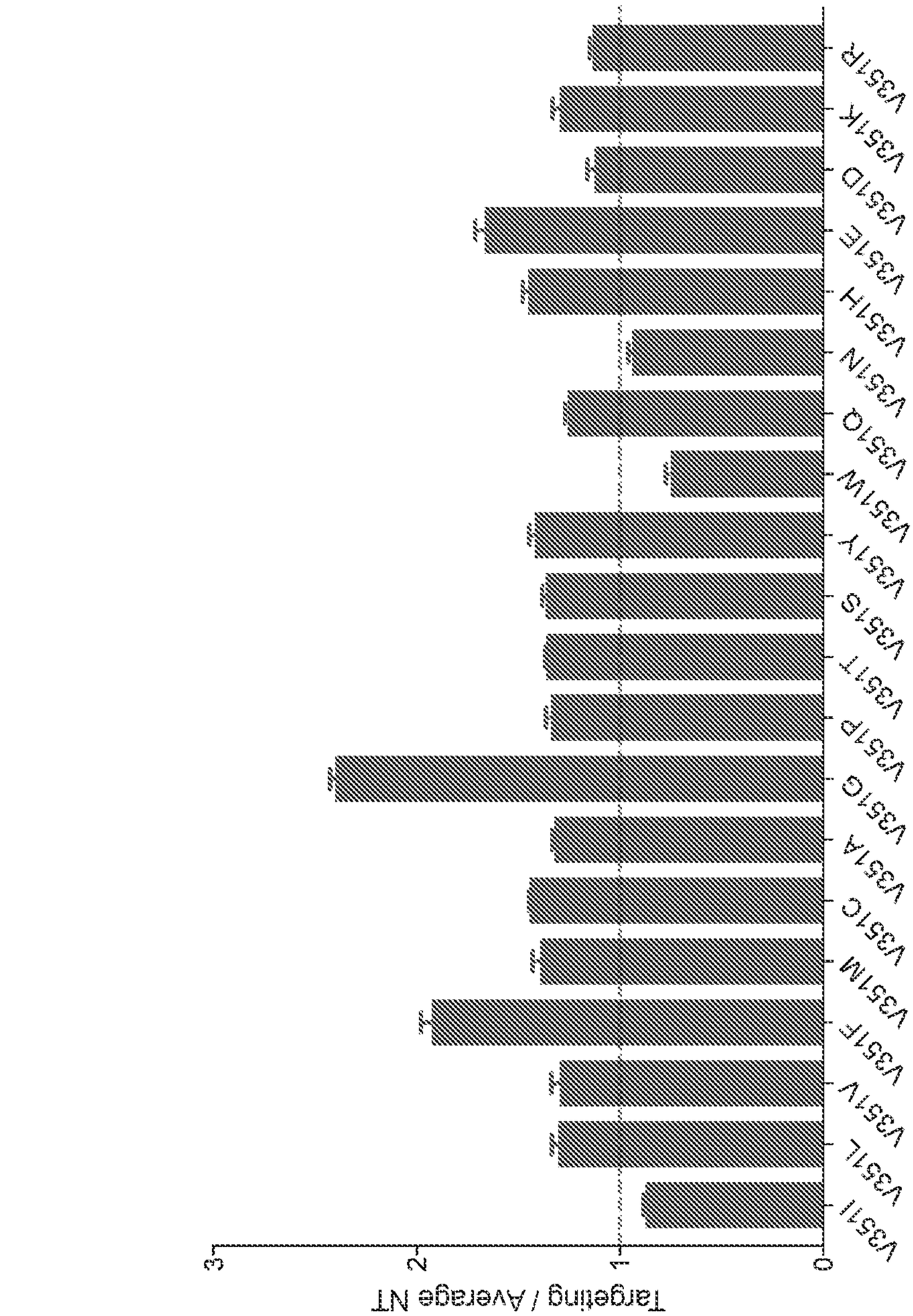

V35A
V35C
V35D
V35E
V35F
V35G
V35H
V35I
V35K
V35L
V35M
V35N
V35P
V35Q
V35R
V35S
V35T
V35V
V35W
V35Y
0
1
2
3
Targeting / Average NT
Mutant on E488Q background

FIG. 97

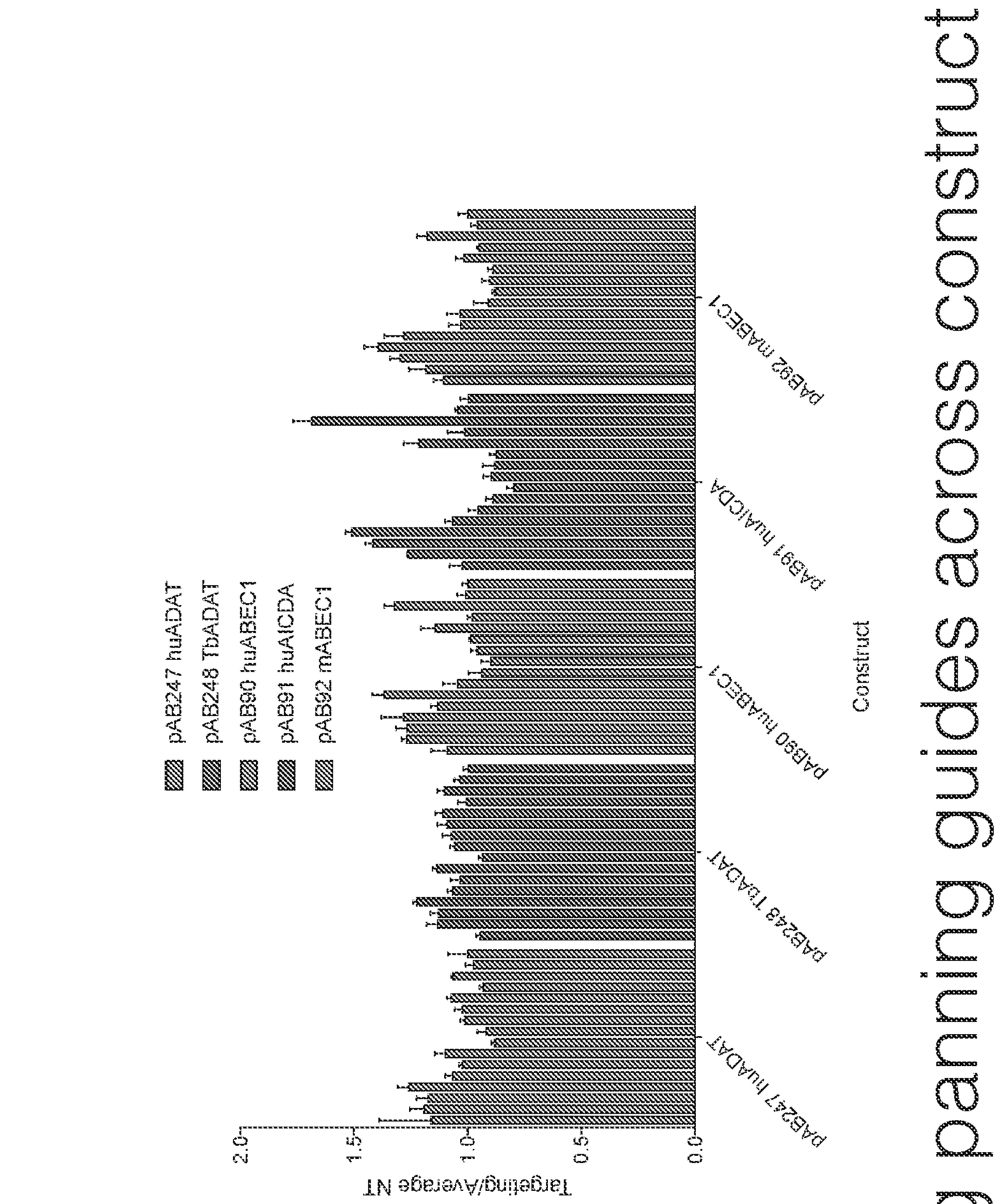

Construct
Targeting/Average NT
2.0
1.5
1.0
0.5
0.0
pAB247 huADAT
pAB248 TbADAT
pAB90 huABEC1
pAB91 huAICDA
pAB92 mABEC1
pAB247 huADAT
pAB248 TbADAT
pAB90 huABEC1
pAB91 huAICDA
pAB92 mABEC1
Testing panning guides across constructs

FIG. 9B

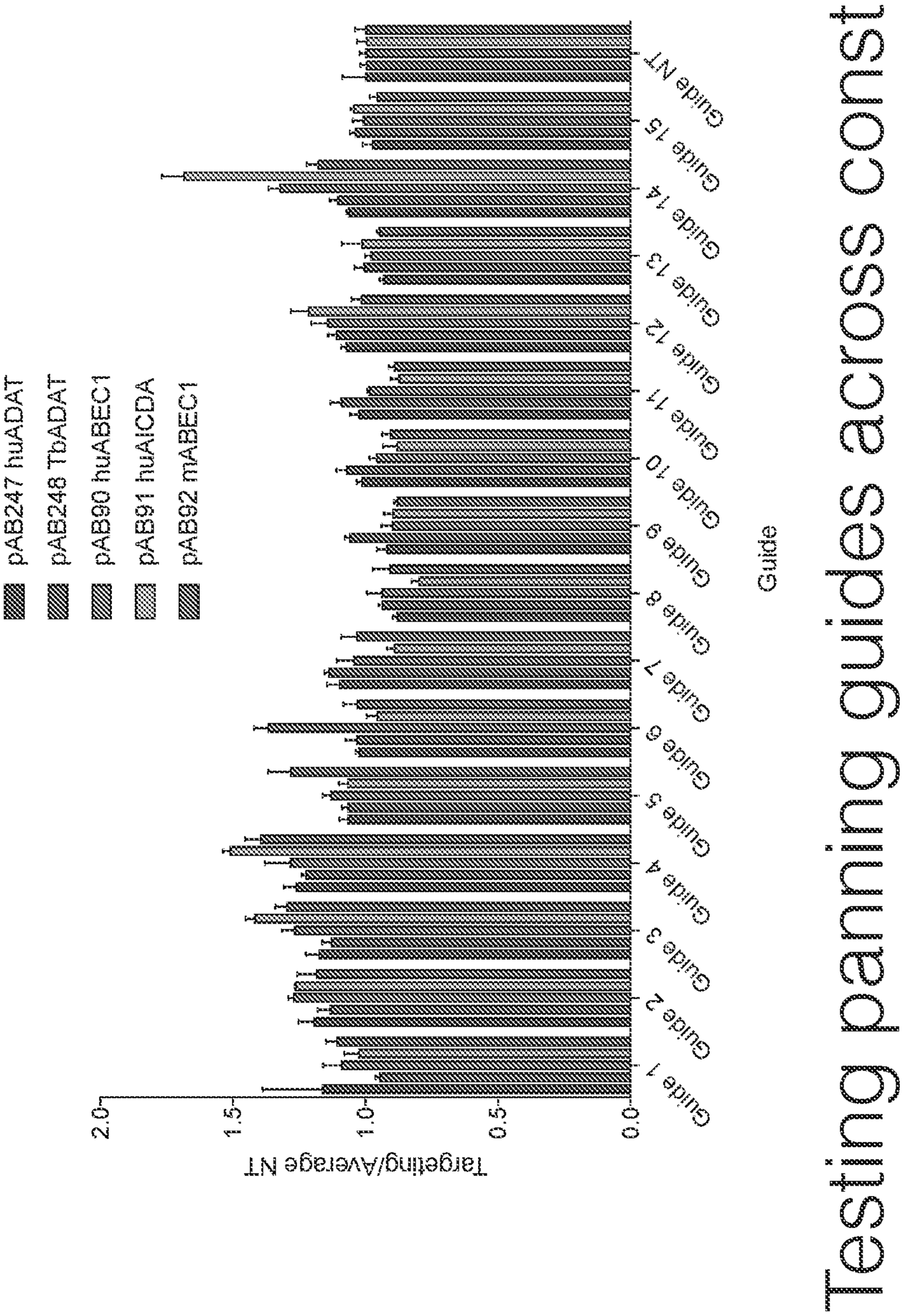

pAB247 huADAT
pAB248 TbADAT
pAB90 huABEC1
pAB91 huAICDA
pAB92 mABEC1
Targeting/Average NT
2.0
1.5
1.0
0.5
0.0
Guide 1
Guide 2
Guide 3
Guide 4
Guide 5
Guide 6
Guide 7
Guide 8
Guide 9
Guide 10
Guide 11
Guide 12
Guide 13
Guide 14
Guide 15
Guide NT
Guide
guides across construct

FIG. 99

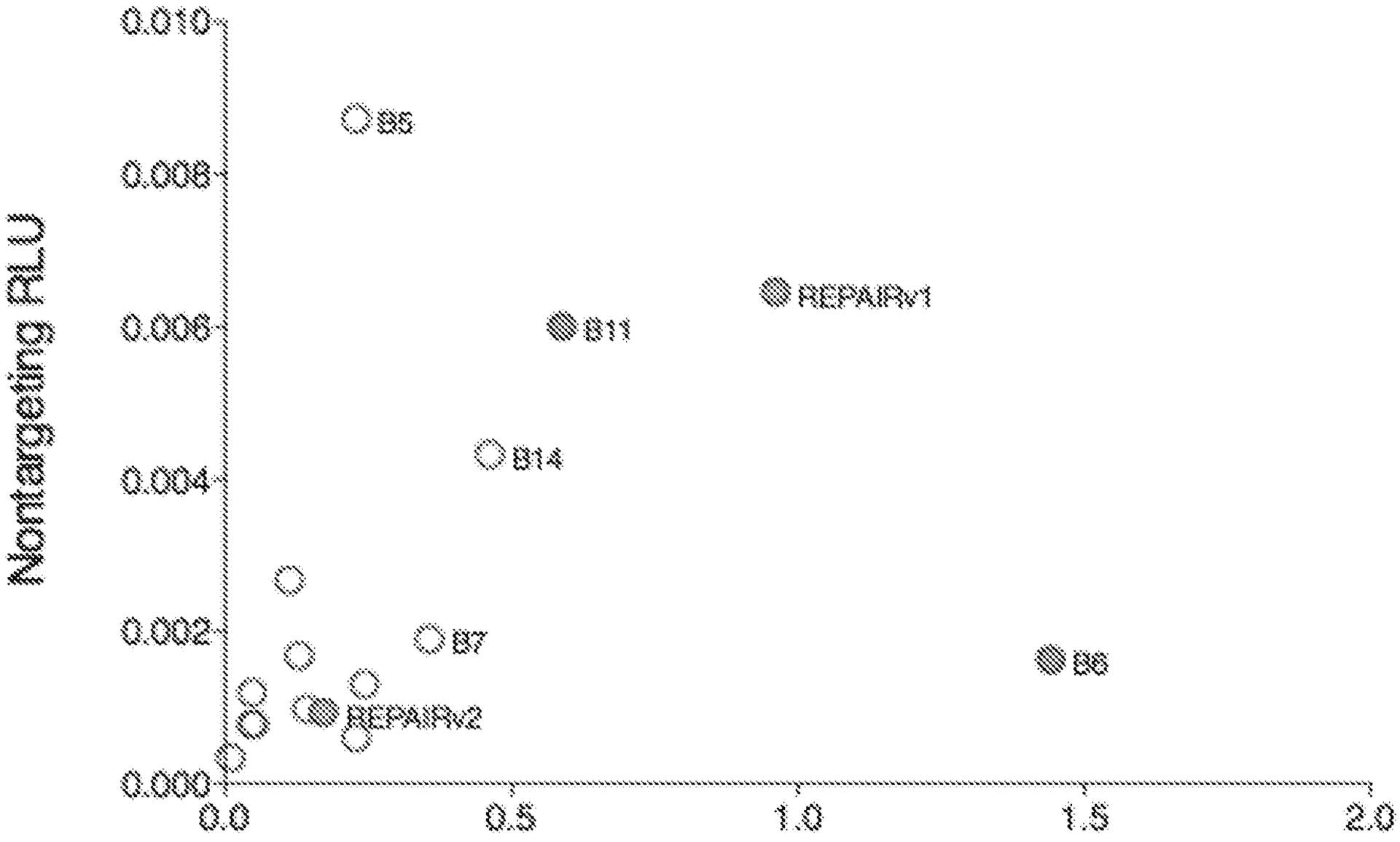

0.010
0.008
0.006
0.004
0.002
0.000
Nontargeting RLU
B5
REPAIRv1
B11
B14
B7
B6
REPAIRv2
0.0
0.5
1.0
1.5
2.0
Targeting RLU

FIG. 100

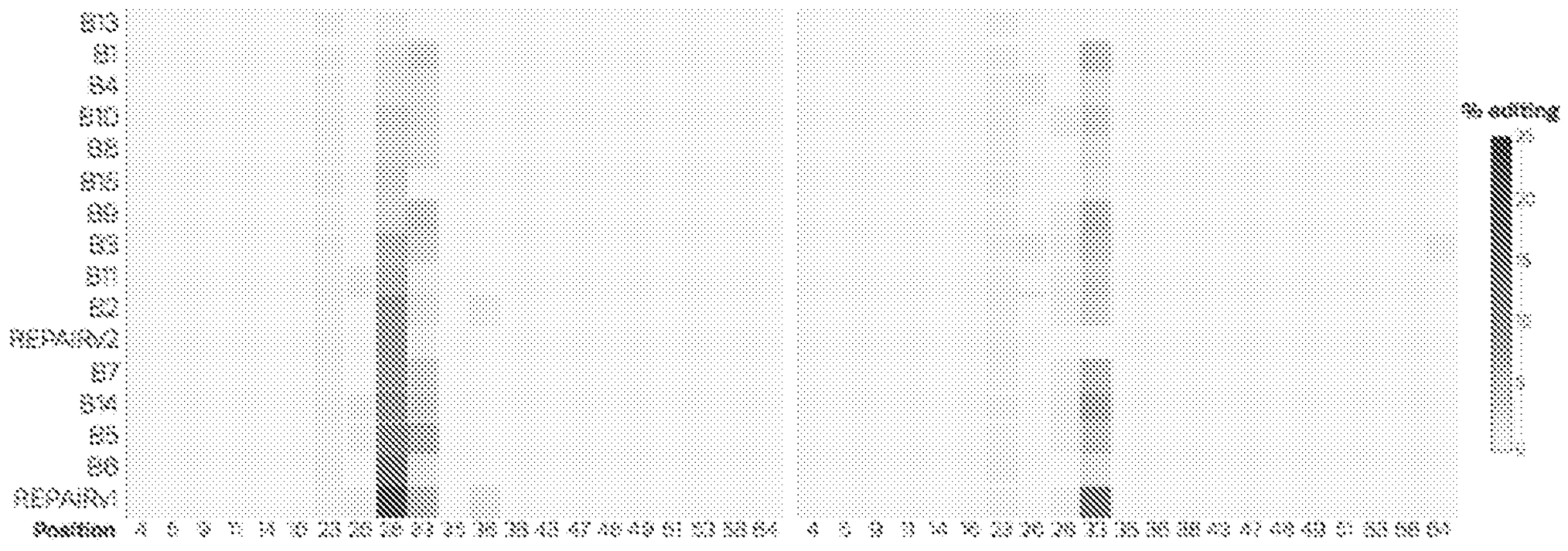

B10
B9
B8
B7
B6
B5
B4
B3
B2
REPAIR2
B7
B6
B5
B6
REPAIR1
Position
% editing
20
15
10
5
0

FIG. 101

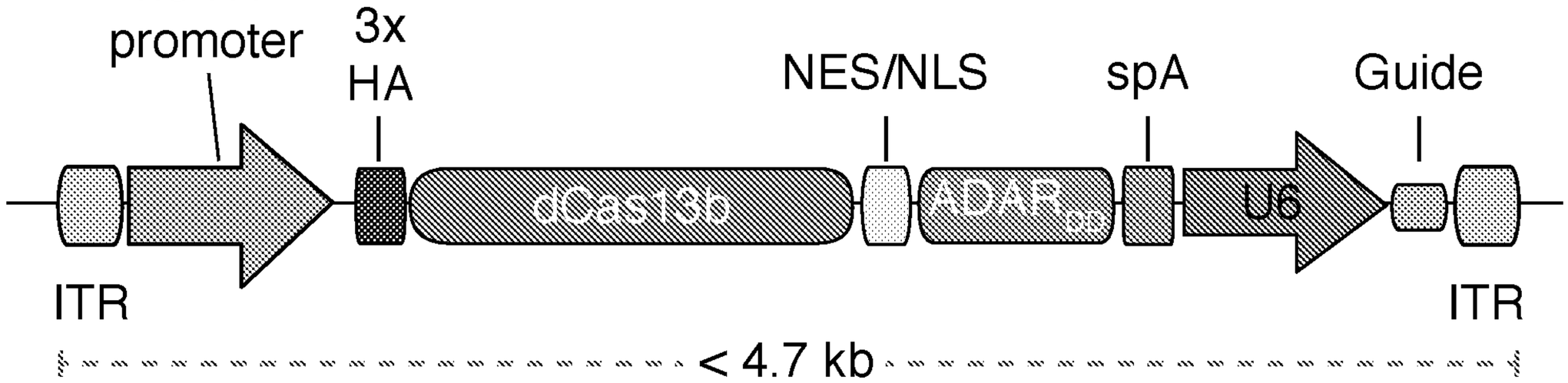

hSYN
promoter
3x
HA
NES/NLS
spA
Guide
dCas13b
ADAR
DD
U6
ITR
ITR
< 4.7 kb

FIG. 102

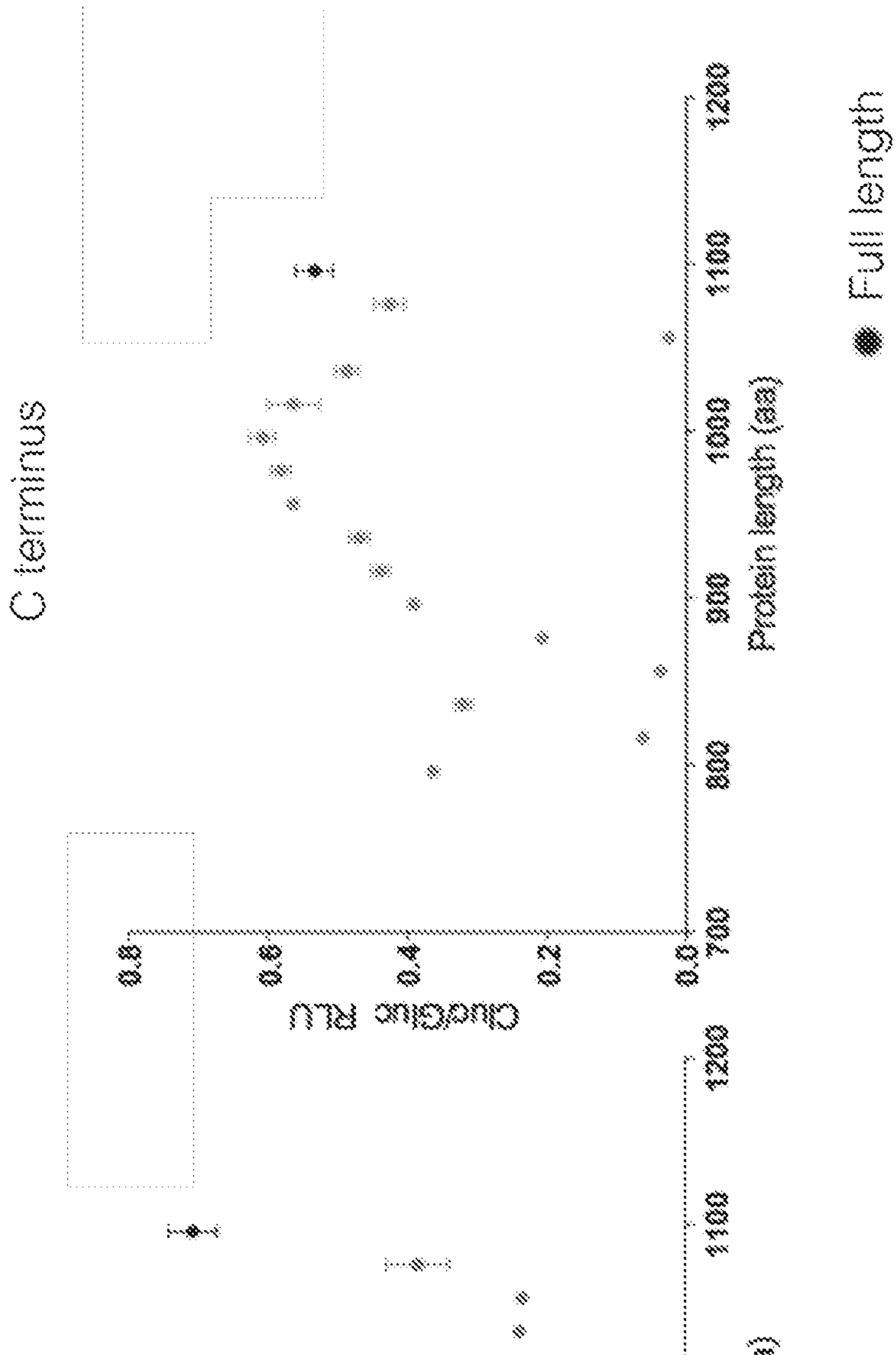


FIG. 103B

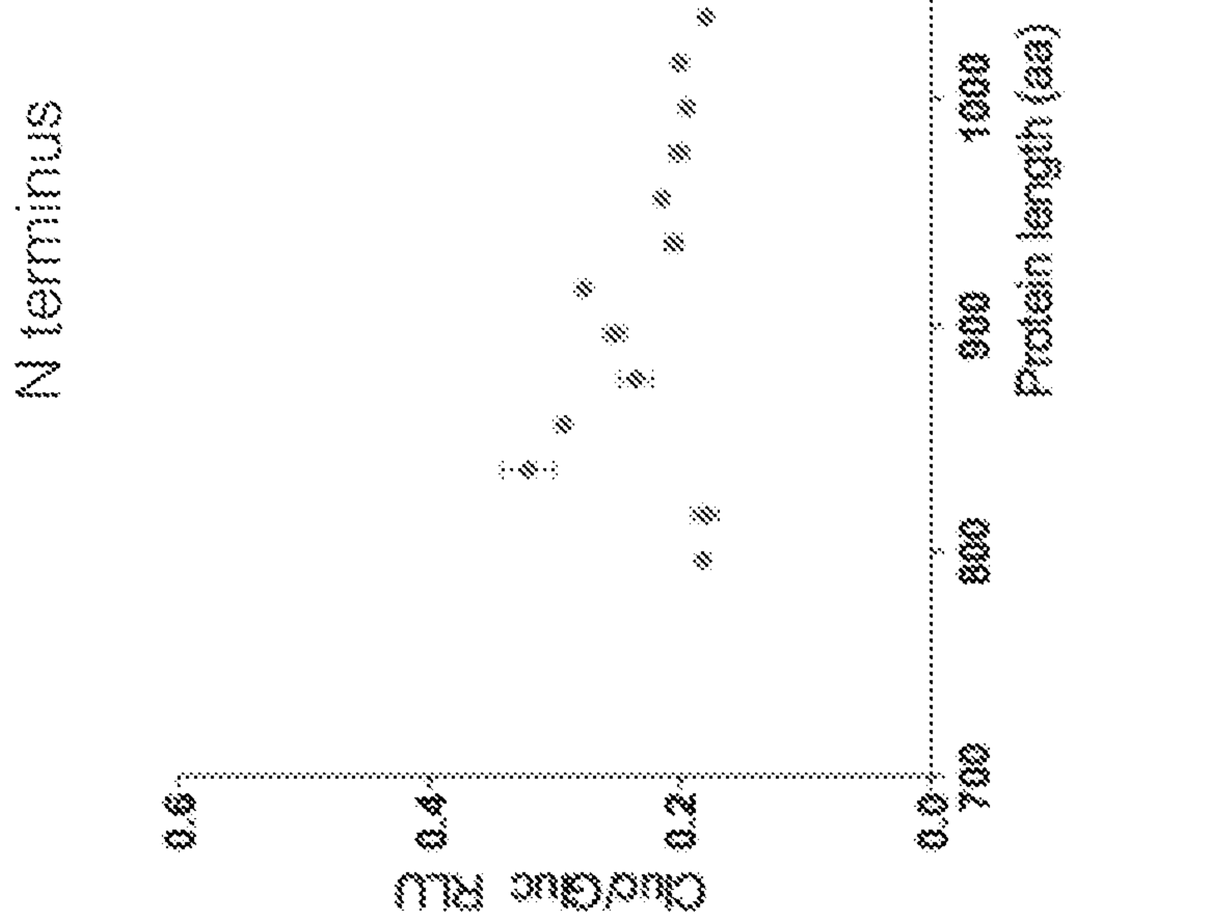


FIG. 103A

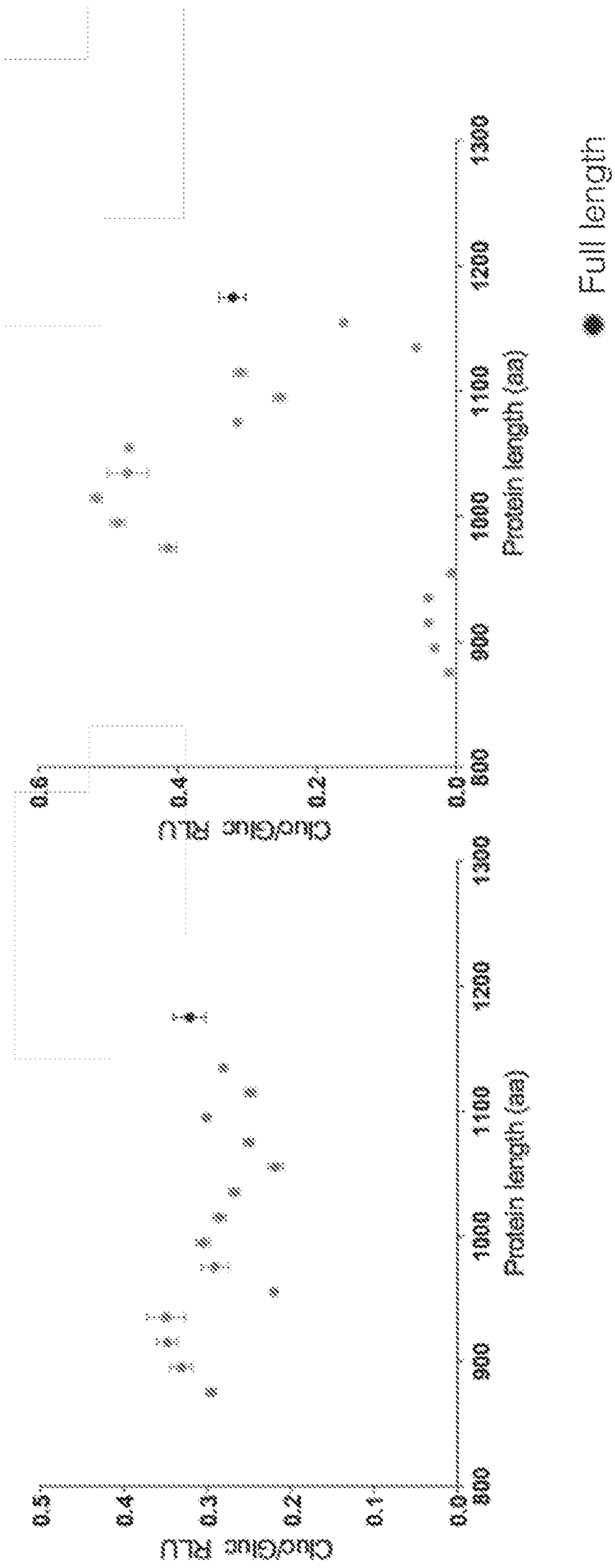


FIG. 104B

FIG. 104A

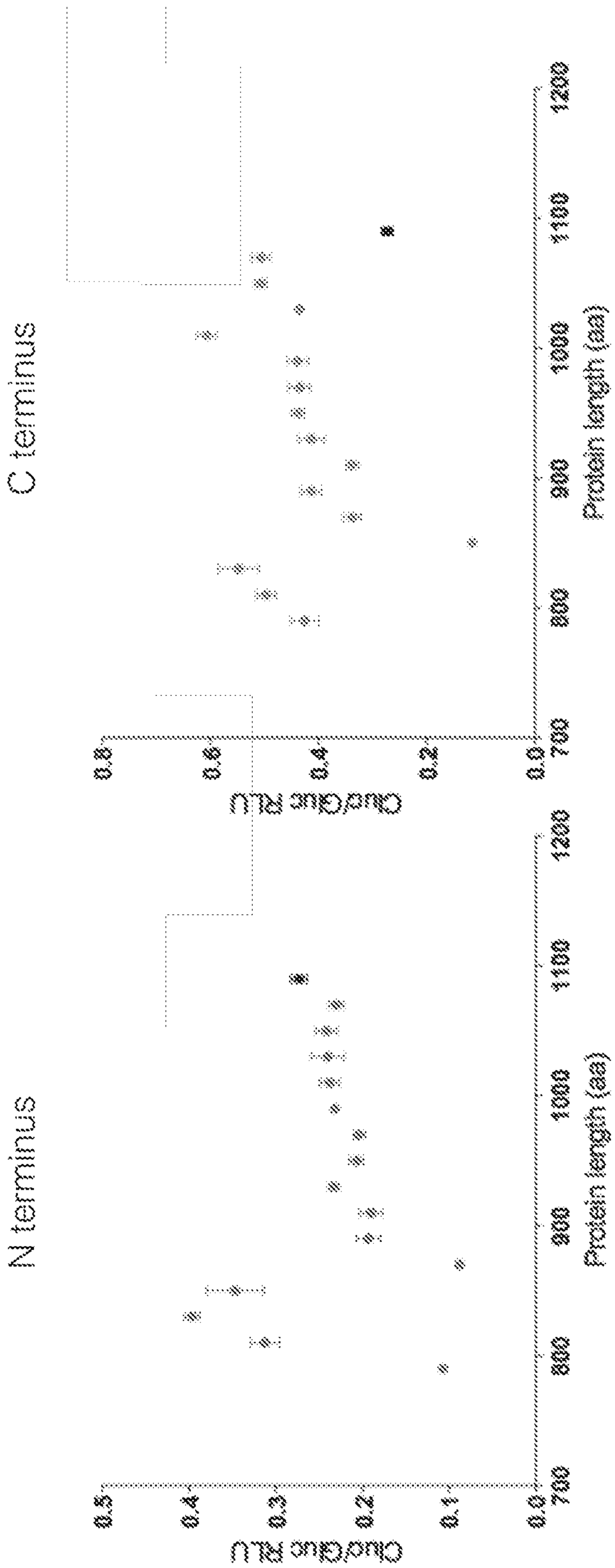


FIG. 105B

FIG. 105A

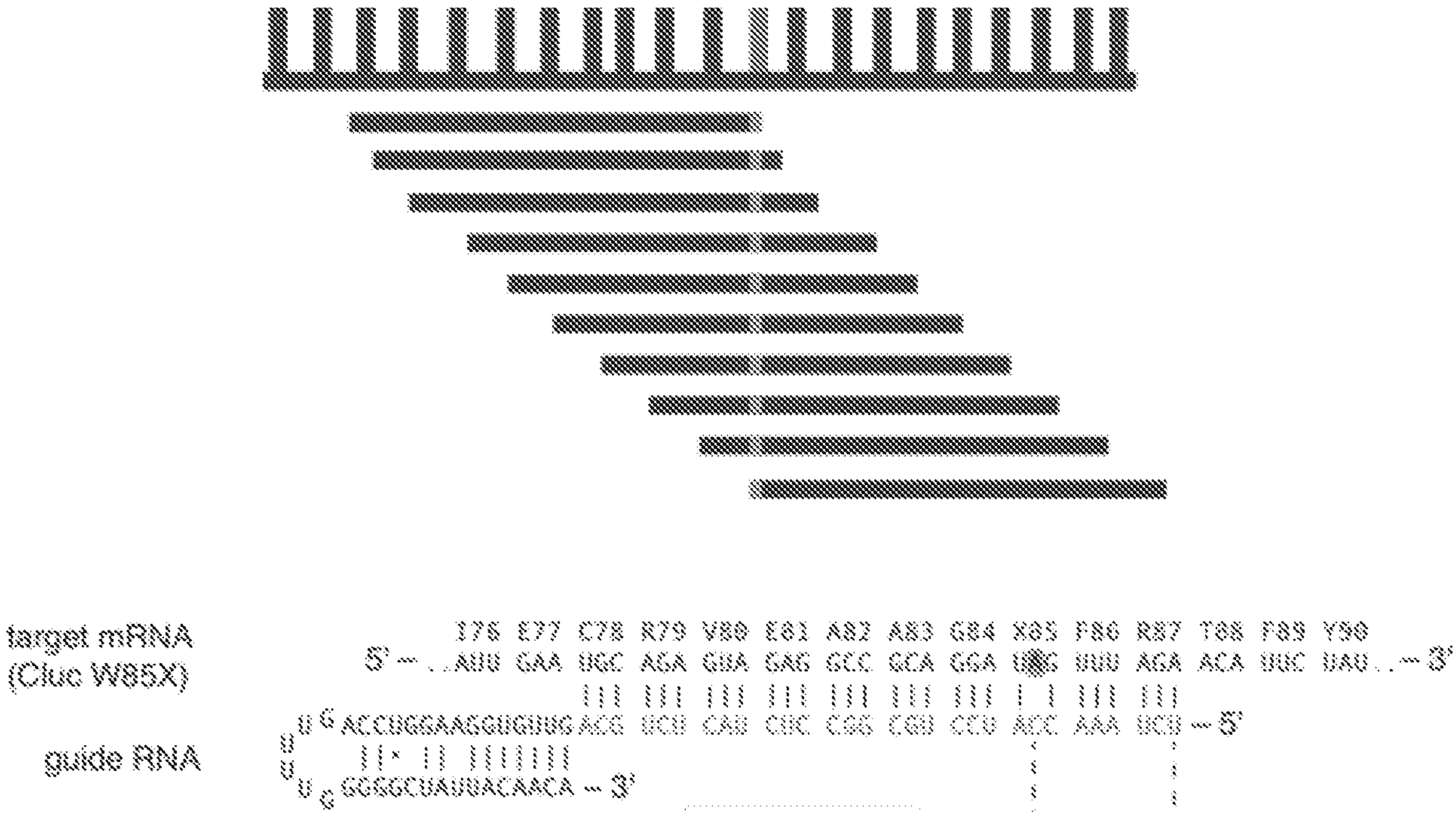


FIG. 106

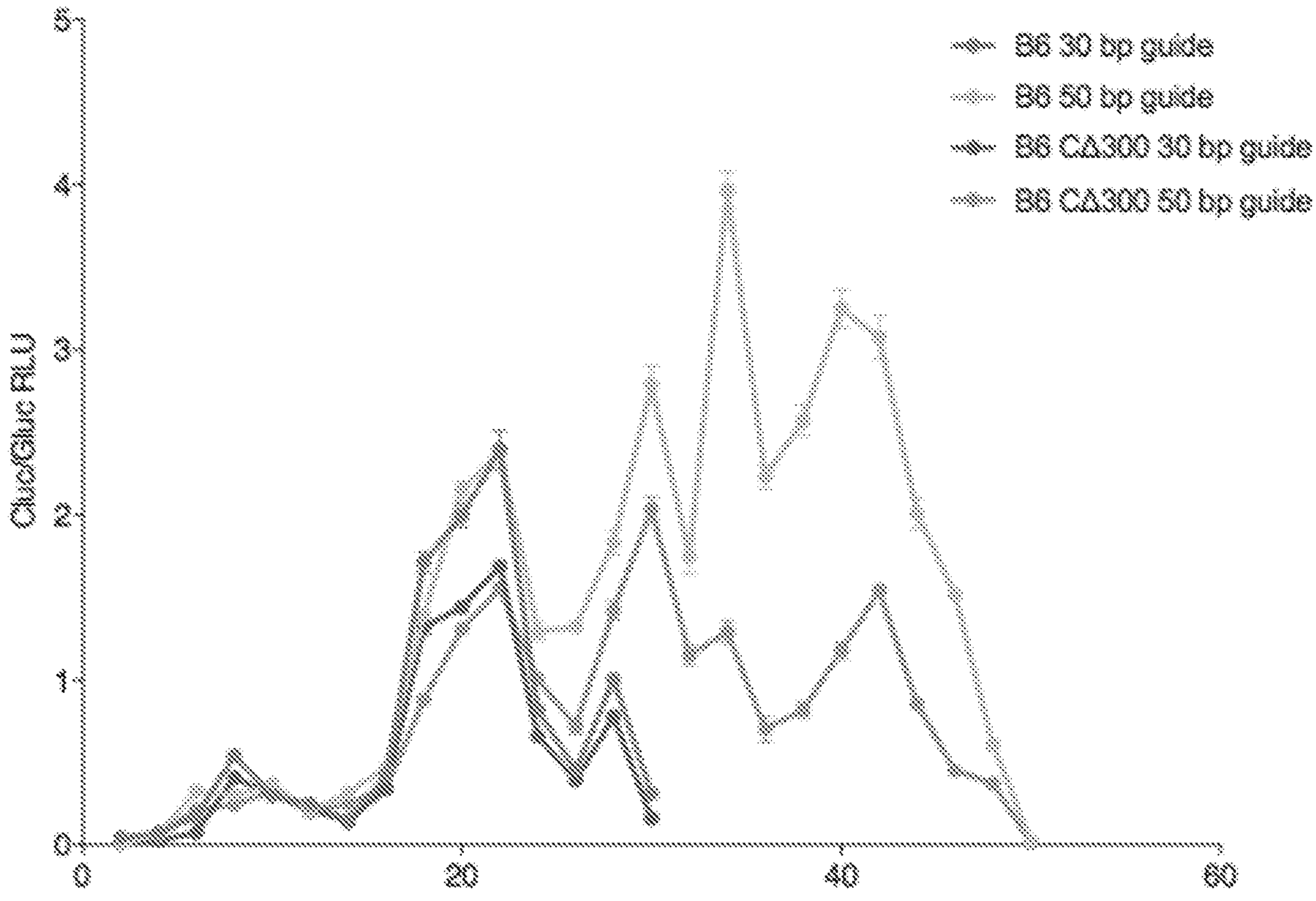

B6 30 bp guide
B6 50 bp guide
B6 CΔ300 30 bp guide
B6 CΔ300 50 bp guide
Cas9cis RLU
0
1
2
3
4
0
20
40
60
Base flip position

FIG. 107

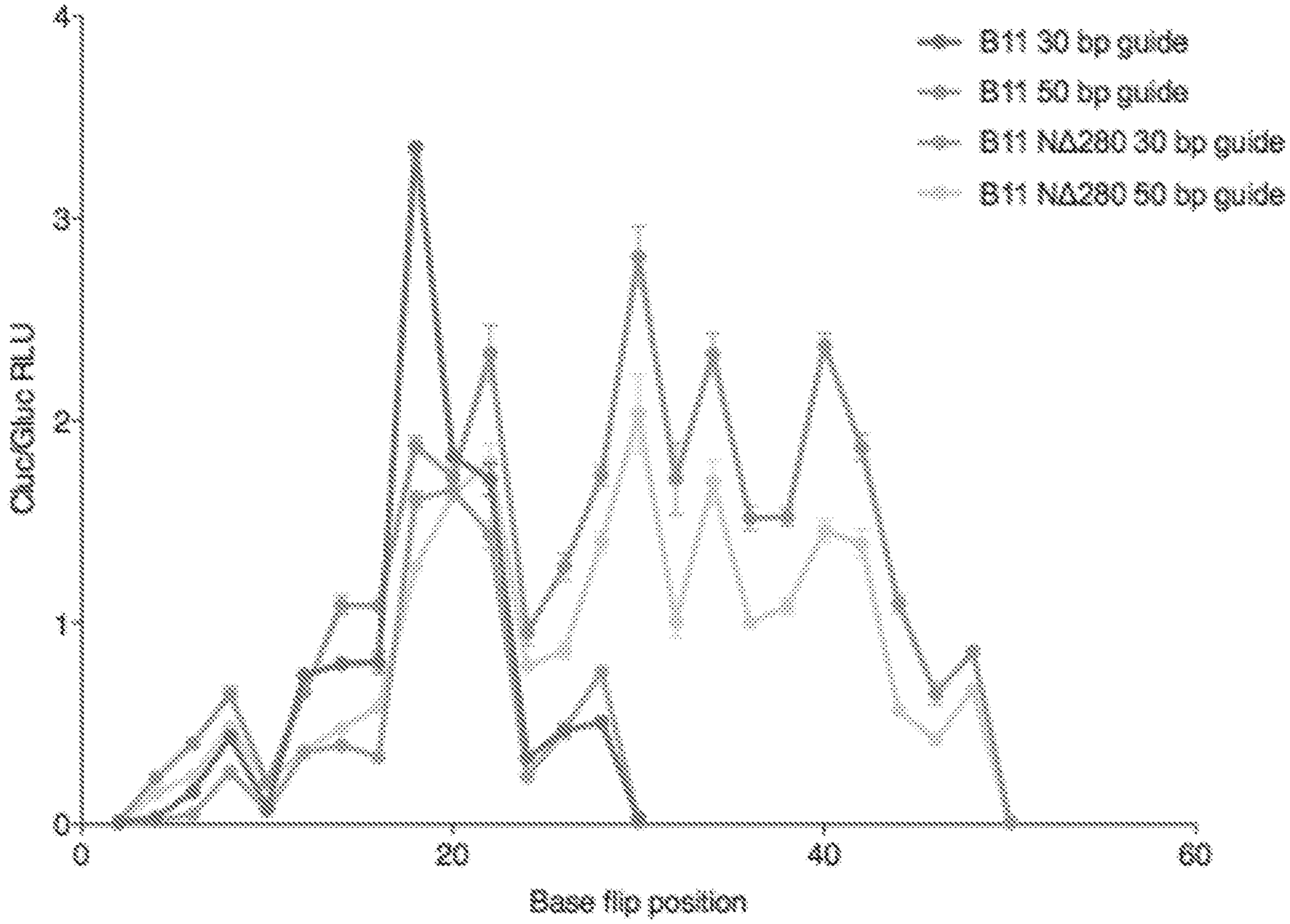

B11 30 bp guide
B11 50 bp guide
B11 NΔ280 30 bp guide
B11 NΔ280 50 bp guide
Cas9/Glue RLU
0
1
2
3
4
0
20
40
60
Base flip position

FIG. 108

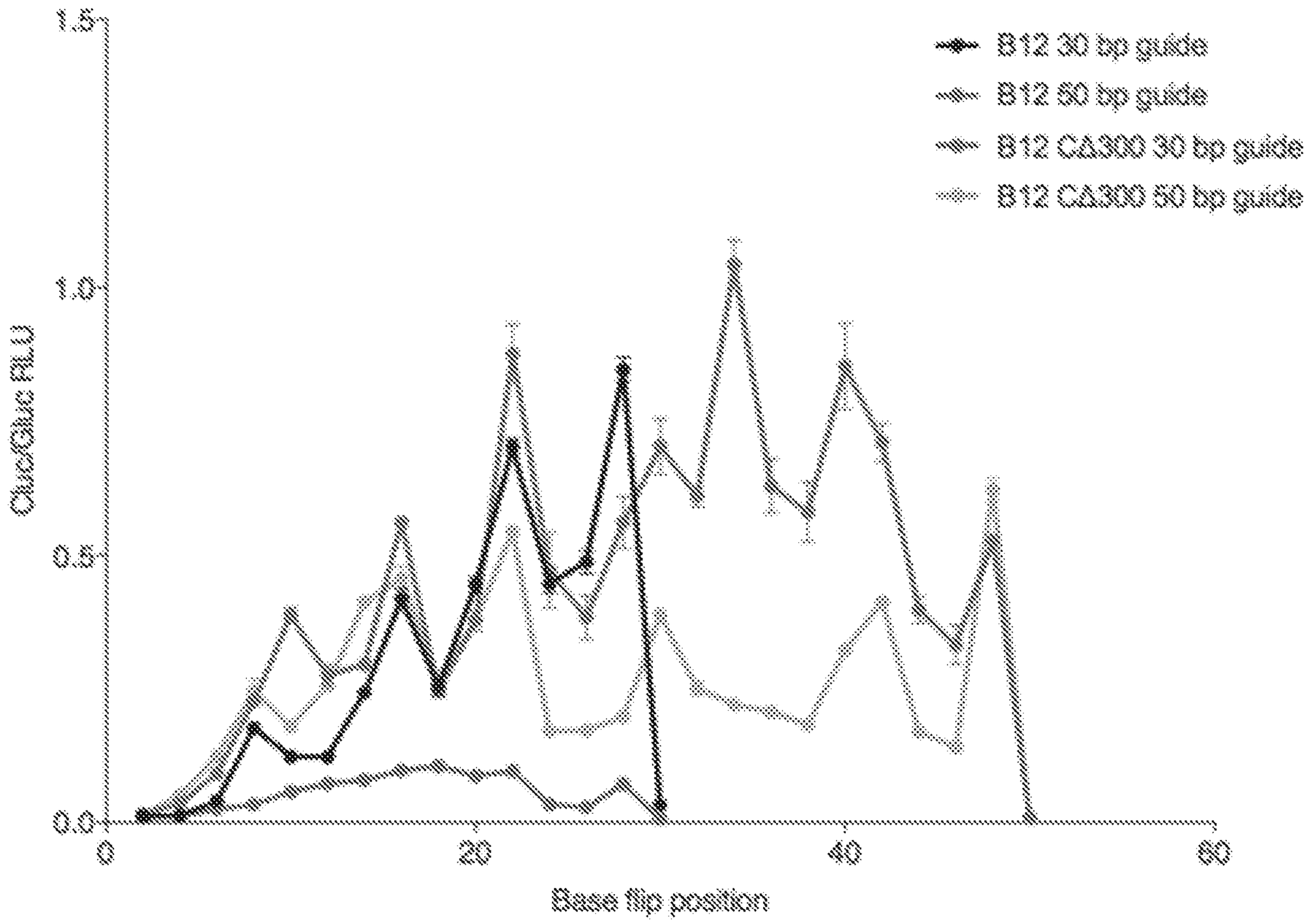

B12 30 bp guide
B12 60 bp guide
B12 CA300 30 bp guide
B12 CA300 60 bp guide
Cas9cis (a.u.)
1.5
1.0
0.5
0.0
0
20
40
60
Base flip position

FIG. 109

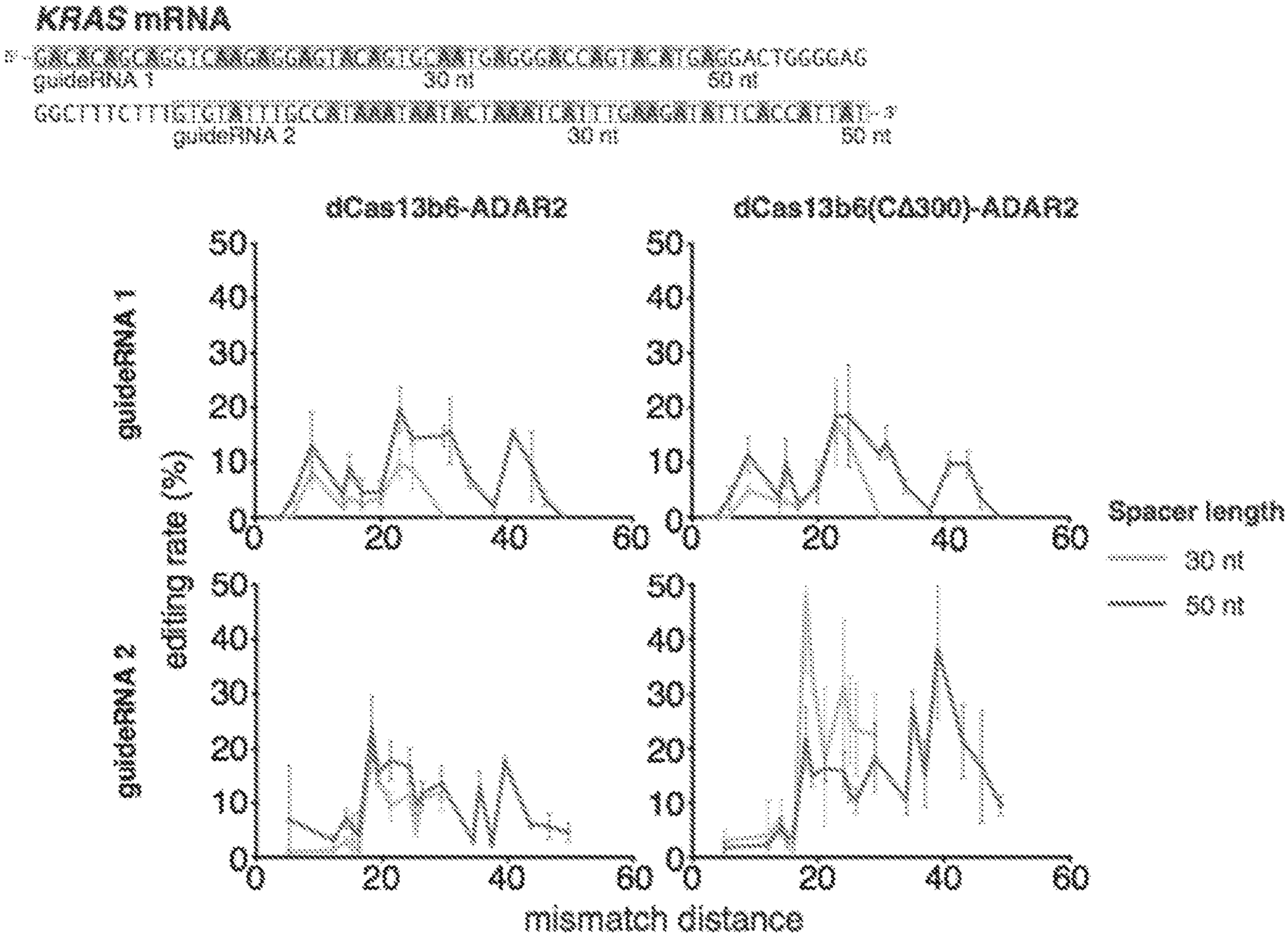

KRAS mRNA
guideRNA 1
30 nt
50 nt
guideRNA 2
30 nt
50 nt
dCas13b6-ADAR2
dCas13b6(CΔ300)-ADAR2
guideRNA 1
guideRNA 2
editing rate (%)
mismatch distance
Spacer length
30 nt
50 nt

FIG. 110

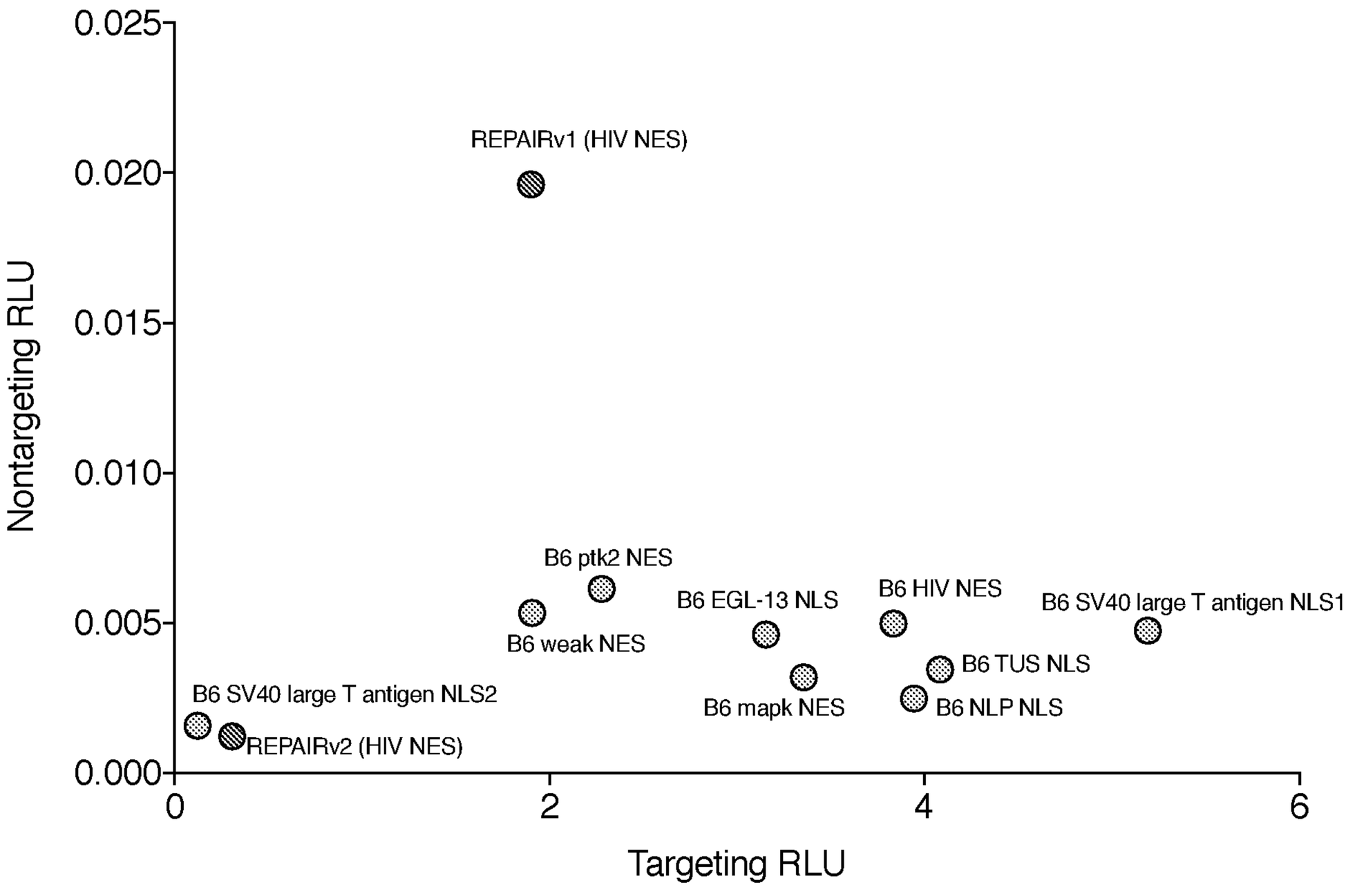

0.025
0.020
0.015
0.010
0.005
0.000
Nontargeting RLU
0
2
4
6
Targeting RLU
REPAIRv1 (HIV NES)
B6 ptk2 NES
B6 weak NES
B6 EGL-13 NLS
B6 HIV NES
B6 SV40 large T antigen NLS1
B6 TUS NLS
B6 NLP NLS
B6 mapk NES
B6 SV40 large T antigen NLS2
REPAIRv2 (HIV NES)

FIG. 111

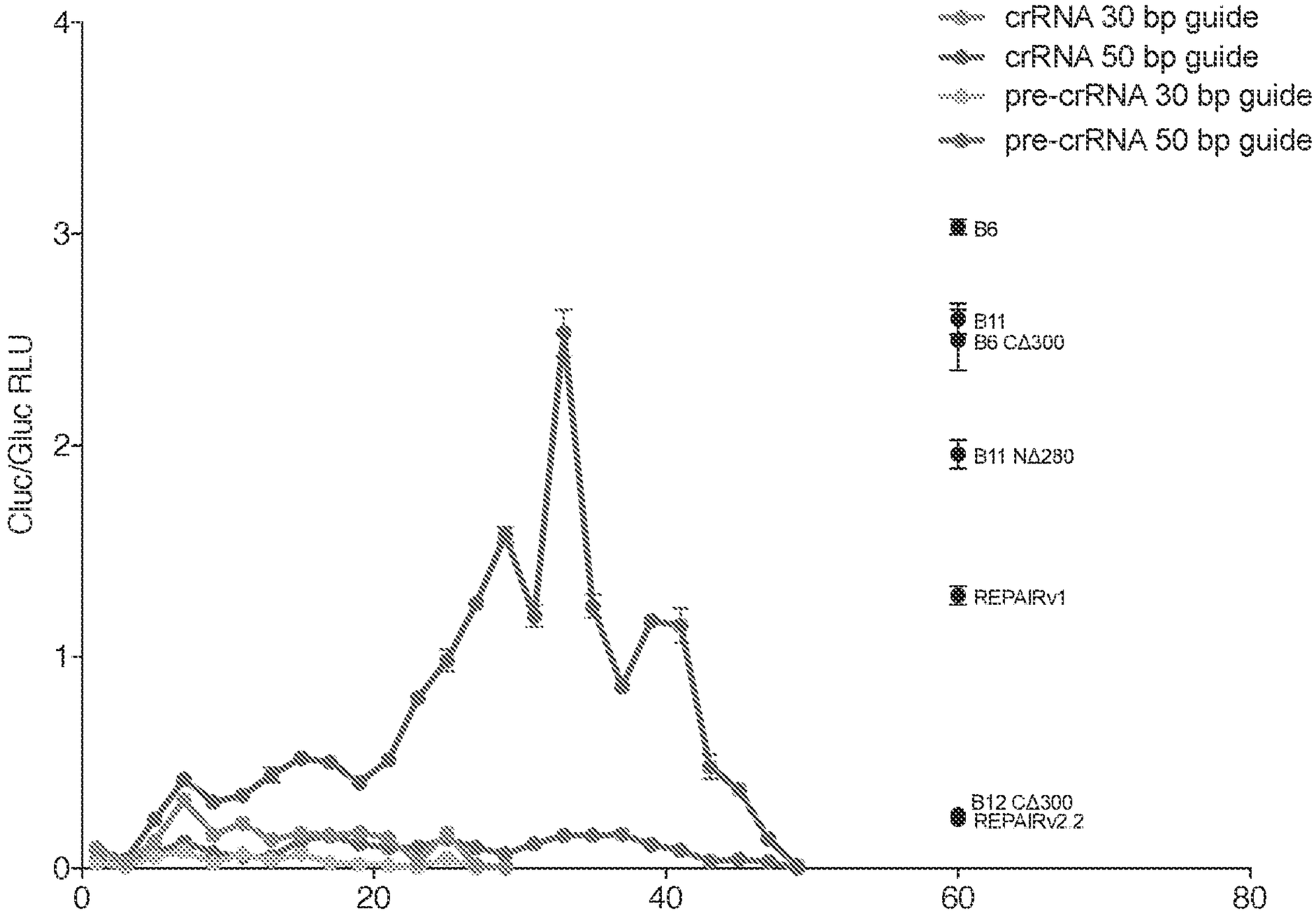

crRNA 30 bp guide
crRNA 50 bp guide
pre-crRNA 30 bp guide
pre-crRNA 50 bp guide
B6
B11
B6 CΔ300
B11 NΔ280
REPAIRv1
B12 CΔ300
REPAIRv2.2
Cluc/Gluc RLU
0
1
2
3
4
0
20
40
60
80
Base flip position

FIG. 112

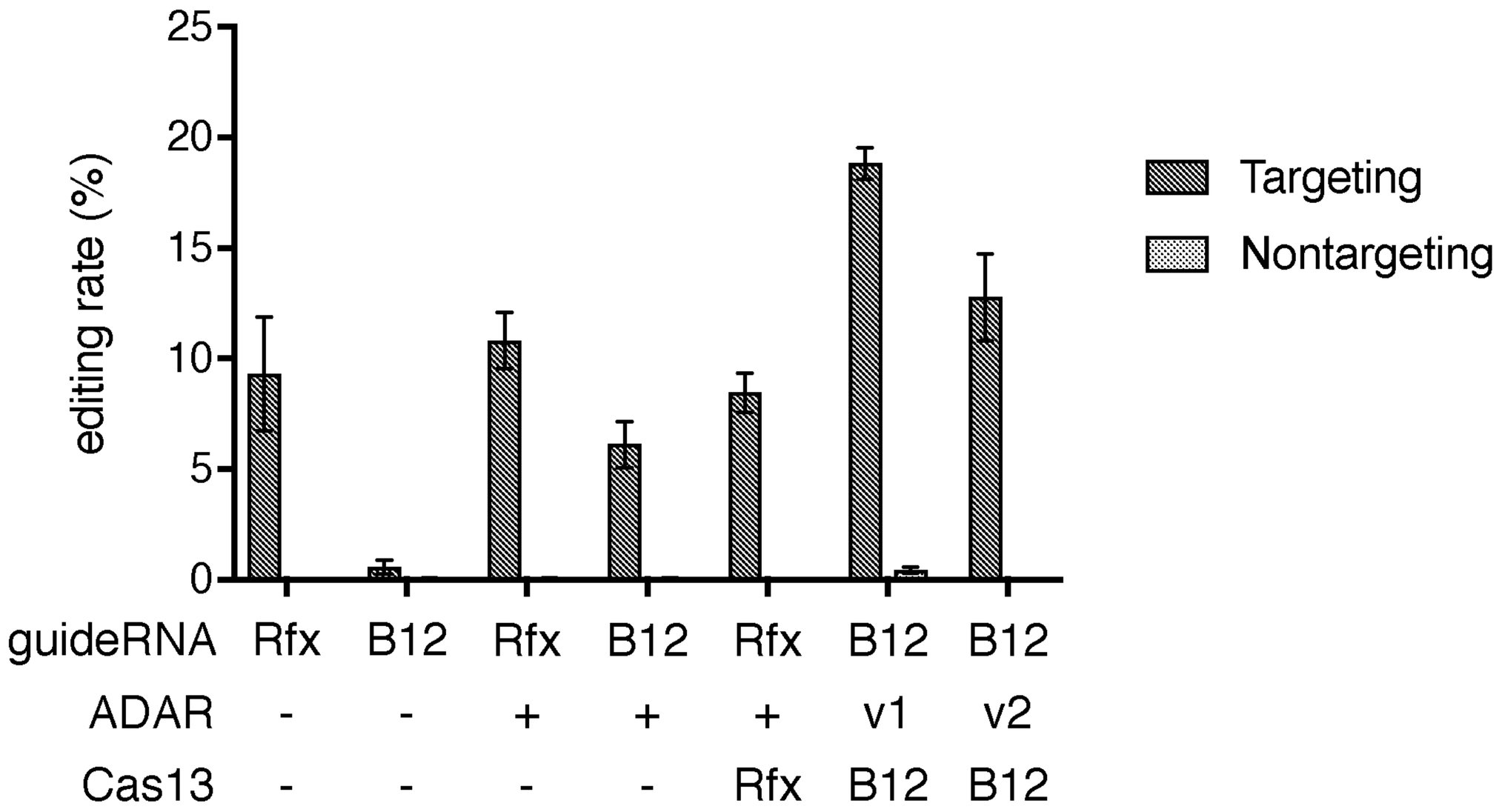
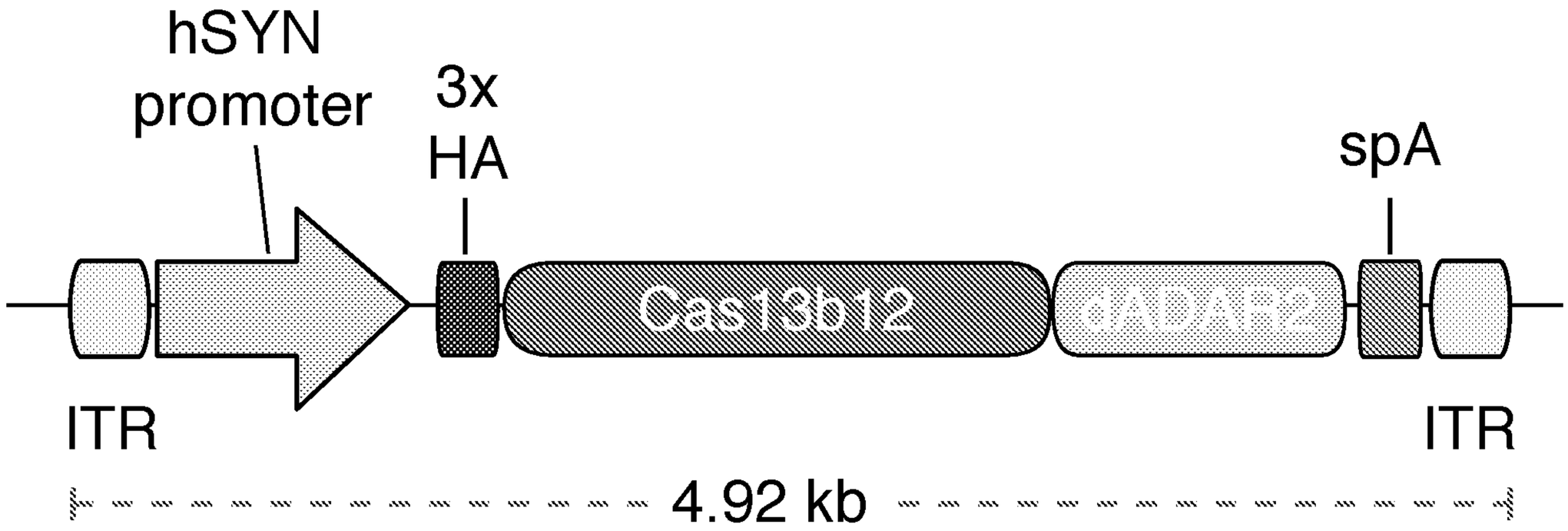

editing rate (%)
25
20
15
10
5
0
Targeting
Nontargeting
guideRNA Rfx B12 Rfx B12 Rfx B12 B12
ADAR - - + + + v1 v2
Cas13 - - - - Rfx B12 B12

FIG. 113


hSYN
promoter
3x
HA
spA
Cas13b12
dADAR2
ITR
ITR
4.92 kb

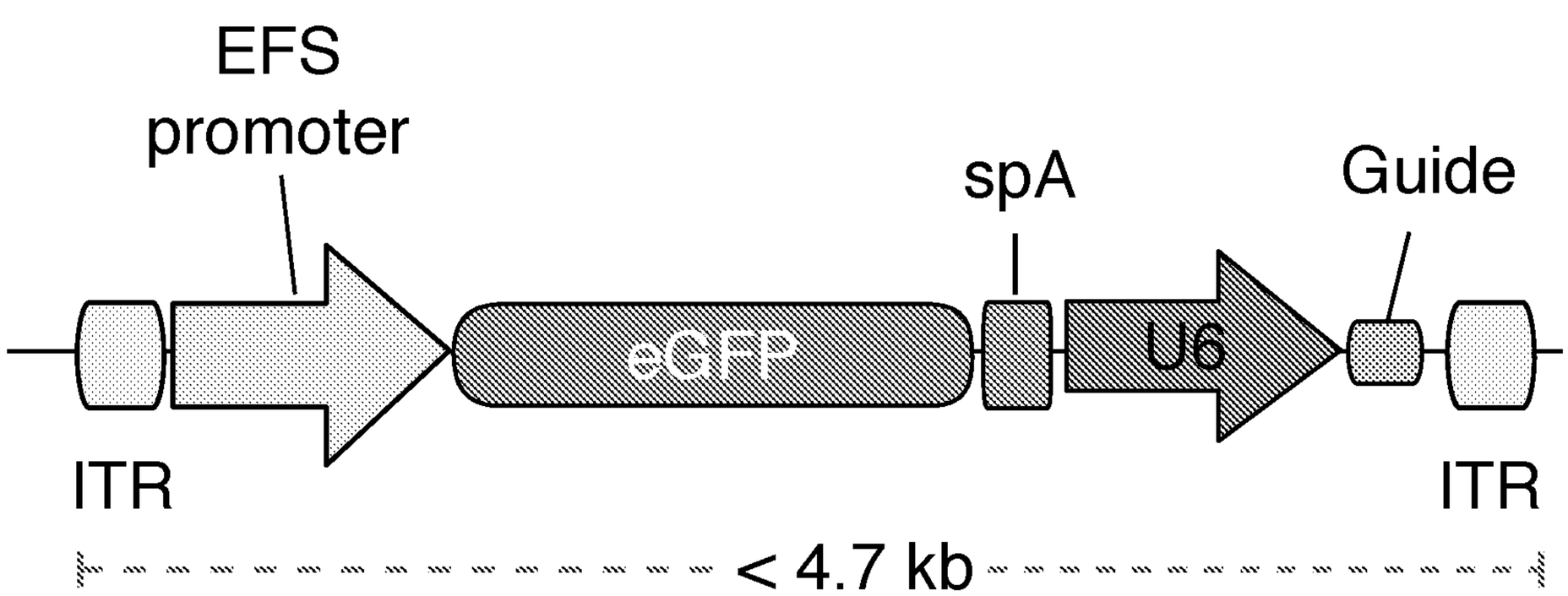

EFS
promoter
spA
Guide
eGFP
U6
ITR
ITR
< 4.7 kb

FIG. 114

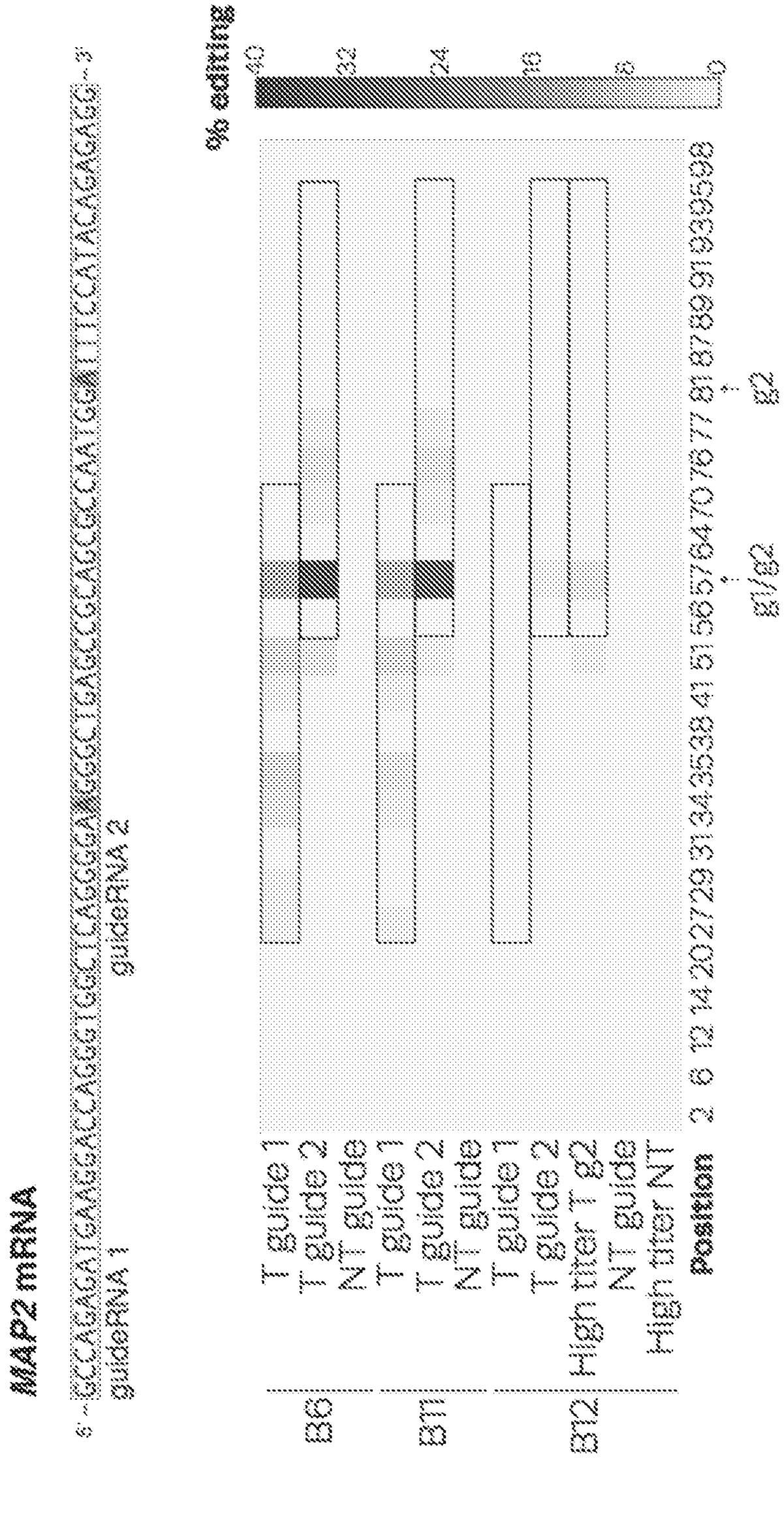

MAP2 mRNA
5'-GCCAGAGATGAAGGACAGGGTGGCCTCAGGGGAAGGGGCTGAGCCGAGCCAGCGGCAATGGAT
guideRNA 1
guideRNA 2
% editing
40
32
24
16
8
0
B6
T guide 1
T guide 2
NT guide
T guide 1
BTI
T guide 2
NT guide
T guide 1
T guide 2
B12 High titer T &2
NT guide
High titer NT
Position
2 6 10 12 14 20 27 29 31 34 35 38 41 51 56 57 64 70 76 77 81 87 89 91 93 95 98
B1/B2
B2

FIG. 115

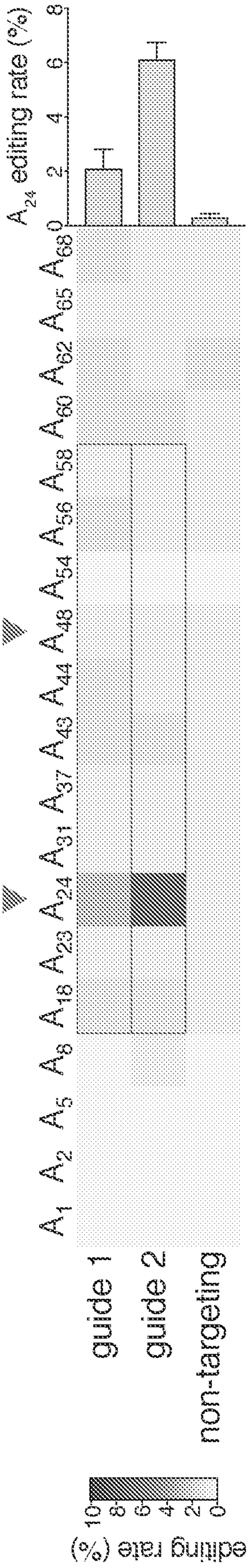

editing rate (%)
0
2
4
6
8
A24
A68
g1/g2
A43
A44
A48
A54
A56
A58
A60
A62
A65
g2
A1
A2
A5
A8
A18
A23
A24
A31
A37
A43
1
2
3
guide
guide
non-targeting
editing rate (%)
0
2
4
6
8
10

FIG. 116

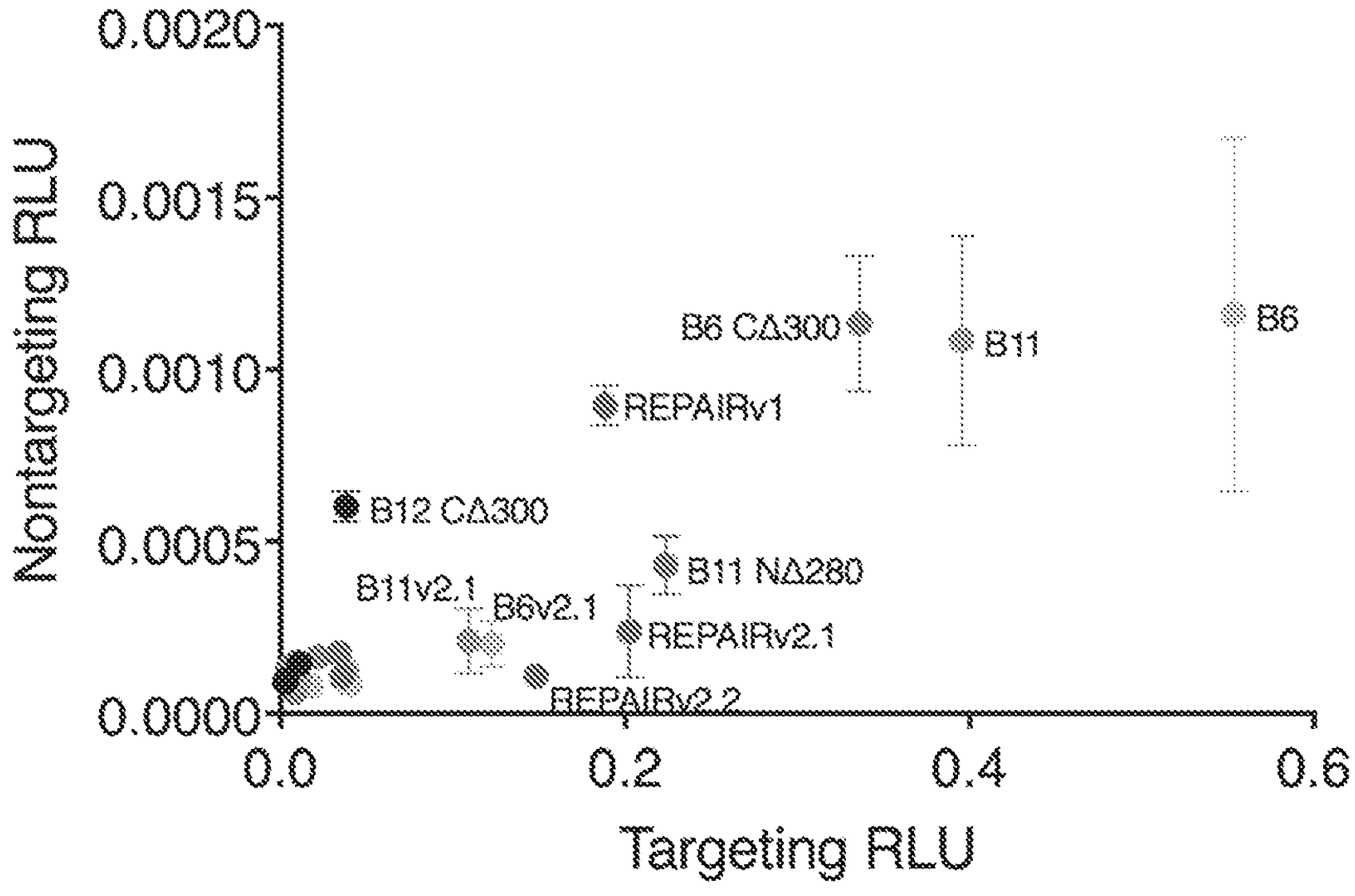

Nontargeting RLU
0.0020
0.0015
0.0010
0.0005
0.0000
0.0
0.2
0.4
0.6
Targeting RLU
B6 CΔ300
B11
B6
REPAIRv1
B12 CΔ300
B11 NΔ280
B11v2.1
B6v2.1
REPAIRv2.1
REPAIRv2.2

FIG. 117

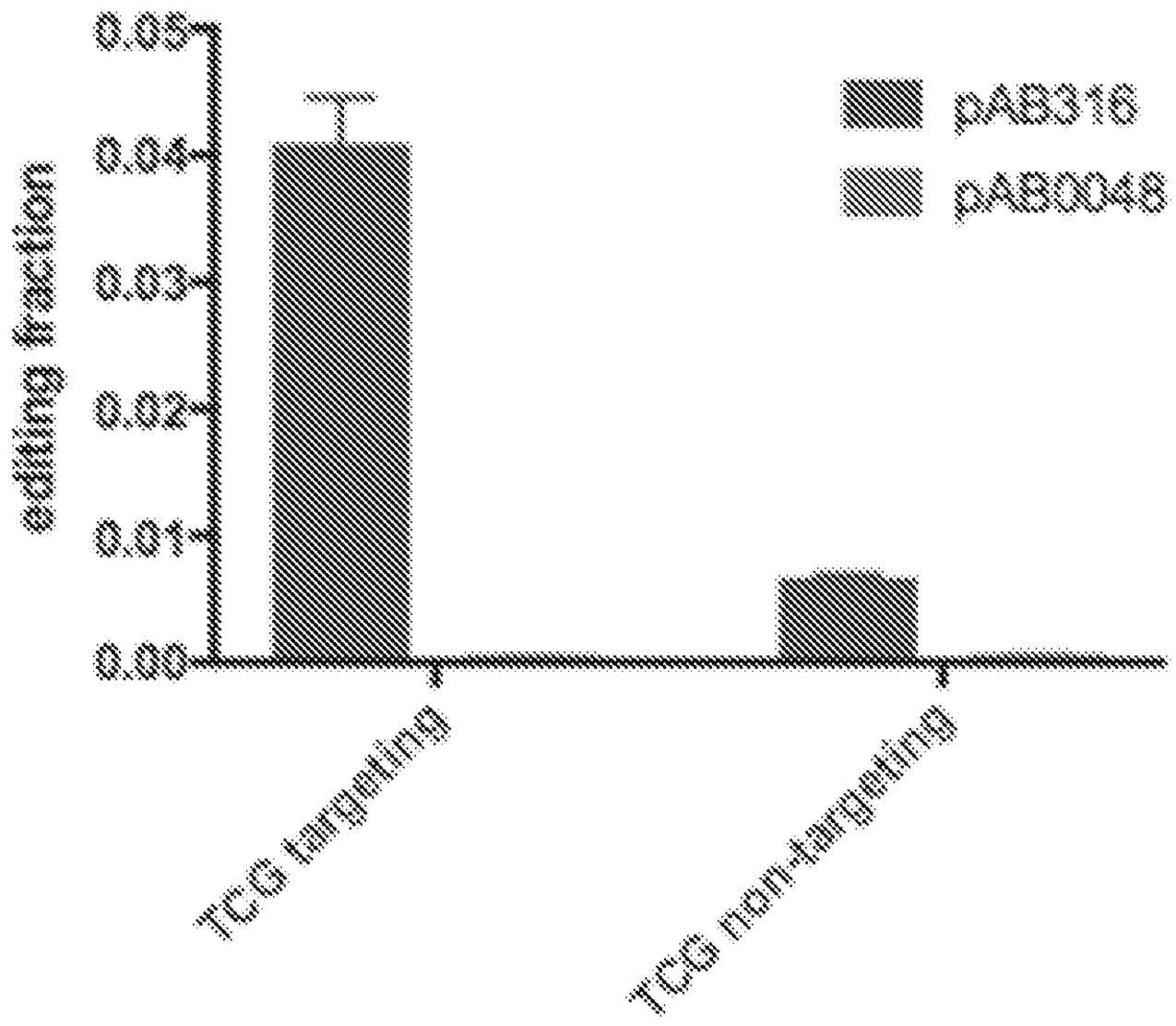

0.05
0.04
0.03
0.02
0.01
0.00
editing fraction
TCG targeting
TCG non-targeting
pAB316
pAB0048

FIG. 118

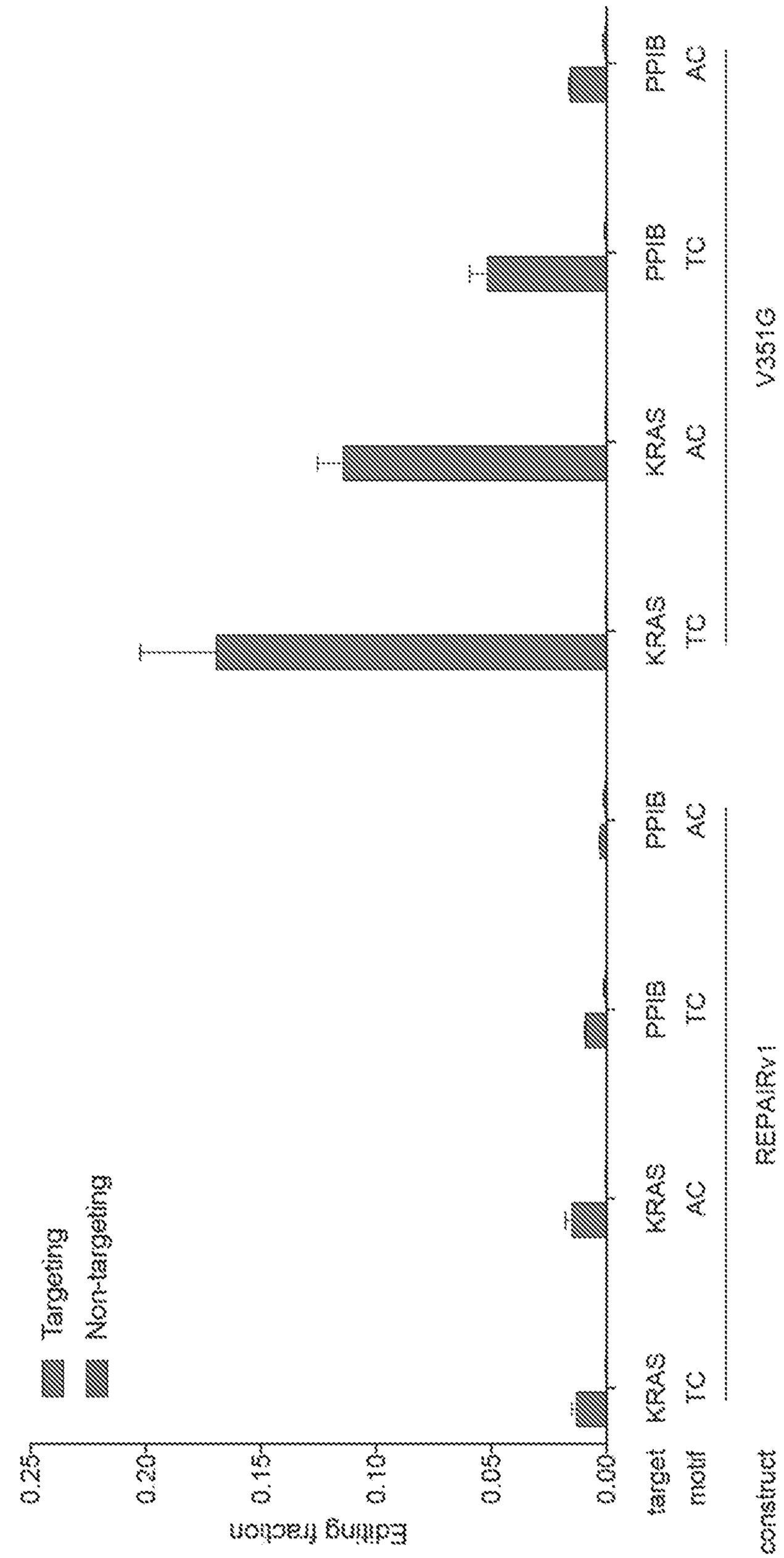

Targeting
Non-targeting
deltaq fraction
0.25
0.20
0.15
0.10
0.05
0.00
target
motif
construct
KRAS
TC
KRAS
AC
PPIB
TC
KRAS
AC
PPIB
TC
PPIB
AC
REPAIRv1
V351G

FIG. 19

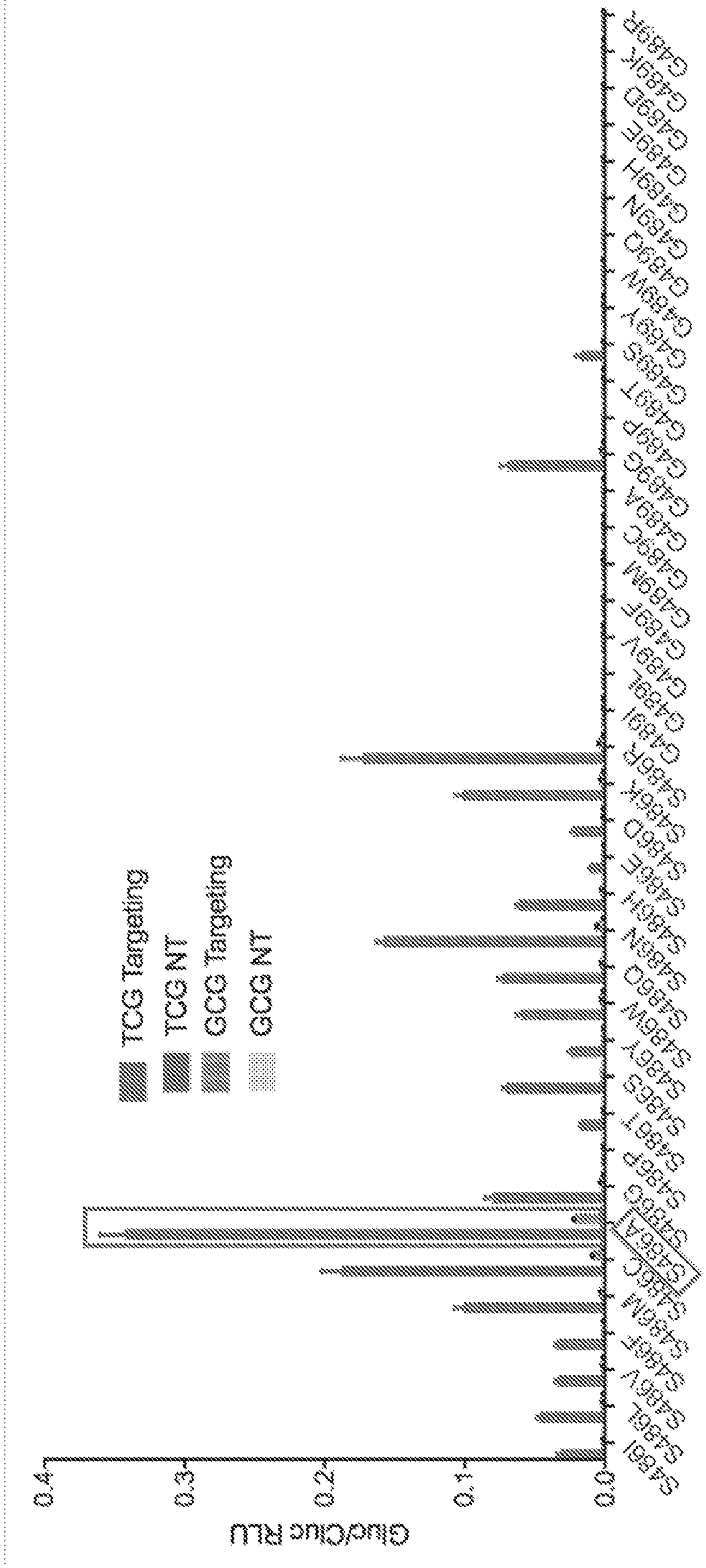

TCG Targeting
TCG NT
GCG Targeting
GCG NT
Glucose RLU
0.4
0.3
0.2
0.1
0.0
Mutant

FIG. 120

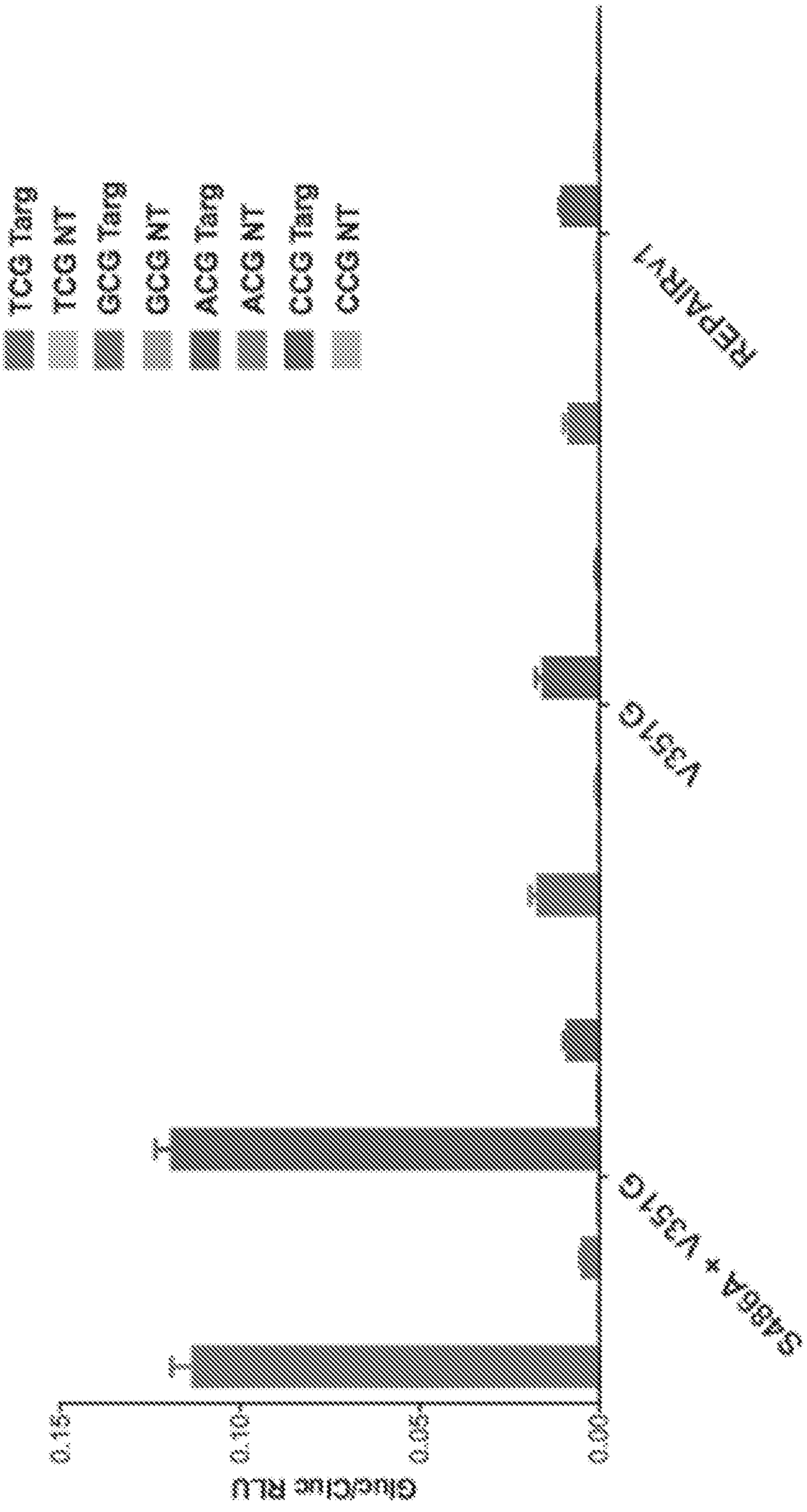

TCG Targ
TCG NT
GCG Targ
GCG NT
ACG Targ
ACG NT
CCG Targ
CCG NT
Targeted
GTCGA
GTCGA
GTCGA
GTCGA * VS2
ADAR Mutant
Glucose Rlu
0.15
0.10
0.05
0.00

FIG. 121

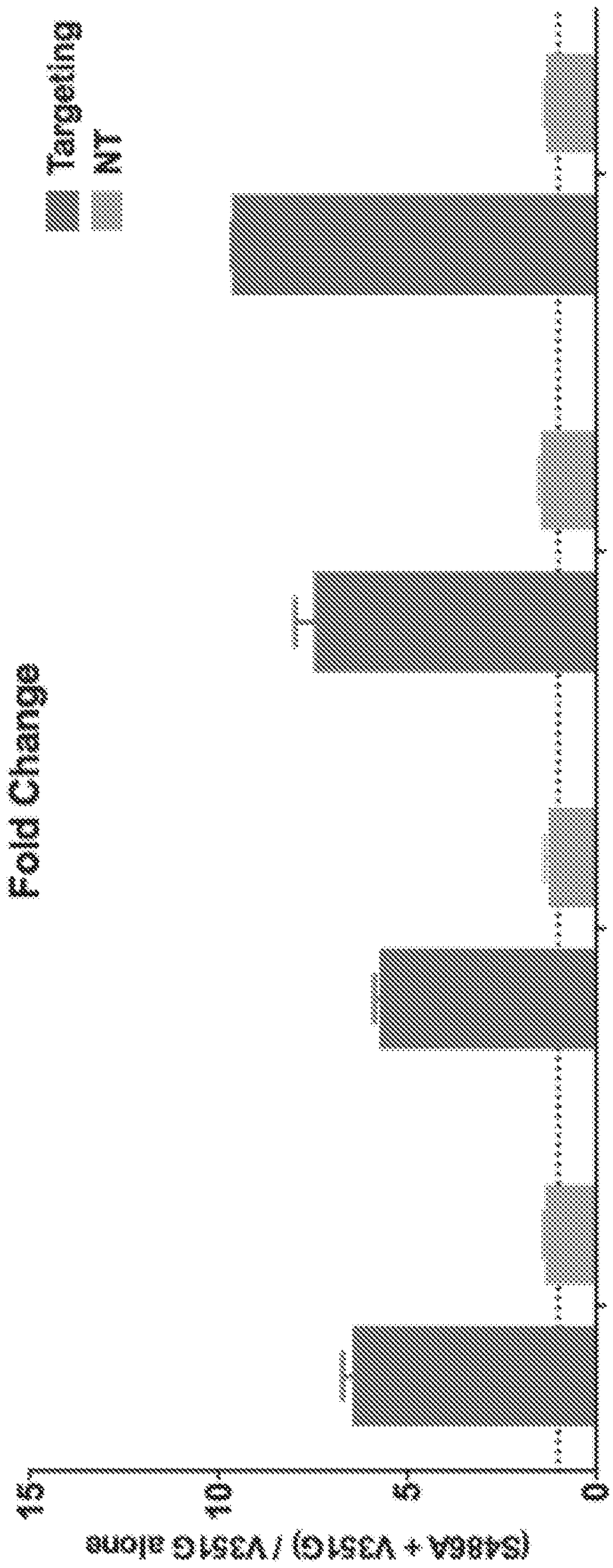

(S486A + V351G) / V351G alone
0
2
4
Fold Change in Image
Motif
C3
C4
C5
C3
C3
targeting
Kit

FIG. 122

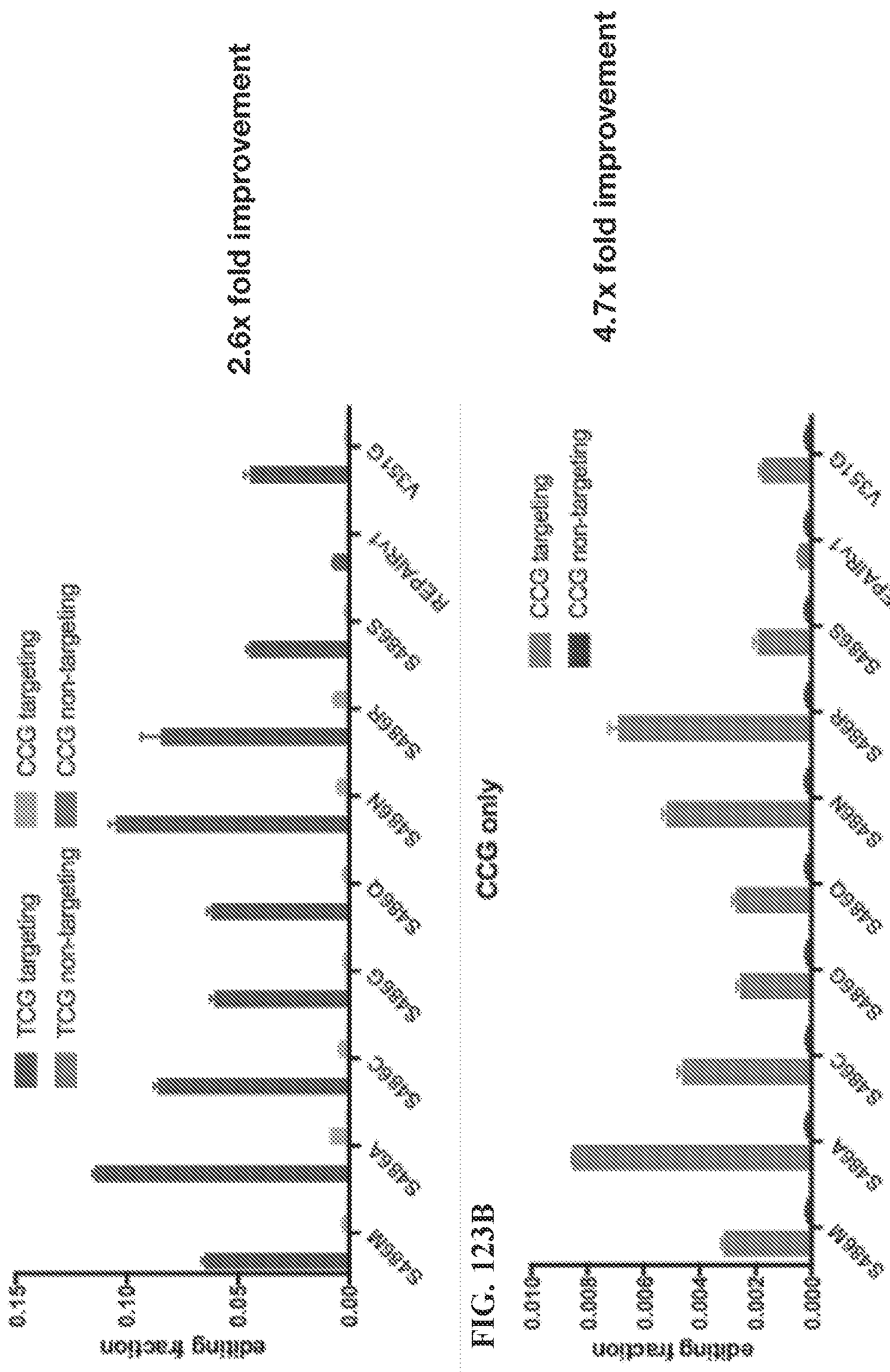


FIG. 123A

FIG. 123B

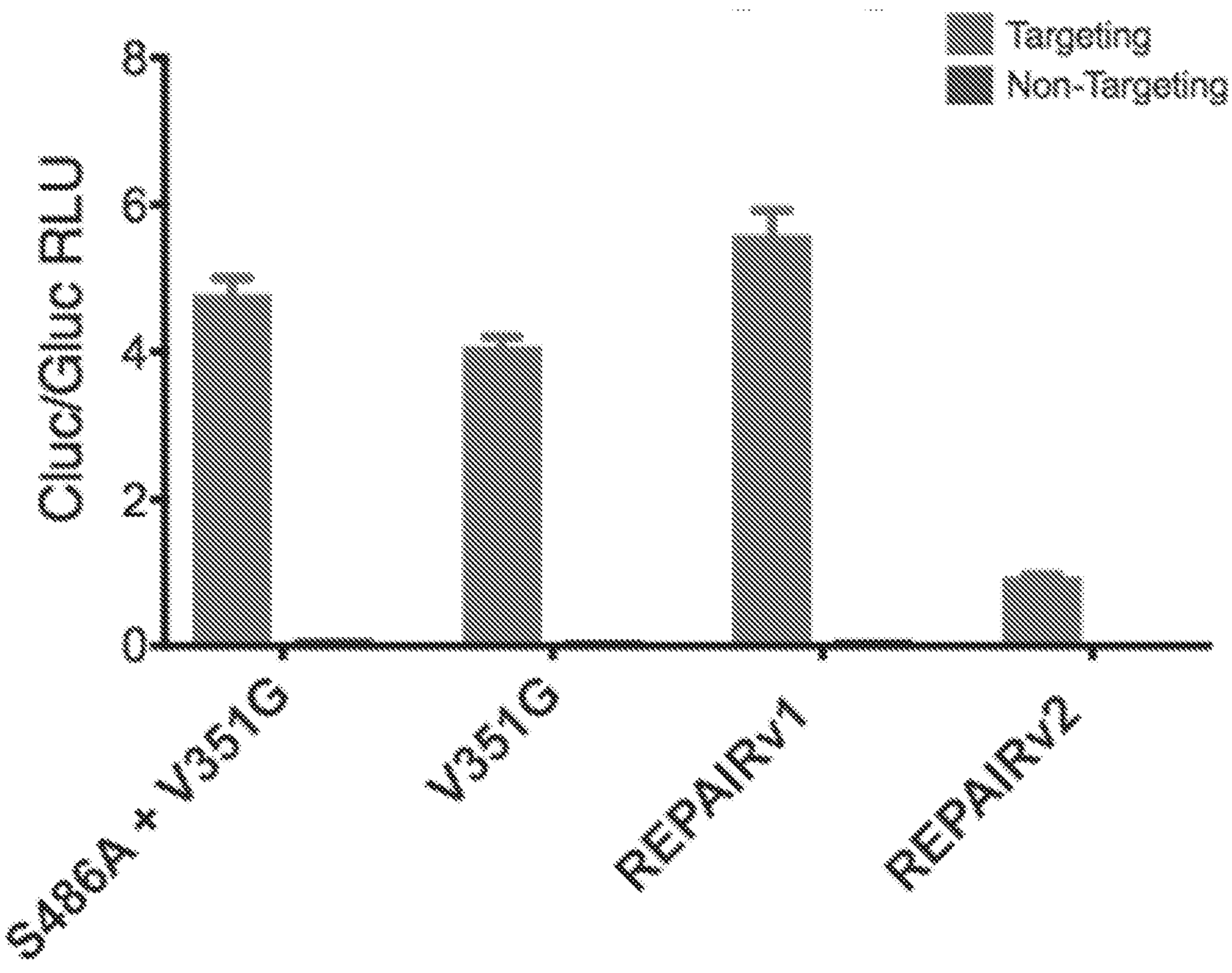

Targeting
Non-Targeting
Cluc/Gluc RLU
0
2
4
6
8
S486A + V351G
V351G
REPAIRv1
REPAIRv2

FIG. 124

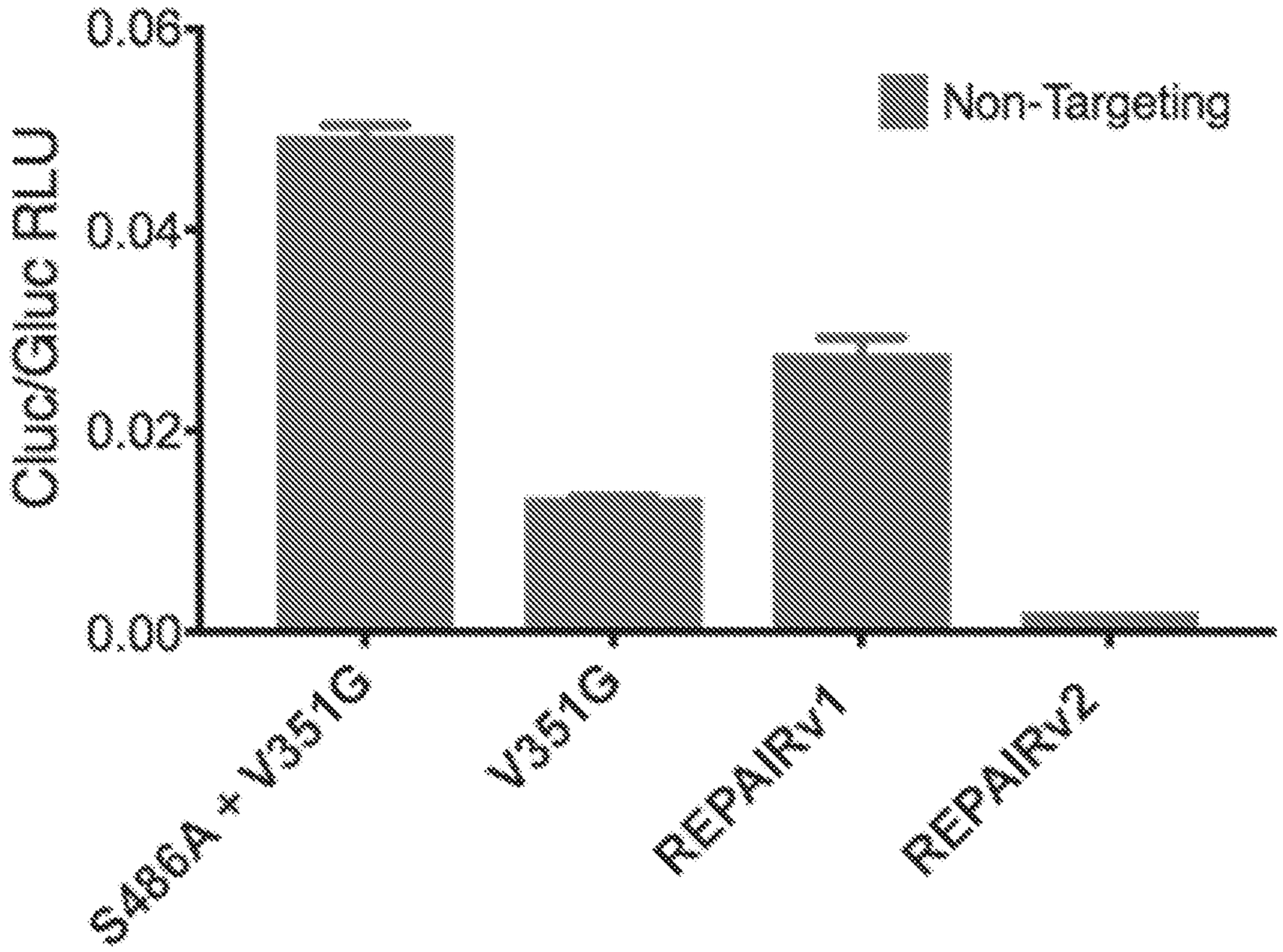

Cluc/Gluc RLU
0.06
0.04
0.02
0.00
Non-Targeting
S486A + V351G
V351G
REPAIRv1
REPAIRv2

FIG. 125

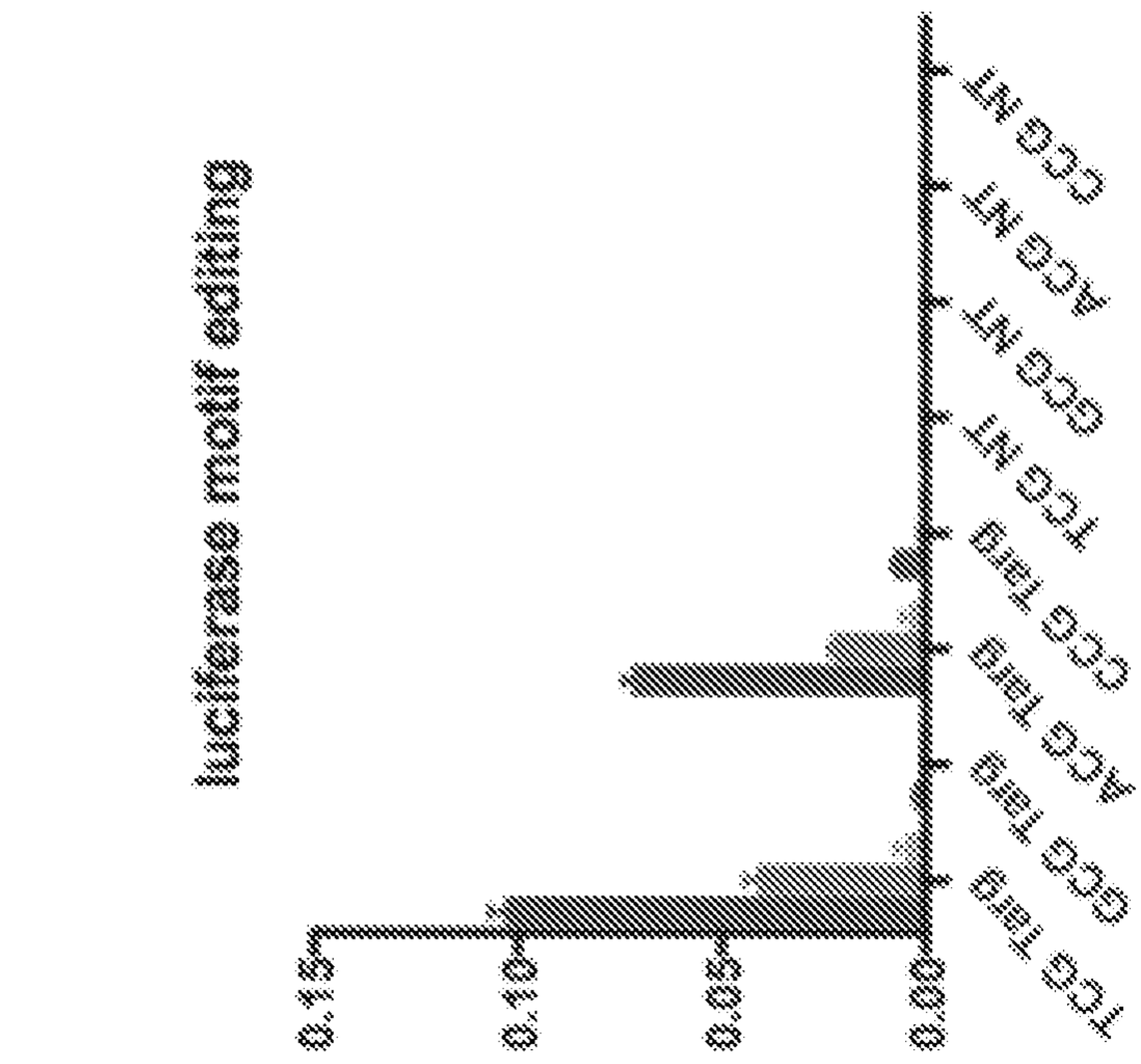


FIG. 126A

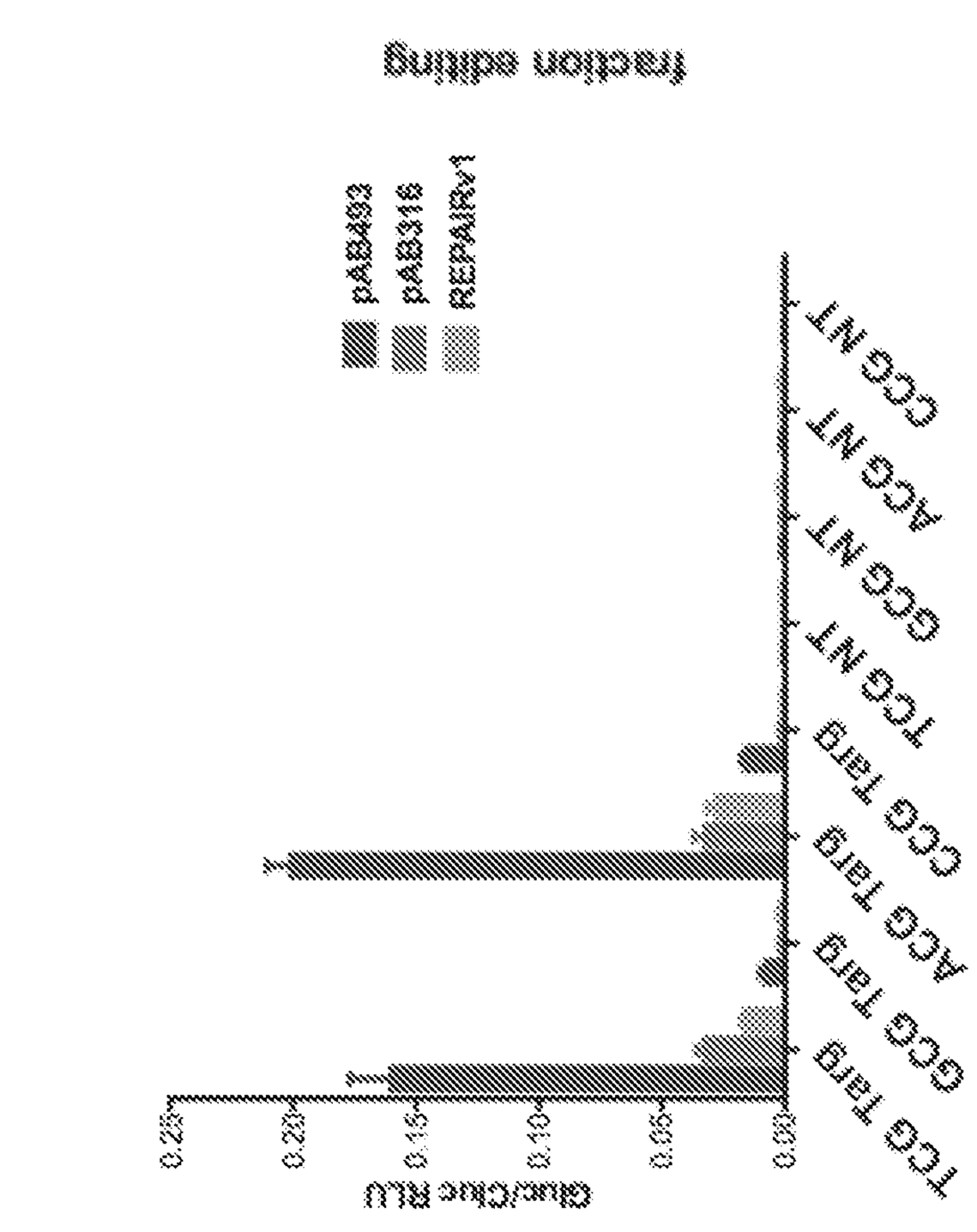


FIG. 126B

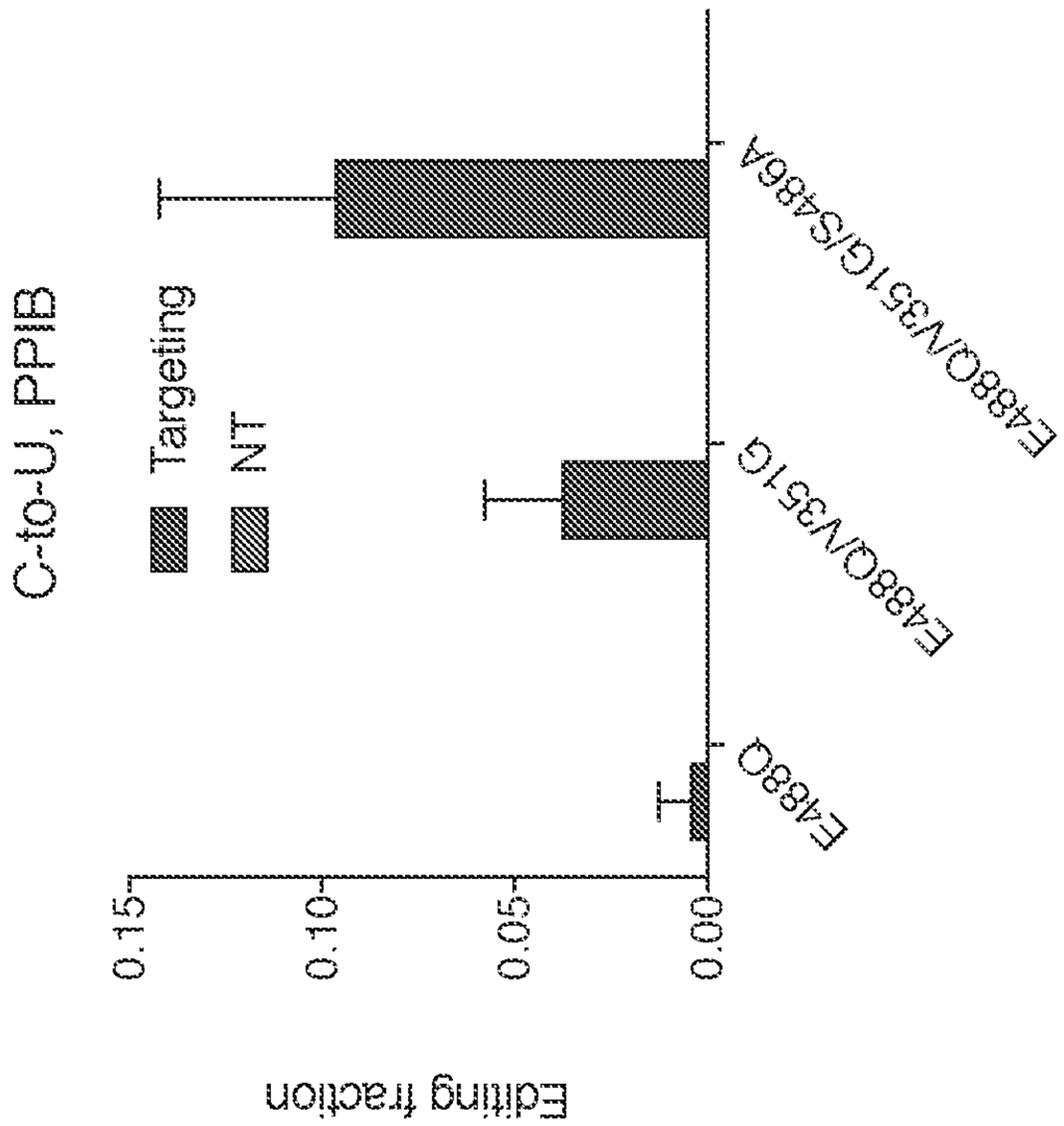


FIG. 127B

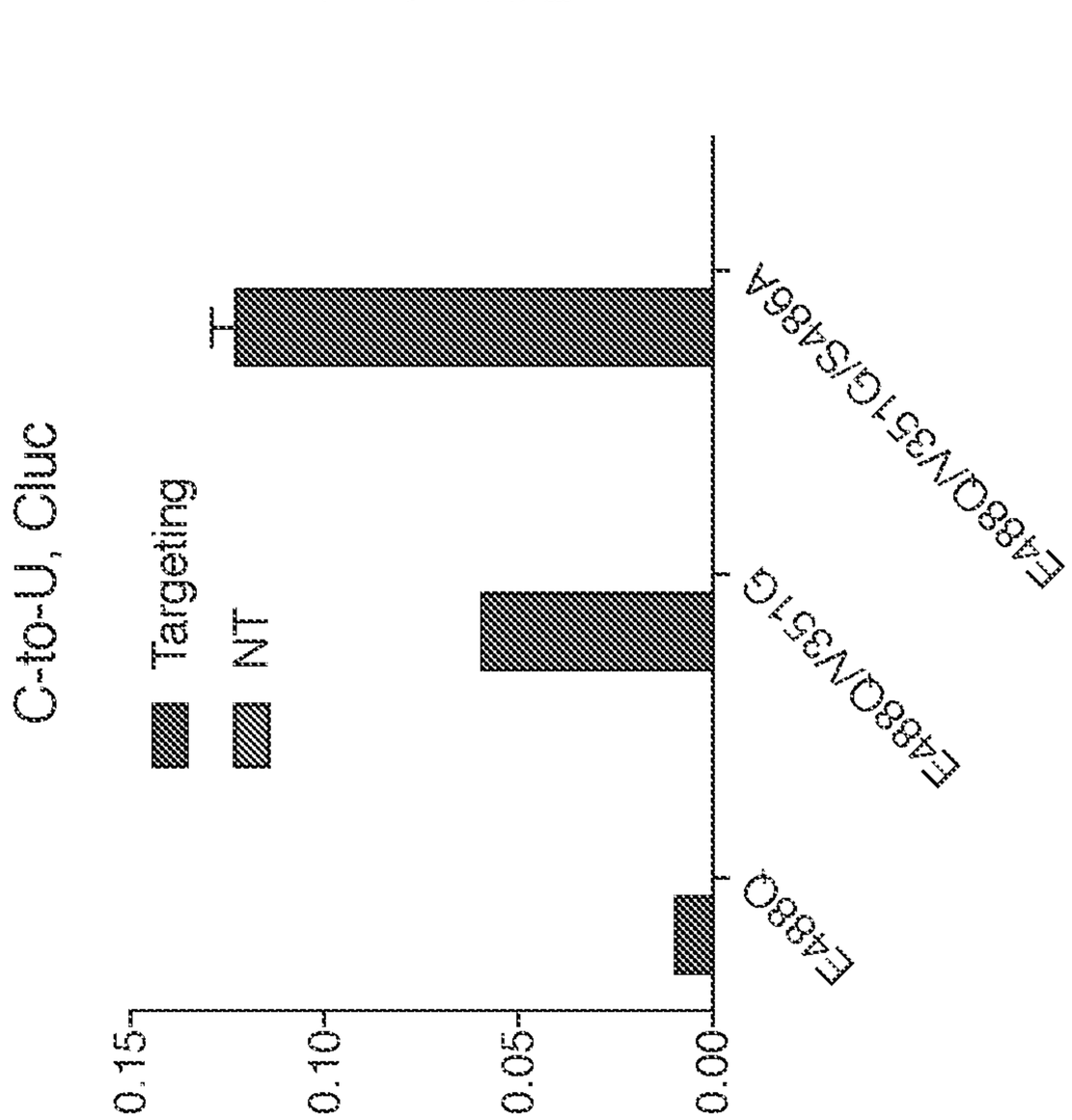


FIG. 127A

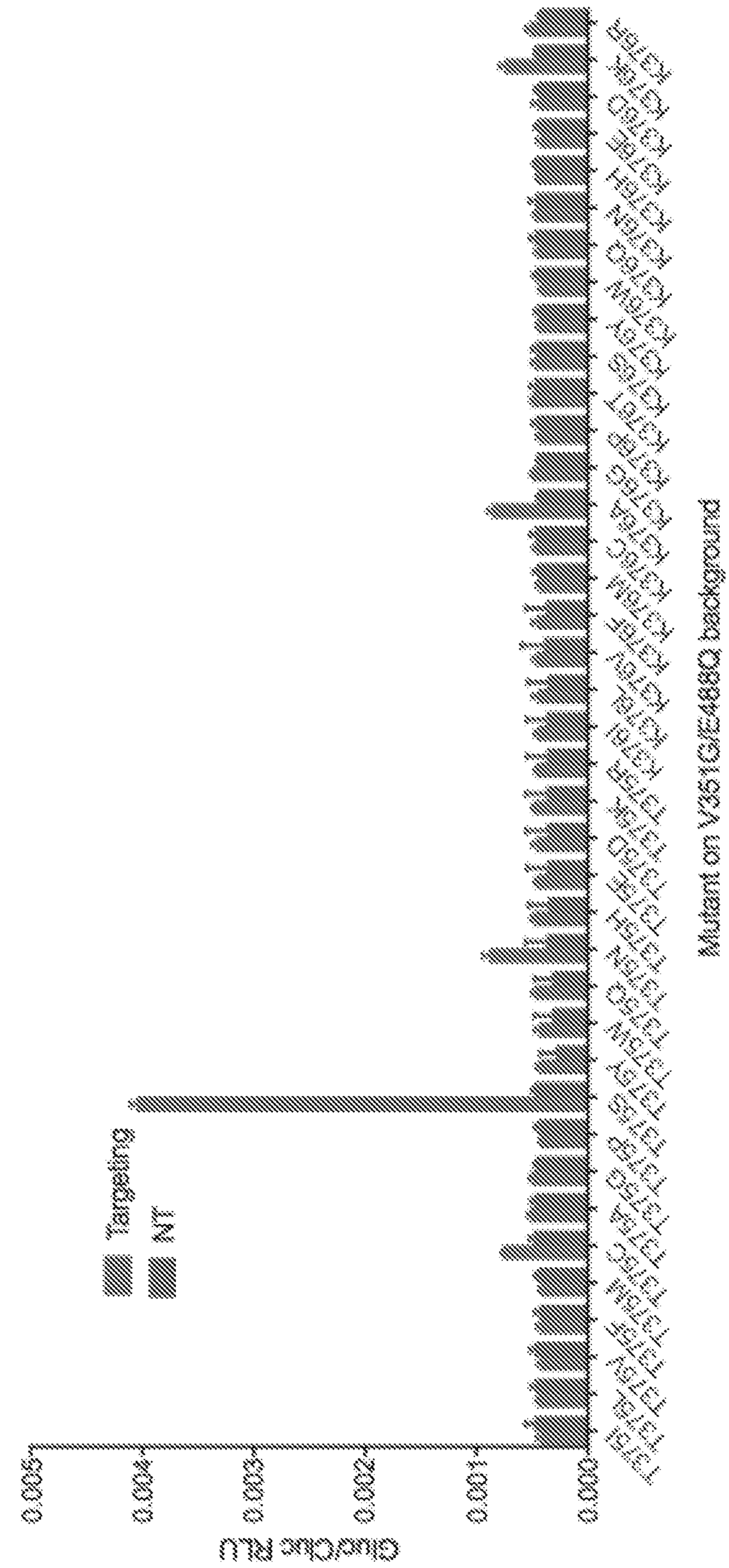

Targeting
NT
0.05
0.04
0.03
0.02
0.01
0.00
Mutant on V351G/E548Q background
Glucose RLU

FIG. 128

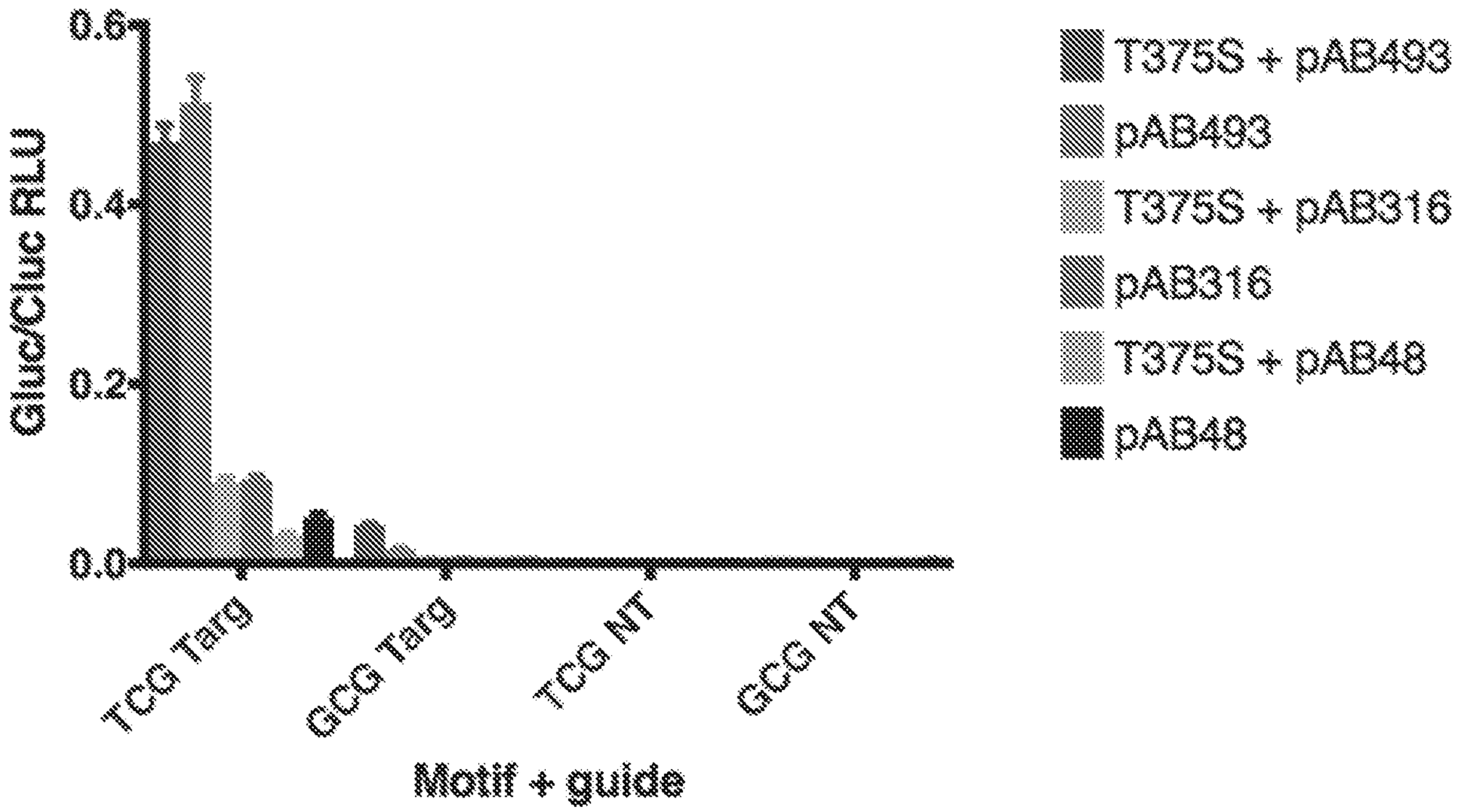

T375S_undil
0.6
0.4
0.2
0.0
Gluc/Cluc RLU
TCG Targ
GCG Targ
TCG NT
GCG NT
Motif + guide
T375S + pAB493
pAB493
T375S + pAB316
pAB316
T375S + pAB48
pAB48

FIG. 129

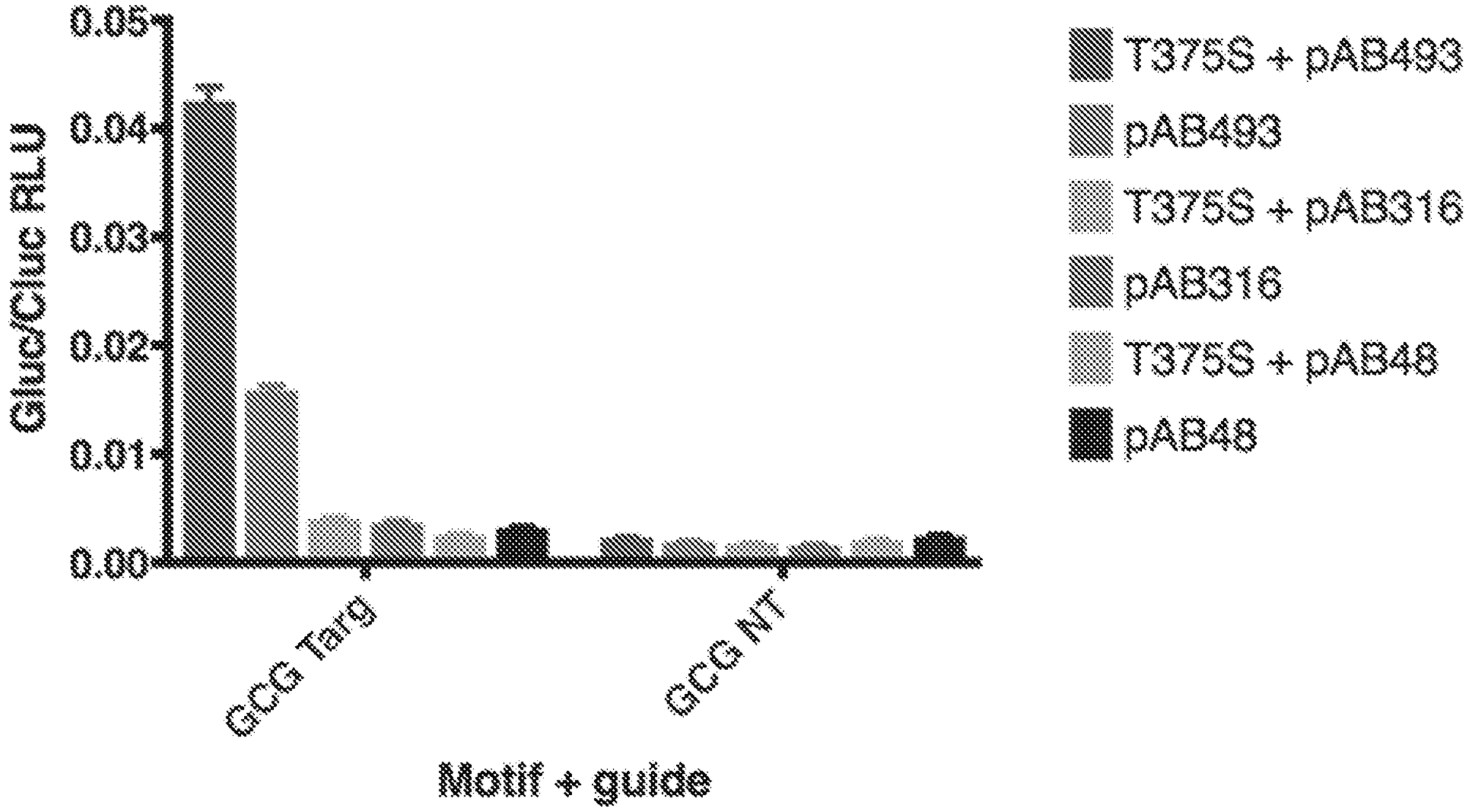

T375S_undil_GCG
Gluc/Cluc RLU
0.05
0.04
0.03
0.02
0.01
0.00
GCG Targ
GCG NT
Motif + guide
T375S + pAB493
pAB493
T375S + pAB316
pAB316
T375S + pAB48
pAB48

FIG. 130

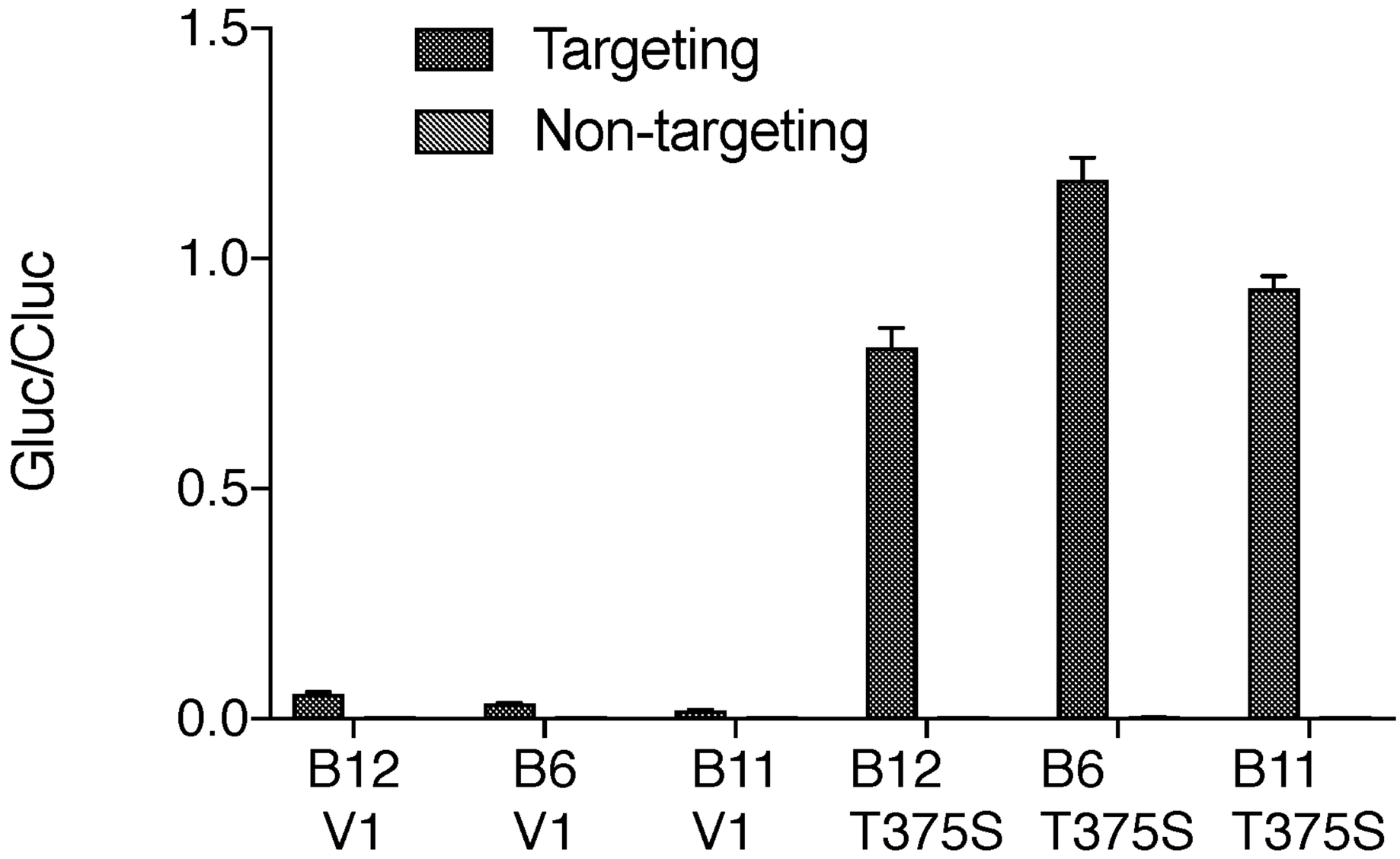

Targeting
Non-targeting
Gluc/Cluc
1.5
1.0
0.5
0.0
B12 V1
B6 V1
B11 V1
B12 T375S
B6 T375S
B11 T375S

FIG. 131

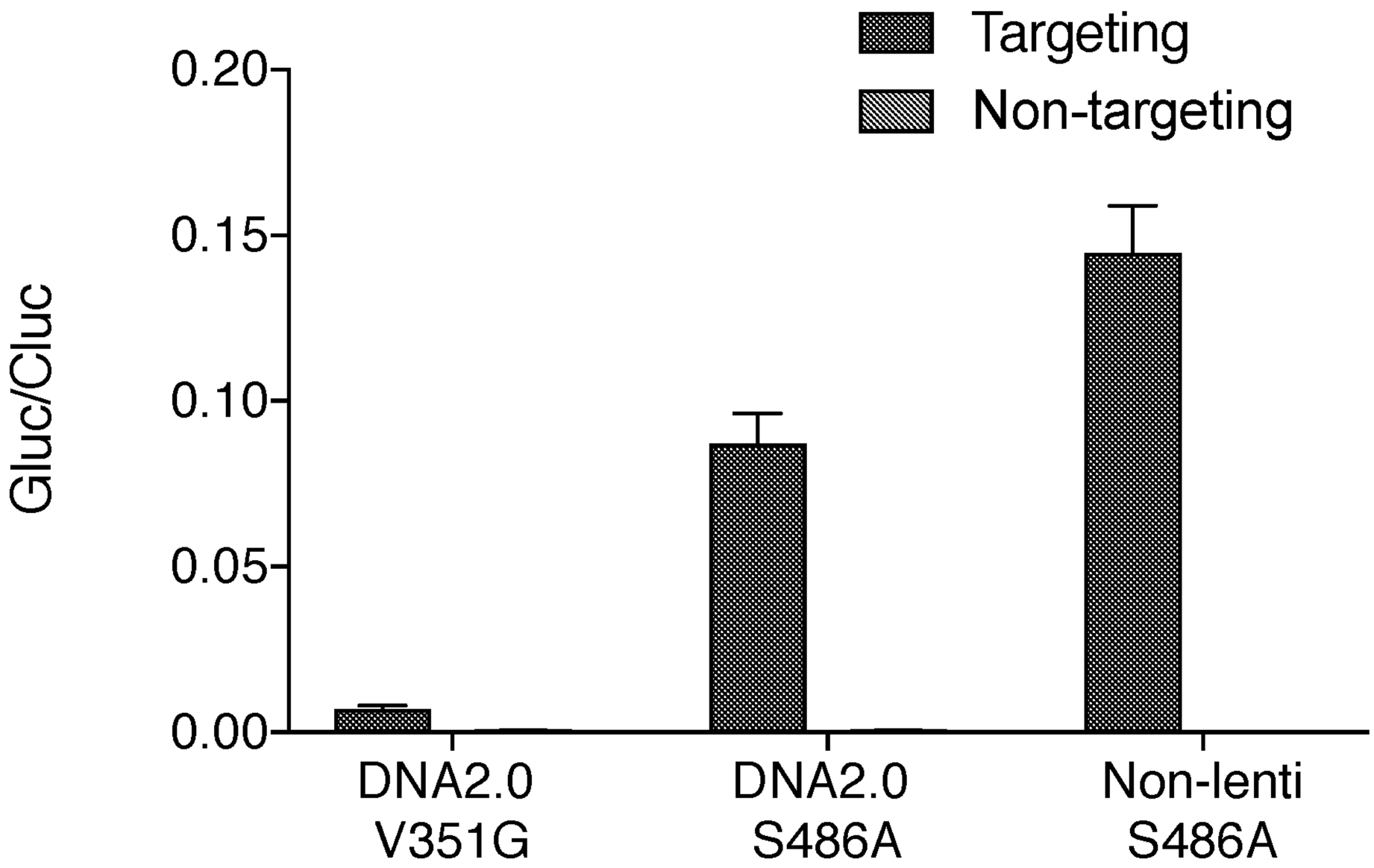

Targeting
Non-targeting
Gluc/Cluc
0.20
0.15
0.10
0.05
0.00
DNA2.0 V351G
DNA2.0 S486A
Non-lenti S486A

FIG. 132

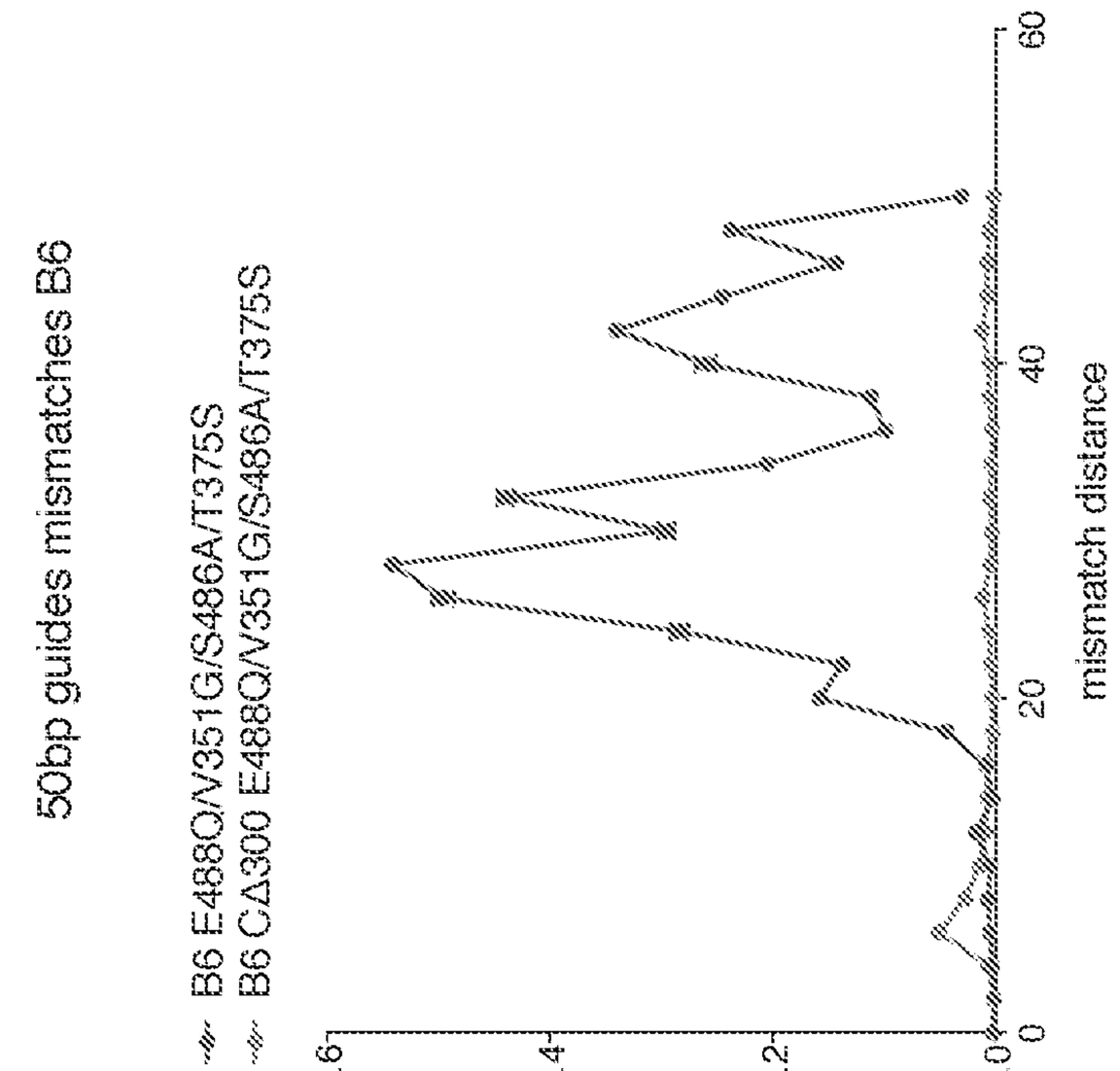


FIG. 133B

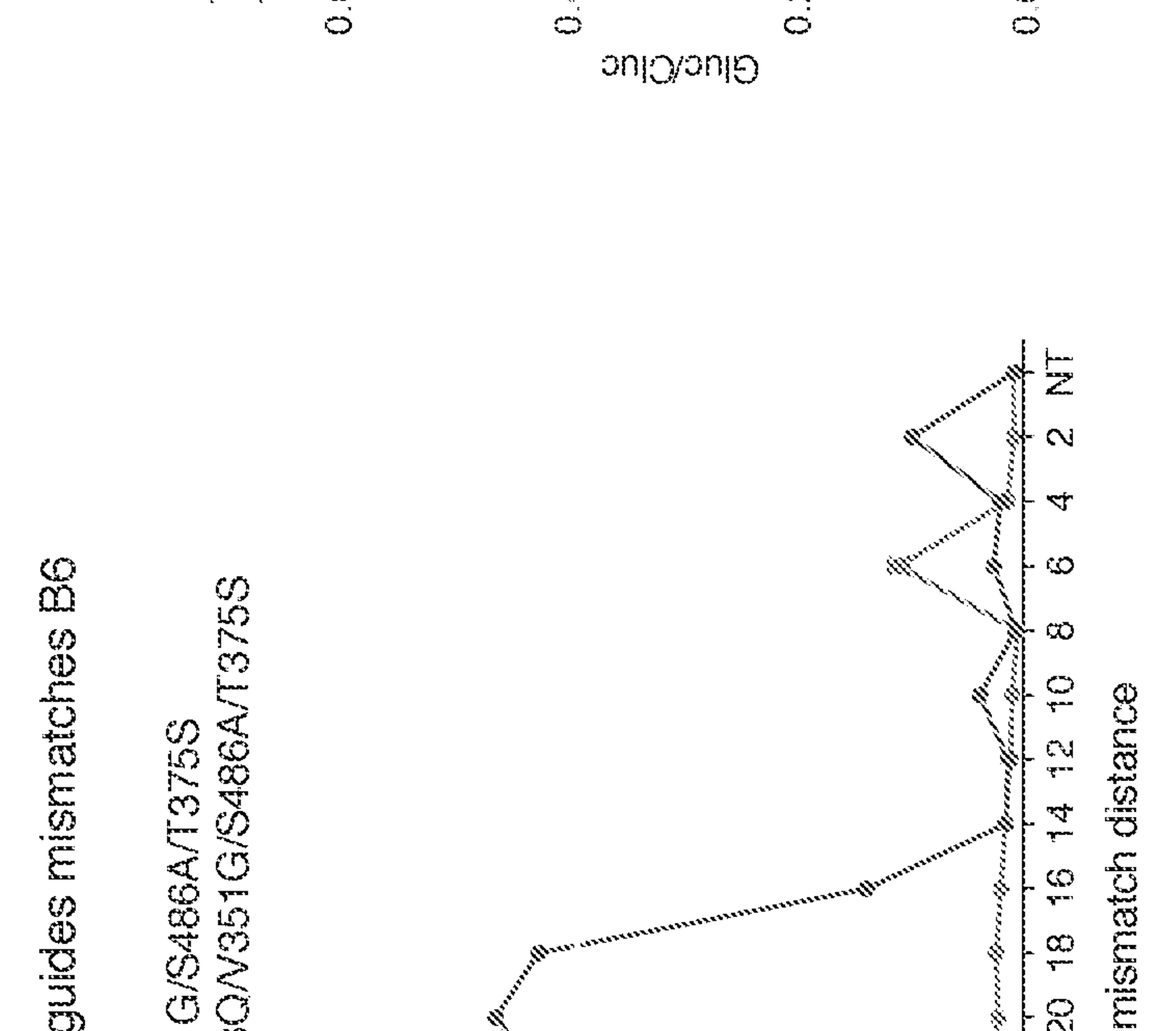


FIG. 133A

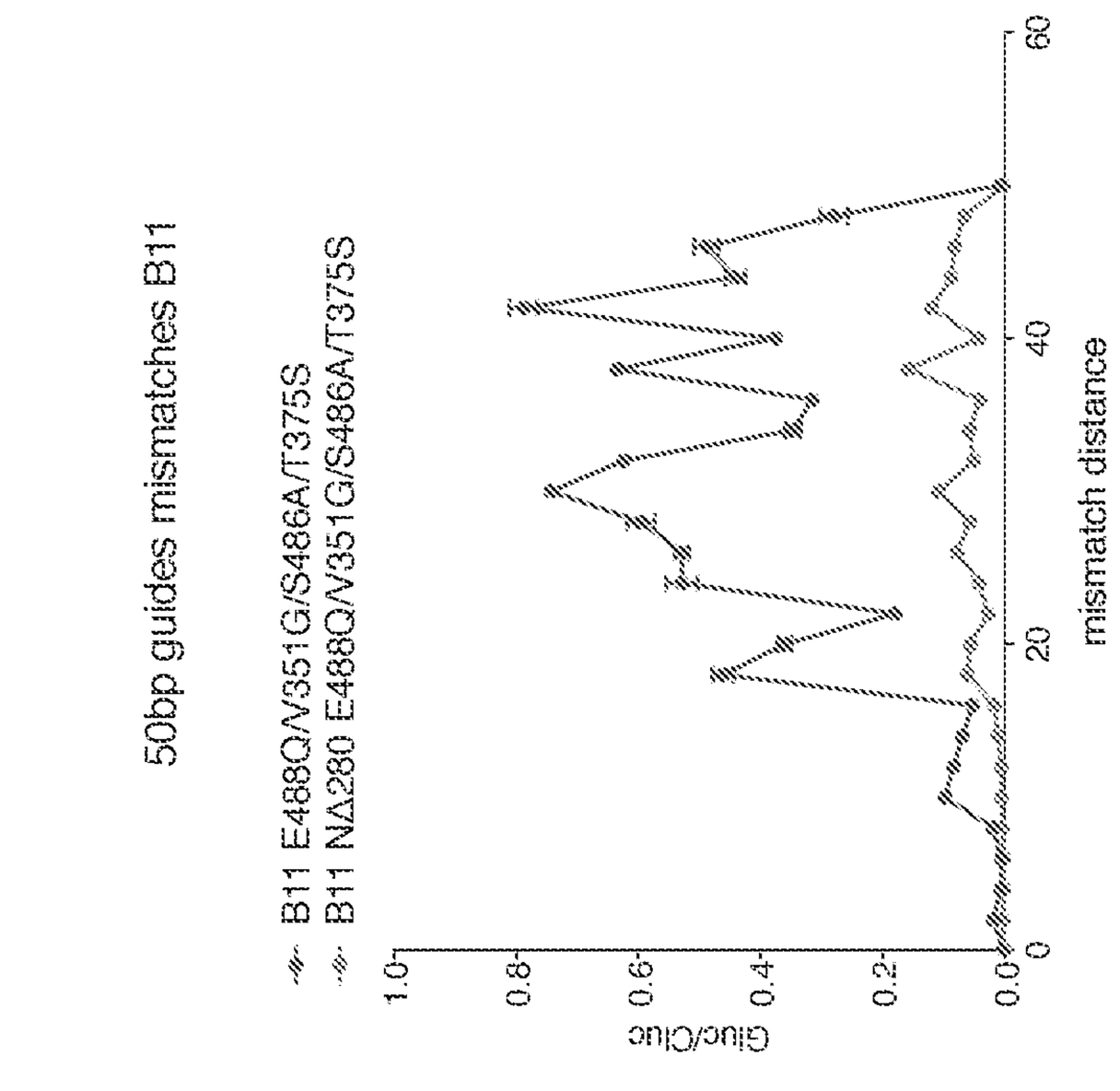


FIG. 134B

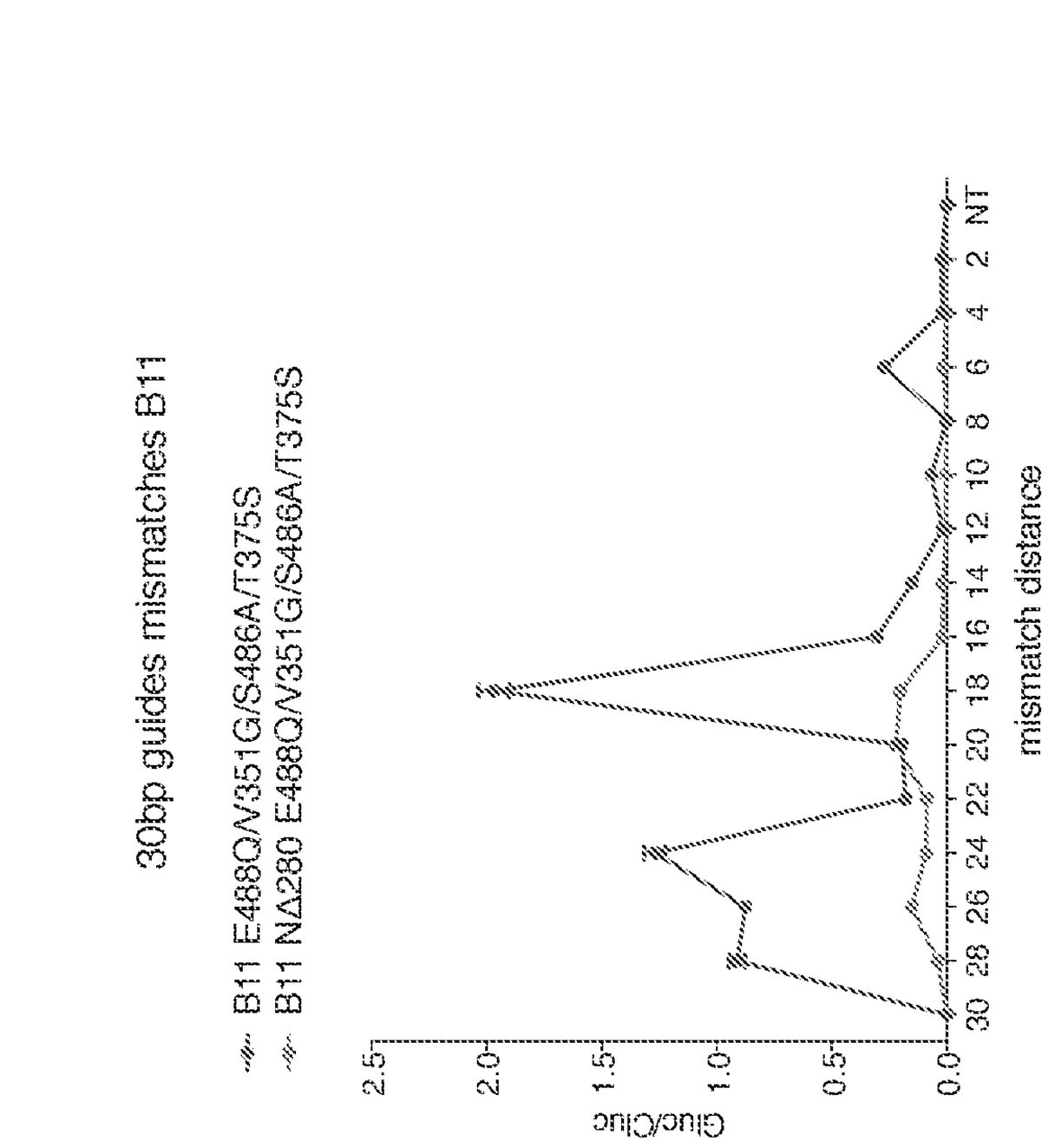


FIG. 134A

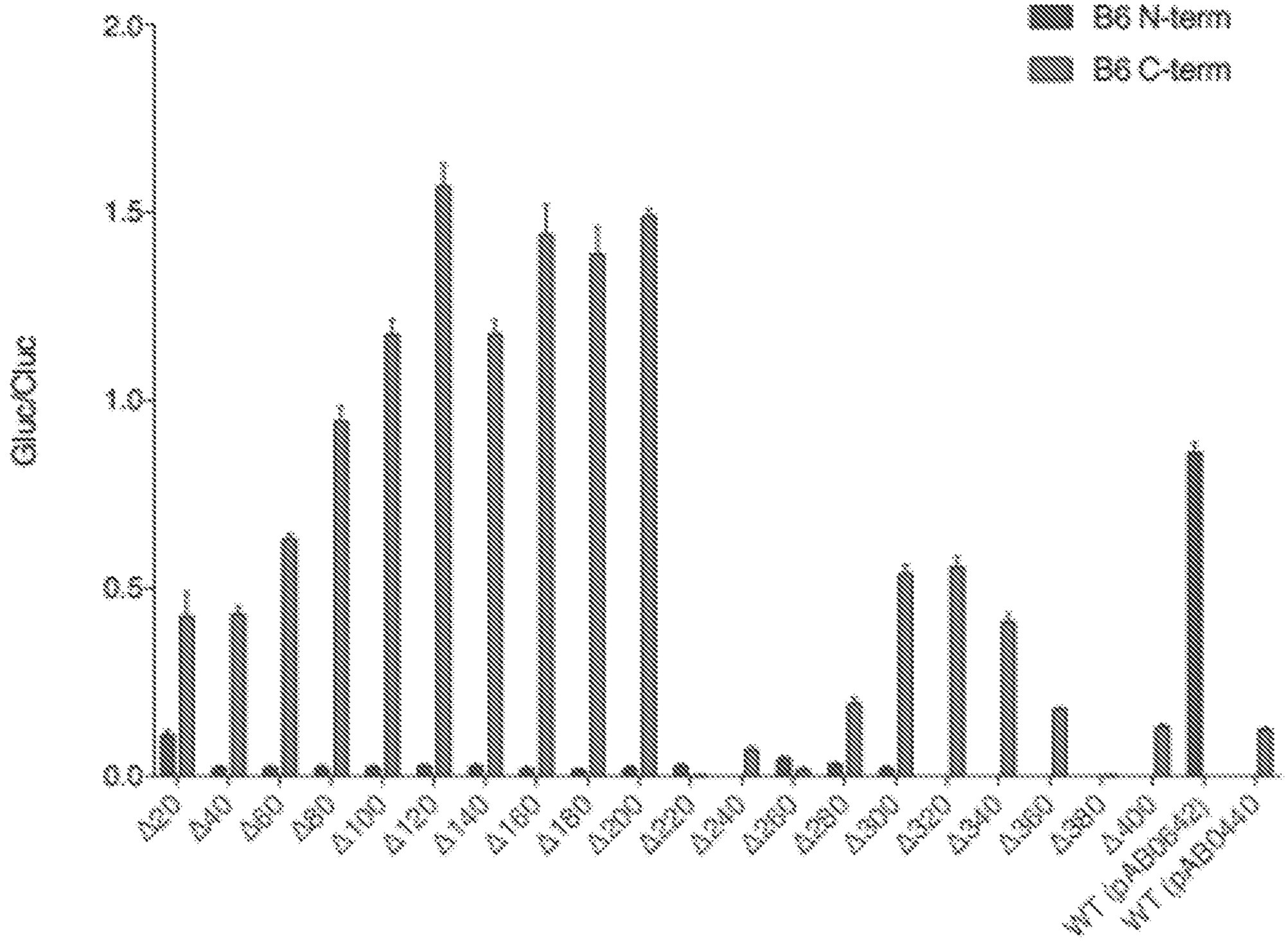

B6
B6 N-term
B6 C-term
20
Glu/Glu
0.0
0.5
1.0
1.5
0:20
0:40
0:60
0:80
0:100
0:120
0:140
0:160
0:180
0:200
0:220
0:240
0:260
0:280
0:300
0:320
0:340
0:360
0:380
0:400
WT (pAB6B5E2)
WT (pAB6B4E0)

FIG. 135

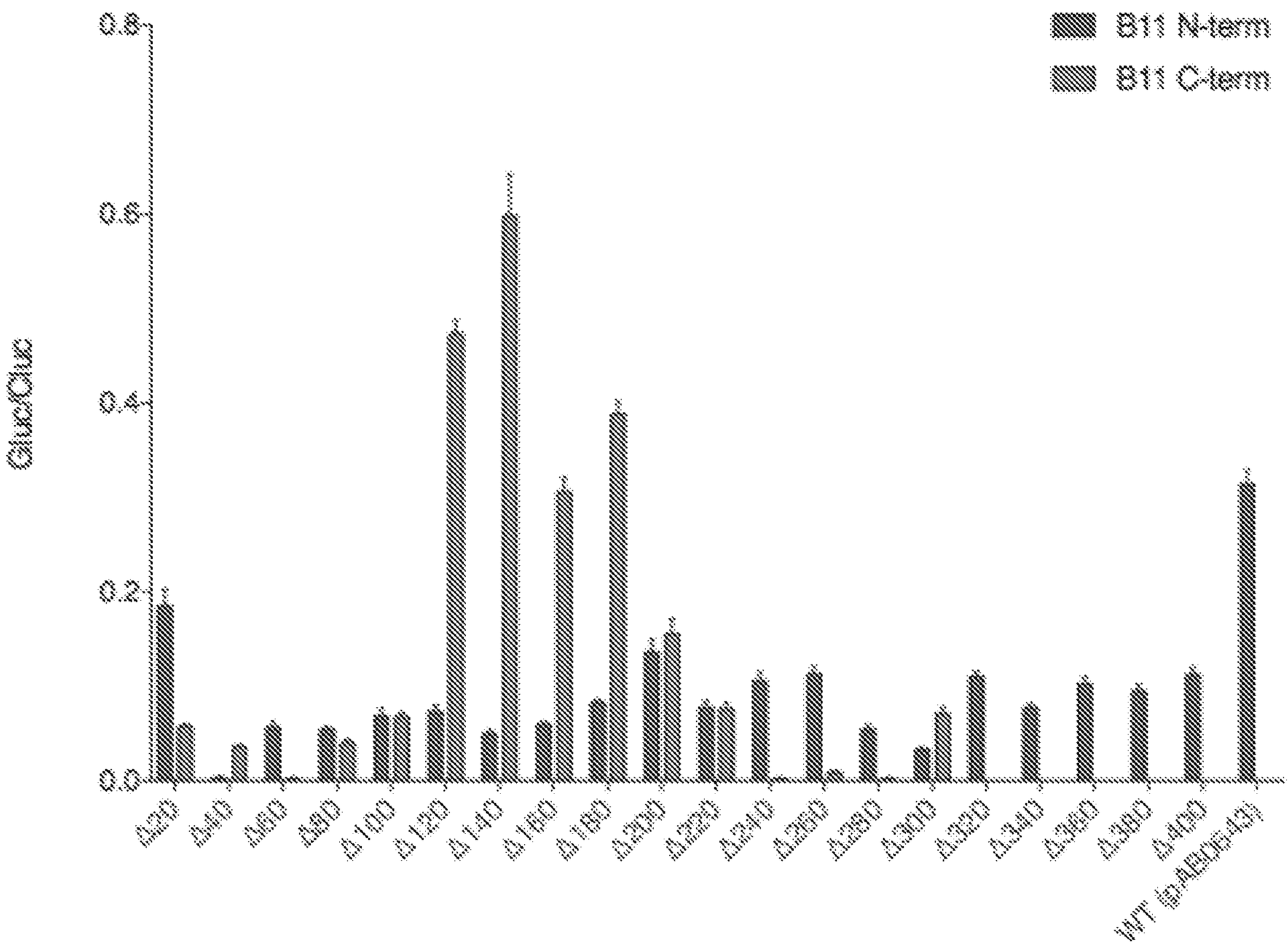

B11
B11 N-term
B11 C-term
GlucOluc
0.0
0.2
0.4
0.6
0.8
0.20
0.40
0.60
0.80
0.100
0.120
0.140
0.160
0.180
0.200
0.220
0.240
0.260
0.280
0.300
0.320
0.340
0.360
0.380
0.400
WT (pAB05-03)

FIG. 136

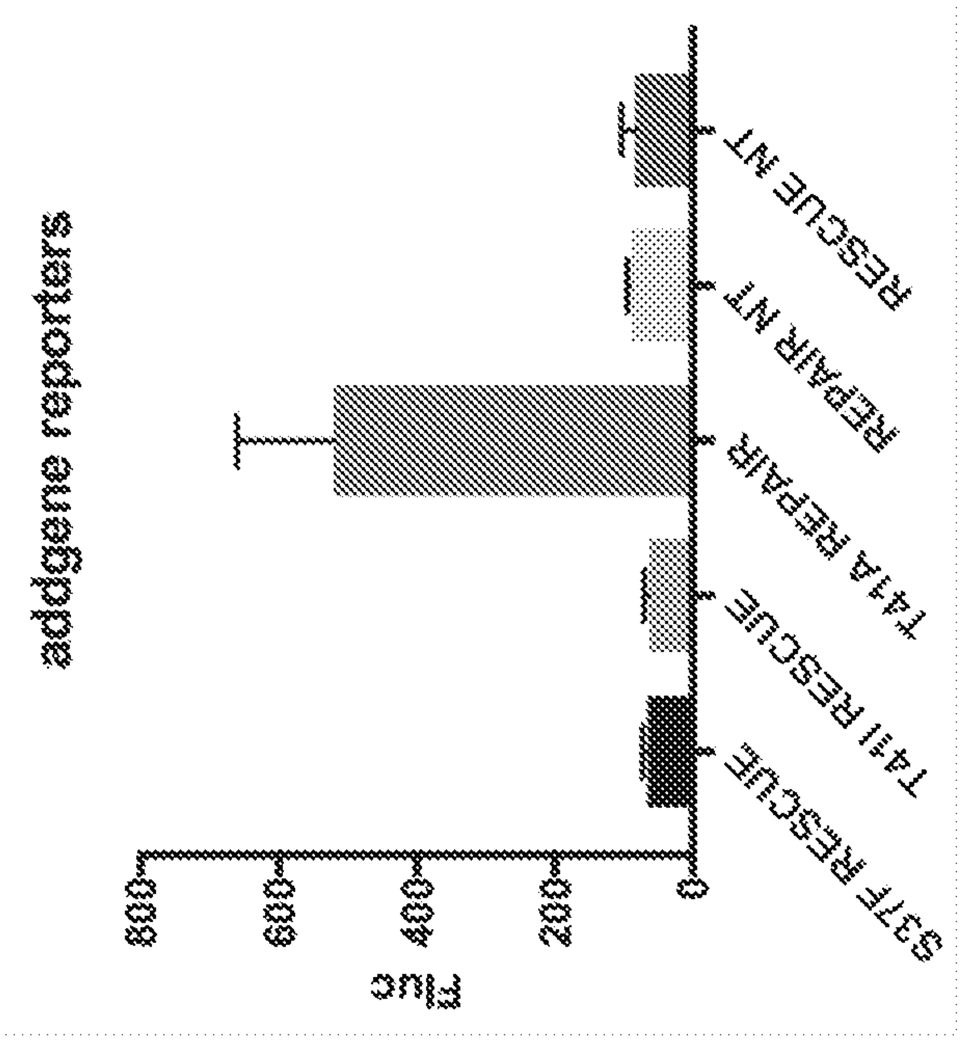


FIG. 137B

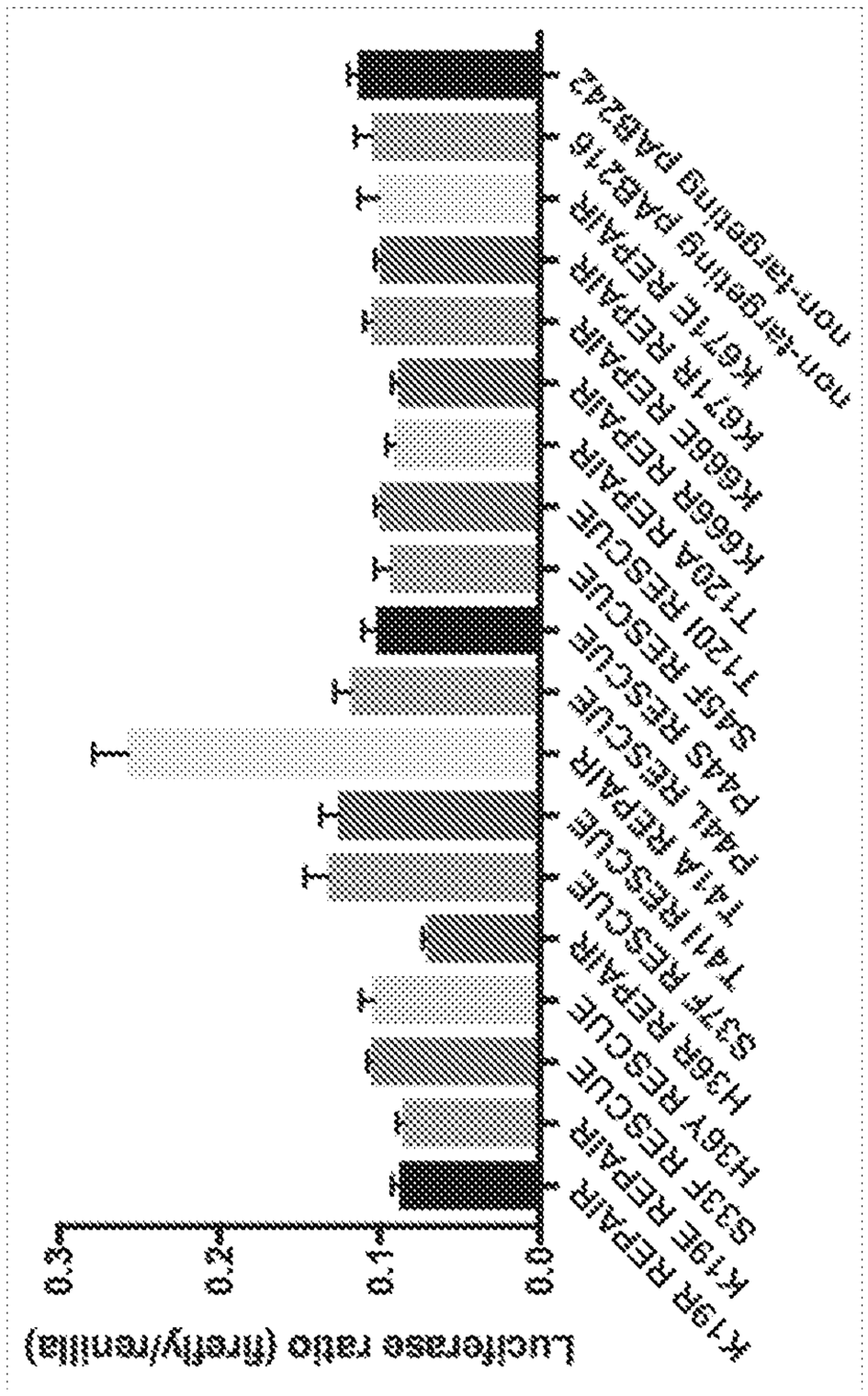


FIG. 137A

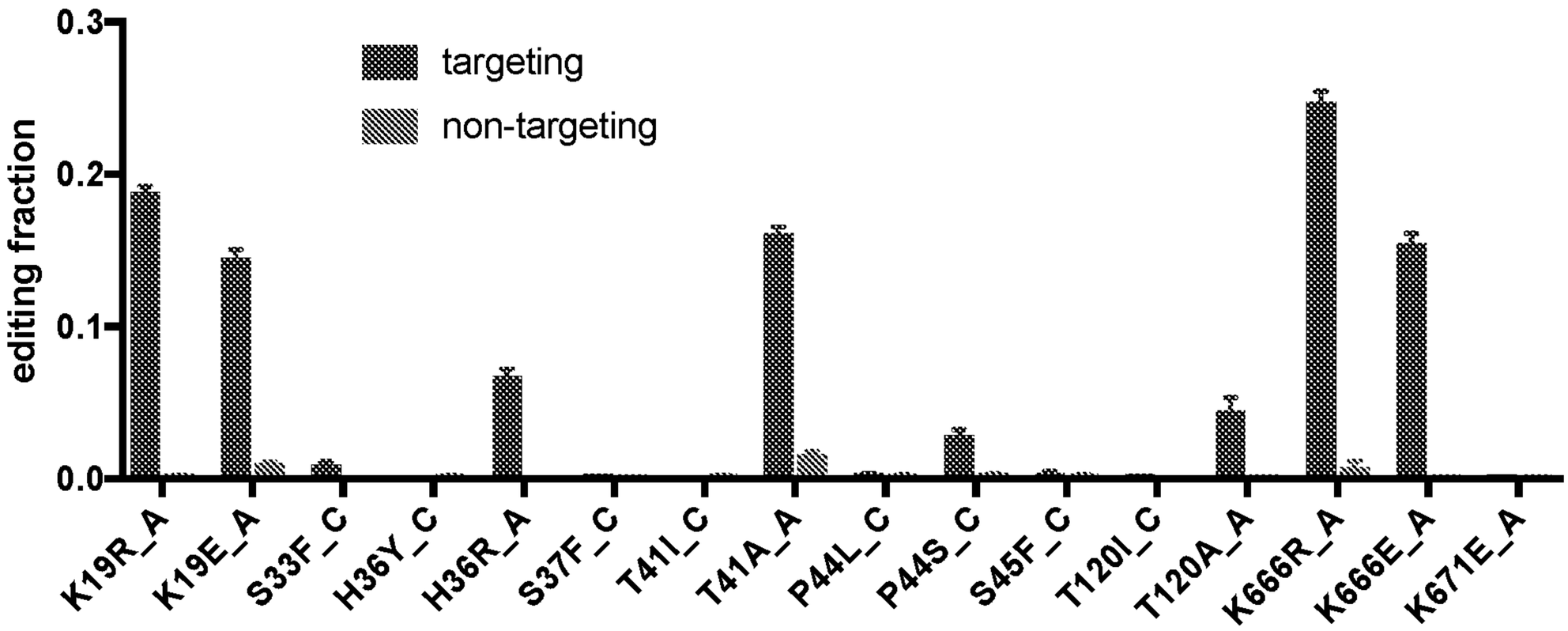

0.3
0.2
0.1
0.0
editing fraction
targeting
non-targeting
K19R_A
K19E_A
S33F_C
H36Y_C
H36R_A
S37F_C
T41I_C
T41A_A
P44L_C
P44S_C
S45F_C
T120I_C
T120A_A
K666R_A
K666E_A
K671E_A

FIG. 138

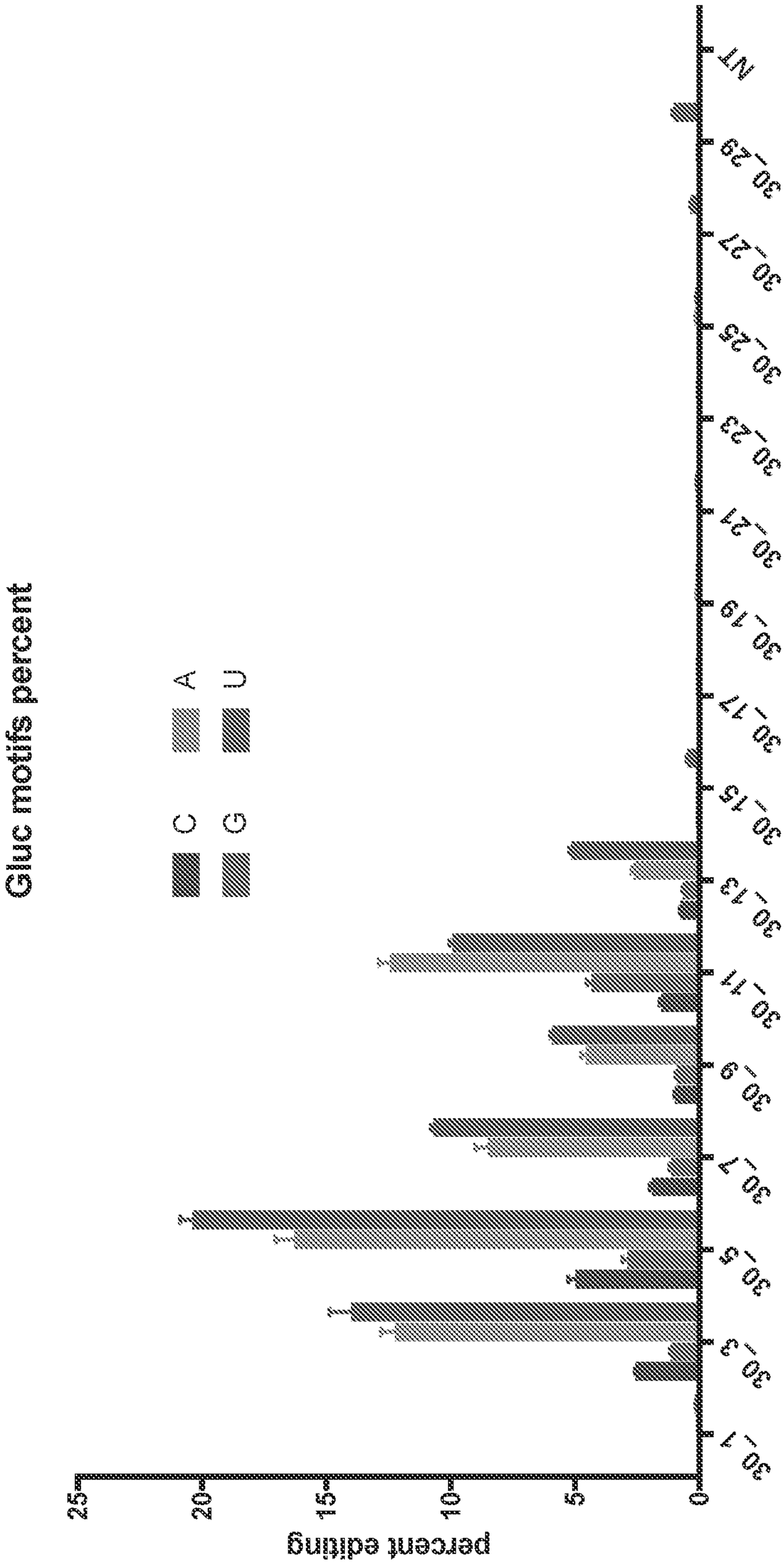

Gluc motifs percent
A
C
G
U
percent editing
0
5
10
15
20
25
NT
30_1
30_3
30_5
30_7
30_9
30_11
30_13
30_15
30_17
30_19
30_21
30_23
30_25
30_27
30_29

FIG. 139

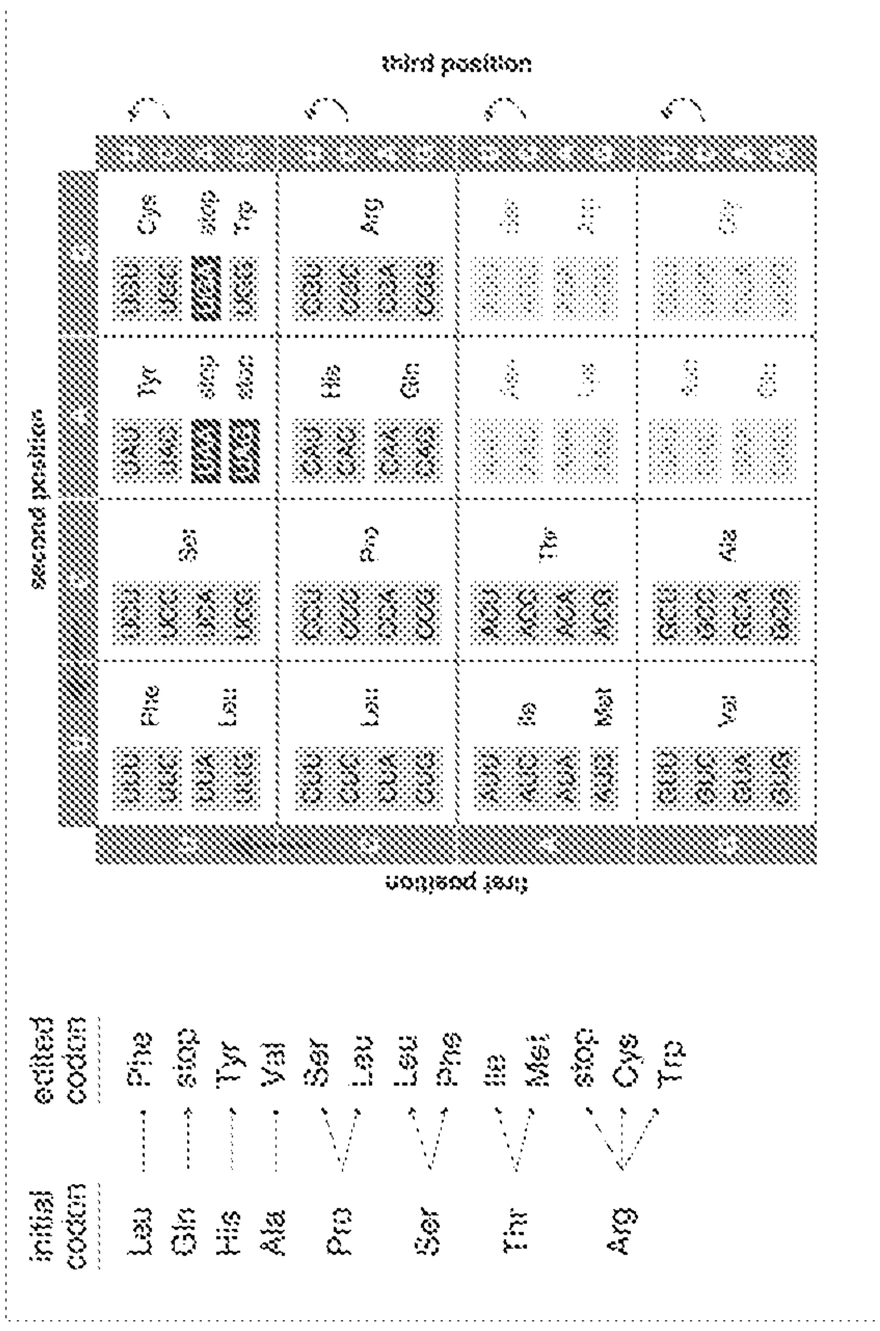


FIG. 40B

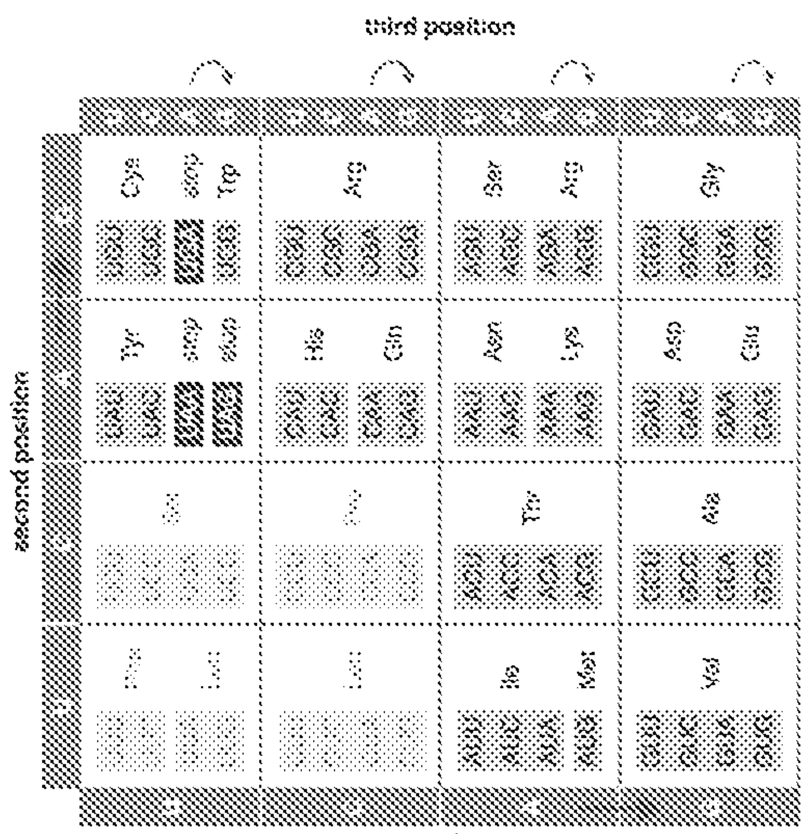


FIG. 40A

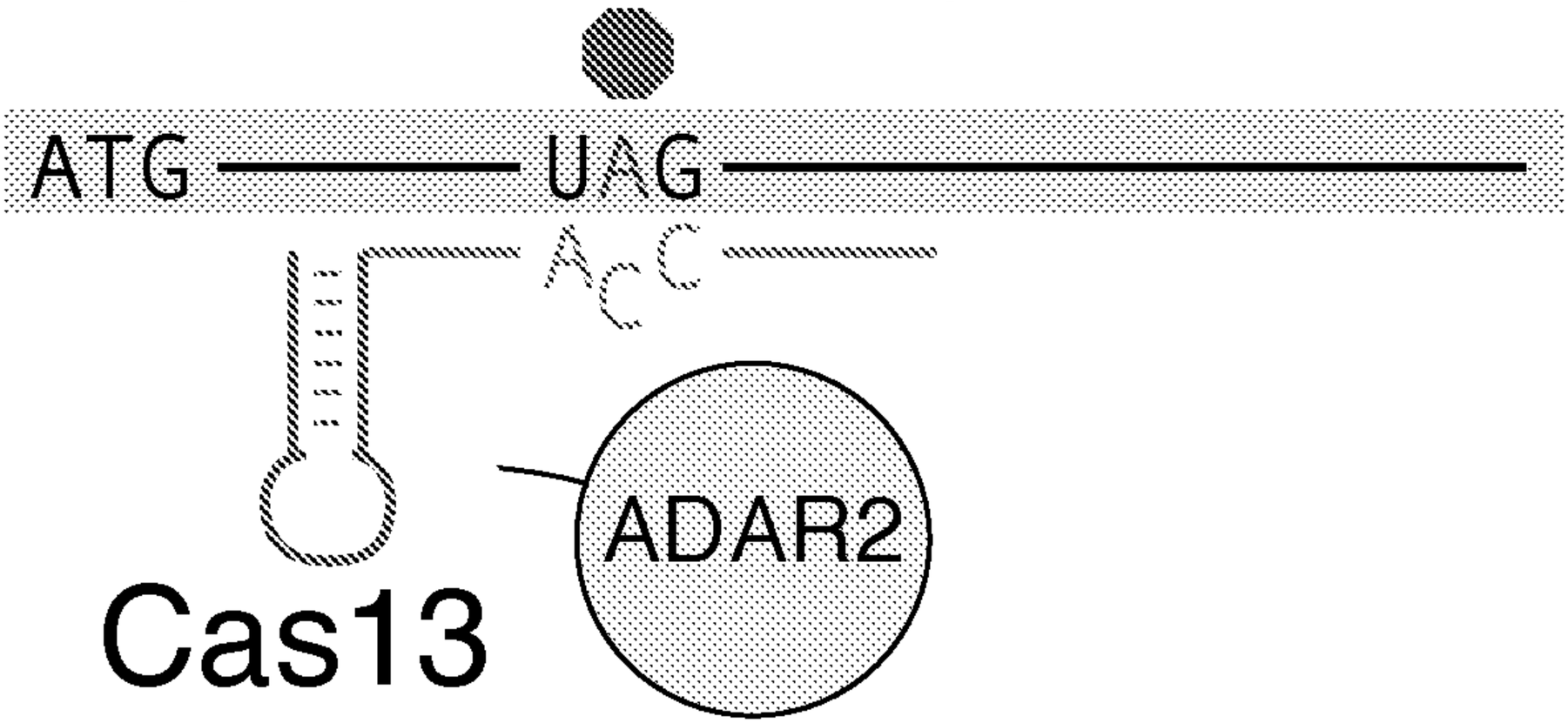


FIG. 1