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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
07 November 2024 (07.11.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/229449 A2

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 15/74* (2006.01)

(21) International Application Number:
PCT/US2024/027887

(22) International Filing Date:
03 May 2024 (03.05.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

63/499,912	03 May 2023 (03.05.2023)	US
63/519,790	15 August 2023 (15.08.2023)	US
63/611,049	15 December 2023 (15.12.2023)	US

(71) Applicant: **METAGENOMI, INC.** [US/US]; 1545 Park Ave, Emeryville, California 94608 (US).

(72) Inventors: **THOMAS, Brian C.**; 1545 Park Ave, Emeryville, California 94608 (US). **BROOKS, Alan**; 1545 Park Ave, Emeryville, California 94608 (US). **BROWN, Christopher**; 1545 Park Ave, Emeryville, California 94608 (US). **BUTTERFIELD, Cristina Noel**; 1545 Park Ave, Emeryville, California 94608 (US). **CASTELLE, Cindy**; 1545 Park Ave, Emeryville, California 94608 (US). **FLAMMER, Steven**; 1545 Park Ave, Emeryville, California 94608 (US). **KREY, Whitney**; 1545 Park Ave, Emeryville, California 94608 (US). **LAMOTHE, Rebecca**; 1545 Park Ave, Emeryville, California 94608 (US). **LIN, Jyun-Liang**; 1545 Park Ave, Emeryville, California 94608 (US). **MACKILLOP, Michael**; 1545 Park Ave, Emeryville, California 94608 (US). **RALLAPALLI, Kartik**; 1545 Park Ave, Emeryville, California 94608 (US). **TAN, Jin Wen**; 1545 Park Ave, Emeryville, California 94608 (US). **TEMOCHE-DIAZ, Morayma**; 1545 Park Ave, Emeryville, California 94608 (US). **HO, Joanne**; 1545 Park Ave, Emeryville, California 94608 (US). **MUYSSON, Jared**; 1545 Park Ave, Emeryville, California 94608 (US).

(74) Agent: **AMAYA, Catherine** et al.; Goodwin Procter LLP, IP DOCKETING DEPT./ 7TH FL, 100 Northern Avenue, Boston, Massachusetts 02210 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: BASE EDITING ENYZMES

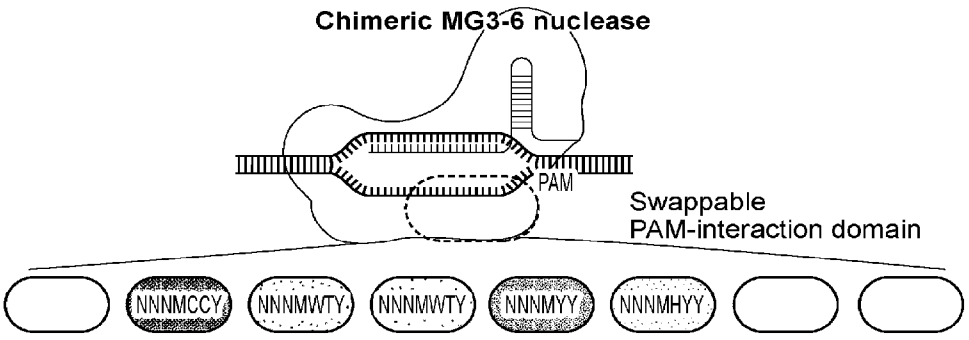

FIG. 29B

(57) Abstract: The present disclosure provides for endonuclease enzymes having distinguishing domain features, as well as methods of using such enzymes or variants thereof.

WO 2024/229449 A2

[Continued on next page]

Declarations under Rule 4.17:

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

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

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

BASE EDITING ENZYMES

CROSS-REFERENCE

[0001] This application claims the benefit of and priority to U.S. Provisional Patent Application No. 63/499,912 filed May 3, 2023, U.S. Provisional Patent Application No. 63/519,790 filed August 15, 2023, and U.S. Provisional Patent Application No. 63/611,049 filed December 15, 2023, each of which is incorporated by reference in its entirety herein.

BACKGROUND

[0002] Cas enzymes along with their associated Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) guide ribonucleic acids (RNAs) appear to be a pervasive (~45% of bacteria, ~84% of archaea) component of prokaryotic immune systems, serving to protect such microorganisms against non-self nucleic acids, such as infectious viruses and plasmids by CRISPR-RNA guided nucleic acid cleavage. While the deoxyribonucleic acid (DNA) elements encoding CRISPR RNA elements may be relatively conserved in structure and length, their CRISPR-associated (Cas) proteins are highly diverse, containing a wide variety of nucleic acid-interacting domains. While CRISPR DNA elements have been observed as early as 1987, the programmable endonuclease cleavage ability of CRISPR complexes has only been recognized relatively recently, leading to the use of recombinant CRISPR systems in diverse DNA manipulation and gene editing applications.

SEQUENCE LISTING

[0003] The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. The XML copy, created on May 3, 2024 and is 2,926,539 bytes in size.

SUMMARY

[0004] Described herein, in certain embodiments, are engineered base editing systems comprising an engineered base editing system comprising: a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415; and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

[0005] In some embodiments, the base editor comprises a sequence having at least 95% sequence identity to any one of SEQ ID NOs: 1654-1703, and 2021-2023.

[0006] In some embodiments, the base editor comprises a sequence having 100% sequence identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023.

[0007] Described herein, in certain embodiments, are engineered base editing system comprising: a base editor encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757, and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

[0008] In some embodiments, base editor is encoded by a nucleic acid sequence having at least 90% identity to any one of SEQ ID NOs: 1727-1757.

[0009] In some embodiments, the base editor is encoded by a nucleic acid sequence having 100% identity to any one of SEQ ID NOs: 1727-1757.

[0010] In some embodiments, the base editor comprises a deaminase. In some embodiments, the deaminase binds non-covalently to the endonuclease. In some embodiments, the deaminase is covalently linked to the endonuclease. In some embodiments, the deaminase is fused to the endonuclease. In some embodiments, the engineered guide polynucleotide is a single guide nucleic acid. In some embodiments, the engineered guide polynucleotide is a dual guide nucleic acid. In some embodiments, the engineered guide polynucleotide is RNA.

[0011] In some embodiments, the endonuclease binds non-covalently to the engineered guide polynucleotide. In some embodiments, the endonuclease is covalently linked to the engineered guide polynucleotide.

[0012] In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019.

[0013] In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 90% sequence identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019.

[0014] In some embodiments, the engineered guide polynucleotide comprises a sequence having 100% sequence identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019.

[0015] In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019.

[0016] In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1431-1454, 1704, and 2010-2019.

[0017] In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1890-1976.

[0018] In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1977-2009.

[0019] In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1479-1483 and 1758-1889.

[0020] In some embodiments, the base editor comprises a nickase domain.

[0021] In some embodiments, the nickase comprises an aspartate to alanine mutation at residue 9 relative to SEQ ID NO: 70, residue 13 relative to SEQ ID NOs: 71, 72, or 74, residue 12 relative to SEQ ID NO: 73, residue 17 relative to SEQ ID NO: 75, residue 23 relative to SEQ ID NO: 76, or residue 10 relative to SEQ ID NO: 597, or any combination thereof.

[0022] In some embodiments, the base editor further comprises a uracil DNA glycosylase inhibitor sequence. In some embodiments, the base editor further comprises a FAM72A sequence. In some embodiments, the FAM72A sequence has at least 80% identity to SEQ ID NO: 1121.

[0023] Described herein, in certain embodiments, are nucleic acids encoding engineered base editing systems described herein.

[0024] Described herein, in certain embodiments, are vectors comprising nucleic acids disclosed herein. In some embodiments, the the vector is a plasmid, a minicircle, a CELiD, an adeno-associated virus (AAV) derived virion, a lentivirus, or an adenovirus.

[0025] Described herein, in certain embodiments, are cells comprising the engineered systems described herein, the nucleic acids described herein, or the vectors described herein. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the cell is a mammalian cell. In some embodiments, the cell is an immortalized cell. In some embodiments, the cell is an insect cell. In some embodiments, the cell is a yeast cell. In some embodiments, the cell is a plant cell. In some embodiments, the cell is a fungal cell. In some embodiments, the cell is a prokaryotic cell. In some embodiments, the cell is an A549, HEK-293, HEK-293T, BHK, CHO, HeLa, MRC5, Sf9, Cos-1, Cos-7, Vero, BSC 1, BSC 40, BMT 10, WI38, HeLa, Saos, C2C12, L cell, HT1080, HepG2, Huh7, K562, primary cell, or a derivative thereof. In some embodiments, the cell is an engineered cell. In some embodiments, the cell is a stable cell.

[0026] Described herein, in certain embodiments, are methods for modifying a target nucleic acid sequence, comprising: contacting the target nucleic acid sequence using an engineered base editing system disclosed herein. In some embodiments, modifying the target nucleic acid sequence comprises converting an adenine to a guanine in the target nucleic acid sequence. In some embodiments, modifying the target nucleic acid sequence comprises converting a cytosine to a uracil in the target nucleic acid sequence. In some embodiments, the target nucleic acid sequence comprises deoxyribonucleic acid (DNA). In some embodiments, the target nucleic acid sequence comprises ribonucleic acid (RNA). In some embodiments, the target nucleic acid sequence comprises genomic DNA, viral DNA, viral RNA, or bacterial DNA.

[0027] In some embodiments, the target nucleic acid sequence is modified in vitro. In some embodiments, target nucleic acid sequence is modified in vivo. In some embodiments, the target nucleic acid sequence is modified ex vivo. In some embodiments, the target nucleic acid sequence is modified within a cell. In some embodiments, the cell is a prokaryotic cell, a bacterial cell, a eukaryotic cell, a fungal cell, a plant cell, an animal cell, a mammalian cell, a rodent cell, a primate cell, a human cell, or a primary cell.

[0028] Described herein, in certain embodiments, are methods of modifying a nucleic acid encoding ANGPTL3 comprising contacting the nucleic acid sequence encoding ANGPTL3 with an engineered base editing system, said base editing system comprising:a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415, or encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757; and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1479-1483 and 1758-1889.

[0029] Described herein, in certain embodiments, are methods of modifying a nucleic acid encoding APOA1 comprising contacting the nucleic acid sequence encoding APOA1 with an engineered base editing system, said base editing system comprising: a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415, or encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757; and an engineered guide polynucleotide which

forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1431-1454, 1704, and 2010-2019.

[0030] Described herein, in certain embodiments, are methods of modifying a nucleic acid encoding BCL11A comprising contacting the nucleic acid sequence encoding BCL11A with an engineered base editing system, said base editing system comprising a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415, or encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757, and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1890-1976.

[0031] Described herein, in certain embodiments, are methods of modifying a nucleic acid encoding PAH comprising contacting the nucleic acid sequence encoding PAH with an engineered base editing system, said base editing system comprising a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415, or encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757, and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1977-2009.

[0032] Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative embodiments of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

BRIEF DESCRIPTION OF THE DRAWINGS

[0033] The novel features of the disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and the accompanying drawings (also "Figure" and "FIG." herein), of which:

[0034] FIG. 1A – FIG. 1K depict dose dependence of A→G editing activity by ABE07 at 11 distinct target sites through transfection in primary mouse hepatocytes. Barplots show the average editing by ABE07 at adenines within the protospacer editing window accessible to the base editor. Target adenines are numbered with respect to the PAM-distal position. Values and error bars represent mean ± s.e.m. of n = 4 different biological replicates. The varying concentrations of ABE07 mRNA that were transfected are highlighted in the legend.

[0035] FIG. 2 depicts a schematic diagram highlighting the domain architecture of the homodimeric and heterodimeric ABEs. Two copies of MG68-4 variants were inserted into the MG3-6/3-8 nuclease chassis allowing for PID swapping to diversify the PAM accessible to these ABE enzymes.

[0036] FIG. 3A - FIG. 3B depict boxplots showing comparison of A→G editing by ABE variants at 31 distinct gRNAs targeting 3 different genes in Hepa1-6 cells. **FIG. 3A** depicts boxplots showing the mean observed editing, and **FIG. 3B** depicts boxplots showing the max observed editing by individual ABE variants, with data corresponding to n = 2 biologically independent replicates for each target guide. Each individual data point represents a unique target locus.

[0037] FIGs. 4A - 4B depict boxplots showing comparison of C→G promiscuous editing by ABE variants at 31 distinct gRNAs targeting 3 different genes in Hepa1-6 cells. **FIG. 4A** depicts boxplots showing the mean observed editing, and **FIG. 4B** depicts boxplots showing the max observed editing by individual ABE variants, with data corresponding to n = 2 biologically independent replicates for each target guide. Each individual data point represents a unique target locus.

[0038] FIG. 5 depicts boxplots showing the comparison of unwanted indel formation by ABE variants at 31 distinct gRNAs targeting 3 different genes in Hepa1-6 cells. Boxplots show the mean observed indel formation by individual ABE variants, with data corresponding to n = 2 biologically independent replicates for each target guide. Each individual data point represents a unique target locus.

[0039] FIG. 6 depicts a heatmap showing the guide-wise and variant-wise breakdown of maximum A→G editing activity observed across the highly edited gRNAs targeting 3 different genes in Hepa1-6 cells. The guides are ranked according to the observed A→G editing activity in decreasing order from left to right. ABE07-77 is highlighted along with the guides chosen for further *in vivo* investigation.

[0040] FIG. 7 depicts in vivo A→G editing activity for the ABE07-77 variant across 6 distinct gRNA targeting. Boxplots depict the max observed editing by ABE07-77. Individual data points represent the individual mice. Values and error bars represent mean ± s.e.m. of n = 4 different mice.

[0041] FIG. 8 depicts in vivo indel formation activity of the ABE07-77 variant across 6 distinct gRNA targeting. Barplots depict the indels formed by ABE07-77 and corresponding parent nuclease, MG3-6/3-8. Individual data points represent the individual mice. Values and error bars represent mean ± s.e.m. of n = 4 different mice.

[0042] FIG. 9 depicts in vivo A→G editing activity of ABE07-77 variant across the protospacer at the highest edited target loci. Barplots depict the average editing by ABE07-77 at adenines within the protospacer editing window accessible to the base editor. Target adenines are numbered with respect to the PAM-distal position. Individual data points represent the individual mice. Values and error bars represent mean ± s.e.m. of n = 4 different mice.

[0043] FIG. 10 depicts a graph showing that max C to T editing for 139-52-V2, 139-52-V13, 139-52-V14, 139-52-V17, 139-86v12, and 152-6v13 across all 5 guides is 26.2%, 58.2%, 22.7%, 16.8%, 12.6%, 50.9% respectively. When compared to the hyperactive positive control variant: 139-52-V2, 139-52-V13, 139-52-V14, 139-52-V17, 139-86v12, and 152-6v13 edited at 159.2%, 102.1%, 353.7%, 137.8%, 76.3%, and 309.1% of the maximum positive control editing respectively.

[0044] FIG. 11 depicts bar plots showing -1 nucleotide in the 5' position of the cytidine being deaminated has been shown to be important for binding and deamination reaction by the cytidine deaminase. Using NGS data, the -1 nucleotide preference for each of our 6 engineered CDA variants were measured. To do this, the number of reads of the four highest edited cytidine sites for each guide that edited >1% (and their -1 nucleotide identity) were tabulated. This was done separately for all five guides after which the preference across all 5 guides was averaged. The resulting graph represents the relative -1 nucleotide preference each CDA has for the type of cytidine it prefers to deaminate across the five guides targeting the HEK293 engineered site.

[0045] FIG. 12A and **FIG. 12B** depict graphs showing the comparison of on target A→G editing by oligomeric ABE variants at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells. Boxplots depict the Mean observed editing (**FIG. 12A**) and Max observed editing (**FIG. 12B**) by

individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data points each represent a unique target locus.

[0046] FIG. 13A and **FIG. 13B** depict graphs showing the comparison of promiscuous C editing by oligomeric ABE variants at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells. Boxplots depict the Mean observed C editing (**FIG. 13A**) and Max observed C editing (**FIG. 13B**) by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data points each represent a unique target locus.

[0047] FIG. 14A and **FIG. 14B** depict graphs showing the comparison of on target A→G editing by engineered ABE variants at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells. Boxplots depict the Mean observed editing (**FIG. 14A**) and Max observed editing (**FIG. 14B**) by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data points each represent a unique target locus.

[0048] FIG. 15A and **FIG. 15B** depict graphs showing the comparison of promiscuous C editing by engineered ABE variants at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells. Boxplots depict the Mean observed C editing (**FIG. 15A**) and Max observed C editing (**FIG. 15B**) by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data each points represent a unique target locus.

[0049] FIG. 16 depicts a graph showing the comparison of unwanted indel formation by engineered ABE variants at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells. Boxplots depict the mean observed indel formation (indel percentage) by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data points each represent a unique target locus.

[0050] FIG. 17A - FIG. 17C depict graphs showing the preferred sequence context of the edited adenines. Frequency distribution of bases surrounding all editable adenines within the 15 distinct gRNA tested in this study (**FIG. 17A**), the highly edited (greater than 30% editing activity) adenines by ABE15 (**FIG. 17B**), and the less edited (less than 30% editing activity) adenines by ABE15 (**FIG. 17C**).

[0051] FIG. 18A and **FIG. 18B** depict graphs showing the comparison of on target A→G editing by engineered ABE variants at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells. Boxplots depict the Mean observed editing (**FIG. 18A**) and Max observed editing (**FIG. 18B**) by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data points each represent a unique target locus.

[0052] FIG. 19A and **FIG. 19B** depict graphs showing the comparison of promiscuous C editing by engineered ABE variants at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells. Boxplots depict the Mean observed C editing (**FIG. 19A**) and Max observed C editing (**FIG.**

19B) by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data points each represent a unique target locus.

[0053] FIG. 20A and **FIG. 20B** depict graphs showing the of unwanted indel formation engineered ABE variants at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells and preferred sequence editing context for engineered ABE23. **FIG. 20A** depicts boxplots showing the mean observed indel formation by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data points each represent a unique target locus. **FIG. 20B** shows the frequency distribution of bases surrounding the highly edited (greater than 30% editing activity) adenines by ABE23.

[0054] FIG. 21A and **FIG. 21B** depict graphs showing the comparison of on target A→G editing by engineered D109Q ABE variants and variants with additional proline mutations designed to relax the sequence preference at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells. Boxplots depict the Mean observed editing (**FIG. 21A**) and Max observed editing (**FIG. 21B**) by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data points each represent a unique target locus. D109Q mutations designed to curb promiscuous C deamination increased the mean and max A to G editing.

[0055] FIG. 22A and **FIG. 22B** depict graphs showing the comparison of promiscuous C editing by engineered D109Q ABE variants and variants with additional proline mutations designed to relax the sequence preference at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells. Boxplots depict the Mean observed C editing (**FIG. 22A**) and Max observed C editing (**FIG. 22B**) by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide.

[0056] FIG. 23 depicts a graph showing the comparison of unwanted indel formation and editing context for engineered D109Q variants and variants with additional proline mutations designed to relax the sequence preference at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells. Boxplots depict the mean observed indel formation by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data points each represent a unique target locus.

[0057] FIG. 24A and **FIG. 24B** depict graphs showing the comparison of on target A→G editing by top performing ABEs at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells. Boxplots depict the Mean observed editing (**FIG. 24A**) and Max observed editing (**FIG. 24B**) by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data points each represent a unique target locus.

[0058] FIG. 25A and **FIG. 25B** depict graphs showing the comparison of promiscuous C editing by top performing ABEs at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells.

Boxplots depict the Mean observed C editing (**FIG. 25A**) and Max observed C editing (**FIG. 25B**) by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data points each represent a unique target locus.

[0059] FIG. 26 depicts a graph showing the comparison of unwanted indel formation for top performing ABEs at 15 distinct gRNA targeting 2 different genes in Hepa1-6 cells. Boxplots depict the mean observed indel formation by individual ABE variants, with data corresponding to n= 2 biologically independent replicates for each target guide. Individual data points each represent a unique target locus.

[0060] FIGs. 27A-FIG. 27C depict small base editor fusion constructs, including a schematic of the small base editors indicating the relative position of the deaminase domain with respect to the nickase domain (**FIG. 27A**), the structural alignment of the computationally-predicted structure of MG3-6_3-8 ABE (light grey) with the predicted structure of MG34-29 nuclease (dark grey) (**FIG. 27B**), and an alternative view in which the MG3-6_3-8 nickase is omitted from the alignment shown in **FIG. 27B** to uncover a potential loop within the MG34-29 structure prediction to fuse the MG68-4 deaminase (**FIG. 27C**).

[0061] FIGs. 28A-FIG. 28D depict graphs demonstrating efficient A→G editing activity by small base editors in immortalized human K562 cells**.** Barplots depict the maximum observed A→G editing by MG34-29 ABEs and a positive control for the hAPOA1 target guide, ABE33 (**FIG. 28A**) and MG102-39 ABEs (**FIG. 28C**) across a panel of multiple guides. The average editing observed across the highly edited target loci, AAVS1 E7, AAVS1 C7 (**FIG. 28B**), and TRAC C11 (**FIG. 28D**) are depicted as barplots. Target adenines are numbered with respect to the PAM-distal position in the respective ABE variant. Values and error bars represent mean ± s.e.m. of n= 2 biologically independent replicates for each target guide.

[0062] FIGs. 29A-FIG. 29C depict PAM-interacting domain (PID) engineering that was used to develop a suite of chimeric MG3-6 base editors with extensive genome targetability. **FIG. 29A** shows the percentage of genomic adenines in the hg38 human reference genome that are targetable by SpCas9 ABEs. **FIG. 29B** depicts a schematic diagram of the PID-swappable chimeric nuclease platform highlighting the diversity and range of PAMs accessible through PID-swapping. **FIG. 29C** depicts the percentage of genomic adenines in the hg38 human reference genome that are targetable using the MG3-6 PID-swappable ABEs platform.

[0063] FIGs. 30A-FIG. 30D depict a schematic of the experimental design for high-throughput chimeric ABE testing. **FIG. 30A** shows an editing window of two highly active MG3-6_3-8 ABEs**.** Each dot represents a unique editing event observed across over 15 unique guides by ABE07 (**SEQ ID NO: 1411**) and ABE33 (**SEQ ID NO: 1673**). The Gaussian curves on the top panel indicate the estimated editing window for MG3-6_3-8 ABEs that was used to design the

tiling guides used throughout this experiment. **FIG. 30B** shows a scheme depicting base editing of the splice sites using ABEs, which can lead to intron retention (splice donor disruption) or exon skipping (splice acceptor disruption) and hence be utilized for targeted knockout of genes. Plate layout for **(FIG. 30C)** high-throughput pooled screen and **(FIG. 30D)** deconvolution screen were used to identify highly active ABEs and guide combinations that yield efficient targeted knockouts.

[0064] FIGs. 31A- FIG. 31B depict targeted knockout of hANGPTL3 using PID-swapped chimeric ABEs. **FIG. 31A** shows a graph of the comparison of theoretical splice site targetability of reference ABEs and MG3-6_3-8 chimeric ABEs. **FIG. 31B** shows the results of pooled screening of compatible chimeric ABEs across the splice sites of hANGPTL3 gene as tested in K562 cells. The heatmap values indicate the percentage A→G editing observed at the splice site adenines by the corresponding chimeric ABE variants.

[0065] FIGs. 32A- FIG. 32D show targeted knockout of hANGPTL3 using PID-swapped chimeric ABEs. Barplots of the A→G editing at the splice site adenines used to deconvolute the active guides and chimeric ABE combinations at the hANGPTL3 exon 1 splice donor (FIG. 32A), exon 4 splice acceptor and splice donor **(FIG. 32B)**, exon 5 splice donor **(FIG. 32C)**, and exon 7 splice acceptor **(FIG. 32D)**.

[0066] FIGs. 33A- FIG. 33C depict targeted enhancer site disruption of GATA1-binding site in hBCL11A using PID-swapped chimeric ABEs. **FIG. 33A** shows graph of the comparison of theoretical targetability GATA1 enhancer binding site of reference ABEs and MG3-6_3-8 chimeric ABEs. **FIG. 33B** shows the results of pooled screening of compatible chimeric ABEs across the various enhancer factor binding sites of BCL11A gene tested in K562 cells. The heatmap values indicate the percentage A→G editing observed at the GATA site adenine by the corresponding chimeric ABE variants. **FIG. 33C** shows a barplot of the A→G editing at the GATA1 adenines used to deconvolute the active guides and chimeric ABE combinations at DHS+58 enhancer binding site.

[0067] FIGs. 34A- FIG. 34D depict targeted correction of hPAH SNVs using PID-swapped chimeric ABEs. **FIG. 34A** shows a table of the most common pathogenic SNVs reported in hPAH and their responsiveness to Sapropterin, adapted from *Regier DS, Greene CL. Phenylalanine Hydroxylase Deficiency.* **FIGs. 34B-34C** show pooled screening of compatible chimeric ABEs across the SNVs sites of hPAH gene as tested in modified cell line. The heatmap values indicate the percentage A→G editing observed at the target adenines as well as the bystander A's by the corresponding chimeric ABE variants for the following SNVs: 1222C>T (p.R408W) SNV **(FIG. 34B)**, 1066-11G>A **(FIG. 34C)**, and 1315+1G>A **(FIG. 34D)**.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

[0068] The Sequence Listing filed herewith provides exemplary polynucleotide and polypeptide sequences for use in methods, compositions and systems according to the disclosure. Below are exemplary descriptions of sequences therein.

[0069] SEQ ID NOs: 1-47 show the full-length peptide sequences of MG66 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0070] SEQ ID NOs: 48-49 show the full-length peptide sequences of MG67 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0071] SEQ ID NOs: 50-51 show the full-length peptide sequences of MG68 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0072] SEQ ID NOs: 52-56 show the sequences of uracil DNA glycosylase inhibitors suitable for the engineered nucleic acid editing systems described herein.

[0073] SEQ ID NOs: 57-66 show the sequences of reference deaminases.

[0074] SEQ ID NO: 67 shows the sequence of a reference uracil DNA glycosylase inhibitor.

[0075] SEQ ID NO: 68 shows the sequence of an adenine base editor.

[0076] SEQ ID NO: 69 shows the sequence of a cytosine base editor.

[0077] SEQ ID NOs: 70-78 show the full-length peptide sequences of MG nickases suitable for the engineered nucleic acid editing systems described herein.

[0078] SEQ ID NOs: 79-87 shows the protospacer and PAM used in *in vitro* nickase assays described herein.

[0079] SEQ ID NOs: 88-96 show the peptide sequences of single guide RNA used in *in vitro* nickase assays described herein.

[0080] SEQ ID NOs: 97-156 show the sequences of spacers when targeting *E. coli* lacZ.

[0081] SEQ ID NOs: 157-176 show the sequences of primers when conducting site directed mutagenesis.

[0082] SEQ ID NOs: 177-178 show the sequences of primers for lacZ sequencing.

[0083] SEQ ID NOs: 179-342 show the sequences of primers used during amplification.

[0084] SEQ ID NOs: 343-345 show the sequences of primers for lacZ sequencing.

[0085] SEQ ID NOs: 346-359 show the sequences of primers used during amplification.

[0086] SEQ ID NOs: 360-368 show protospacer adjacent motifs suitable for the engineered nucleic acid editing systems described herein.

[0087] SEQ ID NOs: 369-384 show nuclear localization sequences (NLS's) suitable for the engineered nucleic acid editing systems described herein.

[0088] SEQ ID NOs: 385-443 show the full-length peptide sequences of MG68 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0089] SEQ ID NOs: 444-447 show the full-length peptide sequences of MG121 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0090] SEQ ID NOs: 448-475 show the full-length peptide sequences of MG68 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0091] SEQ ID NOs: 476 and 477 show sequences of adenine base editors.

[0092] SEQ ID NOs: 478-482 show sequences of cytosine base editors.

[0093] SEQ ID NOs: 483-487 show the sequences of plasmids suitable for encoding the engineered nucleic acid editing systems described herein.

[0094] SEQ ID NOs: 488 and 489 show the sgRNA scaffold sequences for MG15-1 and MG34-1.

[0095] SEQ ID NOs: 490-522 show the sequences of spacers used to target genomic loci in *E. coli* and HEK293T cells.

[0096] SEQ ID NOs: 523-585 show the sequences of primers used during amplification and Sanger sequencing.

[0097] SEQ ID NOs: 584-585 show the sequences of primers used during amplification.

[0098] SEQ ID NO: 586 shows the sequence of an adenine base editor.

[0099] SEQ ID NO: 587 shows the sequence of a cytosine base editor.

[0100] SEQ ID NOs: 588-589 show sequences of adenine base editors.

[0101] SEQ ID NOs: 590-593 show the full-length peptide sequences of linkers suitable for the engineered nucleic acid editing systems described herein.

[0102] SEQ ID NO: 594 shows the sequence of a cytosine deaminase.

[0103] SEQ ID NO: 595 shows the sequence of an adenosine deaminase.

[0104] SEQ ID NO: 596 shows the sequence of an MG34 active effector suitable for the engineered nucleic acid editing systems described herein.

[0105] SEQ ID NO: 597 shows the sequence of an MG34 nickase suitable for the engineered nucleic acid editing systems described herein.

[0106] SEQ ID NO: 598 shows the sequence of an MG34 PAM.

[0107] SEQ ID NOs: 599-638 show the full-length peptide sequences of MG138 cytidine deaminases suitable for the engineered nucleic acid editing systems described herein.

[0108] SEQ ID NOs: 639-659 show the full-length peptide sequences of MG139 cytidine deaminases suitable for the engineered nucleic acid editing systems described herein.

[0109] SEQ ID NOs: 660-662 show the full-length peptide sequences of MG141 cytidine deaminases suitable for the engineered nucleic acid editing systems described herein.

[0110] SEQ ID NOs: 663-664 show the full-length peptide sequences of MG142 cytidine deaminases suitable for the engineered nucleic acid editing systems described herein.

[0111] SEQ ID NOs: 665-675 show the full-length peptide sequences of MG93 cytidine deaminases suitable for the engineered nucleic acid editing systems described herein.

[0112] SEQ ID NOs: 676-678 show sequences of adenine base editors.

[0113] SEQ ID NOs: 679-680 show the sgRNA scaffold sequences for MG34-1 and SpCas9.

[0114] SEQ ID NOs: 681-689 show spacer sequences used to target genomic loci in guide RNAs.

[0115] SEQ ID NOs: 690-707 show sequences of primers used to amplify genomic targets of adenine bae editors (ABE) for next generation sequencing (NGS) analysis.

[0116] SEQ ID NO: 708 shows the sequence of a blasticidin (BSD) resistance cassette.

[0117] SEQ ID NOs: 709-719 show spacer sequences used to target genomic loci in guide RNAs.

[0118] SEQ ID NOs: 720-726 show the sequences of plasmids suitable for encoding the engineered nucleic acid editing systems described herein.

[0119] SEQ ID NOs: 728-729 show sequences of adenine base editors.

[0120] SEQ ID NOs: 730-736 show spacer sequences used to target genomic loci in guide RNAs.

[0121] SEQ ID NOs: 737-738 show the sequences of plasmids suitable for encoding the engineered nucleic acid editing systems described herein.

[0122] SEQ ID NOs: 739-740 show sequences of cytidine base editors.

[0123] SEQ ID NO: 741 shows the sequence of a plasmid suitable for encoding the A1CF gene.

[0124] SEQ ID NO: 742 shows the sequence of an RNA used to test CDAs for RNA activity.

[0125] SEQ ID NO: 743 shows the sequence of a labelled primer for poisoned primer extension assay used to test CDAs for RNA activity.

[0126] SEQ ID NOs: 744-827 show the full-length peptide sequences of MG139 cytidine deaminases suitable for the engineered nucleic acid editing systems described herein.

[0127] SEQ ID NO: 828 shows the full-length peptide sequence of an MG93 cytidine deaminase suitable for the engineered nucleic acid editing systems described herein.

[0128] SEQ ID NO: 829 shows the full-length peptide sequence of an MG142 cytidine deaminase suitable for the engineered nucleic acid editing systems described herein.

[0129] SEQ ID NOs: 830-835 show the full-length peptide sequences of MG152 cytidine deaminases suitable for the engineered nucleic acid editing systems described herein.

[0130] SEQ ID NOs: 836-860 show sequences of adenine base editors.

[0131] SEQ ID NOs: 861-864 show spacer sequences used to target genomic loci in guide RNAs.

[0132] SEQ ID NOs: 865-872 show sequences of primers used to amplify genomic targets of adenine bae editors (ABE) for next generation sequencing (NGS) analysis.

[0133] SEQ ID NOs: 873-875 show the sequences of plasmids suitable for encoding the engineered nucleic acid editing systems described herein.

[0134] SEQ ID NO: 876 shows the sgRNA scaffold sequence for MG34-1.

[0135] SEQ ID NOs: 877-916 show sequences of cytosine base editors.

[0136] SEQ ID NOs: 917-931 show the sequences of sgRNAs suitable for the engineered nucleic acid editing systems described herein.

[0137] SEQ ID NOs: 932-961 show sequences of primers used to amplify genomic targets of adenine base editors (ABE) for next generation sequencing (NGS) analysis.

[0138] SEQ ID NO: 962 shows a site engineered in mammalian cell line with 5 PAMs compatible with Cas9 and MG3-6 editing.

[0139] SEQ ID NOs: 963-967 show the sequences of sgRNAs suitable for the engineered nucleic acid editing systems described herein.

[0140] SEQ ID NOs: 968-969 show sequences of cytosine base editors.

[0141] SEQ ID NO: 970 shows the full-length peptide sequence of an MG139 cytidine deaminase suitable for the engineered nucleic acid editing systems described herein.

[0142] SEQ ID NOs: 971-977 show the full-length peptide sequences of MG93 cytidine deaminases suitable for the engineered nucleic acid editing systems described herein.

[0143] SEQ ID NOs: 978-981 show the full-length peptide sequences of MG138 cytidine deaminases suitable for the engineered nucleic acid editing systems described herein.

[0144] SEQ ID NO: 982 shows the full-length peptide sequence of MG142 cytidine deaminase suitable for the engineered nucleic acid editing systems described herein.

[0145] SEQ ID NO: 983-1014 shows the full-length peptide sequence of MG128 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0146] SEQ ID NO: 1015-1026 shows the full-length peptide sequence of MG129 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0147] SEQ ID NO: 1027-1031 shows the full-length peptide sequence of MG130 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0148] SEQ ID NO: 1032-1040 shows the full-length peptide sequence of MG131 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0149] SEQ ID NO: 1041-1043 shows the full-length peptide sequence of MG132 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0150] SEQ ID NO: 1044-1057 shows the full-length peptide sequence of MG133 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0151] SEQ ID NO: 1058-1061 shows the full-length peptide sequence of MG134 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0152] SEQ ID NO: 1062-1069 shows the full-length peptide sequence of MG135 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0153] SEQ ID NO: 1070-1081 shows the full-length peptide sequence of MG136 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0154] SEQ ID NO: 1082-1098 shows the full-length peptide sequence of MG137 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0155] SEQ ID NOs: 1099-1105 show the sequences of sgRNAs suitable for the engineered nucleic acid editing systems described herein.

[0156] SEQ ID NOs: 1106-1111 show the sequences of MG35 PAMs.

[0157] SEQ ID NO: 1112 shows the DNA sequence of a gene encoding the ABE-MG35-1 adenine base editor.

[0158] SEQ ID NO: 1113 shows the protein sequence of the ABE-MG35-1 adenine base editor.

[0159] SEQ ID NO: 1114 shows the nucleotide sequence of a plasmid encoding a Cas9-based cytosine base editor (CBE).

[0160] SEQ ID NO: 1115 shows the nucleotide sequence of a plasmid encoding Fam72a.

[0161] SEQ ID NOs: 1116-1117 show the sequences of Cas9-CBE target sites.

[0162] SEQ ID NOs: 1118-1119 show the sequences of NGS amplicons.

[0163] SEQ ID NO: 1120 shows the full-length peptide sequence of an MG35 nuclease.

[0164] SEQ ID NO: 1121 shows the full-length peptide sequence of Fam72A.

[0165] SEQ ID NOs: 1122-1127 shows the full-length peptide sequences of MG35 nucleases.

[0166] SEQ ID NOs: 1128-1160 shows the full-length peptide sequences of MG3-6/3-8 adenine base editors.

[0167] SEQ ID NOs: 1161-1186 shows the full-length peptide sequences of MG34-1 adenine base editors.

[0168] SEQ ID NOs: 1187-1195 show the sequences of sgRNAs suitable for the engineered nucleic acid editing systems described herein.

[0169] SEQ ID NOs: 1196-1204 show spacer sequences used to target genomic loci in guide RNAs.

[0170] SEQ ID NO: 1205 shows the nucleotide sequence of a plasmid encoding an MG3-6/3-8 adenine base editor.

[0171] SEQ ID NO: 1206 shows the nucleotide sequence of a plasmid encoding an sgRNA suitable for an MG3-6/3-8 adenine base editor described herein.

[0172] SEQ ID NO: 1207 shows the nucleotide sequence of a plasmid encoding an MG34-1 adenine base editor.

[0173] SEQ ID NOs: 1208-1269 show the full-length peptide sequences of MG93 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0174] SEQ ID NOs: 1270-1296 show the full-length peptide sequences of MG139 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0175] SEQ ID NOs: 1297-1311 show the full-length peptide sequences of MG152 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0176] SEQ ID NOs: 1312-1313 show the full-length peptide sequences of MG138 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0177] SEQ ID NOs: 1314-1315 show the full-length peptide sequences of MG139 deaminases suitable for the engineered nucleic acid editing systems described herein.

[0178] SEQ ID NOs: 1316-1319 show the nucleotide sequences of 5'-FAM-labeled ssDNAs.

[0179] SEQ ID NOs: 1320-1321 show the nucleotide sequences of Cy5.5-labeled ssDNAs.

[0180] SEQ ID NOs: 1322-1355 show sequences of cytidine base editors.

[0181] SEQ ID NOs: 1356-1362 show the full-length peptide sequences of MG34-1 adenine base editors.

[0182] SEQ ID NOs: 1363-1415 show the full-length peptide sequences of MG3-6/3-8 adenine base editors.

[0183] SEQ ID NOs: 1416-1417 show the nucleotide sequences of sgRNAs suitable for use with MG34-1 adenine base editors described herein.

[0184] SEQ ID NO: 1418 shows the nucleotide sequence of an sgRNA suitable for use with MG3-6/3-8 adenine base editors described herein.

[0185] SEQ ID NOs: 1419-1420 show the DNA sequences of target sites suitable for targeting by MG34-1 adenine base editors described herein.

[0186] SEQ ID NO: 1421 shows a DNA sequence of a target site suitable for targeting by MG3-6/3-8 adenine base editors described herein.

[0187] SEQ ID NO: 1422 shows the nucleotide sequence of a plasmid suitable for expression of an MG34-1 adenine base editor described herein.

[0188] SEQ ID NO: 1423 shows the nucleotide sequence of a plasmid suitable for expression of an MG3-6/3-8 adenine base editor described herein.

[0189] SEQ ID NO: 1424 shows the full-length peptide sequence of an MG35-1 adenine base editor.

[0190] SEQ ID NO: 1425-1426 show the nucleotide sequences of plasmids suitable for expression of MG35-1 adenine base editors and sgRNAs described herein.

[0191] SEQ ID NOs: 1427-1428 show the nucleotide sequences of sgRNAs suitable for use with MG35-1 adenine base editors described herein.

[0192] SEQ ID NOs: 1429-1430 show the DNA sequences of target sites suitable for targeting by MG35-1 adenine base editors described herein.

[0193] SEQ ID NOs: 1431-1454 show the nucleotide sequences of sgRNAs engineered to function with an MG3-6/3-8 adenine base editor in order to target APOA1.

[0194] SEQ ID NOs: 1455-1478 show the DNA sequences of APOA1 target sites.

[0195] SEQ ID NOs: 1479-1483 show the nucleotide sequences of sgRNAs engineered to function with an MG3-6/3-8 adenine base editor in order to target ANGPTL3.

[0196] SEQ ID NOs: 1484-1488 show the DNA sequences of ANGPTL3 target sites.

[0197] SEQ ID NOs: 1489-1490 show the nucleotide sequences of sgRNAs engineered to function with an MG3-6/3-8 adenine base editor in order to target TRAC.

[0198] SEQ ID NOs: 1491-1492 show the DNA sequences of TRAC sites.

[0199] SEQ ID NOs: 1493-1516 show the nucleotide sequences of NGS primers suitable for use in assessing base editing of APOA1.

[0200] SEQ ID NOs: 1517-1521 show the nucleotide sequences of NGS primers suitable for use in assessing base editing of ANGPTL3.

[0201] SEQ ID NOs: 1522-1523 show the nucleotide sequences of NGS primers suitable for use in assessing base editing of TRAC.

[0202] SEQ ID NOs: 1524-1547 show the nucleotide sequences of NGS primers suitable for use in assessing base editing of APOA1.

[0203] SEQ ID NOs: 1548-1552 show the nucleotide sequences of NGS primers suitable for use in assessing base editing of ANGPTL3.

[0204] SEQ ID NOs: 1553-1554 show the nucleotide sequences of NGS primers suitable for use in assessing base editing of TRAC.

[0205] SEQ ID NO: 1555 shows the nucleotide sequence of a plasmid suitable for use in mRNA production.

[0206] SEQ ID NOs: 1556-1562 show the full-length peptide sequences of MG131 adenine deaminase variants.

[0207] SEQ ID NOs: 1563-1566 show the full-length peptide sequences of MG134 adenine deaminase variants.

[0208] SEQ ID NOs: 1567-1574 show the full-length peptide sequences of MG135 adenine deaminase variants.

[0209] SEQ ID NOs: 1575-1589 show the full-length peptide sequences of MG137 adenine deaminase variants.

[0210] SEQ ID NOs: 1590-1599 show the full-length peptide sequences of MG68 adenine deaminase variants.

[0211] SEQ ID NOs: 1600-1602 show the full-length peptide sequences of MG132 adenine deaminase variants.

[0212] SEQ ID NOs: 1603-1616 show the full-length peptide sequences of MG133 adenine deaminase variants.

[0213] SEQ ID NOs: 1617-1624 show the full-length peptide sequences of MG136 adenine deaminase variants.

[0214] SEQ ID NOs: 1625-1633 show the full-length peptide sequences of MG129 adenine deaminase variants.

[0215] SEQ ID NOs: 1634-1638 show the full-length peptide sequences of MG130 adenine deaminase variants.

[0216] SEQ ID NOs: 1639-1644 show the full-length peptide sequences of MG34-1 adenine base editors.

[0217] SEQ ID NOs: 1698-1703 show the full-length peptide sequences of MG34 adenine base editors.

[0218] SEQ ID NOs: 1645-1646 show the nucleotide sequences of ssDNA substrates suitable for testing adenine deaminase activity *in vitro*.

[0219] SEQ ID NOs: 1647-1653 show linker sequences for deaminase systems described herein.

[0220] SEQ ID NOs: 1654-1658, and 1665-1694 show the full-length peptide sequences of MG3-6/3-8 chimera base editors.

[0221] SEQ ID NOs: 1659 and 1661-1664 show the full-length peptide sequences of MG139 cytidine deaminases suitable for the engineered nucleic acid editing systems described herein.

[0222] SEQ ID NO: 1660 shows the full-length peptide sequence of an MG152 cytidine deaminase suitable for the engineered nucleic acid editing systems described herein.

[0223] SEQ ID NOs: 1695-1697 show the full-length peptide sequences of MG102 adenine base editors.

[0224] SEQ ID NO: 1704 shows the full-length nucleotide sequence of an hApoA1_1 guide RNA.

[0225] SEQ ID NOs: 1705-1710 show the full-length nucleotide sequences of MG34 effector chemically synthesized/modified sgRNAs.

[0226] SEQ ID NOs: 1711-1719 show the full-length nucleotide sequences of TRAC target sites.

[0227] SEQ ID NOs: 1720-1725 show the full-length nucleotide sequences of AAVS1 target sites.

[0228] SEQ ID NO: 1726 shows the full-length nucleotide sequence of the hApoA1 target site.

[0229] SEQ ID NOs: 1727-1728, and 1743 show the nucleotide sequence of MG3-6/3-8 adenine base editor.

[0230] SEQ ID NOs: 1729 and 1744 show the nucleotide sequence of MG3-6/3-8/3-4 adenine base editor.

[0231] SEQ ID NOs: 1730 and 1745 shows the nucleotide sequence of MG3-6/3-8/3-6 adenine base editor.

[0232] SEQ ID NOs: 1731 and 1746 show the nucleotide sequence of MG3-6/3-8/3-7 adenine base editor.

[0233] SEQ ID NOs: 1732 and 1747 show the nucleotide sequence of MG3-6/3-8/3-22 adenine base editor.

[0234] SEQ ID NOs: 1733 and 1748 show the nucleotide sequence of MG3-6/3-8/3-24 adenine base editor.

[0235] SEQ ID NOs: 1734 and 1749 show the nucleotide sequence of MG3-6/3-8/3-38 adenine base editor.

[0236] SEQ ID NOs: 1735 and 1750 show the nucleotide sequence of MG3-6/3-8/3-89 adenine base editor.

[0237] SEQ ID NOs: 1736 and 1751 show the nucleotide sequence of MG3-6/3-8/3-90 adenine base editor.

[0238] SEQ ID NOs: 1737 and 1752 show the nucleotide sequence of MG3-6/3-8/3-92 adenine base editor.

[0239] SEQ ID NOs: 1738 and 1753 show the nucleotide sequence of MG3-6/3-8/3-93 adenine base editor.

[0240] SEQ ID NOs: 1739 and 1754 show the nucleotide sequence of MG3-6/3-8/3-95 adenine base editor.

[0241] SEQ ID NOs: 1740 and 1755 show the nucleotide sequence of MG3-6/3-8/3-104 adenine base editor.

[0242] SEQ ID NOs: 1741 and 1756 show the nucleotide sequence of MG3-6/3-8/150-2 adenine base editor.

[0243] SEQ ID NOs: 1742 and 1757 show the nucleotide sequence of MG3-6/3-8/150-9 adenine base editor.

[0244] SEQ ID NOs: 1758-1889 show the nucleotide sequence of MG3-6 ABE hANGPTL3 guide.

[0245] SEQ ID NOs: 1890-1976 show the nucleotide sequence of MG3-6 ABE hBCL11A guide.

[0246] SEQ ID NOs: 1977-2009 show the nucleotide sequence of MG3-6 ABE hPAH guide.

[0247] SEQ ID NOs: 2010-2019 show the nucleotide sequence of MG3-6 ABE APOA1 guide.

[0248] SEQ ID NO: 2020 shows the nucleotide sequence of engineered therapeutic sequence.

[0249] SEQ ID NO: 2021 shows the protein sequence of MG3-6/3-8/3-4 adenine base editor.

[0250] SEQ ID NO: 2022 shows the protein sequence of MG3-6/3-8/3-7 adenine base editor.

[0251] SEQ ID NO: 2023 shows the protein sequence of MG3-6/3-8/3-104 adenine base editor.

[0252] SEQ ID NO: 2024-2043 show the amino acid sequences of nuclear localization signals (NLS).

DETAILED DESCRIPTION

[0253] While various embodiments of the disclosure have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions may occur to those skilled in the art without departing from the disclosure. It should be understood that various alternatives to the embodiments of the disclosure described herein may be employed.

[0254] The practice of some methods disclosed herein employ, unless otherwise indicated, techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA. See for example Sambrook and Green, Molecular Cloning: A Laboratory Manual, 4th Edition (2012); the series Current Protocols in Molecular Biology (F. M. Ausubel, et al. eds.); the series Methods In Enzymology (Academic Press, Inc.), PCR 2: A Practical Approach (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) Antibodies, A Laboratory Manual, and Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, 6th Edition (R.I. Freshney, ed. (2010)).

[0255] As used herein, the singular forms "a", "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. Furthermore, to the extent that the terms "including", "includes", "having", "has", "with", or variants thereof are used in either the detailed description or the claims, such terms are intended to be inclusive in a manner similar to the term "comprising."

[0256] The term "about" or "approximately" means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, i.e., the limitations of the measurement system. For example, "about" can mean within one or more than one standard deviation, per the practice in the art. Alternatively, "about" can mean a range of up to 20%, up to 15%, up to 10%, up to 5%, or up to 1% of a given value.

[0257] The term "nucleotide," as used herein, refers to a base-sugar-phosphate combination. Contemplated nucleotides include naturally occurring nucleotides and synthetic nucleotides.

Nucleotides are monomeric units of a nucleic acid sequence (*e.g.*, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA)). The term nucleotide includes ribonucleoside triphosphates adenosine triphosphate (ATP), uridine triphosphate (UTP), cytosine triphosphate (CTP), guanosine triphosphate (GTP) and deoxyribonucleoside triphosphates such as dATP, dCTP, dITP, dUTP, dGTP, dTTP, or derivatives thereof. Such derivatives include, for example, [αS]dATP, 7-deaza-dGTP and 7-deaza-dATP, and nucleotide derivatives that confer nuclease resistance on the nucleic acid molecule containing them. The term nucleotide as used herein encompasses dideoxyribonucleoside triphosphates (ddNTPs) and their derivatives. Illustrative examples of ddNTPs include, but are not limited to, ddATP, ddCTP, ddGTP, ddITP, and ddTTP. A nucleotide may be unlabeled or detectably labeled, such as using moieties comprising optically detectable moieties (*e.g.*, fluorophores) or quantum dots. Detectable labels include, for example, radioactive isotopes, fluorescent labels, chemiluminescent labels, bioluminescent labels, and enzyme labels. Fluorescent labels of nucleotides include but are not limited fluorescein, 5-carboxyfluorescein (FAM), 2′7′-dimethoxy-4′5-dichloro-6-carboxyfluorescein (JOE), rhodamine, 6-carboxyrhodamine (R6G), N,N,N′,N′-tetramethyl-6-carboxyrhodamine (TAMRA), 6-carboxy-X-rhodamine (ROX), 4-(4′dimethylaminophenylazo) benzoic acid (DABCYL), Cascade Blue, Oregon Green, Texas Red, Cyanine and 5-(2′-aminoethyl)aminonaphthalene-1-sulfonic acid (EDANS). Specific examples of fluorescently labeled nucleotides include [R6G]dUTP, [TAMRA]dUTP, [R110]dCTP, [R6G]dCTP, [TAMRA]dCTP, [JOE]ddATP, [R6G]ddATP, [FAM]ddCTP, [R110]ddCTP, [TAMRA]ddGTP, [ROX]ddTTP, [dR6G]ddATP, [dR110]ddCTP, [dTAMRA]ddGTP, and [dROX]ddTTP available from Perkin Elmer, Foster City, Calif; FluoroLink DeoxyNucleotides, FluoroLink Cy3-dCTP, FluoroLink Cy5-dCTP, FluoroLink Fluor X-dCTP, FluoroLink Cy3-dUTP, and FluoroLink Cy5-dUTP available from Amersham, Arlington Heights, IL; Fluorescein-15-dATP, Fluorescein-12-dUTP, Tetramethyl-rodamine-6-dUTP, IR770-9-dATP, Fluorescein-12-ddUTP, Fluorescein-12-UTP, and Fluorescein-15-2′-dATP available from Boehringer Mannheim, Indianapolis, Ind.; and Chromosome Labeled Nucleotides, BODIPY-FL-14-UTP, BODIPY-FL-4-UTP, BODIPY-TMR-14-UTP, BODIPY-TMR-14-dUTP, BODIPY-TR-14-UTP, BODIPY-TR-14-dUTP, Cascade Blue-7-UTP, Cascade Blue-7-dUTP, fluorescein-12-UTP, fluorescein-12-dUTP, Oregon Green 488-5-dUTP, Rhodamine Green-5-UTP, Rhodamine Green-5-dUTP, tetramethylrhodamine-6-UTP, tetramethylrhodamine-6-dUTP, Texas Red-5-UTP, Texas Red-5-dUTP, and Texas Red-12-dUTP available from Molecular Probes, Eugene, Oreg. The term nucleotide encompasses chemically modified nucleotides. An exemplary chemically-modified nucleotide is biotin-dNTP. Non-limiting examples of biotinylated dNTPs include, biotin-dATP (*e.g.*, bio-N6-ddATP, biotin-14-

dATP), biotin-dCTP (*e.g.*, biotin-11-dCTP, biotin-14-dCTP), and biotin-dUTP (*e.g.*, biotin-11-dUTP, biotin-16-dUTP, biotin-20-dUTP).

[0258] The terms "polynucleotide," "oligonucleotide," and "nucleic acid" are used interchangeably to refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof, either in single-, double-, or multi-stranded form. Contemplated polynucleotides include a gene or fragment thereof. Exemplary polynucleotides include, but are not limited to, DNA, RNA, coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, cell-free polynucleotides including cell-free DNA (cfDNA) and cell-free RNA (cfRNA), nucleic acid probes, and primers. In a polynucleotide when referring to a T, a T means U (Uracil) in RNA and T (Thymine) in DNA. A polynucleotide can be exogenous or endogenous to a cell and/or exist in a cell-free environment. The term polynucleotide encompasses modified polynucleotides (*e.g.*, altered backbone, sugar, or nucleobase). If present, modifications to the nucleotide structure are imparted before or after assembly of the polymer. Non-limiting examples of modifications include: 5-bromouracil, peptide nucleic acid, xeno nucleic acid, morpholinos, locked nucleic acids, glycol nucleic acids, threose nucleic acids, dideoxynucleotides, cordycepin, 7-deaza-GTP, fluorophores (*e.g.*, rhodamine or fluorescein linked to the sugar), thiol-containing nucleotides, biotin-linked nucleotides, fluorescent base analogs, CpG islands, methyl-7-guanosine, methylated nucleotides, inosine, thiouridine, pseudouridine, dihydrouridine, queuosine, and wyosine. The sequence of nucleotides may be interrupted by non-nucleotide components.

[0259] The terms "peptide," "polypeptide," and "protein" are used interchangeably herein to refer to a polymer of at least two amino acid residues joined by peptide bond(s). This term does not connote a specific length of polymer, nor is it intended to imply or distinguish whether the peptide is produced using recombinant techniques, chemical or enzymatic synthesis, or is naturally occurring. The terms apply to naturally occurring amino acid polymers as well as amino acid polymers comprising at least one modified amino acid. In some cases, the polymer is interrupted by non-amino acids. The terms include amino acid chains of any length, including full length proteins, and proteins with or without secondary or tertiary structure (*e.g.*, domains). The terms also encompass an amino acid polymer that has been modified, for example, by disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, oxidation, and any other manipulation such as conjugation with a labeling component. The terms "amino acid" and "amino acids," as used herein, refer to natural and non-natural amino acids, including, but

not limited to, modified amino acids. Modified amino acids include amino acids that have been chemically modified to include a group or a chemical moiety not naturally present on the amino acid. The term "amino acid" includes both D-amino acids and L-amino acids.

[0260] As used herein, the "non-native" refers to a nucleic acid or polypeptide sequence that is non-naturally occurring. Non-native refers to a non-naturally occurring nucleic acid or polypeptide sequence that comprises modifications such as mutations, insertions, or deletions. The term non-native encompasses fusion nucleic acids or polypeptides that encodes or exhibits an activity (*e.g.,* enzymatic activity, methyltransferase activity, acetyltransferase activity, kinase activity, ubiquitinating activity, etc.) of the nucleic acid or polypeptide sequence to which the non-native sequence is fused. A non-native nucleic acid or polypeptide sequence includes those linked to a naturally-occurring nucleic acid or polypeptide sequence (or a variant thereof) by genetic engineering to generate a chimeric nucleic acid or polypeptide sequence encoding a chimeric nucleic acid or polypeptide.

[0261] As used herein, "operably linked", "operable linkage", "operatively linked", or grammatical equivalents thereof refer to an arrangement of genetic elements, *e.g.,* a promoter, an enhancer, a polyadenylation sequence, etc., wherein an operation (*e.g.,* movement or activation) of a first genetic element has some effect on the second genetic element. The effect on the second genetic element can be, but need not be, of the same type as operation of the first genetic element. For example, two genetic elements are operably linked if movement of the first element causes an activation of the second element. For instance, a regulatory element, which may comprise promoter and/or enhancer sequences, is operatively linked to a coding region if the regulatory element helps initiate transcription of the coding sequence. There may be intervening residues between the regulatory element and coding region so long as this functional relationship is maintained.

[0262] A "functional fragment" of a DNA or protein sequence refers to a fragment that retains a biological activity (either functional or structural) that is substantially similar to a biological activity of the full-length DNA or protein sequence. A biological activity of a DNA sequence includes its ability to influence expression in a manner attributed to the full-length sequence.

[0263] The terms "engineered," "synthetic," and "artificial" are used interchangeably herein to refer to an object that has been modified by human intervention. For example, the terms refer to a polynucleotide or polypeptide that is non-naturally occurring. An engineered peptide has, but does not require, low sequence identity (*e.g.,* less than 50% sequence identity, less than 25% sequence identity, less than 10% sequence identity, less than 5% sequence identity, less than 1% sequence identity) to a naturally occurring human protein. For example, VPR and VP64 domains are synthetic transactivation domains. Non-limiting examples include the following: a nucleic

acid modified by changing its sequence to a sequence that does not occur in nature; a nucleic acid modified by ligating it to a nucleic acid that it does not associate with in nature such that the ligated product possesses a function not present in the original nucleic acid; an engineered nucleic acid synthesized *in vitro* with a sequence that does not exist in nature; a protein modified by changing its amino acid sequence to a sequence that does not exist in nature; an engineered protein acquiring a new function or property. An "engineered" system comprises at least one engineered component.

[0264] The term "tracrRNA" or "tracr sequence" means trans-activating CRISPR RNA. tracrRNA interacts with the CRISPR (cr) RNA to form guide (g) RNA in type II and subtype V-B CRISPR-Cas systems. If the tracrRNA is engineered, it may have about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% sequence identity and/or sequence similarity to a wild type exemplary tracrRNA sequence (*e.g.*, a tracrRNA from *S. pyogenes*, *S. aureus*). tracrRNA may refer to a modified form of a tracrRNA that can comprise a nucleotide change such as a deletion, insertion, or substitution, variant, mutation, or chimera. The term tracrRNA encompasses a nucleic acid that can be at least about 60% identical to a wild type exemplary tracrRNA (*e.g.*, a tracrRNA from *S. pyogenes*, *S. aureus*, etc) sequence over a stretch of at least 6 contiguous nucleotides. For example, a tracrRNA sequence has at least about 60% identical, at least about 65% identical, at least about 70% identical, at least about 75% identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, or 100 % identical to a wild type exemplary tracrRNA (*e.g.*, a tracrRNA from *S. pyogenes*, *S. aureus*, etc) sequence over a stretch of at least 6 contiguous nucleotides. Type II tracrRNA sequences can be predicted on a genome sequence by identifying regions with complementarity to part of the repeat sequence in an adjacent CRISPR array.

[0265] As used herein, a "guide nucleic acid" or "guide polynucleotide" refers to a nucleic acid that may hybridize to a target nucleic acid and thereby directs an associated nuclease to the target nucleic acid. A guide nucleic acid is, but is not limited to, RNA (guide RNA or gRNA), DNA, or a mixture of RNA and DNA. A guide nucleic acid can include a crRNA or a tracrRNA or a combination of both. The term guide nucleic acid encompasses an engineered guide nucleic acid and a programmable guide nucleic acid to specifically bind to the target nucleic acid. A portion of the target nucleic acid may be complementary to a portion of the guide nucleic acid. The strand of a double-stranded target polynucleotide that is complementary to and hybridizes with the guide nucleic acid is the complementary strand. The strand of the double-stranded target polynucleotide that is complementary to the complementary strand, and therefore is not complementary to the guide nucleic acid is called noncomplementary strand. A guide nucleic

acid having a polynucleotide chain is a "single guide nucleic acid." A guide nucleic acid having two polynucleotide chains is a "double guide nucleic acid." If not otherwise specified, the term "guide nucleic acid" is inclusive, referring to both single guide nucleic acids and double guide nucleic acids. A guide nucleic acid may comprise a segment referred to as a "nucleic acid-targeting segment" or a "nucleic acid-targeting sequence," or a "spacer." A nucleic acid-targeting segment can include a sub-segment referred to as a "protein binding segment" or "protein binding sequence" or "Cas protein binding segment."

[0266] The term "sequence identity" or "percent identity" in the context of two or more nucleic acids or polypeptide sequences, generally refers to two (e.g., in a pairwise alignment) or more (e.g., in a multiple sequence alignment) sequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a local or global comparison window, as measured using a sequence comparison algorithm. Suitable sequence comparison algorithms for polypeptide sequences include, e.g., BLASTP using parameters of a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix setting gap costs at existence of 11, extension of 1, and using a conditional compositional score matrix adjustment for polypeptide sequences longer than 30 residues; BLASTP using parameters of a wordlength (W) of 2, an expectation (E) of 1000000, and the PAM30 scoring matrix setting gap costs at 9 to open gaps and 1 to extend gaps for sequences of less than 30 residues (these are the default parameters for BLASTP in the BLAST suite available at https://blast.ncbi.nlm.nih.gov); CLUSTALW with parameters of ; the Smith-Waterman homology search algorithm with parameters of a match of 2, a mismatch of -1, and a gap of -1; MUSCLE with default parameters; MAFFT with parameters retree of 2 and maxiterations of 1000; Novafold with default parameters; HMMER hmmalign with default parameters.

[0267] As used herein, the term "RuvC_III domain" refers to a third discontinuous segment of a RuvC endonuclease domain (the RuvC nuclease domain being comprised of three discontiguous segments, RuvC_I, RuvC_II, and RuvC_III). A RuvC domain or segments thereof can generally be identified by alignment to documented domain sequences, structural alignment to proteins with annotated domains, or by comparison to Hidden Markov Models (HMMs) built based on documented domain sequences (*e.g.*, Pfam HMM PF18541 for RuvC_III).

[0268] As used herein, the term "HNH domain" refers to an endonuclease domain having characteristic histidine and asparagine residues. An HNH domain can generally be identified by alignment to documented domain sequences, structural alignment to proteins with annotated domains, or by comparison to Hidden Markov Models (HMMs) built based on documented domain sequences (*e.g.*, Pfam HMM PF01844 for domain HNH).

[0269] As used herein, the term "base editor" refers to an enzyme that catalyzes the conversion of one target base or base pair into another (e.g., A:T to G:C, C:G to T:A) without requiring the creation and repair of a double-strand break. An exemplary base editor is a deaminase. In some embodiments, the base editor comprises a deaminase and a nuclease that is deficient in nuclease activity. In some embodiments, the base editor comprises a deaminase and a catalytically inactive nuclease. In some embodiments, the base editor comprises a fusion of a deaminase and a catalytically inactive nuclease.

[0270] As used herein, the term "deaminase" refers to a protein or enzyme that catalyzes a deamination reaction (i.e., a reaction that removes an amino group). Deaminases include adenosine deaminases, which catalyze the hydrolytic deamination of adenine or adenosine (e.g., an engineered adenosine deaminase that deaminates adenosine in DNA), and cytidine (or cytosine) deaminases, which catalyze the hydrolytic deamination of cytidine (or cytosine) or deoxycytidine to uridine (or uracil) or deoxyuridine, respectively. The deaminase or deaminase domain can be a naturally-occurring deaminase or deaminase domain from an organism, such as a human, chimpanzee, gorilla, monkey, cow, dog, rat, mouse, or bacterium (e.g., *E. coli*), a variant of a naturally-occurring deaminase or deaminase domain, or a non-naturally occurring deaminase or deaminase domain.

[0271] The term "optimally aligned" in the context of two or more nucleic acids or polypeptide sequences, generally refers to two (e.g., in a pairwise alignment) or more (e.g., in a multiple sequence alignment) sequences that have been aligned to maximal correspondence of amino acids residues or nucleotides, for example, as determined by the alignment producing a highest or "optimized" percent identity score.

[0272] As used herein, the term "complex" refers to a joining of at least two components. The two components may each retain the properties/activities they had prior to forming the complex or gain properties as a result of forming the complex. The joining includes, but is not limited to, covalent bonding, non-covalent bonding (i.e., hydrogen bonding, ionic interactions, Van der Waals interactions, and hydrophobic bond), use of a linker, fusion, or any other suitable method. Contemplated components of the complex include polynucleotides, polypeptides, or combinations thereof. For example, a complex comprises an endonuclease and a guide polynucleotide.

[0273] Included in the current disclosure are variants of any of the enzymes described herein with one or more conservative amino acid substitutions. Such conservative substitutions can be made in the amino acid sequence of a polypeptide without disrupting the three-dimensional structure or function of the polypeptide. Conservative substitutions can be accomplished by substituting amino acids with similar hydrophobicity, polarity, and R chain length for one

another. Additionally, or alternatively, by comparing aligned sequences of homologous proteins from different species, conservative substitutions can be identified by locating amino acid residues that have been mutated between species (e.g., non-conserved residues) without altering the basic functions of the encoded proteins. Such conservatively substituted variants may include variants with at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to any one of the endonuclease protein sequences described herein. In some embodiments, such conservatively substituted variants are functional variants. Such functional variants can encompass sequences with substitutions such that the activity of one or more critical active site residues or guide RNA binding residues of the endonuclease are not disrupted.

[0274] Also included in the current disclosure are variants of any of the enzymes described herein with substitution of one or more catalytic residues to decrease or eliminate activity of the enzyme (e.g., decreased-activity variants). In some embodiments, a decreased activity variant as a protein described herein comprises a disrupting substitution of at least one, at least two, or all three catalytic residues. In some embodiments, any of the endonucleases described herein can comprise a nickase mutation. In some embodiments, any of the endonucleases described herein can comprise a RuvC domain lacking nuclease activity. In some embodiments, any of the endonucleases described herein can be configured to cleave one strand of a double-stranded target deoxyribonucleic acid. In some embodiments, any of the endonucleases described herein can comprise can be configured to lack endonuclease activity or be catalytically dead.

[0275] Conservative substitution tables providing functionally similar amino acids are available from a variety of references (see, for e.g., Creighton, Proteins: Structures and Molecular Properties (W H Freeman & Co.; 2nd edition (December 1993)). The following eight groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Glycine (G);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
7) Serine (S), Threonine (T); and
8) Cysteine (C), Methionine (M).

Overview

[0276] The discovery of new CRISPR enzymes with unique functionality and structure may offer the potential to further disrupt deoxyribonucleic acid (DNA) editing technologies, improving speed, specificity, functionality, and ease of use. Relative to the predicted prevalence of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems in microbes and the sheer diversity of microbial species, comparatively few functionally characterized CRISPR enzymes exist in the literature. This is partly because a huge number of microbial species may not be readily cultivated in laboratory conditions. Metagenomic sequencing from natural environmental niches that represent large numbers of microbial species may offer the potential to drastically increase the number of new CRISPR systems documented and speed the discovery of new oligonucleotide editing functionalities. A recent example of the fruitfulness of such an approach is demonstrated by the 2016 discovery of CasX/CasY CRISPR systems from metagenomic analysis of natural microbial communities.

[0277] CRISPR systems are RNA-directed nuclease complexes that have been described to function as an adaptive immune system in microbes. In their natural context, CRISPR systems occur in CRISPR (clustered regularly interspaced short palindromic repeats) operons or loci, which generally comprise two parts: (i) an array of short repetitive sequences (30-40bp) separated by equally short spacer sequences, which encode the RNA-based targeting element; and (ii) ORFs encoding the nuclease polypeptide directed by the RNA-based targeting element alongside accessory proteins/enzymes. Efficient nuclease targeting of a particular target nucleic acid sequence generally requires both (i) complementary hybridization between the first 6-8 nucleic acids of the target (the target seed) and the crRNA guide; and (ii) the presence of a protospacer-adjacent motif (PAM) sequence within a defined vicinity of the target seed (the PAM usually being a sequence not commonly represented within the host genome). Depending on the exact function and organization of the system, CRISPR systems are commonly organized into 2 classes, 5 types and 16 subtypes based on shared functional characteristics and evolutionary similarity (see FIG. 1).

[0278] Class 1 CRISPR systems have large, multisubunit effector complexes, and comprise Types I, III, and IV.

[0279] Type I CRISPR systems are considered of moderate complexity in terms of components. In Type I CRISPR systems, the array of RNA-targeting elements is transcribed as a long precursor crRNA (pre-crRNA) that is processed at repeat elements to liberate short, mature crRNAs that direct the nuclease complex to nucleic acid targets when they are followed by a suitable short consensus sequence called a protospacer-adjacent motif (PAM). This processing

occurs via an endoribonuclease subunit (Cas6) of a large endonuclease complex called Cascade, which also comprises a nuclease (Cas3) protein component of the crRNA-directed nuclease complex. Type I nucleases function primarily as DNA nucleases.

[0280] Type III CRISPR systems may be characterized by the presence of a central nuclease, known as Cas10, alongside a repeat-associated mysterious protein (RAMP) that comprises Csm or Cmr protein subunits. Like in Type I systems, the mature crRNA is processed from a pre-crRNA using a Cas6-like enzyme. Unlike type I and II systems, type III systems appear to target and cleave DNA-RNA duplexes (such as DNA strands being used as templates for an RNA polymerase).

[0281] Type IV CRISPR systems possess an effector complex that comprises a highly reduced large subunit nuclease (csf1), two genes for RAMP proteins of the Cas5 (csf3) and Cas7 (csf2) groups, and, in some cases, a gene for a predicted small subunit; such systems are commonly found on endogenous plasmids.

[0282] Class 2 CRISPR systems generally have single-polypeptide multidomain nuclease effectors, and comprise Types II, V and VI.

[0283] Type II CRISPR systems are considered the simplest in terms of components. In Type II CRISPR systems, the processing of the CRISPR array into mature crRNAs does not require the presence of a special endonuclease subunit, but rather a small trans-encoded crRNA (tracrRNA) with a region complementary to the array repeat sequence; the tracrRNA interacts with both its corresponding effector nuclease (e.g., Cas9) and the repeat sequence to form a precursor dsRNA structure, which is cleaved by endogenous RNAse III to generate a mature effector enzyme loaded with both tracrRNA and crRNA. Type II nucleases are known as DNA nucleases. Type II effectors generally exhibit a structure comprising a RuvC-like endonuclease domain that adopts the RNase H fold with an unrelated HNH nuclease domain inserted within the folds of the RuvC-like nuclease domain. The RuvC-like domain is responsible for the cleavage of the target (e.g., crRNA complementary) DNA strand, while the HNH domain is responsible for cleavage of the displaced DNA strand.

[0284] Type V CRISPR systems are characterized by a nuclease effector (e.g., Cas12) structure similar to that of Type II effectors, comprising a RuvC-like domain. Similar to Type II, most (but not all) Type V CRISPR systems use a tracrRNA to process pre-crRNAs into mature crRNAs; however, unlike Type II systems which requires RNAse III to cleave the pre-crRNA into multiple crRNAs, type V systems are capable of using the effector nuclease itself to cleave pre-crRNAs. Like Type-II CRISPR systems, Type V CRISPR systems are again known as DNA nucleases. Unlike Type II CRISPR systems, some Type V enzymes (e.g., Cas12a) appear to have

a robust single-stranded nonspecific deoxyribonuclease activity that is activated by the first crRNA directed cleavage of a double-stranded target sequence.

[0285] Type VI CRISPR systems have RNA-guided RNA endonucleases. Instead of RuvC-like domains, the single polypeptide effector of Type VI systems (e.g., Cas13) comprises two HEPN ribonuclease domains. Differing from both Type II and V systems, Type VI systems also may not require a tracrRNA in some instances for processing of pre-crRNA into crRNA. Similar to type V systems, however, some Type VI systems (e.g., C2C2) appear to possess robust single-stranded nonspecific nuclease (ribonuclease) activity activated by the first crRNA directed cleavage of a target RNA.

[0286] Because of their simpler architecture, Class 2 CRISPR have been most widely adopted for engineering and development as designer nuclease/genome editing applications.

[0287] One of the early adaptations of such a system for *in vitro* use involved (i) recombinantly-expressed, purified full-length Cas9 (*e.g.*, a Class 2, Type II Cas enzyme) isolated from *S. pyogenes* SF370, (ii) purified mature ~42 nt crRNA bearing a ~20 nt 5' sequence complementary to the target DNA sequence desired to be cleaved followed by a 3' tracr-binding sequence (the whole crRNA being *in vitro* transcribed from a synthetic DNA template carrying a T7 promoter sequence); (iii) purified tracrRNA *in vitro* transcribed from a synthetic DNA template carrying a T7 promoter sequence, and (iv) Mg^{2+}. A later improved, engineered system involved the crRNA of (ii) joined to the 5' end of (iii) by a linker (*e.g.*, GAAA) to form a single fused synthetic guide RNA (sgRNA) capable of directing Cas9 to a target by itself.

[0288] Such engineered systems can be adapted for use in mammalian cells by providing DNA vectors encoding (i) an ORF encoding codon-optimized Cas9 (*e.g.*, a Class 2, Type II Cas enzyme) under a suitable mammalian promoter with a C-terminal nuclear localization sequence (*e.g.*, SV40 NLS) and a suitable polyadenylation signal (*e.g.*, TK pA signal); and (ii) an ORF encoding an sgRNA (having a 5' sequence beginning with G followed by 20 nt of a complementary targeting nucleic acid sequence joined to a 3' tracr-binding sequence, a linker, and the tracrRNA sequence) under a suitable Polymerase III promoter (*e.g.*, the U6 promoter).

Base editing

[0289] Base editing is the conversion of one target base or base pair into another (e.g., A:T to G:C, C:G to T:A) without requiring the creation and repair of a double-strand break. The base editing may be achieved with the help of DNA and RNA base editors that allow the introduction of point mutations at specific sites, in either DNA or RNA. Generally, DNA base editors may comprise a fusion of a catalytically inactive nuclease and a catalytically active base-modification enzyme that acts on single-stranded DNAs (ssDNAs). RNA base editors may comprise of

similar, RNA-specific enzymes. Base editing may increase the efficiency of gene modification, while reducing the off-target and random mutations in the DNA.

[0290] DNA base editors are engineered ribonucleoprotein complexes that act as tools for single base substitution in cells and organism. They may be created by fusing an engineered base-modification enzyme and a catalytically deficient CRISPR endonuclease variant that cannot cut dsDNA, but it is able to unfold the dsDNA in a protospacer adjacent motif (PAM) sequence-dependent manner, such that a guide RNA can find its complementary target to indicate a ssDNA scission site. The guide RNA anneals to the complementary DNA, displacing a fragment of ssDNA and directing the CRISPR 'scissors' to the base modification site. The cellular repair machinery will repair the nicked non-edited strand using information from the complementary edited template.

[0291] So far, two types of DNA editors, cytosine base (CBEs) and adenine base editors (ABEs) have been developed. However, recent findings indicate that off-target modifications are present in DNA, and that many off-target modifications are also introduced into RNA by DNA base editors.

MG Base Editors

[0292] Described herein, in certain embodiments, are engineered systems comprising: (a) a base editor; (b) an endonuclease configured to bind the base editor and is deficient in nuclease activity; and (c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence. Described herein, in certain embodiments, are engineered base editing systems comprising an engineered base editing system comprising: a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415; and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence. Described herein, in certain embodiments, are engineered base editing systems comprising an engineered base editing system comprising: a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415; and an engineered guide polynucleotide which forms a complex with an endonuclease deficient in nuclease activity of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

[0293] Described herein, in certain embodiments, are engineered base editing systems comprising an engineered base editing system comprising: a base editor comprising a sequence having at least 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023; and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence. Described herein, in certain embodiments, are engineered base editing systems comprising an engineered base editing system comprising: a base editor comprising a sequence having at least 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023; and an engineered guide polynucleotide which forms a complex with an endonuclease deficient in nuclease activity of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

[0294] In some embodiments, the base editor comprises a sequence with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703. In some embodiments, the base editor comprises a sequence having at least about 70% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703. In some embodiments, the base editor comprises a sequence having at least about 75% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703. In some embodiments, the base editor comprises a sequence having at least about 80% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703. In some embodiments, the base editor comprises a sequence having at least about 85% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703. In some embodiments, the base editor comprises a sequence having at least about 90% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703. In some embodiments, the base editor comprises a sequence having at least about 95% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703. In some embodiments, the base editor comprises a sequence having at least about 96% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443,

444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703. In some embodiments, the base editor comprises a sequence having at least about 97% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703. In some embodiments, the base editor comprises a sequence having at least about 98% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703. In some embodiments, the base editor comprises a sequence having at least about 99% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703. In some embodiments, the base editor comprises a sequence having 100% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703.

[0295] In some embodiments, the base editor comprises a sequence with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 70% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 75% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 80% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 85% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 86% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 87% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 88% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 89% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 90% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 95% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 96% identity to any one of SEQ ID NOs: 1654-1703 and 2021-

2023. In some embodiments, the base editor comprises a sequence having at least about 97% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 98% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having at least about 99% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023. In some embodiments, the base editor comprises a sequence having 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023.

[0296] Described herein, in certain embodiments, are engineered base editing system comprising: a base editor encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757, and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence. Described herein, in certain embodiments, are engineered base editing system comprising: a base editor encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757, and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor deficient in nuclease activity and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

[0297] In some embodiments, the base editor is encoded by a nucleic acid having a sequence with at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 1727-1757. In some embodiments, the base editor is encoded by a nucleic acid sequence having at least about 70% identity to any one of SEQ ID NOs: 1727-1757. In some embodiments, the base editor comprises a sequence having at least about 75% identity to any one of SEQ ID NOs: 1727-1757. In some embodiments, the base is encoded by a nucleic acid having a sequence with at least about 80% identity to any one of SEQ ID NOs: 1727-1757. In some embodiments, the base editor is encoded by a nucleic acid having a sequence with at least about 85% identity to any one of SEQ ID NOs: 1727-1757. In some embodiments, the base editor is encoded by a nucleic acid having a sequence with at least about 90% identity to any one of SEQ ID NOs: 1727-1757. In some embodiments, the base editor is encoded by a nucleic acid having a sequence with at least about 95% identity to any one of SEQ ID NOs: 1727-1757. In some embodiments, the base editor is encoded by a nucleic acid having a sequence with at least about 96% identity to any one of SEQ ID NOs: 1727-1757. In some embodiments, the base editor is encoded by a nucleic acid having a sequence with at least about 97% identity to any one of SEQ ID NOs: 1727-1757. In some embodiments,

the base editor is encoded by a nucleic acid having a sequence with at least about 98% identity to any one of SEQ ID NOs: 1727-1757. In some embodiments, the base editor is encoded by a nucleic acid having a sequence with at least about 99% identity to any one of SEQ ID NOs: 1727-1757. In some embodiments, the base editor is encoded by a nucleic acid having a sequence having 100% identity to any one of SEQ ID NOs: 1727-1757.

[0298] In some embodiments, the base editor comprises a deaminase. In some embodiments, the deaminase binds non-covalently to the endonuclease. In some embodiments, the deaminase is covalently linked to the endonuclease. In some embodiments, the deaminase is fused to the endonuclease.

[0299] In some embodiments, the base editor is an adenine deaminase. In some embodiments, the adenosine deaminase comprises a sequence with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NOs: 50-51, 57, 385-443, 448-475, 595, 1356-1415, 1424, 1556-1644, 1654-1658, and 1665-1703. In some embodiments, the adenosine deaminase comprises a sequence having at least about 70% identity to any one of SEQ ID NOs: 50-51, 57, 385-443, 448-475, 595, 1356-1415, 1424, 1556-1644, 1654-1658, and 1665-1703. In some embodiments, the adenosine deaminase comprises a sequence having at least about 75% identity to any one of SEQ ID NOs: 50-51, 57, 385-443, 448-475, 595, 1356-1415, 1424, 1556-1644, 1654-1658, and 1665-1703. In some embodiments, the adenosine deaminase comprises a sequence having at least about 80% identity to any one of SEQ ID NOs: 50-51, 57, 385-443, 448-475, 595, 1356-1415, 1424, 1556-1644, 1654-1658, and 1665-1703. In some embodiments, the adenosine deaminase comprises a sequence having at least about 85% identity to any one of SEQ ID NOs: 50-51, 57, 385-443, 448-475, 595, 1356-1415, 1424, 1556-1644, 1654-1658, and 1665-1703. In some embodiments, the adenosine deaminase comprises a sequence having at least about 90% identity to any one of SEQ ID NOs: 50-51, 57, 385-443, 448-475, 595, 1356-1415, 1424, 1556-1644, 1654-1658, and 1665-1703. In some embodiments, the adenosine deaminase comprises a sequence having at least about 95% identity to any one of SEQ ID NOs: 50-51, 57, 385-443, 448-475, 595, 1356-1415, 1424, 1556-1644, 1654-1658, and 1665-1703. In some embodiments, the adenosine deaminase comprises a sequence having at least about 96% identity to any one of SEQ ID NOs: 50-51, 57, 385-443, 448-475, 595, 1356-1415, 1424, 1556-1644, 1654-1658, and 1665-1703. In some embodiments, the adenosine deaminase comprises a sequence having at least about 97% identity to any one of SEQ ID NOs: 50-51, 57, 385-443, 448-475, 595, 1356-1415, 1424, 1556-1644, 1654-1658, and 1665-1703. In some embodiments, the adenosine deaminase comprises a sequence having at least

about 98% identity to any one of SEQ ID NOs: 50-51, 57, 385-443, 448-475, 595, 1356-1415, 1424, 1556-1644, 1654-1658, and 1665-1703. In some embodiments, the adenosine deaminase comprises a sequence having at least about 99% identity to any one of SEQ ID NOs: 50-51, 57, 385-443, 448-475, 595, 1356-1415, 1424, 1556-1644, 1654-1658, and 1665-1703. In some embodiments, the adenosine deaminase comprises a sequence having 100% identity to any one of SEQ ID NOs: 50-51, 57, 385-443, 448-475, 595, 1356-1415, 1424, 1556-1644, 1654-1658, and 1665-1703.

[0300] In some embodiments, the base editor is a cytosine deaminase. In some embodiments, the cytosine deaminase comprises a sequence with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 1-49, 58-66, 444-447, 594, 744-835, 970-982, and 1659-1664. In some embodiments, the cytosine deaminase comprises a sequence having at least about 70% identity to any one of SEQ ID NOs: 1-49, 58-66, 444-447, 594, 744-835, 970-982, and 1659-1664. In some embodiments, the cytosine deaminase comprises a sequence having at least about 75% identity to any one of SEQ ID NOs: 1-49, 58-66, 444-447, 594, 744-835, 970-982, and 1659-1664. In some embodiments, the cytosine deaminase comprises a sequence having at least about 80% identity to any one of SEQ ID NOs: 1-49, 58-66, 444-447, 594, 744-835, 970-982, and 1659-1664. In some embodiments, the cytosine deaminase comprises a sequence having at least about 85% identity to any one of SEQ ID NOs: 1-49, 58-66, 444-447, 594, 744-835, 970-982, and 1659-1664. In some embodiments, the cytosine deaminase comprises a sequence having at least about 90% identity to any one of SEQ ID NOs: 1-49, 58-66, 444-447, 594, 744-835, 970-982, and 1659-1664. In some embodiments, the cytosine deaminase comprises a sequence having at least about 95% identity to any one of SEQ ID NOs: 1-49, 58-66, 444-447, 594, 744-835, 970-982, and 1659-1664. In some embodiments, the cytosine deaminase comprises a sequence having at least about 96% identity to any one of SEQ ID NOs: 1-49, 58-66, 444-447, 594, 744-835, 970-982, and 1659-1664. In some embodiments, the cytosine deaminase comprises a sequence having at least about 97% identity to any one of SEQ ID NOs: 1-49, 58-66, 444-447, 594, 744-835, 970-982, and 1659-1664. In some embodiments, the cytosine deaminase comprises a sequence having at least about 98% identity to any one of SEQ ID NOs: 1-49, 58-66, 444-447, 594, 744-835, 970-982, and 1659-1664. In some embodiments, the cytosine deaminase comprises a sequence having at least about 99% identity to any one of SEQ ID NOs: 1-49, 58-66, 444-447, 594, 744-835, 970-982, and 1659-1664. In some embodiments, the cytosine deaminase comprises a sequence having 100% identity to any one of SEQ ID NOs: 1-49, 58-66, 444-447, 594, 744-835, 970-982, and 1659-1664.

[0301] In some embodiments, the base editor comprises one or more modifications. In some embodiments, the base editor comprises a substitution at least one of residues T2, D7, E10, M13, W24, G32, K38, G45, G51, A63, E66, R75, C91, G93, H97, A107, E108, D109, P110, H124, A126, H129, F150, or S165, or any combination thereof relative to SEQ ID NO: 50 when optimally aligned. In some embodiments, the substitution comprises W24G, G51V, E108D, P110H, F150P, D7G, E10G, or H129N, or any combination thereof, relative to SEQ ID NO: 50 when optimally aligned. In some embodiments, the substitution comprises $T2X_1$, $D7X_1$, $E10X_1$, $M13X_4$, $W24X_1$, $G32X_1$, $K38X_2$, $G45X_2$, $G51X_5$, $A63X_7$, $E66X_5$, $E66X_2$, R75H, C91R, $G93X_6$, $H97X_6$, $H97X_5$, $A107X_5$, $E108X_2$, D109N, P110H, $H124X_6$, $A126X_2$, H129R, H129N, F150P, F150S, $S165X_5$, or any combination thereof relative to SEQ ID NO: 50 when optimally aligned, wherein X_1 is A or G; X_2 is D or E; X_3 is N or Q; X_4 is R or K; X_5 is I, L, M, or V; X_6 is F, Y, or W; and X_7 is S or T.

Endonucleases

[0302] Described herein, in certain embodiments, are endonucleases deficient in nuclease activity. In some embodiments, the endonuclease comprises RuvC domain and an HNH domain. In some embodiments, the RuvC domain lacks nuclease activity. In some embodiments, the endonuclease comprises a nickase mutation. In some embodiments, the endonuclease is derived from an uncultivated microorganism. In some embodiments, the endonuclease is a class 2, type II endonuclease. In some embodiments, the endonuclease is configured to cleave one strand of a target nucleic acid (e.g., DNA).

[0303] In some embodiments, the endonuclease is not a Cas9 endonuclease, a Cas14 endonuclease, a Cas12a endonuclease, a Cas12b endonuclease, a Cas 12c endonuclease, a Cas12d endonuclease, a Cas12e endonuclease, a Cas13a endonuclease, a Cas13b endonuclease, a Cas13c endonuclease, or a Cas 13d endonuclease. In some embodiments, the endonuclease has less than 80% identity to a Cas9 endonuclease.

[0304] In some embodiments, the endonuclease comprises a sequence with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127. In some embodiments, the endonuclease comprises a sequence having at least about 70% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127. In some embodiments, the endonuclease comprises a sequence having at least about 75% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127. In some embodiments, the endonuclease comprises a sequence having at least about 80% identity to

any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127. In some embodiments, the endonuclease comprises a sequence having at least about 85% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127. In some embodiments, the endonuclease comprises a sequence having at least about 90% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127. In some embodiments, the endonuclease comprises a sequence having at least about 95% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127. In some embodiments, the endonuclease comprises a sequence having at least about 96% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127. In some embodiments, the endonuclease comprises a sequence having at least about 97% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127. In some embodiments, the endonuclease comprises a sequence having at least about 98% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127. In some embodiments, the endonuclease comprises a sequence having at least about 99% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127. In some embodiments, the endonuclease comprises a sequence having 100% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127.

[0305] In some embodiments, the endonuclease is configured to bind to a protospacer adjacent motif (PAM) sequence comprising any one of SEQ ID NOs: 360-368 and 598. In some embodiments, the endonuclease is configured to bind to a protospacer adjacent motif (PAM) sequence comprising any one of SEQ ID NOs: 360, 362, and 368.

Guide Polynucleotides

[0306] In some embodiments, the engineered system disclosed herein comprises an engineered guide polynucleotide, *e.g.*, a guide ribonucleic acid (gRNA), a single gRNA, or a dual guide RNA.

[0307] In some embodiments, the engineered guide polynucleotide (*e.g.*, engineered guide RNA) is configured to form a complex with the engineered endonuclease. In some embodiments, the engineered guide polynucleotide comprises a spacer sequence. In some embodiments, the spacer sequence is configured to hybridize to a target nucleic acid sequence. In some embodiments, the endonuclease is configured to bind to a protospacer adjacent motif (PAM) sequence.

[0308] In some embodiments, the guide polynucleotide comprises a sequence is encoded by a sequence having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%

identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide is encoded by a sequence having at least about 80% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide is encoded by a sequence having at least about 85% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide is encoded by a sequence having at least about 90% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide is encoded by a sequence having at least about 95% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide is encoded by a sequence having at least about 96% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide is encoded by a sequence having at least about 97% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019In some embodiments, the guide polynucleotide is encoded by a sequence having at least about 98% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide is encoded by a sequence having at least about 99% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide is encoded by a sequence having 100% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019.

[0309] In some embodiments, the guide polynucleotide hybridizes or targets a sequence complementary to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019 or a sequence having at least 90%, 95%, 97%, 98%, or 99% sequence identity to

any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide hybridizes or targets a sequence complementary to a sequence having at least about 80% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide hybridizes or targets a sequence complementary to a sequence having at least about 85% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide hybridizes or targets a sequence complementary to a sequence having at least about 90% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide hybridizes or targets a sequence complementary to a sequence having at least about 95% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide hybridizes or targets a sequence complementary to a sequence having at least about 96% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide hybridizes or targets a sequence complementary to a sequence having at least about 97% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide hybridizes or targets a sequence complementary to a sequence having at least about 98% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide hybridizes or targets a sequence complementary to a sequence having at least about 99% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019. In some embodiments, the guide polynucleotide hybridizes or targets a sequence complementary to a sequence having 100% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019.

[0310] In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019.

[0311] In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1431-1454, 1704, and 2010-2019. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 85% sequence identity to any one of SEQ ID NOs: 1431-1454, 1704, and 2010-2019. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 90% sequence identity to any one of SEQ ID NOs: 1431-1454, 1704, and 2010-2019. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 95% sequence identity to any one of SEQ ID NOs: 1431-1454, 1704, and 2010-2019. In some embodiments, the engineered guide polynucleotide comprises a sequence having 100% sequence identity to any one of SEQ ID NOs: 1431-1454, 1704, and 2010-2019.

[0312] In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1890-1976. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 85% sequence identity to any one of SEQ ID NOs: 1890-1976. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 90% sequence identity to any one of SEQ ID NOs: 1890-1976. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 95% sequence identity to any one of SEQ ID NOs: 1890-1976. In some embodiments, the engineered guide polynucleotide comprises a sequence having 100% sequence identity to any one of SEQ ID NOs: 1890-1976.

[0313] In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1977-2009. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 85% sequence identity to any one of SEQ ID NOs: 1977-2009. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 90% sequence identity to any one of SEQ ID NOs: 1977-2009. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 95% sequence identity to any one of SEQ ID NOs: 1977-2009. In some embodiments, the engineered guide polynucleotide comprises a sequence having 100% sequence identity to any one of SEQ ID NOs: 1977-2009.

[0314] In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1479-1483 and 1758-1889. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 85% sequence identity to any one of SEQ ID NOs: 1479-1483 and 1758-1889. In some embodiments,

the engineered guide polynucleotide comprises a sequence having at least 90% sequence identity to any one of SEQ ID NOs: 1479-1483 and 1758-1889. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 95% sequence identity to any one of SEQ ID NOs: 1479-1483 and 1758-1889. In some embodiments, the engineered guide polynucleotide comprises a sequence having 100% sequence identity to any one of SEQ ID NOs: 1479-1483 and 1758-1889.

[0315] In some embodiments, the guide polynucleotide comprises a sequence complementary to a eukaryotic, fungal, plant, mammalian, or human genomic polynucleotide sequence. In some embodiments, the guide polynucleotide comprises a sequence complementary to a eukaryotic genomic polynucleotide sequence. In some embodiments, the guide polynucleotide comprises a sequence complementary to a fungal genomic polynucleotide sequence. In some embodiments, the guide polynucleotide comprises a sequence complementary to a plant genomic polynucleotide sequence. In some embodiments, the guide polynucleotide comprises a sequence complementary to a mammalian genomic polynucleotide sequence. In some embodiments, the guide polynucleotide comprises a sequence complementary to a human genomic polynucleotide sequence.

[0316] In some embodiments, the guide polynucleotide is 30-250 nucleotides in length. In some embodiments, the guide polynucleotide is 42-44 nucleotides in length. In some embodiments, the guide polynucleotide is 42 nucleotides in length. In some embodiments, the guide polynucleotide is 43 nucleotides in length. In some embodiments, the guide polynucleotide is 44 nucleotides in length. In some embodiments, the guide polynucleotide is 85-245 nucleotides in length. In some embodiments, the guide polynucleotide is more than 90 nucleotides in length. In some embodiments, the guide polynucleotide is less than 245 nucleotides in length. In some embodiments, the guide RNA is 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, or more than 240 nucleotides in length. In some embodiments, the guide RNA is about 30 to about 40, about 30 to about 50, about 30 to about 60, about 30 to about 70, about 30 to about 80, about 30 to about 90, about 30 to about 100, about 30 to about 120, about 30 to about 140, about 30 to about 160, about 30 to about 180, about 30 to about 200, about 30 to about 220, about 30 to about 240, about 50 to about 60, about 50 to about 70, about 50 to about 80, about 50 to about 90, about 50 to about 100, about 50 to about 120, about 50 to about 140, about 50 to about 160, about 50 to about 180, about 50 to about 200, about 50 to about 220, about 50 to about 240, about 100 to about 120, about 100 to about 140, about 100 to about 160, about 100 to about 180, about 100 to about 200, about 100 to about 220, about 100 to about 240, about 160 to about 180, about 160 to about 200, about 160 to about 220, or about 160 to about 240 nucleotides.

[0317] In some embodiments, the engineered guide polynucleotide comprises synthetic nucleotides or modified nucleotides. In some embodiments, the engineered guide polynucleotide comprises one or more inter-nucleoside linkers modified from the natural phosphodiester. In some embodiments, all of the inter-nucleoside linkers of the engineered guide polynucleotide, or contiguous nucleotide sequence thereof, are modified. For example, in some embodiments, the inter nucleoside linkage comprises Sulphur (S), such as a phosphorothioate inter-nucleoside linkage.

[0318] In some embodiments, the engineered guide polynucleotide comprises modifications to a ribose sugar or nucleobase. In some embodiments, the engineered guide polynucleotide comprises one or more nucleosides comprising a modified sugar moiety, wherein the modified sugar moiety is a modification of the sugar moiety when compared to the ribose sugar moiety found in deoxyribose nucleic acid (DNA) and RNA. In some embodiments, the modification is within the ribose ring structure. Exemplary modifications include, but are not limited to, replacement with a hexose ring (HNA), a bicyclic ring having a biradical bridge between the C2 and C4 carbons on the ribose ring (*e.g.*, locked nucleic acids (LNA)), or an unlinked ribose ring which typically lacks a bond between the C2 and C3 carbons (*e.g.*, UNA). In some embodiments, the sugar-modified nucleosides comprise bicyclohexose nucleic acids or tricyclic nucleic acids. In some embodiments, the modified nucleosides comprise nucleosides where the sugar moiety is replaced with a non-sugar moiety, for example peptide nucleic acids (PNA) or morpholino nucleic acids.

[0319] In some embodiments, the engineered guide polynucleotide comprises one or more modified sugars. In some embodiments, the sugar modifications comprise modifications made by altering the substituent groups on the ribose ring to groups other than hydrogen, or the 2'-OH group naturally found in DNA and RNA nucleosides. In some embodiments, substituents are introduced at the 2', 3', 4', or 5' positions, or combinations thereof. In some embodiments, nucleosides with modified sugar moieties comprise 2' modified nucleosides, *e.g.*, 2' substituted nucleosides. A 2' sugar modified nucleoside, in some embodiments, is a nucleoside that has a substituent other than -H or -OH at the 2' position (2' substituted nucleoside) or comprises a 2' linked biradical, and comprises 2' substituted nucleosides and LNA (2'-4' biradical bridged) nucleosides. Examples of 2'-substituted modified nucleosides comprise, but are not limited to, 2'-O-alkyl-RNA, 2'-O-methyl-RNA, 2'-alkoxy-RNA, 2'-O-methoxyethyl-RNA (MOE), 2'-amino-DNA, 2'-Fluoro-RNA, and 2'-F-ANA nucleosides. In some embodiments, the modification in the ribose group comprises a modification at the 2' position of the ribose group. In some embodiments, the modification at the 2' position of the ribose group is selected from the group consisting of 2'-O-methyl, 2'-fluoro, 2'-deoxy, and 2'-O-(2-methoxyethyl).

[0320] In some embodiments, the engineered guide polynucleotide comprises one or more modified sugars. In some embodiments, the engineered guide polynucleotide comprises only modified sugars. In certain embodiments, the engineered guide polynucleotide comprises greater than about 10%, 25%, 50%, 75%, or 90% modified sugars. In some embodiments, the modified sugar is a bicyclic sugar. In some embodiments, the modified sugar comprises a 2'-O-methoxyethyl group. In some embodiments, the engineered guide polynucleotide comprises both inter-nucleoside linker modifications and nucleoside modifications.

[0321] In some embodiments, the engineered guide polynucleotide comprises a hairpin comprising at least 8 base-paired ribonucleotides. In some embodiments, the engineered guide polynucleotide comprises a hairpin comprising at least 9 base-paired ribonucleotides. In some embodiments, the engineered guide polynucleotide comprises a hairpin comprising at least 10 base-paired ribonucleotides. In some embodiments, the engineered guide polynucleotide comprises a hairpin comprising at least 11 base-paired ribonucleotides. In some embodiments, the engineered guide polynucleotide comprises a hairpin comprising at least 12 base-paired ribonucleotides.

[0322] In some embodiments, the engineered guide polynucleotide comprises a DNA-targeting segment. In some embodiments, the DNA-targeting segment comprises a nucleotide sequence that is complementary to a target sequence. In some embodiments, the target sequence is in a target DNA molecule. In some embodiments, the engineered guide polynucleotide comprises a protein-binding segment. In some embodiments, the protein-binding segment comprises two complementary stretches of nucleotides. In some embodiments, the two complementary stretches of nucleotides hybridize to form a double-stranded RNA (dsRNA) duplex. In some embodiments, the two complementary stretches of nucleotides are covalently linked to one another with intervening nucleotides.

Base Editing Systems

[0323] Described herein, in certain embodiments, are engineered systems comprising: (a) a base editor; (b) an endonuclease configured to bind the base editor and is deficient in nuclease activity; and (c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence. In a polynucleotide when referring to a T, a T means U (Uracil) in RNA and T (Thymine) in DNA.

[0324] Described herein, in certain embodiments, are engineered base editing systems comprising an engineered base editing system comprising: a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ

ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415; and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence. Described herein, in certain embodiments, are engineered base editing systems comprising an engineered base editing system comprising: a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415; and an engineered guide polynucleotide which forms a complex with an endonuclease deficient in nuclease activity of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

[0325] Described herein, in certain embodiments, are engineered base editing systems comprising an engineered base editing system comprising: a base editor comprising a sequence having at least 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023; and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence. Described herein, in certain embodiments, are engineered base editing systems comprising an engineered base editing system comprising: a base editor comprising a sequence having at least 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023; and an engineered guide polynucleotide which forms a complex with an endonuclease deficient in nuclease activity of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

[0326] In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 70% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease; and c) an engineered guide polynucleotide. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 75% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease; and c) an engineered guide polynucleotide. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 80% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease; and c) an engineered guide polynucleotide. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 85% identity to

any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease; and c) an engineered guide polynucleotide. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 90% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease; and c) an engineered guide polynucleotide. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 95% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease; and c) an engineered guide polynucleotide. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 96% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease; and c) an engineered guide polynucleotide. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 97% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease; and c) an engineered guide polynucleotide. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 98% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease; and c) an engineered guide polynucleotide. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 99% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease; and c) an engineered guide polynucleotide. In some embodiments, the engineered system comprises a) a base editor comprising 100% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease; and c) an engineered guide polynucleotide.

[0327] In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 70% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and

comprising a spacer sequence configured to hybridize to a target nucleic acid sequence. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 75% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 80% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 85% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 90% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 95% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 96% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer

sequence configured to hybridize to a target nucleic acid sequence. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 97% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 98% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 99% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence. In some embodiments, the engineered system comprises a) a base editor comprising 100% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence.

[0328] In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 70% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 70% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 75% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595,

599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 75% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 80% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 80% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 85% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 85% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 90% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 90% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 95% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595,

599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 95% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 96% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 96% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 97% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 97% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 98% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 98% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 99% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595,

599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 99% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising 100% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising 100% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the guide polynucleotide hybridizes or targets a sequence complementary to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710 or a sequence having at least 90%, 95%, 97%, 98%, or 99% sequence identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710.

[0329] In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 70% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor and comprising a sequence having at least about 70% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 70% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 75% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor and comprising a

sequence having at least about 75% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 75% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 80% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor and comprising a sequence having at least about 80% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 80% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 85% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor and comprising a sequence having at least about 85% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, or 1122-1127; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 85% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 90% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor and comprising a sequence having at least about 90% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 90% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-

1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 95% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor and comprising a sequence having at least about 95% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 95% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 96% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor and comprising a sequence having at least about 96% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 96% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 97% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor and comprising a sequence having at least about 97% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 97% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 98% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor and comprising a sequence

having at least about 98% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 98% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising a sequence having at least about 99% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor and comprising a sequence having at least about 99% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising a sequence having at least about 99% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the engineered system comprises a) a base editor comprising 100% identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703; b) an endonuclease configured to bind the base editor and comprising a sequence having 100% identity to any one of SEQ ID NOs: 70-78, 596, 597, 1120, and 1122-1127; and c) an engineered guide polynucleotide configured to form a complex with the endonuclease and comprising a spacer sequence configured to hybridize to a target nucleic acid sequence, the engineered guide polynucleotide comprising 100% identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710. In some embodiments, the guide polynucleotide hybridizes or targets a sequence complementary to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710 or a sequence having at least 90%, 95%, 97%, 98%, or 99% sequence identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, and 1704-1710.

[0330] In some embodiments, the endonuclease comprises a sequence at least 70%, at least 80%, or at least 90% identical to SEQ ID NO: 70; the guide RNA structure comprises a sequence at least 70%, at least 80%, or at least 90% identical to at least one of SEQ ID NO: 88; and the endonuclease is configured to bind to a PAM comprising SEQ ID NO: 360.

[0331] In some embodiments, the endonuclease comprises a sequence at least 70%, at least 80%, or at least 90% identical to SEQ ID NO: 71; the guide RNA structure comprises a sequence at least 70%, at least 80%, or at least 90% identical to at least one of SEQ ID NO: 89; and the endonuclease is configured to bind to a PAM comprising SEQ ID NO: 361.

[0332] In some embodiments, the endonuclease comprises a sequence at least 70%, at least 80%, or at least 90% identical to SEQ ID NO: 73; the guide RNA structure comprises a sequence at least 70%, at least 80%, or at least 90% identical to at least one of SEQ ID NO: 91; and the endonuclease is configured to bind to a PAM comprising SEQ ID NO: 363.

[0333] In some embodiments, the endonuclease comprises a sequence at least 70%, at least 80%, or at least 90% identical to SEQ ID NO: 75, or a variant thereof; the guide RNA structure comprises a sequence at least 70%, at least 80%, or at least 90% identical to at least one of SEQ ID NO: 93; and the endonuclease is configured to bind to a PAM comprising SEQ ID NO: 365.

[0334] In some embodiments, the endonuclease comprises a sequence at least 70%, at least 80%, or at least 90% identical to SEQ ID NO: 76, or a variant thereof; the guide RNA structure comprises a sequence at least 70%, at least 80%, or at least 90% identical to at least one of SEQ ID NO: 94; and the endonuclease is configured to bind to a PAM comprising SEQ ID NO: 366.

[0335] In some embodiments, the endonuclease comprises a sequence at least 70%, at least 80%, or at least 90% identical to SEQ ID NO: 77, or a variant thereof; the guide RNA structure comprises a sequence at least 70%, at least 80%, or at least 90% identical to at least one of SEQ ID NO: 95; and the endonuclease is configured to bind to a PAM comprising SEQ ID NO: 367.

[0336] In some embodiments, the endonuclease comprises a sequence at least 70%, at least 80%, or at least 90% identical to SEQ ID NO: 78, or a variant thereof; the guide RNA structure comprises a sequence at least 70%, at least 80%, or at least 90% identical to at least one of SEQ ID NO: 96; and the endonuclease is configured to bind to a PAM comprising SEQ ID NO: 368.

[0337] In some embodiments, the base editor comprises an adenine deaminase. In some embodiments, the adenine deaminase comprises SEQ ID NO: 57. In some embodiments, the base editor comprises a cytosine deaminase. In some embodiments, the cytosine deaminase comprises SEQ ID NO: 58.

[0338] In some embodiments, the endonuclease or base editor comprises one or more modifications in a nickase domain. In some embodiments, the nickase domain comprises an aspartate to alanine mutation at residue 9 relative to SEQ ID NO: 70, residue 13 relative to SEQ ID NOs: 71, 72, or 74, residue 12 relative to SEQ ID NO: 73, residue 17 relative to SEQ ID NO: 75, residue 23 relative to SEQ ID NO: 76, or residue 10 relative to SEQ ID NO: 597, or any combination thereof. In some embodiments, the endonuclease or base editor comprises a substitution of 109N and at least one other substitution comprising any one of 24R, 37L, 49A,

52L, 83S, 85F, 107V, 110S, 112R, 120N, 123N, 124Y, 147C, 148Y, 148R, 150Y, 156V, 157F, 158N, 166I, or 129N, or any combination thereof relative to SEQ ID NO: 386 when optimally aligned. the endonuclease or base editor comprises at least one substitution of a wild-type amino acid for a non-wild-type amino acid comprising any one of W90A, W90F, W90H, W90Y, Y120F, Y120H, Y121F, Y121H, Y121Q, Y121A, Y121D, Y121W, H122Y, H122F, H122I, H122A, H122W, H122D, Y121T, R33A, R34A, R34K, H122A, R33A, R34A, R52A, N57G, H122A, E123A, E123Q, W127F, W127H, W127Q, W127A, W127D, R39A, K40A, H128A, N63G, R58A, H121F, H121Y, H121Q, H121A, H121D, H121W, R33A, K34A, H122A, H121A, R52A, P26R, P26A, N27R, N27A, W44A, W45A, K49G, S50G, R51G, R121A, I122A, N123A, Y88F, Y120F, P22R, P22A, K23A, K41R, K41A, E54A, E54A, E55A, K30A, K30R, M32A, M32K, Y117A, K118A, I119A, I119H, R120A, R121A, P46A, P46R, N29A, R27A, or N50G, or any combination thereof. In some embodiments, the nickase comprises an aspartate to alanine mutation at residue 9 relative to SEQ ID NO: 70, residue 13 relative to SEQ ID NOs: 71, 72, or 74, residue 12 relative to SEQ ID NO: 73, residue 17 relative to SEQ ID NO: 75, residue 23 relative to SEQ ID NO: 76, or residue 10 relative to SEQ ID NO: 597, or any combination thereof.

[0339] In some embodiments, the engineered system further comprises a uracil DNA glycosylase inhibitor. In some embodiments, the uracil DNA glycosylase inhibitor comprises a sequence with at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 52-56 and SEQ ID NO: 67. In some embodiments, the uracil DNA glycosylase inhibitor comprises a sequence having at least about 70% identity to any one of SEQ ID NOs: 52-56 and 67. In some embodiments, the uracil DNA glycosylase inhibitor comprises a sequence having at least about 75% identity to any one of SEQ ID NOs: 52-56 and 67. In some embodiments, the uracil DNA glycosylase inhibitor comprises a sequence having at least about 80% identity to any one of SEQ ID NOs: 52-56 and 67. In some embodiments, the uracil DNA glycosylase inhibitor comprises a sequence having at least about 85% identity to any one of SEQ ID NOs: 52-56 and 67. In some embodiments, the uracil DNA glycosylase inhibitor comprises a sequence having at least about 90% identity to any one of SEQ ID NOs: 52-56 or 67. In some embodiments, the uracil DNA glycosylase inhibitor comprises a sequence having at least about 95% identity to any one of SEQ ID NOs: 52-56 and 67. In some embodiments, the uracil DNA glycosylase inhibitor comprises a sequence having at least about 96% identity to any one of SEQ ID NOs: 52-56 and 67. In some embodiments, the uracil DNA glycosylase inhibitor comprises a sequence having at least about 97% identity to any one of SEQ ID NOs: 52-56 and

67. In some embodiments, the uracil DNA glycosylase inhibitor comprises a sequence having at least about 98% identity to any one of SEQ ID NOs: 52-56 and 67. In some embodiments, the uracil DNA glycosylase inhibitor comprises a sequence having at least about 99% identity to any one of SEQ ID NOs: 52-56 and 67. In some embodiments, the uracil DNA glycosylase inhibitor comprises a sequence having 100% identity to any one of SEQ ID NOs: 52-56 and 67.

[0340] In some embodiments, the base editor binds non-covalently to the endonuclease. In some embodiments, the base editor is covalently linked to the endonuclease. In some embodiments, the base editor is fused to the endonuclease at the N-terminus or at the C-terminus. In some embodiments, the base editor is fused to the endonuclease.

[0341] In some embodiments, the endonuclease is covalently coupled linked to the base editor or covalently linked to the base editor through a linker. In some embodiments, the linker comprises a sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SGGSSGGSSGSETPGTSESATPESSGGSSGGS, SGSETPGTSESATPESA, GSGGS, SGSETPGTSESATPES, SGGSS, or GAAA. In some embodiments, the linker comprises a sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 1647-1653.

Table 1: Linkers for deaminase systems described herein

Length	Sequence	SEQ ID NO
7 AAs	PAPAPAP	1647
14 AAs	KLGGGAPAVGGGPK	1648
15 AAs	GGGGSGGGGSGGGGS	1649
XTEN (17 aa)	SGSETPGTSEASTPESA	1650
26 AAs	GGGGSGGGGSEAAAAKGGGGSGGGGS	1651
32 AAs	GGGGSGGGGSEAAAAKEAAAAKGGGGSGGGGS	1652
44 AAs	KGKGKGMGAGTLSTDKGESLGIKYEEGQSHRPTNPNASR MAQKV	1653

[0342] In some embodiments, the system further comprises a source of Mg^{2+}.

[0343] In some embodiments, the endonuclease comprises one or more nuclear localization sequences (NLSs) proximal to an N- or C-terminus of the endonuclease. In some embodiments,

the base editor comprises one or more nuclear localization sequences (NLSs) proximal to an N- or C-terminus of the endonuclease. The NLS can comprise any of the sequences in Table 2 below, or a combination thereof.

[0344] In some embodiments, the NLS comprises a sequence of any one of SEQ ID NOs: 369-384 and 2024-2053, or a sequence having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having at least about 80% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having at least about 85% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having at least about 90% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having at least about 91% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having at least about 92% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having at least about 93% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having at least about 94% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having at least about 95% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having at least about 96% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having at least about 97% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having at least about 98% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having at least about 99% identity to any one of SEQ ID NOs: 369-384 and 2024-2053. In some embodiments, the NLS comprises a sequence having 100% identity to any one of SEQ ID NOs: 369-384 and 2024-2053.

Table 2: Example NLS Sequences that can be used with Effectors According to the Disclosure

Source	NLS amino acid sequence	SEQ ID NO:
SV40	PKKKRKV	369
nucleoplasmin bipartite NLS	KRPAATKKAGQAKKKK	370

Source	NLS amino acid sequence	SEQ ID NO:
c-myc NLS	PAAKRVKLD	371
c-myc NLS	RQRRNELKRSP	372
hRNPA1 M9 NLS	NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKP RNQGGY	373
Importin-alpha IBB domain	RMRIZFKNKGKDTAELRRRRVEVSVELRKAKK DEQILKRRNV	374
Myoma T protein	VSRKRPRP	375
Myoma T protein	PPKKARED	376
p53	PQPKKKPL	377
mouse c-abl IV	SALIKKKKKMAP	378
influenza virus NS1	DRLRR	379
influenza virus NS1	PKQKKRK	380
Hepatitis virus delta antigen	RKLKKKIKKL	381
mouse Mx1 protein	REKKKFLKRR	382
human poly(ADP-ribose) polymerase	KRKGDEVDGVDEVAKKKSKK	383
steroid hormone receptor (human) glucocorticoid	RKCLQAGMNLEARKTKK	384
5' SV40 NLS	MAPKKKRKVGGGGS	2024
3' SV40 1 NLS	SGGAPKKKRKV	2025
3' SV40 2 NLS	SGGAPKKKRKV	2026
3' SV40 3 NLS	SGGAPKKKRKVSGGAPKKKRKVSGGAPKKKR KV	2027
5' Nucleoplasmin NLS	MKRPAATKKAGQAKKKKGGGGS	2028
3' Nucleoplasmin 1 NLS	SGGKRPAATKKAGQAKKKK	2029
3' Nucleoplasmin 2 NLS	SGGKRPAATKKAGQAKKKKSGGKRPAATKKA GQAKKKK	2030
3' Nucleoplasmin 3 NLS	SGGKRPAATKKAGQAKKKKSGGKRPAATKKA GQAKKKKSGGKRPAATKKAGQAKKKK	2031
5' c-Myc NLS	MPAAKRVKLDGGGGS	2032
3' c-Myc 1 NLS	SGGPAAKRVKLD	2033
3' c-Myc 2 NLS	SGGPAAKRVKLDSGGPAAKRVKLD	2034
3' c-Myc 3 NLS	SGGPAAKRVKLDSGGPAAKRVKLDSGGPAAK RVKLD	2035
Synthetic 5' Class 2-1 NLS	MQAAKRPRTTGGGGS	2036
Synthetic 3' Class 2-1 1 NLS	SGGQAAKRPRTT	2037
Synthetic 3' Class 2-1 2 NLS	SGGQAAKRPRTTSGGQAAKRPRTT	2038
Synthetic 5' Class 2-2 NLS	MRAAKRPRTTGGGGS	2039
Synthetic 3' Class 2-2 1 NLS	SGGRAAKRPRTT	2040
Synthetic 3' Class 2-2 2 NLS	SGGRAAKRPRTTSGGRAAKRPRTT	2041
Synthetic 5' Class 2-2 NLS	MRAAKRPRTTGGGGS	2042
Synthetic 3' Class 2-2 1 NLS	SGGRAAKRPRTT	2043

Source	NLS amino acid sequence	SEQ ID NO:
Synthetic 3' Class 2-2 2 NLS	SGGRAAKRPRTTSGGRAAKRPRTT	2044
Synthetic 5' Class 3 NLS	MAAAKRSWSMAFGGGGS	2045
Synthetic 3' Class 3 1 NLS	MSGGAAAKRSWSMAF	2046
Synthetic 3' Class 3 2 NLS	MSGGAAAKRSWSMAFMSGGAAAKRSWSMAF	2047
Synthetic 5' Class 4 NLS	MRAAKRKYFAVGGGGS	2048
Synthetic 3' Class 4 1 NLS	MSGGRAAKRKYFAV	2049
Synthetic 3' Class 4 2 NLS	MSGGRAAKRKYFAVMSGGRAAKRKYFAV	2050
5' HCV NLS	MPKPQRKTKRGGGGS	2051
3' HCV NLS	SGGPPRKKRTVV	2052
3' hRNP A1 NLS	SGGFGNYNNQSSNFGPMKGGNFGGRSSGPY	2053

Cells

[0345] Described herein, in certain embodiments, is a cell comprising the systems described herein.

[0346] In some embodiments, the cell is a eukaryotic cell (*e.g.*, a plant cell, an animal cell, a protist cell, or a fungi cell), a mammalian cell (a Chinese hamster ovary (CHO) cell, baby hamster kidney (BHK), human embryo kidney (HEK), mouse myeloma (NS0), or human retinal cells), an immortalized cell (*e.g.*, a HeLa cell, a COS cell, a HEK-293T cell, a MDCK cell, a 3T3 cell, a PC12 cell, a Huh7 cell, a HepG2 cell, a K562 cell, a N2a cell, or a SY5Y cell), an insect cell (*e.g.*, a *Spodoptera frugiperda* cell, a *Trichoplusia ni* cell, a *Drosophila melanogaster* cell, a S2 cell, or a *Heliothis virescens* cell), a yeast cell (*e.g.*, a *Saccharomyces cerevisiae* cell, a Cryptococcus cell, or a Candida cell), a plant cell (*e.g.*, a parenchyma cell, a collenchyma cell, or a sclerenchyma cell), a fungal cell (*e.g.*, a *Saccharomyces cerevisiae* cell, a Cryptococcus cell, or a Candida cell), or a prokaryotic cell (*e.g.*, a *E. coli* cell, a streptococcus bacterium cell, a streptomyces soil bacteria cell, or an archaea cell). In some embodiments, the cell is a eukaryotic cell. In some embodiments, the cell is a mammalian cell. In some embodiments, the cell is an immortalized cell. In some embodiments, the cell is an insect cell. In some embodiments, the cell is a yeast cell. In some embodiments, the cell is a plant cell. In some embodiments, the cell is a fungal cell. In some embodiments, the cell is a prokaryotic cell.

[0347] In some embodiments, the cell is an A549, HEK-293, HEK-293T, BHK, CHO, HeLa, MRC5, Sf9, Cos-1, Cos-7, Vero, BSC 1, BSC 40, BMT 10, WI38, HeLa, Saos, C2C12, L cell, HT1080, HepG2, Huh7, K562, a primary cell, or derivative thereof.

[0348] In some embodiments, the present disclosure provides a cell (*e.g.*, host cell) comprising a vector described herein. In some embodiments, the cell expresses the engineered system described herein or components thereof. In some embodiments, the cell is a human cell. In some

embodiments, the cell is genome edited *ex vivo*. In some embodiments, the cell is genome edited *in vivo*.

[0349] Described herein, in some embodiments, are host cells comprising an open reading frame encoding a heterologous endonuclease and a heterologous base editor having at least 75% sequence identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703. In some embodiments, said heterologous base editor comprises a sequence having at least about 75%, at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to any one of SEQ ID NOs: 1-51, 57-66, 385-443, 444-475, 594-595, 599-675, 744-835, 970-1098, 1128-1186, 1208-1315, 1356-1415, 1424, 1556-1644, and 1654-1703.

[0350] In some embodiments, the host cell is a bacterial cell. In some embodiments, the bacterial cell is *Bifidobacterium longum, Bifidobacterium lactis, Bifidobacterium animalis, Bifidobacterium breve, Bifidobacterium infantis, Bifidobacterium adolescentis, Lactobacillus acidophilus, Lactobacillus casei, Lactobacillus paracasei, Lactobacillus salivarius, Lactobacillus reuteri, Lactobacillus rhamnosus, Lactobacillus johnsonii, Lactobacillus plantarum, Lactobacillus fermentum, Lactococcus lactis, Streptococcus thermophilus, Lactococcus lactis, Lactococcus diacetylactis, Lactococcus cremoris, Lactobacillus bulgaricus, Lactobacillus helveticus, Lactobacillus delbrueckii,* or *Escherichia coli*. In some embodiments, the host cell is an *E. coli* cell. In some embodiments, the *E. coli* cell is a λDE3 lysogen or a BL21(DE3) strain. In some embodiments, the *E. coli* cell has an *ompT lon* genotype.

[0351] In some embodiments, the cell is within a cochlea. In some embodiments, the cell is within an embryo. In some embodiments, the embryo is a two-cell embryo. In some embodiments, the embryo is a mouse embryo.

Lipid nanoparticles

[0352] Lipid nanoparticles as described herein can be 4-component lipid nanoparticles. Such nanoparticles can be configured for delivery of RNA or other nucleic acids (e.g. synthetic RNA, mRNA, or *in vitro*-synthesized mRNA) and can be generally formulated as described in WO2012135805A2. Such nanoparticles can generally comprise: (a) a cationic lipid (e.g. 98N12-5 (TETA5-LAP), DLin DMA, DLin-K-DMA (2,2-Dilinoleyl-4-dimethylaminomethyl-[1,3]-dioxolane), DLin-KC2-DMA, DLin-MC3-DMA, or C12-200), (b) a neutral lipid (e.g. DSPC or

DOPE), (c) a sterol (e.g. cholesterol or a cholesterol analog), and (d) a PEG-modified lipid (e.g. PEG-DMG).

[0353] The cationic lipid referred to herein as "C12-200" is disclosed by Love *et al.*, Proc Natl Acad Sci USA. 2010 107:1864-1869 and Liu and Huang, Molecular Therapy. 2010 669-670. Cationic lipid formulations can include particles comprising either 3 or 4 or more components in addition to polynucleotide, primary construct, or RNA (e.g. mRNA). As an example, formulations with certain cationic lipids include, but are not limited to, 98N12-5, and may contain 42% lipidoid, 48% cholesterol, and 10% PEG (C14 or greater alkyl chain length). As another example, formulations with certain lipidoids include, but are not limited to, C12-200 and may contain 50% cationic lipid, 10% disteroylphosphatidyl choline, 38.5% cholesterol, and 1.5% PEG-DMG.

[0354] In some embodiments, lipid nanoparticles are formulated as described in US10709779B2. In some embodiments, the cationic lipid nanoparticle comprises a cationic lipid, a PEG-modified lipid, a sterol, and a non-cationic lipid. In some embodiments, the cationic lipid is selected from the group consisting of 98N12-5 (TETA5-LAP), DLin DMA, DLin-K-DMA (2,2-Dilinoleyl-4-dimethylaminomethyl-[1,3]-dioxolane), DLin-KC2-DMA, DLin-MC3-DMA, and C12-200. In some embodiments, the cationic lipid nanoparticle has a molar ratio of about 20-60% cationic lipid, about 5-25% non-cationic lipid, about 25-55% sterol, and about 0.5-15% PEG-modified lipid. In some embodiments, the cationic lipid nanoparticle comprises a molar ratio of about 50% cationic lipid, about 1.5% PEG-modified lipid, about 38.5% cholesterol, and about 10% non-cationic lipid. In some embodiments, the cationic lipid nanoparticle comprises a molar ratio of about 55% cationic lipid, about 2.5% PEG-modified lipid, about 32.5% cholesterol, and about 10% non-cationic lipid. In some embodiments, the cationic lipid is an ionizable cationic lipid, the non-cationic lipid is a neutral lipid, and the sterol is a cholesterol. In some embodiments, the cationic lipid nanoparticle has a molar ratio of 50:38.5:10:1.5 of cationic lipid: cholesterol: PEG2000-DMG:DSPC or DMG:DOPE. In some embodiments, lipid nanoparticles as described herein can comprise cholesterol, 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 1,1'-((2-(4-(2-((2-(bis(2-hydroxydodecyl)amino)ethyl)(2-hydroxydodecyl)amino)ethyl)piperazin-1-yl)ethyl)azanediyl)bis(dodecan-2-ol) (C12-200), and DMG-PEG-2000 at molar ratios of 47.5:16:35:1.5.

Delivery and Vectors

[0355] Disclosed herein, in some embodiments, are nucleic acids encoding an engineered system described herein comprising a base editor, an endonuclease, and an engineered guide

polynucleotide or components thereof (*e.g.*, a base editor, an endonuclease, or an engineered guide polynucleotide).

[0356] In some embodiments, the nucleic acid encoding the engineered system or components thereof is a DNA, for example a linear DNA, a plasmid DNA, or a minicircle DNA. In some embodiments, the nucleic acid encoding the engineered system is an RNA, for example a mRNA.

[0357] In some embodiments, the nucleic acid encoding the engineered system or components thereof is delivered by a nucleic acid-based vector. In some embodiments, the nucleic acid-based vector is a plasmid (*e.g.*, circular DNA molecules that can autonomously replicate inside a cell), cosmid (*e.g.*, pWE or sCos vectors), artificial chromosome, human artificial chromosome (HAC), yeast artificial chromosomes (YAC), bacterial artificial chromosome (BAC), P1-derived artificial chromosomes (PAC), phagemid, phage derivative, bacmid, or virus. In some embodiments, the nucleic acid-based vector is selected from the list consisting of: pSF-CMV-NEO-NH2-PPT-3XFLAG, pSF-CMV-NEO-COOH-3XFLAG, pSF-CMV-PURO-NH2-GST-TEV, pSF-OXB20-COOH-TEV-FLAG(R)-6His, pCEP4 pDEST27, pSF-CMV-Ub-KrYFP, pSF-CMV-FMDV-daGFP, pEF1a-mCherry-N1 vector, pEF1a-tdTomato vector, pSF-CMV-FMDV-Hygro, pSF-CMV-PGK-Puro, pMCP-tag(m), pSF-CMV-PURO-NH2-CMYC, pSF-OXB20-BetaGal,pSF-OXB20-Fluc, pSF-OXB20, pSF-Tac, pRI 101-AN DNA, pCambia2301,pTYB21, pKLAC2, pAc5.1/V5-His A, and pDEST8.

[0358] In some embodiments, the nucleic acid-based vector comprises a promoter. In some embodiments, an open reading frame is operably linked to the promoter. In some embodiments, the promoter is selected from the group consisting of a mini promoter, an inducible promoter, a constitutive promoter, and derivatives thereof. In some embodiments, the promoter is selected from the group consisting of CMV, CBA, EF1a, CAG, PGK, TRE, U6, UAS, T7, Sp6, lac, araBad, trp, Ptac, p5, p19, p40, Synapsin, CaMKII, GRK1, and derivatives thereof. In some embodiments the promoter is a U6 promoter. In some embodiments, the promoter is a CAG promoter.

[0359] In some embodiments, the open reading frame is operably linked to a T7 promoter sequence, a T7-lac promoter sequence, a lac promoter sequence, a tac promoter sequence, a trc promoter sequence, a ParaBAD promoter sequence, a PrhaBAD promoter sequence, a T5 promoter sequence, a *cspA* promoter sequence, an $araP_{BAD}$ promoter, a strong leftward promoter from phage lambda (pL promoter), or any combination thereof.

[0360] In some embodiments, the open reading frame comprises a sequence encoding an affinity tag linked in-frame to a sequence encoding said base editor. In some embodiments, the affinity tag is an immobilized metal affinity chromatography (IMAC) tag. In some embodiments, the IMAC tag is a polyhistidine tag. In some embodiments, the affinity tag is a myc tag, a human

influenza hemagglutinin (HA) tag, a maltose binding protein (MBP) tag, a glutathione S-transferase (GST) tag, a streptavidin tag, a FLAG tag, or any combination thereof.

[0361] In some embodiments, the affinity tag is linked in-frame to said sequence encoding said base editor via a linker sequence encoding a protease cleavage site. In some embodiments, the protease cleavage site is a tobacco etch virus (TEV) protease cleavage site, a PreScission® protease (PSP) cleavage site, a Thrombin cleavage site, a Factor Xa cleavage site, an enterokinase cleavage site, or any combination thereof.

[0362] In some embodiments, the open reading frame is codon-optimized for expression in said host cell. In some embodiments, the open reading frame is provided on a vector. In some embodiments, the open reading frame is integrated into a genome of said host cell.

[0363] In some embodiments, the nucleic acid-based vector is a virus. In some embodiments, the virus is an alphavirus, a parvovirus, an adenovirus, an AAV, a baculovirus, a Dengue virus, a lentivirus, a herpesvirus, a poxvirus, an anellovirus, a bocavirus, a vaccinia virus, or a retrovirus. In some embodiments, the virus is an alphavirus. In some embodiments, the virus is a parvovirus. In some embodiments, the virus is an adenovirus. In some embodiments, the virus is an AAV. In some embodiments, the virus is a baculovirus. In some embodiments, the virus is a Dengue virus. In some embodiments, the virus is a lentivirus. In some embodiments, the virus is a herpesvirus. In some embodiments, the virus is a poxvirus. In some embodiments, the virus is an anellovirus. In some embodiments, the virus is a bocavirus. In some embodiments, the virus is a vaccinia virus. In some embodiments, the virus is or a retrovirus.

[0364] In some embodiments, the AAV is AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAV13, AAV14, AAV15, AAV16, AAV-rh8, AAV-rh10, AAV-rh20, AAV-rh39, AAV-rh74, AAV-rhM4-1, AAV-hu37, AAV-Anc80, AAV-Anc80L65, AAV-7m8, AAV-PHP-B, AAV-PHP-EB, AAV-2.5, AAV-2tYF, AAV-3B, AAV-LK03, AAV-HSC1, AAV-HSC2, AAV-HSC3, AAV-HSC4, AAV-HSC5, AAV-HSC6, AAV-HSC7, AAV-HSC8, AAV-HSC9, AAV-HSC10, AAV-HSC11, AAV-HSC12, AAV-HSC13, AAV-HSC14, AAV-HSC15, AAV-TT, AAV-DJ/8, AAV-Myo, AAV-NP40, AAV-NP59, AAV-NP22, AAV-NP66, AAV-HSC16, or a derivative thereof. In some embodiments, the herpesvirus is HSV type 1, HSV-2, VZV, EBV, CMV, HHV-6, HHV-7, or HHV-8.

[0365] In some embodiments, the virus is AAV1 or a derivative thereof. In some embodiments, the virus is AAV2 or a derivative thereof. In some embodiments, the virus is AAV3 or a derivative thereof. In some embodiments, the virus is AAV4 or a derivative thereof. In some embodiments, the virus is AAV5 or a derivative thereof. In some embodiments, the virus is AAV6 or a derivative thereof. In some embodiments, the virus is AAV7 or a derivative thereof. In some embodiments, the virus is AAV8 or a derivative thereof. In some embodiments, the virus

is AAV9 or a derivative thereof. In some embodiments, the virus is AAV10 or a derivative thereof. In some embodiments, the virus is AAV11 or a derivative thereof. In some embodiments, the virus is AAV12 or a derivative thereof. In some embodiments, the virus is AAV13 or a derivative thereof. In some embodiments, the virus is AAV14 or a derivative thereof. In some embodiments, the virus is AAV15 or a derivative thereof. In some embodiments, the virus is AAV16 or a derivative thereof. In some embodiments, the virus is AAV-rh8 or a derivative thereof. In some embodiments, the virus is AAV-rh10 or a derivative thereof. In some embodiments, the virus is AAV-rh20 or a derivative thereof. In some embodiments, the virus is AAV-rh39 or a derivative thereof. In some embodiments, the virus is AAV-rh74 or a derivative thereof. In some embodiments, the virus is AAV-rhM4-1 or a derivative thereof. In some embodiments, the virus is AAV-hu37 or a derivative thereof. In some embodiments, the virus is AAV-Anc80 or a derivative thereof. In some embodiments, the virus is AAV-Anc80L65 or a derivative thereof. In some embodiments, the virus is AAV-7m8 or a derivative thereof. In some embodiments, the virus is AAV-PHP-B or a derivative thereof. In some embodiments, the virus is AAV-PHP-EB or a derivative thereof. In some embodiments, the virus is AAV-2.5 or a derivative thereof. In some embodiments, the virus is AAV-2tYF or a derivative thereof. In some embodiments, the virus is AAV-3B or a derivative thereof. In some embodiments, the virus is AAV-LK03 or a derivative thereof. In some embodiments, the virus is AAV-HSC1 or a derivative thereof. In some embodiments, the virus is AAV-HSC2 or a derivative thereof. In some embodiments, the virus is AAV-HSC3 or a derivative thereof. In some embodiments, the virus is AAV-HSC4 or a derivative thereof. In some embodiments, the virus is AAV-HSC5 or a derivative thereof. In some embodiments, the virus is AAV-HSC6 or a derivative thereof. In some embodiments, the virus is AAV-HSC7 or a derivative thereof. In some embodiments, the virus is AAV-HSC8 or a derivative thereof. In some embodiments, the virus is AAV-HSC9 or a derivative thereof. In some embodiments, the virus is AAV-HSC10 or a derivative thereof. In some embodiments, the virus is AAV-HSC11 or a derivative thereof. In some embodiments, the virus is AAV-HSC12 or a derivative thereof. In some embodiments, the virus is AAV-HSC13 or a derivative thereof. In some embodiments, the virus is AAV-HSC14 or a derivative thereof. In some embodiments, the virus is AAV-HSC15 or a derivative thereof. In some embodiments, the virus is AAV-TT or a derivative thereof. In some embodiments, the virus is AAV-DJ/8 or a derivative thereof. In some embodiments, the virus is AAV-Myo or a derivative thereof. In some embodiments, the virus is AAV-NP40 or a derivative thereof. In some embodiments, the virus is AAV-NP59 or a derivative thereof. In some embodiments, the virus is AAV-NP22 or a derivative thereof. In some embodiments, the virus is AAV-NP66 or a derivative thereof. In some embodiments, the virus is AAV-HSC16 or a derivative thereof.

[0366] In some embodiments, the virus is HSV-1 or a derivative thereof. In some embodiments, the virus is HSV-2 or a derivative thereof. In some embodiments, the virus is VZV or a derivative thereof. In some embodiments, the virus is EBV or a derivative thereof. In some embodiments, the virus is CMV or a derivative thereof. In some embodiments, the virus is HHV-6 or a derivative thereof. In some embodiments, the virus is HHV-7 or a derivative thereof. In some embodiments, the virus is HHV-8 or a derivative thereof.

[0367] In some embodiments, the nucleic acid encoding the engineered system, the endonuclease, or the engineered guide polynucleotide is delivered by a non-nucleic acid-based delivery system (*e.g.*, a non-viral delivery system). In some embodiments, the non-viral delivery system is a liposome. In some embodiments, the nucleic acid is associated with a lipid. The nucleic acid associated with a lipid, in some embodiments, is encapsulated in the aqueous interior of a liposome, interspersed within the lipid bilayer of a liposome, attached to a liposome via a linking molecule that is associated with both the liposome and the nucleic acid, entrapped in a liposome, complexed with a liposome, dispersed in a solution containing a lipid, mixed with a lipid, combined with a lipid, contained as a suspension in a lipid, contained or complexed with a micelle, or otherwise associated with a lipid. In some embodiments, the nucleic acid is comprised in a lipid nanoparticle (LNP).

[0368] In some embodiments, the engineered system, the endonuclease, or the engineered guide polynucleotide is introduced into the cell in any suitable way, either stably or transiently. In some embodiments, the engineered system, the endonuclease, or the engineered guide polynucleotide is transfected into the cell. In some embodiments, the cell is transduced or transfected with a nucleic acid construct that encodes the engineered system, the endonuclease, or the engineered guide polynucleotide. For example, a cell is transduced (*e.g.*, with a virus encoding the engineered system, the endonuclease, or the engineered guide polynucleotide), or transfected (*e.g.*, with a plasmid encoding the engineered system, the endonuclease, or the engineered guide polynucleotide) with a nucleic acid that encodes the engineered system, the endonuclease, or the engineered guide polynucleotide, or the translated engineered system or endonuclease. In some embodiments, the transduction is a stable or transient transduction. In some embodiments, cells expressing the engineered system, the endonuclease, or the engineered guide polynucleotide are transduced or transfected with one or more gRNA molecules. In some embodiments, a plasmid expressing the engineered system, the endonuclease, or the engineered guide polynucleotide is introduced into cells through electroporation, transient (*e.g.*, lipofection) and stable genome integration (*e.g.*, piggybac) and viral transduction (for example lentivirus or AAV) or other methods known to those of skill in the art. In some embodiments, the engineered system, the endonuclease, or the engineered guide polynucleotide is introduced into the cell as one or more

polypeptides. In some embodiments, delivery is achieved through the use of RNP complexes. Delivery methods to cells for polypeptides and/or RNPs are known in the art, for example by electroporation or by cell squeezing.

[0369] Exemplary methods of delivery of nucleic acids include lipofection, nucleofection, electroporation, stable genome integration (*e.g.*, piggybac), microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in *e.g.*, U.S. Pat. Nos. 5,049,386; 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (*e.g.*, Transfectam™, Lipofectin™ and SF Cell Line 4D-Nucleofector X Kit™ (Lonza)). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of WO 91/17424 and WO 91/16024. In some embodiments, the delivery is to cells (*e.g., in vitro* or *ex vivo* administration) or target tissues (*e.g., in vivo* administration). In some embodiments, the nucleic acid is comprised in a liposome or a nanoparticle that specifically targets a host cell.

[0370] Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US 2003/0087817.

Methods of Use

[0371] Described herein, in certain embodiments, are methods of modifying a target nucleic acid with the base editor or engineered systems described herein, disrupting a gene locus using the base editor or engineered systems described herein, or manufacturing a base editor or engineered systems described herein.

[0372] In some embodiments, the engineered system comprises an adenine deaminase base editor, the nucleotide is an adenine, and modifying the target nucleic acid locus comprises converting the adenine to a guanine. In some embodiments, the engineered system comprises a cytidine deaminase base editor and a uracil DNA glycosylase inhibitor, the nucleotide is a cytosine, and modifying the target nucleic acid locus comprises converting the cytosine to a uracil.

[0373] In some embodiments, the methods are used to introduce a modification in the genome of a cell. In some embodiments, the target nucleic acid is modified *in vitro*. In some embodiments, the target nucleic acid sequence is modified *in vivo*. In some embodiments, the target nucleic acid sequence is modified *ex vivo*.

[0374] In some embodiments, the target nucleic acid comprises genomic DNA, viral DNA, or bacterial DNA. In some embodiments, the target nucleic acid is within a cell. In some embodiments, the cell is a prokaryotic cell, a bacterial cell, a eukaryotic cell, a fungal cell, a plant

cell, an animal cell, a mammalian cell, a rodent cell, a primate cell, or a human cell. In some embodiments, the cell is within an animal.

[0375] In some embodiments, the target nucleic acid comprises DNA. In some embodiments, the DNA comprises a first strand comprising a sequence complementary to a sequence of the engineered guide polynucleotide and a second strand comprising a PAM. In some embodiments, the PAM is directly adjacent to the 3' end of the sequence complementary to the sequence of the engineered guide polynucleotide. In some embodiments, the PAM comprises a sequence selected from the group consisting of SEQ ID NOs: 360-368 or 598.

[0376] In some embodiments, the present disclosure provides a method of modifying a target nucleic acid (e.g., gene) locus. In some embodiments, the method comprises delivering to the target nucleic acid locus the engineered system described herein. In some embodiments, the endonuclease is configured to form a complex with the engineered guide polynucleotide. In some embodiments, the complex is configured such that upon binding of the complex to the target nucleic acid locus, the complex modifies the target nucleic acid locus.

[0377] In some embodiments, delivery of the engineered system to the target nucleic acid locus comprises delivering the nucleic acid described herein or the vector described herein. In some embodiments, delivery of the engineered system to the target nucleic acid locus comprises delivering a nucleic acid comprising an open reading frame encoding the base editor and the endonuclease. In some embodiments, the nucleic acid comprises a promoter. In some embodiments, the open reading frame encoding the base editor and the endonuclease is operably linked to the promoter.

[0378] In some embodiments, delivery of the engineered system to the target nucleic acid locus comprises delivering a capped mRNA containing the open reading frame encoding the base editor and the endonuclease. In some embodiments, delivery of the engineered system to the target nucleic acid locus comprises delivering a translated polypeptide. In some embodiments, delivery of the engineered system to the target nucleic acid locus comprises delivering a deoxyribonucleic acid (DNA) encoding the engineered guide RNA operably linked to a ribonucleic acid (RNA) pol III promoter.

[0379] In some embodiments, the target gene is TRAC. In some embodiments, the gRNA comprises a sequence having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the

gRNA comprises a sequence having at least about 70% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 75% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 80% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 85% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 90% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 91% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 92% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 93% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 94% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 95% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 96% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 97% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 98% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having at least about 99% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490. In some embodiments, the gRNA comprises a sequence having 100% identity to SEQ ID NO: 1489 or SEQ ID NO: 1490.

[0380] In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 70% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 75% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 80% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 85% identity to any one of SEQ ID NOs: 1491-1492 and 1711-

1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 90% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 91% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 92% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 93% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 94% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 95% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 96% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 97% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 98% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having at least about 99% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719. In some embodiments, the gRNA hybridizes to a TRAC sequence having 100% identity to any one of SEQ ID NOs: 1491-1492 and 1711-1719.

[0381] In some embodiments, the target gene is AAVS1. In some embodiments, the gRNA comprises a sequence having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 70% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 75% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 80% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 85% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 90% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 91% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 92% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a

sequence having at least about 93% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 94% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 95% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 96% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 97% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 98% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having at least about 99% identity to SEQ ID NOs: 1705-1710. In some embodiments, the gRNA comprises a sequence having 100% identity to SEQ ID NOs: 1705-1710.

[0382] In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 70% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 75% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 80% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 85% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 90% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 91% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 92% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 93% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 94% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 95% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 96% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 97% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at

least about 98% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having at least about 99% identity to any one of SEQ ID NOs: 1720-1725. In some embodiments, the gRNA hybridizes to a AAVS1 sequence having 100% identity to any one of SEQ ID NOs: 1720-1725.

[0383] In some embodiments, the target gene is hApoA1. In some embodiments, the gRNA comprises a sequence having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 70% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 75% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 80% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 85% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 90% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 91% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 92% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 93% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 94% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 95% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 96% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 97% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 98% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having at least about 99% identity to SEQ ID NOs: 1704. In some embodiments, the gRNA comprises a sequence having 100% identity to SEQ ID NOs: 1704.

[0384] In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID

NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 70% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 75% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 80% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 85% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 90% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 91% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 92% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 93% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 94% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 95% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 96% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 97% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 98% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having at least about 99% identity to any one of SEQ ID NOs: 1726. In some embodiments, the gRNA hybridizes to a hApoA1 sequence having 100% identity to any one of SEQ ID NOs: 1726.

[0385] Described herein, in certain embodiments, are methods of modifying a nucleic acid encoding ANGPTL3 comprising contacting the nucleic acid sequence encoding ANGPTL3 with an engineered base editing system, said base editing system comprising:a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415, or encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757; and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1479-1483 and 1758-1889.

[0386] Described herein, in certain embodiments, are methods of modifying a nucleic acid encoding APOA1 comprising contacting the nucleic acid sequence encoding APOA1 with an engineered base editing system, said base editing system comprising: a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415, or encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757; and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1431-1454, 1704, and 2010-2019.

[0387] Described herein, in certain embodiments, are methods of modifying a nucleic acid encoding BCL11A comprising contacting the nucleic acid sequence encoding BCL11A with an engineered base editing system, said base editing system comprising a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415, or encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757, and an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence. In some embodiments, the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1890-1976.

[0388] In some embodiments, the endonuclease induces a single-stranded break or a double-stranded break at or proximal to the target locus. In some embodiments, the endonuclease induces a staggered single stranded break within or 5' to said target locus. In some embodiments, the endonuclease does not induce a break at or proximal to the target locus.

[0389] In some embodiments, the present disclosure provides methods of manufacturing or producing a base editor. In some embodiments, the method comprises cultivating the cell. In some embodiments, the methods of producing a base editor, comprising cultivating the host cell described herein in compatible growth medium. In some embodiments, the methods further comprise inducing expression of said base editor by addition of an additional chemical agent or an increased amount of a nutrient. In some embodiments, the chemical agent is Isopropyl β-D-1-thiogalactopyranoside (IPTG). In some embodiments, the nutrient is lactose. In some

embodiments, the methods further comprise isolating said host cell after said cultivation and lysing said host cell to produce a protein extract. In some embodiments, the methods further comprise subjecting said protein extract to IMAC, or ion-affinity chromatography. In some embodiments, the methods further comprise cleaving said IMAC affinity tag by contacting a protease corresponding to said protease cleavage site to said base editor. In some embodiments, the methods further comprise performing subtractive IMAC affinity chromatography to remove said affinity tag from a composition comprising said base editor.

[0390] Systems of the present disclosure may be used for various applications, such as, for example, nucleic acid editing (*e.g.*, gene editing), binding to a nucleic acid molecule (*e.g.*, sequence-specific binding). Such systems may be used, for example, for addressing (*e.g.*, removing or replacing) a genetically inherited mutation that may cause a disease in a subject, inactivating a gene in order to ascertain its function in a cell, as a diagnostic tool to detect disease-causing genetic elements (*e.g.*, via cleavage of reverse-transcribed viral RNA or an amplified DNA sequence encoding a disease-causing mutation), as deactivated enzymes in combination with a probe to target and detect a specific nucleotide sequence (*e.g.*, sequence encoding antibiotic resistance int bacteria), to render viruses inactive or incapable of infecting host cells by targeting viral genomes, to add genes or amend metabolic pathways to engineer organisms to produce valuable small molecules, macromolecules, or secondary metabolites, to establish a gene drive element for evolutionary selection, to detect cell perturbations by foreign small molecules and nucleotides as a biosensor.

Kits

[0391] In some embodiments, this disclosure provides kits comprising one or more nucleic acid constructs encoding the various components of the engineered system described herein, *e.g.*, comprising a nucleotide sequence encoding the components of the engineered editing system capable of modifying a target DNA sequence. In some embodiments, the nucleotide sequence comprises a heterologous promoter that drives expression of the engineered system components.

[0392] In some embodiments, any of the engineered editing systems disclosed herein is assembled into a pharmaceutical, diagnostic, or research kit to facilitate its use in therapeutic, diagnostic, or research applications. A kit may include one or more containers housing any of the vectors disclosed herein and instructions for use.

[0393] The kit may be designed to facilitate use of the methods described herein by researchers and can take many forms. Each of the compositions of the kit, where applicable, may be provided in liquid form (*e.g.*, in solution), or in solid form, (*e.g.*, a dry powder). In certain cases, some of the compositions may be constitutable or otherwise processable (*e.g.*, to an active form), for

example, by the addition of a suitable solvent or other species (for example, water or a cell culture medium), which may or may not be provided with the kit. As used herein, "instructions" can define a component of instruction and/or promotion, and typically involve written instructions on or associated with packaging of the disclosure. Instructions also can include any oral or electronic instructions provided in any manner such that a user will clearly recognize that the instructions are to be associated with the kit, for example, audiovisual (*e.g.*, videotape, DVD, etc.), Internet, and/or web-based communications, etc. The written instructions, in some embodiments, are in a form prescribed by a governmental agency regulating the manufacture, use, or sale of pharmaceuticals or biological products, which instructions can also reflect approval by the agency of manufacture, use, or sale for animal administration.

EXAMPLES

[0394] The following examples are given for the purpose of illustrating various embodiments of the disclosure and are not meant to limit the present disclosure in any fashion. The present examples, along with the methods described herein, are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the disclosure. Changes therein and other uses which are encompassed within the spirit of the disclosure as defined by the scope of the claims will occur to those skilled in the art.

Example 1 – Base editor ABE07 targets APOA1 and ANGPTL3 genes in primary mouse hepatocytes across a range of doses

[0395] Hypercholesterolemia is a metabolic disorder characterized by elevated blood plasma levels of low-density lipoprotein (LDL) which can lead to atherosclerosis, heart attack, and stroke. Proteins that regulate plasma lipoprotein levels are primarily expressed in hepatocytes in the liver and are encoded by the APOA1 and ANGPTL3 genes. Knockdown of these genes can enable treatment of human lipoprotein metabolism disorders such as hypercholesterolemia and can be achieved by precise introduction of mutations in their coding sequence through base editing.

[0396] In this example, eleven guides targeting mouse APOA1 gene and one guide targeting the ANGPTL3 gene (**Table 3**) were co-transfected with mRNA of ABE07 (SEQ ID NO: 1411) via lipofection in primary mouse hepatocytes. A range of mRNA doses were administered and resulted in A-to-G conversion in eleven guides across a range of four mRNA doses (**FIGs. 1A-1K**).

[0397] *Primary cell culture, transfections, next generation sequencing, and base editing activity analysis*

[0398] Primary mouse hepatocytes were plated at 100,000 viable cells per well in Collagen I pre-coated plates. After 24-hours overnight incubation in Hepatocyte Basal Media at 37 °C and 5% CO2, ABE07 mRNA (0.21, 0.42, 0.625, 1.25 μg) was co-transfected with a 1:20 molar ratio of mRNA:guide RNA via lipofection. At three days post-transfection, genomic DNA was harvested from cells. PCR primers appropriate for use in NGS-based DNA sequencing were generated, optimized, and used to amplify the individual target sequences for each guide RNA. The amplicons were sequenced and analyzed to measure percent A-to-G conversion at relevant spacer positions for each guide.

Table 3: sgRNAs and targeting spacer sequences used in primary mouse hepatocytes.

Target Site Name	Spacer DNA sequence	sgRNA sequence
Apoa1 A4	TCTCCTGGAAAACTGGGAC ACT	mU*mC*mU*rCrCrUrGrGrArArArArCrUrGrGrGrArCrArC rUrGrUrUrGrArGrArArArUrCrGrArArArGrArUrUrCrUrUrAr ArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCr UrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGr GrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 A5	CTGGGATAACCTGGAGAAA GAA	mC*mU*mG*rGrGrArUrArArCrCrUrGrGrArGrArArArGrA rArGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrAr ArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCr UrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGr GrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 A8	CCTACCTTGAACGAGTACC ACA	mC*mC*mU*rArCrCrUrUrGrArArCrGrArGrUrArCrCrArCr ArGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrArA rUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrU rUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrG rArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 C5	TTGGGTGAGACAGGAGATG AAC	mU*mU*mG*rGrGrUrGrArGrArCrArGrGrArGrArUrGrArA rCrGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrAr ArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCr UrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGr GrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 D7	CAGCGAACAGATGCGCGAG AGC	mC*mA*mG*rCrGrArArCrArGrArUrGrCrGrCrGrArGrArG rCrGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrAr ArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCr UrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGr GrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 D11	ACTATGGCGCAGGTCCTCC AGC	mA*mC*mU*rArUrGrGrCrGrCrArGrGrUrCrCrUrCrCrArGr CrGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrArA rUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrU rUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrG rArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 E1	AGCAAGATGAACCCCAGTC CCA	mA*mG*mC*rArArGrArUrGrArArCrCrCrCrArGrUrCrCrCr ArGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrArA rUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrU rUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrG rArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 F4	AGGAACGGCTGGGCCCATT GAC	mA*mG*mG*rArArCrGrGrCrUrGrGrGrCrCrCrArUrUrGrA rCrGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrAr ArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCr UrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGr GrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 F12	CTGGTGTGGTACTCGTTCAA GG	mC*mU*mG*rGrUrGrUrGrGrUrArCrUrCrGrUrUrCrArArG rGrGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrAr ArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCr UrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGr GrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU

Target Site Name	Spacer DNA sequence	sgRNA sequence
Apoa1 G4	GGAACGGCTGGGCCCATTG ACT	mG*mG*mA*rArCrGrGrCrUrGrGrGrCrCrCrArUrUrGrArC rUrGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrAr ArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCr UrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGr GrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 G6	AGGGAGACTGTCCCCTGTG GCT	mA*mG*mG*rGrArGrArCrUrGrUrCrCrCrCrUrGrUrGrGrC rUrGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrAr ArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCr UrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGr GrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Angptl3 C12	ACTATTAAACCAAGAAACT CCC	mA*mC*mU*rArUrUrArArArCrCrArArGrArArArCrUrCrC rCrGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrAr ArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCr UrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGr GrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
(r =native ribose base, m = 2'-O methyl modified base, * = phosphorothioate bond)		

[0399] *Results*

[0400] NGS analysis of genomic DNA harvested from primary mouse hepatocytes at three days post-transfection revealed base editing activity in eleven APOA1 guides (A5, A8, C5, D7, D11, E1, F4, and F12) and one ANGPTL3 guide (C12). Maximum A-to-G conversion of 70.3% was achieved with APOA1 guide F12 transfected with the lowest mRNA dose of 0.21 μg ABE07 at spacer position A11 (**FIGs. 1A-1K**). The second lowest mRNA dose of 0.42 μg resulted in maximum A-to-G conversion of 64.4%, also with APOA1 guide F12 at spacer position A11. The highest two mRNA doses of 0.625 and 1.25 μg resulted in a maximum of 55.0% and 35.9% A-to-G conversion, respectively, in APOA1 guide F4 at spacer position A5.

[0401] Maximum A-to-G conversion was often observed toward the middle of the mRNA dose range within a given spacer position, resulting in an inverted U-shaped dose-effect curve. This shape is well-demonstrated in spacer position A7 of the APOA1 G6 guide results graph (**FIGs. 1A-1K**), where mRNA doses 0.21, 0.42, 0.625, and 1.25 μg result in A-to-G conversion of 47.4%, 54.6%, 54.4%, and 35.9%, respectively. Capturing the range of base editing activity within a spacer position with relatively small deviations between data points at a given dose suggest a well-characterized and robust base editing tool ready for downstream optimization and in vivo delivery.

Example 2 – Engineering and optimizing ABE variants through extensive guide screening in Hepa1-6 cells

[0402] Mutations V83S, T112R, H129N, and A155R, along with the D109N mutation, were inserted into 3-68_DIV30_HT/HM nickase chassis (where 3-68, DIV30, and HT/HM, stand for MG3-6/3-8 nickase, domain inlaid version 30, and heterodimer/homodimer, respectively) to generate new dimeric ABE variants **(SEQ ID NO: 1654-1658)**. In the homodimer configuration, two copies of the mutant were fused with a linker while in the heterodimer, the wild type deaminase was fused via linker to the beneficial mutant **(FIG. 2)**. Further details of these variants are summarized in **Table 4**.

[0403] In this example, these ABE variants were screened alongside the homodimeric D109N variant, ABE07 **(SEQ ID NO: 1411)** and the heterodimeric D109N variant, ABE07-74 **(SEQ ID NO: 1654)** over 31 pre-characterized genetic loci in the Hepa1-6 cell line **(Table 5; SEQ ID NO: 1455-1478 and 1484-1488)**.

[0404] *mRNA production, cell culture, transfections, next generation sequencing, and base editing analyses for ABE screens*

[0405] The mRNA corresponding to ABE variants listed in **Table 4** were produced. Following mRNA production, these ABE variants were nucleofected into Hepa1-6 cells along with chemically-synthesized sgRNA targeting the loci listed in **Table 5**. Amplicons were sequenced and analyzed to measure gene editing.

Table 4: Design of ABE variants tested in Example 2.

ABE variant	Architecture	Mutations in deaminase domain
ABE07	Homodimeric	MG68-4(D109N)—MG68-4(D109N)
ABE07-74	Heterodimeric	MG68-4(wt)—MG68-4(D109N)
ABE07-75	Heterodimeric	MG68-4(wt)—MG68-4(V83S+D109N+A155R)
ABE07-76	Heterodimeric	MG68-4(wt)—MG68-4(D109N+T112R+H129N)
ABE07-77	Heterodimeric	MG68-4(wt)—MG68-4(D109N+T112R+A155R)
ABE07-78	Heterodimeric	MG68-4(wt)—MG68-4(V83S/D109N/T112R/H129N)

[0406] *Results*

[0407] The editing across all spacers and all ABE variants averaged at 11.19%, with the ABE07 homodimeric construct outperforming its heterodimeric homolog ABE07-74 **(FIG. 3A, Table 5)**. However, the incorporation of additional mutations in ABE07-75, ABE07-76, ABE07-77, and ABE07-78 allowed these heterodimeric variants to perform on par with the less evolved homodimeric ABE07 **(FIG. 3A, Table 6)**. Specifically, upon comparing the maximum observed

A-to-G base conversion at the different genomic targets, the ABE07-77 (36.13% across all 31 tested guides) and ABE07-78 (36.66% across all 31 tested guides) demonstrated equivalent editing efficiencies to the ABE07 variant (36.88% across all 31 tested guides; **FIG. 3B, Table 6**). Additionally, these variants performed slightly better than ABE07 when comparing their mean editing efficiencies across the entire guide panel: ABE07 (10.36% across all 31 tested guides) versus ABE07-77 (13.07% across all 31 tested guides) and ABE07-78 (14.22% across all 31 tested guides; **FIG. 3A, Table 6**).

Table 5: Mean A-to-G Editing activity at 31 genomic loci targeted in the Hepa1-6 screening experiment.

Guide name	Spacer	ABE07	ABE07-74	ABE07-75	ABE07-76	ABE07-77	ABE07-78
mApoa1 BE F12	CTGGTGTG GTACTCGT TCAAGG	31.57±0.05	28.12±0.67	27.17±0.45	29.45±0.33	-	27.58±0.44
mApoa1 BE D11	ACTATGGC GCAGGTC CTCCAGC	25.96±0.87	21.79±0.31	25.96±0.91	25.2±0.61	31.21±1.9	27.4±1.98
mApoa1 BE C5	TTGGGTGA GACAGGA GATGAAC	14.82±0.27	12.71±0.59	24.72±0.55	21.15±0.21	25.8±0.93	27.47±0.87
mApoa1 BE A4	TCTCCTGG AAAACTG GGACACT	30.04±0.4	25.81±1.79	37.1±0.73	36.26±1.12	47.67±2.07	40.22±1.09
mApoa1 BE F4	AGGAACG GCTGGGC CCATTGAC	19.03±0.63	14.83±1.05	18.08±0.1	20.75±0.15	24.1±3.13	17.54±0.67
mApoa1 BE A5	CTGGGAT AACCTGG AGAAAGA A	15.74±0.34	11.38±0.48	18.01±0.51	15.37±0.14	20.12±1.9	19.26±1.65
mApoa1 BE E12	CCTGGTGT GGTACTCG TTCAAG	12.9±0.93	6.72±0.25	16.98±0.61	14.81±0.25	18.42±3.09	19.97±2.21
mApoa1 BE A11	AGCATGG GCATCAG ACTATGGC	6.27±8.83	8.56±0.42	18.51±0.3	12.27±0.36	21.42±2.57	20.17±2.46
mApoa1 BE B4	CTCCTGGA AAACTGG GACACTC	17.69±0.27	3.73±0.27	14.79±0.27	14.93±0.12	16.1±3.56	20.13±0.87
mApoa1 BE G4	GGAACGG CTGGGCCC ATTGACT	12.78±1.11	6.07±0.22	12.44±0.39	10.39±0.21	13.32±2.4	11.49±1.08
mApoa1 BE B2	GCCACAG GGGACAG TCTCCCTT	19.6±0.6	10.15±0	18.7±22.18	27.97±0.75	37.18±4.98	41.09±2.44

Guide name	Spacer	ABE07	ABE07-74	ABE07-75	ABE07-76	ABE07-77	ABE07-78
mApoa1 BE D7	CAGCGAACAGATGCGCGAGAGC	9.4±0.05	2.24±2.38	10.52±1.91	10.01±0.01	10.28±3.52	11.86±1
mApoa1 BE B5	ATTGGGTGAGACAGGAGATGAA	7.69±0.36	2.93±0	8.01±0.07	8.7±0.87	9.29±1.35	10.95±1.37
mApoa1 BE G6	AGGGAGACTGTCCCCTGTGGCT	18.12±1.22	4.11±5.81	11.69±1.36	14.77±0.15	14.03±3.1	14.33±1.02
mApoa1 BE A8	CCTACCTTGAACGAGTACCACA	12.69±0.13	5.5±0.35	13.12±0.44	11.37±0.5	12.33±1.48	14.31±0.87
mApoa1 BE F2	GGCCCAAGGAGGAGGATTCAAA	3.82±5.38	2.77±0.41	8.38±0.18	7.73±0.37	7.57±2.51	9.38±1.01
mApoa1 BE E1	AGCAAGATGAACCCCAGTCCCA	9.38±0.25	2.71±0.25	9.28±0.22	10.18±1.26	8.53±2.67	11.09±0.16
mApoa1 BE B8	CTACCTTGAACGAGTACCACAC	12.99±0.79	4.48±0.15	14.55±0.23	11.73±0.02	14.9±2.67	18.42±1.3
mApoa1 BE H8	CATGCTGGAGACGCTTAAGACC	4.69±0.22	1.65±0.16	7.26±0.14	6.44±0.52	8.3±1.65	11.07±1.19
mApoa1 BE H6	TCGCGACCGCATGCGCACACAC	4.62±0.44	0.7±0.34	1.24±0.04	2.19±0.01	1.66±0.65	1.37±0.01
mApoa1 BE F5	ACGAATTCCAGAAGAAATGGAA	3.49±0.28	0.88±0.01	4.56±0.59	4.46±0.04	6.22±1.33	7.28±0.65
mApoa1 BE H3	CTAGCCTGAATCTCCTGGAAAA	6.27±0.37	0.88±0.11	4.62±0.35	4.32±0.07	3.95±0.65	4.58±1.37
mApoa1 BE H4	TGGGCCCATTGACTCGGGACTT	7.33±0.28	1.64±0.58	6.74±0.47	6.99±0.06	8.45±2.24	9.44±1.06
mApoa1 BE F8	CGAGAAAGCCAGACCTGCGCTG	2.94±4.13	1.57±0.02	14.3±0.71	3.87±4.96	11.6±2.7	17.15±0.92
mAngptl3 BE C12	ACTATTAAACCAAGAAACTCCC	0.03±0.01	0.01±0.01	0.03±0.01	0.02±0	0.03±0.02	0.05±0.05
mAngptl3 BE B2	CGAAACATGGGAAAACTACGAA	1.31±0.13	0.42±0.05	2.36±0.17	1.76±0.31	2.79±0.48	3.39±0.27

Guide name	Spacer	ABE07	ABE07-74	ABE07-75	ABE07-76	ABE07-77	ABE07-78
mAngptl3 BE C1	AGTAATTG CATCCAG AGTGGAT	0.02±0	0.01±0	0.02±0.02	0.01±0	0.01±0	0.02±0
mAngptl3 BE F3	AAGAGAA GACAGCC CTTCAACA	1.48±0.14	0.58±0.13	4.08±0.7	2.17±0.52	3.84±1.26	6.52±0.35
mAngptl3 BE G1	TTTAGCGA ATGGCCTC CTGCAG	2.39±0.2	0.88±0.04	8.35±0.63	2.33±0.77	5.45±1.61	9.81±1.98
mTrac BE E1	ACCAGTTA AAAGATC CTCGGTC	2.3±0.38	0.8±0.1	2.65±0.57	1.97±0.37	2.39±0.76	3.14±0.46
mTrac BE D10	TTCACAAT CCCACCTG GATCTC	3.94±1.39	1.11±0.07	4.69±0.21	3.34±0.98	5.1±2.1	4.27±0.12

Table 6: Max A-to-G Editing activity at 31 genomic loci targeted in the Hepa1-6 screening experiment.

Guide name	Spacer	ABE07	ABE07-74	ABE07-75	ABE07-76	ABE07-77	ABE07-78
mApoa1 BE F12	CTGGTGTG GTACTCGT TCAAGG	94.69±0.15	84.32±1.99	81.49±1.35	88.33±0.97	82.71±1.34	94.69±0.15
mApoa1 BE D11	ACTATGGC GCAGGTC CTCCAGC	87.37±3.37	75.03±1.6	73.8±2.28	75.85±1.57	89.27±3.54	70.7±4.62
mApoa1 BE C5	TTGGGTGA GACAGGA GATGAAC	77.74±2.36	67.88±3.68	87.27±1.02	88.63±0.87	94.59±1.46	90.32±1.39
mApoa1 BE A4	TCTCCTGG AAAACTG GGACACT	81.75±0.19	72.88±3.16	67.12±0.71	76.75±2.18	84.54±2.25	70.66±1.17
mApoa1 BE F4	AGGAACG GCTGGGC CCATTGAC	71.74±2.16	56.04±3.31	51.94±0.64	67.54±0.88	70.33±8.02	51.16±1.67
mApoa1 BE A5	CTGGGAT AACCTGG AGAAAGA A	76.64±0.64	57.38±3.6	65.01±2.2	68.77±0.66	77.67±6.58	69.05±5.55
mApoa1 BE E12	CCTGGTGT GGTACTCG TTCAAG	38.69±2.78	20.13±0.73	50.93±1.82	44.39±0.76	55.24±9.3	59.88±6.65
mApoa1 BE A11	AGCATGG GCATCAG ACTATGGC	25.26±35.66	36.16±1.7	54.08±0.63	43.63±1.73	67.11±6.65	52.07±5.79
mApoa1 BE B4	CTCCTGGA AAACTGG	60.18±1	13.73±1.64	27.88±0.57	36.67±0.81	32.55±6.94	36.68±1.38

Guide name	Spacer	ABE07	ABE07-74	ABE07-75	ABE07-76	ABE07-77	ABE07-78
	GACACTC						
mApoa1 BE G4	GGAACGG CTGGGCCC ATTGACT	47.86±4.31	22.78±0.86	34.8±1.31	35.51±1.04	40.1±6.53	33.07±3.59
mApoa1 BE B2	GCCACAG GGGACAG TCTCCCTT	42.97±1.66	22.66±0	28.92±33.48	52.34±0.54	61.13±7.67	64.64±3.86
mApoa1 BE D7	CAGCGAA CAGATGC GCGAGAG C	44.66±0.04	10.41±10.43	32.64±3.76	35.92±0.26	31.34±10.58	30.35±2.33
mApoa1 BE B5	ATTGGGTG AGACAGG AGATGAA	50.88±2.18	19.64±0.16	40.8±0.4	51.68±5.17	49.92±6.53	52.43±6.68
mApoa1 BE G6	AGGGAGA CTGTCCCC TGTGGCT	53.08±3.26	12±16.96	32.6±3.17	42.19±0.41	38.71±8.64	38.46±2.85
mApoa1 BE A8	CCTACCTT GAACGAG TACCACA	52.14±1.41	22.63±0.92	43.43±1.15	35.95±1.48	38.53±4.7	39.1±2.26
mApoa1 BE F2	GGCCCAA GGAGGAG GATTCAA A	15.66±22.12	11.93±1.25	27.9±0.65	27.57±0.95	23.98±7.81	27.49±3.4
mApoa1 BE E1	AGCAAGA TGAACCCC AGTCCCA	30.68±0.35	8.73±0.78	20.67±0.45	25.82±3.03	19.33±5.69	23.85±0.42
mApoa1 BE B8	CTACCTTG AACGAGT ACCACAC	43.14±2.5	14.59±0.17	38.92±0.37	36.79±0.02	40.01±7.03	45.34±3.15
mApoa1 BE H8	CATGCTGG AGACGCTT AAGACC	22.07±0.76	7.72±0.77	28.16±0.39	28.31±2.44	33.8±6.63	41.46±4.21
mApoa1 BE H6	TCGCGACC GCATGCG CACACAC	21.4±1.9	3.26±1.42	4.03±0.01	9.5±0.11	6.45±2.51	4.31±0.02
mApoa1 BE F5	ACGAATTC CAGAAGA AATGGAA	21.03±1.75	5.28±0.14	21.92±2.24	18.92±0.39	25.41±5.42	26.7±2.53
mApoa1 BE H3	CTAGCCTG AATCTCCT GGAAAA	24.22±2.3	3.34±0.81	15.05±0.97	14.92±0.64	13.48±2.49	14.6±4.1
mApoa1 BE H4	TGGGCCC ATTGACTC GGGACTT	15.05±0.98	3.35±1.27	11.21±0.81	13.38±0.21	15.22±3.87	15.9±1.77
mApoa1 BE E8	CGAGAAA GCCAGAC CTGCGCTG	6.35±8.92	3.38±0.02	21.17±1.27	7.2±9.23	18.67±4.37	27.05±1.77

Guide name	Spacer	ABE07	ABE07-74	ABE07-75	ABE07-76	ABE07-77	ABE07-78
mAngptl3 BE C12	ACTATTAA ACCAAGA AACTCCC	0.12±0.04	0.03±0.01	0.08±0.04	0.06±0.01	0.13±0.09	0.15±0.11
mAngptl3 BE B2	CGAAACA TGGGAAA ACTACGA A	10.48±1.01	3.4±0.33	12.6±0.66	12.32±2.1	15.67±3.05	14.61±1.5
mAngptl3 BE C1	AGTAATTG CATCCAG AGTGGAT	0.05±0.01	0.02±0.01	0.06±0.06	0.02±0	0.04±0.01	0.04±0
mAngptl3 BE F3	AAGAGAA GACAGCC CTTCAACA	9.98±0.94	3.63±0.81	19.47±3.1	12.67±3.35	19.79±6.03	29.1±2.23
mAngptl3 BE G1	TTTAGCGA ATGGCCTC CTGCAG	4.89±0.59	1.78±0.12	15.98±1.49	3.98±1.46	9.61±2.76	14.55±3.06
mTrac BE E1	ACCAGTTA AAAGATC CTCGGTC	5.8±1.08	1.85±0.36	4.42±0.82	3.66±0.74	4.01±1.24	4.35±0.77
mTrac BE D10	TTCACAAT CCCACCTG GATCTC	6.99±2.74	1.93±0.07	7.51±0.59	5.47±1.8	7.29±3.08	5.97±0.23

[0408] When comparing the top performing heterodimeric ABE variants, it was determined that ABE07-78 introduced C-to-G edits across the different guides that were tested. This promiscuous editing by the ABE07-78 variant averaged at 5% and ranged up to 27.78% at one of the sites (**FIGs. 4A-4B**). Additionally, ABE07-78 also introduced indels (insertions and deletions) at a rate much higher (up to 7% at certain sites) than the other ABE variants (**FIG. 5**).

[0409] Thus, based on its high on-target A-to-G editing efficiency (**FIG. 3A – 3B**) and favorable off-target editing activity (**FIGs. 4A-4B**, and **FIG. 5**) the ABE07-77 variant, with the combination of D109N+T112R+A155R mutations in a heterodimeric architecture showed the highest levels of efficiency and specificity relative to other tested variants.

Example 3 – *In vivo* gene editing in liver of mice by the ABE07-77 delivered by systemic administration of lipid nanoparticles

[0410] ABE07-77 was subjected to further testing *in vivo* through the lipid nanoparticle delivery of an mRNA encoding the protein and sgRNA that target the highly edited loci identified in the Hepa1-6 screening experiment and primary mouse hepatocytes (**Table 7**). To inform on any undesirable indel formation by ABE07-77, its parent nuclease, that is MG3-6/3-8, was also included in this study as a test control. Additionally, dose-escalation experiments were performed

to assess the dose-dependence of editing activity of ABE07-77. Lastly, a number of chemical modifications of the native RNA structure were incorporated into these sgRNAs (**Table 7**). These chemical modifications were selected based on their ability to improve the stability of the sgRNA in vitro when incubated in extracts from mammalian cells without negatively impacting editing activity.

[0411] *Preparation of mRNA*

[0412] The mRNA encoding the ABE07-77 was generated by *in vitro* transcription of a linearized plasmid template using T7 RNA polymerase, nucleotides, and enzymes. The DNA sequence that was transcribed into RNA comprised the following elements in order from 5′ to 3′: the T7 RNA polymerase promoter, a 5′ untranslated region (5′ UTR), a nuclear localization signal, a short linker, the coding sequence for the ABE07-77, a short linker, a nuclear localization signal, a 3′ untranslated region, and an approximately 100 nucleotide polyA tail.

[0413] The protein sequence encoded in the synthetic mRNA encoded in this ABE07-77 cassette comprised the following elements from 5′ to 3′: the nuclear localization signal from SV40, a five amino acid linker (GGGGS), the protein coding sequence of the ABE07-77 from which the initiating methionine codon was removed, a 3 amino acid linker (SGG), and the nuclear localization signal from nucleoplasmin. The DNA sequence of the protein coding region of this cassette was modified to reflect the codon usage in humans using a commercially available algorithm. An approximately 100-nucleotide polyA tail was encoded in the plasmid used for in vitro transcription and the mRNA was co-transcriptionally capped. Uridine in the mRNA was replaced with N1-methyl pseudouridine. A similar procedure was performed to prepare the mRNA encoding the MG3-6/3-8 nuclease.

[0414] *Preparation of lipid nanoparticles*

[0415] The lipid nanoparticle (LNP) formulation used to deliver the ABE07-77 mRNA and the guide RNA is based on LNP formulations described in the literature including Kauffman et al. (Nano Lett. 2015, 15, 11, 7300–7306). The four lipid components were dissolved in ethanol and mixed in an appropriate molar ratio to make the lipid working mix. The mRNA and the guide RNA were mixed prior to formulation at a 1:1 mass ratio. RNA was diluted in 100 mM Sodium Acetate (pH 4.0) to make the RNA working stock. The lipid working stock and the RNA working stock were mixed in a microfluidics device at a flow rate ratio of 1:3 and a flow rate of 12 mls/min. The LNP were dialyzed against phosphate buffered saline (PBS) for 2 hours and then concentrated until the reduced volume was achieved. The concentration of RNA in the LNP formulation was measured using the Ribogreen reagent. The diameter and polydispersity (PDI) of the LNP were determined by dynamic light scattering. Representative LNP diameters ranged from 65 nm to 120 nm with PDI of 0.05 to 0.20.

[0416] *Mouse dosing and harvesting*

[0417] LNP for mRNA and sgRNA were mixed at 1:1 mass ratio and injected intravenously into 7-week-old C57Bl6 wild type mice via the tail vein (0.1 mL per mouse) at a total RNA dose of either 1.5 mg or 1 mg RNA per kg body weight (**Table 7**). Seven days post-dosing, all mice in each group were sacrificed. The left liver lobe was collected and flash frozen.

Table 7: sgRNAs and targeting spacer sequences used in mouse study.

Target Site Name	Number of Test Animals	Dosage (mg/kg)	Spacer DNA sequence	sgRNA sequence
mApoa1 A4	4	1.5	TCTCCTGG AAAACTGG GACACT	mU*mC*mU*rCrCrUrGrGrArArArArCrUrGrGrGrA rCrArCrUrGrUrUrGrArGrArArUrCrGrArArArGrAr UrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrC rGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCr CrGrUrUrUrUrCrCrArArUrArGrGrArGrCrGrGrGrC rGrGrUrArUrGrU*mU*mU*mU
mApoa1 A5	4	1.5	CTGGGATA ACCTGGAG AAAGAA	mC*mU*mG*rGrGrArUrArArCrCrUrGrGrArGrAr ArArGrArArGrUrUrGrArGrArArUrCrGrArArArGr ArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrC rCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUr CrCrGrUrUrUrUrCrCrArArUrArGrGrArGrCrGrGrG rCrGrGrUrArUrGrU*mU*mU*mU
mApoa1 C5	4	1.5	TTGGGTGA GACAGGAG ATGAAC	mU*mU*mG*rGrGrUrGrArGrArCrArGrGrArGrArGr UrGrArArCrGrUrUrGrArGrArArUrCrGrArArArGr ArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrC rCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUr CrCrGrUrUrUrUrCrCrArArUrArGrGrArGrCrGrGrG rCrGrGrUrArUrGrU*mU*mU*mU
mApoa1 F4	4	1.5	AGGAACGG CTGGGCCC ATTGAC	mA*mG*mG*rArArCrGrGrCrUrGrGrGrCrCrCrArU rUrGrArCrGrUrUrGrArGrArArUrCrGrArArArGrAr UrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrC rGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCr CrGrUrUrUrUrCrCrArArUrArGrGrArGrCrGrGrGrC rGrGrUrArUrGrU*mU*mU*mU
mApoa1 G4	4	1.5	GGAACGGC TGGGCCCA TTGACT	mG*mG*mA*rArCrGrGrCrUrGrGrGrCrCrCrArUrU rGrArCrUrGrUrUrGrArGrArArUrCrGrArArArGrAr UrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrC rGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCr CrGrUrUrUrUrCrCrArArUrArGrGrArGrCrGrGrGrC rGrGrUrArUrGrU*mU*mU*mU
mAngptl3 C12	4	1.5	ACTATTAA ACCAAGAA ACTCCC	mA*mC*mU*rArUrUrArArArCrCrArArGrArArArC rUrCrCrCrGrUrUrGrArGrArArUrCrGrArArArGrAr UrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrC rGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCr CrGrUrUrUrUrCrCrArArUrArGrGrArGrCrGrGrGrC rGrGrUrArUrGrU*mU*mU*mU

Target Site Name	Number of Test Animals	Dosage (mg/kg)	Spacer DNA sequence	sgRNA sequence
(r =native ribose base, m = 2`-O methyl modified base, * = phosphorothioate bond)				

[0418] *Genomic DNA preparation and editing analysis by Next-Generation sequencing (NGS)*

[0419] The left lateral lobe of the liver (100 mg) was homogenized in the digestion buffer. Genomic DNA was purified from the resulting homogenate using and quantified by measuring the absorbance at 260 nm. Genomic DNA purified from mice injected with PBS buffer alone was used as a control. The region of the APOA1 and ANGPTL3 genes targeted by each specific sgRNA was PCR amplified with DNA polymerase and gene specific primers with adapters complementary to the barcoded primers used for next generation sequencing (NGS) for a total of 29 cycles. The product of this first PCR reaction was PCR amplified using the barcoded primers for NGS for a total of 10 cycles. The resulting product was subjected to NGS, and the results were processed to generate the percentage of sequencing reads that contain insertions or deletions (indels) at the targeted site in the APOA1 and ANGPTL3 genes.

[0420] *Results*

[0421] The NGS analysis of mouse liver genomic DNA showed up to 35% A-to-G conversion in mice dosed with LNP encapsulating ABE07-77 mRNA and mApoa1 A4 gRNA **(FIG. 7)**. Additionally, target sites mApoa1 A5 and mApoa1 C5 showed over 5% average A-to-G editing activity. The indel formation by ABE07-77 remained low at all sites, especially in comparison to the MG3-6/3-8 nuclease, which showed up to 80% indel formation activity at some of the target sites **(FIG. 8)**. The ABE07-77 base editor maintained its relatively large editing window in vivo, as multiple adenines within the protospacer were edited **(FIG. 9)**. Barring the minor differences between the treated groups, ABE07-77 showed robust and reproducible editing outcomes.

Example 4 – Mammalian editing activity of engineered CDAs as CBEs

[0422] In order to test the activity of engineered CDA variants, an engineered cell line was devised with 5 consecutive PAMs compatible with MG3-6 and Cas9. This cell line allows for gRNA tiling to test editing efficiency and find -1 nt preferences of candidate CDA's.

[0423] In order to test the engineered CDA's, the CDA's were cloned in a plasmid backbone containing MG3-6 and an MG uracil glycosylase inhibitor. The CDAs were cloned in the N termini (SEQ ID NOs: 1659-1664). Once the cloning of variant CDAs was confirmed they were transiently transfected into the engineered HEK293T cells using lipofectamine 2000 with associated guides. A total of 6 engineered variants (139-52-V2 (SEQ ID NO: 1271), 139-52-V13

(SEQ ID NO: 1282), 139-52-V14 (SEQ ID NO: 1283), 139-52-V17 (SEQ ID NO: 1314), 139-86v12 (SEQ ID NO: 1296), and 152-6v13 (SEQ ID NO: 1309) were tested in the gRNA tiling experiment described above. Out of the 6 engineered CDA's, all showed editing activity higher than 9% (**FIG. 10**). When the editing activity was normalized per experimental condition relative to a positive control (known high activity CDA: A0A2K5RDN7), it was observed that 5 out of 6 candidates showed higher activity than the A0A2K5RDN7 hyperactive positive control **(FIG. 10)**. 139-52-V2, 139-52-V13, 139-52-V14, 139-52-V17, 139-86v12, and 152-6v13 edited at 159.2%, 102.1%, 353.7%, 137.8%,76.3%, and 309.1% of the maximum A0A2K5RDN7 editing respectively.

[0424] To characterize the -1nt preference, 6 engineered candidates of interest were selected (139-52-V2,139-52-V13, 139-52-V14, 139-52-V17, 139-86v12, and 152-6v13). The -1 nt mammalian cell preference was calculated by selecting the top 4 modified cytosines per guide RNA and calculating the ratio per -1 position. Only cytosines with >1% editing were considered for this analysis. The average ratio for all 5 guides were plotted. The -1nt in vitro preference was plotted by calculating the sum of percentage cleavages (percent cleavage measures percent deamination) per -1 nt preference and then calculating the ratio per -1 nucleotide. The mammalian cell and *in vitro* -1 nt preference is shown in **FIG. 11**. To note is that different CDA families tend to have different -1 nt preferences and their preferences tend to be conserved amongst proteins belonging to the same family. For example, the candidates have different -1 nt preferences: 152-6 WT (SEQ ID NO: 1322) prefers T in the -1 position, whereas 139-52 (WT, SEQ ID NO: 1325; and engineered variants) has a strong preference for C at the -1 position. Having candidates with strong -1 nt preferences is preferable, since having a tighter nt preference improves off target activity. Candidates with different and strong -1 nt preferences allow us to target different loci without risking high off target activity. Candidates were identified with purine preferences: 139-86 whose preference is more G and/or A.

[0425] Altogether, the characterization of engineered CDA candidates working as CBEs in mammalian cells was determined. 6 variants originating from 3 natural active CDAs (MG139-52 (SEQ ID NO: 1325), MG139-86 (SEQ ID NO: 810) and MG152-6 (SEQ ID NO: 1322)) with different -1nt preferences and various levels of activity were tested.

Example 5 – Optimization of ABE architecture

[0426] In the Examples above, either a homodimeric (ABE07) or heterodimeric (ABE07-77, hereafter referred as ABE15) form of architecture (SEQ ID NOs: 1411 and 1657) were utilized. These dimeric states were chosen to emulate the native oligomerization state of the deaminase domain. To reduce the size of the base editor and understand the function of each deaminase

subunit in context of the base editor, several new ABE variants (**Table 9**) which were either monomeric (ABE01, ABE02, ABE55), homodimeric (ABE54, ABE58) or heterodimeric (ABE51, ABE52, ABE57) were constructed. Additionally, in these dimeric variants either the N-terminal or C-terminal deaminase domains were selectively deactivated through the disruption of the catalytic site with a E60A mutation (ABE51, ABE52, ABE54, and ABE57) to assess which of the two deaminase copies is critical for the DNA editing capabilities of the ABE. These variants were tested on a panel of pre-characterized target sites spanning the mouse APOA1 and ANGPTL3 genes **(Table 8)** using the protocols described below.

[0427] *Guide RNA and mRNA production*

[0428] Guide sequences used for screening the activity of novel ABE variants are listed in **Table 8**. These were chosen based on the editing outcomes observed for ABE07-77. The first and last three bases on the guides were modified with 2'-O methyl and phosphorothioate groups. The mRNA corresponding to ABE variants listed in **Tables 9-12** were produced.

Table 8: sgRNAs and targeting spacer sequences used in primary mouse hepatocytes

Target Site Name	Spacer DNA sequence	sgRNA sequence
Angptl3 F3	AAGAGAAGACAGCCCTTCAACA	mA*mA*mG*rArGrArArGrArCrArGrCrCrCrUrUrCrArArCrArGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 A4	TCTCCTGGAAAACTGGGACACT	mU*mC*mU*rCrCrUrGrGrArArArArCrUrGrGrGrArCrArCrUrGrUrGrArGrArArArArCrUrGrGrGrArCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 A5	CTGGGATAACCTGGAGAAAGAA	mC*mU*mG*rGrGrArUrArArCrCrUrGrGrArGrArArArGrArArGrUrUrGrArGrArArArUrCrGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrGrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 A8	CCTACCTTGAACGAGTACCACA	mC*mC*mU*rArCrCrUrUrGrArArCrGrArGrUrArCrCrArCrArGrUrUrGrArGrArArArUrCrGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrGrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 A11	AGCATGGGCATCAGACTATGGC	mA*mG*mC*rArUrGrGrGrCrArUrCrArGrArCrUrArUrGrGrCrGrUrUrGrArGrArArArUrCrGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrGrArGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU

Target Site Name	Spacer DNA sequence	sgRNA sequence
Apoa1 B2	GCCACAGGGGACAGTCTCCCTT	mG*mC*mC*rArCrArGrGrGrGrArCrArGrUrCrUrCrCrCrUrUrGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrArGrGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 B4	CTCCTGGAAAACTGGGACACTC	mC*mU*mC*rCrUrGrGrArArArArCrUrGrGrGrArCrArCrUrCrGrUrUrGrArGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrArGrGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 B5	ATTGGGTGAGACAGGAGATGAA	mA*mU*mU*rGrGrGrUrGrArGrArCrArGrGrArGrArUrGrArArGrUrUrGrArArGrArArUrCrGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrGrArGrGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 B8	CTACCTTGAACGAGTACCACAC	mC*mU*mA*rCrCrUrUrGrArArCrGrArGrUrArCrCrArCrArCrGrUrUrGrArGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrGrArGrGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 C5	TTGGGTGAGACAGGAGATGAAC	mU*mU*mG*rGrGrUrGrArGrArCrArGrGrArGrArUrGrArArCrGrUrUrGrArGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrGrArGrGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 D11	ACTATGGCGCAGGTCCTCCAGC	mA*mC*mU*rArUrGrGrCrGrCrArGrGrUrCrCrUrCrCrArGrCrGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrGrArGrGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 E8	CGAGAAAGCCAGACCTGCGCTG	mC*mG*mA*rGrArArArGrCrCrArGrArCrCrUrGrCrGrCrUrGrGrUrUrGrArGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrGrArGrGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 F2	GGCCCAAGGAGGAGGATTCAAA	mG*mG*mC*rCrCrArArGrGrArGrGrArGrGrArUrUrCrArArArGrUrUrGrArGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrGrArGrGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU
Apoa1 F4	AGGAACGGCTGGGCCCATTGAC	mA*mG*mG*rArArCrGrGrCrUrGrGrGrCrCrCrArUrUrGrArCrGrUrUrGrArGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrGrArGrGrCrGrGrGrCrGrGrUrArUrGrU*mU*mU*mU

Target Site Name	Spacer DNA sequence	sgRNA sequence
Apoa1 F5	ACGAATTCCAGAAGAAATGGAA	mA*mC*mG*rArArUrUrCrCrArGrArArGrArArArUrGrGrArArGrUrUrGrArGrArArUrCrGrArArArGrArUrUrCrUrUrArArUrArArGrGrCrArUrCrCrUrUrCrCrGrArUrGrCrUrGrArCrUrUrCrUrCrArCrCrGrUrCrCrGrUrUrUrUrCrCrArArUrArGrGrArGrCrGrCrGrUrArUrGrU*mU*mU*mU
(r =native ribose base, m = 2'-O methyl modified base, * = phosphorothioate bond)		

[0429] *Mouse cell culture, Transfections, and Data Analysis*

[0430] Hepa1-6 cells were obtained from ATCC. Hepa1-6 cells were nucleofected with 500 ng of mRNA and 150 pmol of chemically synthesized sgRNA by an electroporator. Each nucleofection reaction had 100,000 cells. Immediately after transfection, cells were cultured in DMEM supplemented with 10% FBS and 1X MEM containing non-essential amino acids at 37 °C with 5% CO2 for three days. Genomic DNA was harvested from. Targeted sequences were amplified with NGS primers. Amplicons around 250 bp long were checked by gel, sequenced by NGS, and computationally analyzed to measure gene editing outcomes.

[0431] *Results*

[0432] MG68-4 deaminase (SEQ ID NO: 386) is a putative tRNA adenosine deaminase and natively functions as an obligate dimer, the dimeric constructs outperform their monomeric versions (**FIGs. 12A and 12B**). When comparing the D109N variants, the homodimeric ABE07 (SEQ ID NO: 1411) outperformed the monomeric ABE01 (SEQ ID NO: 1410) which has the MG68-4 (D109N) variant inlaid into the RuvC-III domain as well as ABE02 which has the MG68-4 (D109N) variant inlaid into the REC domain (SEQ ID NO: 1665)**.** Although the addition of the T112R and A155R mutations to D109N mutation seem to enhance the activity of the monomer, a similar trend was also seen for the triplet mutant, where the homodimer ABE58 (SEQ ID NO: 1666) outperformed the monomeric ABE55 (SEQ ID NO: 1667; **FIGs. 12A and 12B**).

[0433] When comparing the selectively deactivated ABE variants, ABE51, ABE52, ABE54, and ABE57 (SEQ ID Nos: 1668-1671), it was observed that only the C-terminal deaminase subunit was actually involved in the genomic DNA editing, and the N-terminal subunit merely aids the activity of the C-terminal subunit. This was evidenced by ABE51, which has a C-terminal deactivated deaminase domain that performed on par with the completely deactivated ABE54, clearly indicating that the N-terminal subunit is not positioned correctly to access the exposed DNA. On the contrary, the N-terminal deactivated variants, ABE52 and ABE57, demonstrated

higher editing activity than the monomeric and sometimes homodimeric ABEs **(FIGs. 12A – 12B, and 13A – 13B)**.

[0434] In conclusion, it is only the C-terminal deaminase domain which is responsible for DNA editing and it may be possible to compensate for the lack of the heterodimeric deaminase domains through the installation of beneficial mutations in the monomeric variant.

Table 9: Designs of oligomeric ABE variants tested in this study

ABE variant	Architecture	Mutations
ABE01	Monomeric	MG68-4 (D109N) – inlaid in RuvC domain
ABE02	Monomeric	MG68-4 (D109N) – inlaid in Rec domain
ABE07	Homodimeric	MG68-4 (D109N) + MG68-4 (D109N)
ABE51	Heterodimeric	MG68-4 (D109N) + dead MG68-4
ABE52	Heterodimeric	dead MG68-4 + MG68-4 (D109N)
ABE54	Homodimeric	dead MG68-4 + dead MG68-4
ABE55	Monomeric	MG68-4 (D109N+T112R+A155R)
ABE57	Heterodimeric	dead MG68-4 + MG68-4 (D109N/T112R/A155R)
ABE58	Homodimeric	MG68-4 (D109N/T112R/A155R) + MG68-4 (D109N/T112R/A155R)

Example 6 – Engineering ABEs with enhanced on-target editing

[0435] Next, the editing activity of the heterodimeric ABE15 variant (SEQ ID NO: 1657) was improved through incorporation of additional mutations in the C-terminal deaminase subunit. Building upon the combinatorial mutagenesis screen, ABE variants ABE32-ABE41(**Table 10**; SEQ ID Nos: 1672-1681) were created and screened them over the guides listed in **Table 8** using the protocols described in Example 5.

[0436] *Results*

[0437] The addition of the L85F mutation to the existing D109N+T112R+A155R mutations in ABE33 (SEQ ID NO: 1665) were found to lead to the most significant increase in both mean and max A:T to G:C editing activity. Specifically, it was observed that ABE33 demonstrated a mean A:T to G:C editing of 32.97% across the 15 tested guides, which is 1.5-fold higher than ABE15 which averages around 22.8% mean A:T to G:C editing activity **(FIGs. 14A – 14B)**. Moreover, ABE33 showed lower promiscuous C deamination compared to ABE15 **(FIGs. 15A – 15B)** while having low indel formation activity **(FIG. 16)**.

[0438] Unlike the addition of L85F, it was determined that that incorporation of A143W (ABE34), A155E (ABE35), E10Y (ABE37), and A126D (ABE40), led to a decrease in the A:T to G:C editing. The addition of V83S (ABE32) significantly increased undesirable promiscuous C deamination and indel formation activity of the ABE without increasing the A:T to G:C editing activity **(FIGs. 14A – 14B, 15A – 15B, and 16)**.

[0439] Overall, it was discovered that the L85F mutation can be successfully added to the ABE15 construct to enhance the editing properties of the resultant ABE33 variant.

Table 10: Designs of engineered ABE variants tested in this study.

ABE variant	Architecture	Mutations
ABE07	Homodimeric	MG68-4 (D109N) + MG68-4 (D109N)
ABE15	Heterodimeric	wt MG68-4 + MG68-4 (D109N/T112R/A155R)
ABE32	Heterodimeric	wt MG68-4 + MG68-4 (V83S+D109N+T112R+A155R)
ABE33	Heterodimeric	wt MG68-4 + MG68-4 (L85F+D109N+T112R+A155R)
ABE34	Heterodimeric	wt MG68-4 + MG68-4 (A143W+D109N+T112R+A155R)
ABE35	Heterodimeric	wt MG68-4 + MG68-4 (A155E+D109N+T112R)
ABE36	Heterodimeric	wt MG68-4 + MG68-4 (A160T+D109N+T112R+A155R)
ABE37	Heterodimeric	wt MG68-4 + MG68-4 (E10Y+D109N+T112R+A155R)
ABE38	Heterodimeric	wt MG68-4 + MG68-4 (D162Q+D109N+T112R+A155R)
ABE39	Heterodimeric	wt MG68-4 + MG68-4 (S147F+D109N+T112R+A155R)
ABE40	Heterodimeric	wt MG68-4 + MG68-4 (A126D+D109N+T112R+A155R)
ABE41	Heterodimeric	wt MG68-4 + MG68-4 (F150C+D109N+T112R+A155R)

Example 7 – Engineering ABEs with relaxed editing context

[0440] While analyzing the editing activity of ABE15 across the 15 guide panel, it was discovered that even the mutated MG68-4 variant had an inherent preference for a -N**A**C- motif which mimics the natural -U**A**CG- motif present in the native substrate of other tRNA adenosine deaminase enzymes. Particularly, even though the guides that were tested **(Table 8)** exhibited an abundance of target adenines (As) flanked by guanines (Gs) on both +1 and -1 positions **(FIG.**

17A), ABE15 tends to preferentially edit As flanked by a +1C (**FIG. 17B**) at rates much higher than As flanked by a +1G (**FIG. 17C**).

[0441] To curb this sequence preference and broaden the editing context of the ABEs, a series of proline variants were designed which allow flexibility in the terminal alpha-helix of the deaminase domain (**Table 11**; SEQ ID Nos: 1682-1685) and screened them over the guides listed in **Table 8** using the protocols described in Example 5.

[0442] *Results*

[0443] Of the proline variants tested, it was observed that simultaneous addition of R153P and R154P mutations on top of the ABE15 mutations, i.e. ABE23 (SEQ ID NO: 1682), led to an overall increase in both mean and max A:T to G:C editing activity (**FIGs. 18A – 18B**) with no significant changes to the C deamination (**FIGs. 19A – 19B**) and indel formation rates (**FIG. 20A**). Specifically, ABE23 demonstrated a mean A:T to G:C editing of 30.8% as compared to 21.4% by ABE15 with analogous increase in the mean of the max observed A:T to G:C editing, 82.1% for ABE23 compared to 73.5% across all the 15 guides tested in this study. Overall, this increased A:T to G:C editing activity of ABE23 is manifested in the relaxation of the stringent +1 C context when compared to ABE15 (**FIG. 20B**).

Table 11: Designs of engineered ABE variants tested in this study

ABE variant	Architecture	Mutations
ABE07	Homodimeric	MG68-4 (D109N) + MG68-4 (D109N)
ABE15	Heterodimeric	wt MG68-4 + MG68-4 (D109N/T112R/A155R)
ABE23	Heterodimeric	wt MG68-4 + MG68-4 (D109N+T112R+A155R+R153P+R154P)
ABE24	Heterodimeric	wt MG68-4 + MG68-4 (D109N+T112R+A155R+R153P)
ABE25	Heterodimeric	wt MG68-4 + MG68-4 (D109N+T112R+A155R+R154P)
ABE26	Heterodimeric	wt MG68-4 + MG68-4 (D109N+T112R+A155R+delR154P)

Example 8 – Engineering ABEs with less promiscuity

[0444] It was observed low levels of cytidine editing with almost all the ABE variants that were tested and described in Example 2, and 5-7 above. This promiscuity has been shown to be reduced by the installation of D109Q mutation.

[0445] Adopting a similar strategy, the undesirable C deamination activity of our ABE was curbed by changing the foundational D109N mutation to Q in several of our top performing

variants (**Table 12**) and screened them over the guides listed in **Table 8** using the protocols described in Example 5.

[0446] Specifically, the D109Q mutations were included in top variants ABE12, ABE13, ABE14, ABE15, and ABE16 (SEQ ID Nos: 1654-1658) to yield ABE18, ABE19, ABE20, ABE21, and ABE22, respectively (SEQ ID Nos: 1686-1690). These mutations were also included in the proline variants tested in **Example 7** to yield ABE28, ABE29, ABE30, and ABE31 (SEQ ID Nos: 1691-1694).

[0447] *Results*

[0448] It was determined that the addition of the D109Q mutation had varied effects on the activity of the ABEs. There was no significant impact of swapping D109N with D109Q in ABE18 versus ABE12, ABE19 versus ABE13, ABE21 versus ABE15 when comparing their mean and max on-target A:T to G:C editing activity. However, this mutational swap resulted in modest enhancement of on-target editing activity of ABE22 and ABE29, and led to significant decrease in activity of ABE28, ABE30, and ABE31(**FIGs. 21A – 21B**). The D109Q mutation had a little to no significant impact on the promiscuous C deamination (**FIGs. 22A – 22B**) and the indel formation rate of these ABE variants (**FIG. 23**).

[0449] Overall, through engineering and screening efforts described in this disclosure key mutations were identified in the deaminase domain of our ABE that yield higher levels of on-target A:T to G:C editing while simultaneously reducing the promiscuous C deamination rates. Particularly, ABE23 and ABE33 emerged as key new ABE variants that have higher on-target editing activity (**FIGs. 24A – 24B**), lower undesirable C deamination rates than their parent ABE15 (**FIGs. 25A – 25B**), and low undesirable indel activity (**FIG. 26**).

Table 12: Designs of engineered ABE variants tested in this study.

ABE variant	Architecture	Mutations
ABE07	Homodimeric	MG68-4 (D109N) + MG68-4 (D109N)
ABE12	Heterodimeric	wt MG68-4 + MG68-4 (D109N)
ABE13	Heterodimeric	wt MG68-4 + MG68-4 (V83S+D109N+A155R)
ABE14	Heterodimeric	wt MG68-4 + MG68-4 (D109N+T112R+H129N)
ABE15	Heterodimeric	wt MG68-4 + MG68-4 (D109N+T112R+A155R)
ABE16	Heterodimeric	wt MG68-4 + MG68-4 (V83S+D109N+T112R+H129N)
ABE18	Heterodimeric	wt MG68-4 + MG68-4 (D109Q)

ABE19	Heterodimeric	wt MG68-4 + MG68-4 (V83S+D109Q+A155R)
ABE20	Heterodimeric	wt MG68-4 + MG68-4 (D109Q+T112R+H129N)
ABE21	Heterodimeric	wt MG68-4 + MG68-4 (D109Q+T112R+A155R)
ABE22	Heterodimeric	wt MG68-4 + MG68-4 (V83S+D109Q+T112R+H129N)
ABE28	Heterodimeric	wt MG68-4 + MG68-4 (D109Q+T112R+A155R+R153P+R154P)
ABE29	Heterodimeric	wt MG68-4 + MG68-4 (D109Q+T112R+A155R+R153P)
ABE30	Heterodimeric	wt MG68-4 + MG68-4 (D109Q+T112R+A155R+R154P)
ABE31	Heterodimeric	wt MG68-4 + MG68-4 (D109Q+T112R+A155R+delR154P)

Example 9 – Novel small base editors display high activity in human cells

[0450] *Small base editor constructs*

[0451] We fused an engineered MG68-4 adenine deaminase, either to the C-terminal, N-terminal, or inlaid within the either the nickase MG102-39 or the nickase MG34-29 to generate adenine base editors ABE75-ABE77 and ABE87-88, respectively (SEQ ID Nos: 1695-1699) (**FIG. 27A** and **Table 8**).

Table 13: Small ABEs tested in this study.

ABE variant	Nickase	Deaminase	Architecture	SEQ ID
ABE75	MG102-39	MG68-4 (D109N, T112R, A155R)	Inlaid	**1695**
ABE76	MG102-39	MG68-4 (D109N, T112R, A155R)	C-terminal	**1696**
ABE77	MG102-39	MG68-4 (D109N, T112R, A155R)	N-terminal	**1697**
ABE87	MG34-29	MG68-4 (D109N, T112R, A155R)	Inlaid	**1698**
ABE88	MG34-29	MG68-4 (D109N, T112R, A155R)	C-terminal	**1699**

[0452] *Guide RNA design and mRNA production*

[0453] Sequences for small ABEs were codon optimized for human expression and each cloned into an expression vector with a T7 promoter and capping initiation sequence, 5' and 3' UTRs, and a polyA tail. The coding sequence contained an N-terminal SV40 nuclear localization signal and a C-terminal nucleoplasmin nuclear localization signal. The expression vector was midi-prepped, linearized with SpeI, cleaned with, and used for in vitro transcription with Hi-T7 polymerase. In vitro transcription reactions contained N1-methylpseudouridine in place of

uridine and had an added capping reagent. The resulting mRNA was cleaned, checked for product size and purity by spectrophotometry and gel electrophoresis and diluted to 250 ng/μL in sterile water for use in nucleofection. ABE33 (SEQ ID NO: 1672) was used as a positive control with a hAPOA1 target guide **(Table 14)**.

[0454] The base editing efficiency of these small ABEs was assessed across target guides chosen based on previously observed activity of the nucleases **(Table 14).** The first and last three bases on these guides were modified with 2`-O methyl and phosphorothioate groups.

[0455] *Mammalian cell culture, Transfections, and Data Analysis*

[0456] K562 cells (ATCC #CCL-243) were cultured in IMDM + L-alanyl-L-glutamine dipeptide media and 10% FBS for 1-2 passages prior to nucleofection. On the day of nucleofection, cells were harvested, counted, washed in 1X PBS, and resuspended in nucleofection buffer according to manufacturer instructions. 120,000 cells were distributed per well and nucleofected with 500 ng of mRNA and 200 pmol of sgRNA using the a nucleofection kit. For some experiments, the amount of guide added varied from 100 to 400 pmol. Cells were added to recovery media and grown for 72 hours before genomic DNA was harvested. Resulting gDNA was diluted 1:3 and used as a template for NGS PCR. Targeted sequences were amplified with NGS primers (IDT). Amplicons around 250 bp long were checked by gel electrophoresis, sequenced by an NGS sequencing machine, and computationally analyzed to measure gene editing outcomes.

[0457] *Results*

[0458] Across the tested panel of guides, high A→G editing was observed for ABEs derived from MG34-29 **(FIG. 28A)**. ABE87, which contains the MG68-4(D109N, T112R, A155R) deaminase inlaid within the MG34-29 nickase **(FIG. 27C)** demonstrated the highest editing of all the tested small ABEs, with a mean editing efficiency of from 74.34 ± 0.02 % at AAVS1 E7 locus and 32.41± 3.77 % at AAVS1 C7 locus **(FIG. 28B)**. The inlaid construct preference has been also observed with other nickases such as MG3-6_3-8 ABEs.

[0459] A similar preference for the inlaid architecture was observed for the small ABEs derived from MG102-39 (ABE75) albeit the editing efficiencies observed were lower **(FIG. 28C)**. ABE75 outperformed its C-terminal and N-terminal variants, with the highest performance of 7.10 ± 1.60 % observed at the TRAC C11 locus **(FIG. 28D)**.

Table 14: Spacer and sgRNA sequences used in testing the small ABEs.

Nuclease Root	Target Site	SEQ ID Reference	Spacer DNA sequence	Nuclease Indel %	Max A→G Editing % as an inlaid ABE
MG34-29	pPE633	861	GAACACAAAGCATAGACTGC	0.04	0.04

Nuclease Root	Target Site	SEQ ID Reference	Spacer DNA sequence	Nuclease Indel %	Max A→G Editing % as an inlaid ABE
	pPE634	862	GGGAAAGACCCAGCATCCGT	0.56	4.09±0.27
	pPE641	864	GGTCGTAGCCAGTCCGAACCC	80.83	2.10±0.41
	pPE635	687	GATGAGAAGGAGAAGTTCTT	0.09	0.32±0.13
	AAVS1 E7	1720	TATCAGGAGACTAGGAAGGAGG	56.11	74.34 ±0.02
	AAVS1 F7	1721	AGACTAGGAAGGAGGAGGCCTA	28.645	3.34±0.56
	AAVS1 B11	1722	CCTGAAGTGGACATAGGGGCCC	23.845	1.04±0.11
	AAVS1 G7	1723	TAGGAAGGAGGAGGCCTAAGGA	19.42	5.86±0.51
	AAVS1 C7	1724	ACCCAATATCAGGAGACTAGGA	13.64	32.41±3.77
	AAVS1 E6	1725	AAGGAATCTGCCTAACAGGAGG	11.37	0.63±0.03
MG102-39	TRAC A9	1711	TCTTGGTTTTACAGATACGAACCT	85.03	1.92±0.19
	TRAC G11	1712	GGCCACTTTCAGGAGGAGGATTCG	73.25	1.85±0.41
	TRAC C11	1713	CAGCCGCAGCGTCATGAGCAGATT	63.46	7.10±1.60
	TRAC B6	1714	CCAGGCCACAGCACTGTTGCTCTT	51.42	0.22±0.02
	TRAC B5	1715	GTCTTCTGGAATAATGCTGTTGTT	48.38	0.03
	TRAC G9	1716	GATTGGGTTCCGAATCCTCCTCCT	46.03	0.22
	TRAC D1	1717	ATTCTGATGTGTATATCACAGACA	36.69	0.9±0.05
	TRAC B11	1718	ACAGCCGCAGCGTCATGAGCAGAT	32.73	0.36±0.03
	TRAC G1	1719	GCTAGACATGAGGTCTATGGACTT	18.19	1.85±0.41
ABE33	hAPOA1	1726	CCTCAAGGCGCGCGGCCAAGCG	NA	93.35±0.34

Example 10 - PAM-interacting domain swaps increase the targetability of ABEs

[0460] Adenine base editors can install precise A:T → G:C edits in a programmable manner, and hence have the capability to correct more than half of the known pathogenic single-nucleotide polymorphism and efficiently knock out protein through splice site disruption. However, due to its programmable nature, the utility of ABEs is highly dependent on the presence of an appropriate PAM downstream of the desired target adenine. In fact, SpCas9-based ABEs (like ABE7.10) can merely access 18% of all the adenines present in the human reference genome due to the lack of NGG in the vicinity of target A (**FIG. 29A**).

[0461] Here, this issue is alleviated by leveraging a repository of PID-swappable MG3-6 nuclease chassis and highly efficient adenine deaminases to generate a suite of ABEs that

collectively offer broad PAM compatibility (**FIG. 29B**). This suite of chimeric ABEs can theoretically target 95% of all the adenines in the human genome (**FIG. 29C**). We utilize these chimeric ABEs for targeted splice site disruption and correction in several different therapeutically relevant contexts and demonstrate that we can achieve highly efficient base editing with several different ABEs and multiple guides on each targeted loci. With its diverse PAMs compatibility, this suite of chimeric ABEs has the potential to provide permanent life-long cures for a broad range of genetic disorders.

[0462] *Methods*

[0463] *Guide RNA Design and mRNA Production*

[0464] The ABEs were codon optimized for human expression and subsequently cloned into an expression vector with a CleanCap T7 promoter, 5' and 3' UTRs, and a polyA tail. The coding sequence contained an N-terminal SV40 nuclear localization signal and a C-terminal nucleoplasmin nuclear localization signal. The expression vectors were midi-prepped, mRNA templates were amplified from plasmids with Q5 High-Fidelity 2X Master Mix, cleaned with HighPrep PCR, and used for in vitro transcription with Hi-T7. In vitro transcription reactions contained N1-methylpseudouridine in place of uridine and had added CleanCap reagent. The resulting mRNA was cleaned with Rneasy, checked for product size and purity by NanoDrop and Tapestation, and diluted to 250 ng/μL in sterile water for use in nucleofection.

Table 15: PID-swapped ABEs tested in this study.

ABE variant	Nickase/PID	Deaminase variant	Architecture	PAM sequence	ABE variant SEQ ID NO:
ABE23/ ABE59	MG3-6_3-8	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Heterodimeric	NNRGNNYN	1727
ABE60	MG3-6_3-8	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNRGNNYN	1728
ABE61	MG3-6_3-8_3-4	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNRMWW	1729
ABE62	MG3-6_3-8_3-6	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNRRKYY	1730
ABE63	MG3-6_3-8_3-7	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNRNYHY	1731
ABE64	MG3-6_3-8_3-22	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNNMYY	1732
ABE65	MG3-6_3-8_3-24	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNRNYHY	1733
ABE66	MG3-6_3-8_3-38	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNNMHTY	1734
ABE67	MG3-6_3-8_3-89	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNRWWYY	1735
ABE68	MG3-6_3-8_3-90	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNRMYY	1736
ABE69	MG3-6_3-8_3-92	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNRGNCR	1737

ABE variant	Nickase/PID	Deaminase variant	Architecture	PAM sequence	ABE variant SEQ ID NO:
ABE70	MG3-6_3-8_3-93	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNNMHYY	1738
ABE71	MG3-6_3-8_3-95	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNNMCMCW	1739
ABE72	MG3-6_3-8_3-104	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNRWYHH	1740
ABE73	MG3-6_3-8_150-2	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNRWYHH	1741
ABE74	MG3-6_3-8_150-9	MG68-4 (D109N, T112R, R153P, R154P, A155R)	Homodimeric	NNRNYHH	1742
ABE33/ ABE100	MG3-6_3-8	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNRGNNYN	1743
ABE101	MG3-6_3-8_3-4	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNRMWW	1744
ABE102	MG3-6_3-8_3-6	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNRRKYY	1745
ABE103	MG3-6_3-8_3-7	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNRNYHY	1746
ABE104	MG3-6_3-8_3-22	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNNMYY	1747
ABE105	MG3-6_3-8_3-24	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNRNYHY	1748
ABE106	MG3-6_3-8_3-38	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNNMHTY	1749
ABE107	MG3-6_3-8_3-89	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNRWWYY	1750
ABE108	MG3-6_3-8_3-90	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNRMYY	1751
ABE109	MG3-6_3-8_3-92	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNRGNCR	1752
ABE110	MG3-6_3-8_3-93	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNNMHYY	1753
ABE111	MG3-6_3-8_3-95	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNNMCMCW	1754
ABE112	MG3-6_3-8_3-104	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNRWYHH	1755
ABE113	MG3-6_3-8_150-2	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNRWYHH	1756
ABE114	MG3-6_3-8_150-9	MG68-4 (L85F/D109N/T112R/A155R)	Heterodimeric	NNRNYHH	1757

[0465] We have determined that the MG3-6 ABEs perform most efficiently at adenine bases which are situated in an editing window spanning from bases 3 to 13 in the protospacer (**FIG. 30A**). Using this editing window, we designed guides tiling across the splice sites of hANGPTL3 (SEQ ID NOs: 1758-1889), the GATA1-binding site of hBCL11A (SEQ ID NO: 1890-1976) and the most frequently occurring SNVs in hPAH (SEQ ID NOs: 1977-2009) (**FIG. 30B**). The first and last three bases on the guides were modified with 2'-O methyl and phosphorothioate groups (IDT).

[0466] *High-throughput Pooled and Arrayed Mammalian Cell Screening*

[0467] K562 cells were obtained from ATCC. On the day of nucleofection, cells were harvested, counted, washed in 1X PBS, and resuspended in SF buffer according to manufacturer instructions in a 4D electroporator. Given that we wanted to maximize the number of chimeric ABEs at each genetic loci, we developed a high-throughput pooled screening approach. The general plate layout for this approach is shown in **FIG. 30C**. In each well, we pooled together 15 pmol of the ten guides that tiled a particular splice site along with 500 ng of mRNA pertaining to a unique chimeric ABE (**FIG. 30B**). With this pooled approach we identified the chimeric ABEs that showed the highest editing efficiency at each splice site. To determine the active guide and chimeric ABE combinations at each splice site we conducted a secondary deconvolution screen. This was done by 'unpooling' the test guides from the most edited guide pools and testing them individually with each active chimeric ABE (**FIG. 30D**). 150 pmol of each test guide was added with 500 ng of mRNA of the chimeric ABE. Each nucleofection reaction had 120,000 cells per guide. Immediately after transfection, cells were cultured in IMDM + GlutaMAX supplemented with 10% and FBS at 37°C with 5% CO2 for three days. Genomic DNA was harvested from cells with Quick Extract. Targeted sequences were amplified with NGS primers. Amplicons around 250 bp long were checked in an Agilent TapeStation D1000 gel, sequenced on an MiSeq machine, and analyzed with CRISPResso2 to measure gene editing outcomes.

[0468] For some experiments, K562 cell lines were engineered with 'disease' state nucleotides using a lentiviral vector encoding a combination of therapeutically relevant targets (about 4.5kb, SEQ ID NO: 2020) at a low MOI. This 4.5kb sequence contained the PAH gene with the relevant SNVs 1222C>T (p.R408W), 1066-11G>A, 1315+1G>A with about 250 nt on either side of the therapeutically relevant target nucleotide. The mutation corresponding to the gene of interest would then be changed to the diseased state, allowing for the correction of a therapeutically relevant edit. In order to amplify the engineered target of interest over the endogenous target, 20 nt of non-native sequences replaced endogenous gene sequence about 125 nt on the 5' and 3' end of the target of interest providing primer binding sites that would result in a 250 nt amplicon for NGS processing. Transduced cells were selected with puromycin from 3 to 10 days post-transduction. These stable cell lines are here on referred to as engineered K562 cells. The same nucleofection approach was used to determine deamination events at the therapeutically relevant sites as described above.

[0469] *Results*

[0470] Over 100 PID-swapped MG3-6-based chimeras were made, each recognizing a unique PAM sequence. By evaluating the PAM targetability of each chimeric nuclease across the human genome as a metric, 15 of these chimeric nucleases were selected and engineered into nickases for ABE construct design. This was done by deactivation of their RuvC nuclease domain and the

insertion of a highly efficient adenine deaminase (MG68-4 with mutations [D109N, T112R, R153P, R154P, A155R] or MG68-4 with mutations [L85F/D109N/T112R/A155R]) to generate a suite of chimeric ABEs (**Table 15,** SEQ ID NOs: 1727-1757).

[0471] Analogous to the observations made when comparing the targetability across the whole genome (**FIGs. 29A-29C**), the suite of PID-swappable MG3-6 chimeric ABEs can theoretically access many more guides for disrupting splice sites (**FIG. 31A),** as well as enhancer binding sites (**FIG. 33A**) and single nucleotide variations (**FIG. 34A**). SpCas9 ABE with its NGG PAM has limited targetability across putative targets of interest, whereas MG3-6 ABEs cumulatively can target these with multiple possible guides. This is also true when comparing them to other well-characterized ABEs with relaxed PAM requirements, such as those for SaCas9, SaKKHCas9, CjCas9, Nme2Cas9, and SauriCas9. Overall, MG3-6 chimeric ABEs offer more options for guides to target genes of interest.

[0472] *hANGPTL3*

[0473] ANGPTL3 is an endogenous inhibitor of lipoprotein lipase (LPL) that is expressed predominantly in the liver and is associated with cholesterol metabolism. Knocking down of ANGPTL3 has been shown to increase LPL levels, which is the main enzyme involved in hydrolysis of triglycerides. Hence, knock out of ANGPTL3 is considered a potential permanent treatment of coronary artery diseases.

[0474] In a pooled screening approach, we screened the splice sites of all seven exons of hANGPTL3 in K562 cells were screened. Multiple chimeric ABEs that disrupted the splice site of this gene were identified **(FIG. 31B)** and deconvolution of the pooled wells demonstrated overall high base editing (**FIGs. 32A-32D**). For example, ABE100 (SEQ ID NO: 1743), which translates into the amino acid sequence ABE33 (SEQ ID NO: 1673) edited the exon 1 splice donor (**FIG. 32A**); while ABE101, ABE103, and ABE112 (SEQ ID NOs: 1744, 1746 and 1755), which translate into the amino acid sequences SEQ ID NOs: 2021-2023, were active at exon 4 splice acceptor and splice donor, exon 5 splice donor and exon 7 splice acceptor (**FIGs. 32B-32D**). Upon deconvolution, it was observed that ABE100 disrupted the splice donor site of exon 1 with an average editing efficiency of 45.9% (**FIG. 32A**). ABE112 (SEQ ID NO: 1755) showed the best editing at ANGPTL3 as it demonstrated 21.9% editing at exon 4 splice acceptor site, 61.5% editing at exon 4 splice donor site, and 25% editing at exon 7 splice acceptor site.

[0475] *hBCL11A*

[0476] To demonstrate the applicability of the chimeric ABE platform for targeted gene silencing through enhancer site disruption, the BCL11A gene was chosen as a target, which is known to play a critical role in β-thalassemia. BCL11A has three well-characterized human BCL11A composite enhancer DHS sites(namely DHS+55, DHS+58, and DHS+62). These enhancer sites

have a consensus GATA motif that binds the GATA1 and TAL1 enhancers, upregulating the expression of BCL11A. Mutating these GATA sites has the potential to suppress BCL11A levels which subsequently leads to an increase in the γ-globin levels and rescues the β-thalassemia disease state. All possible sgRNAs were designed that can target these three GATA sites from both the forward and reverse strands and screened them along with the chimeric ABEs using the previously described high-throughput pooled guide screening approach **(FIG. 33B)**. We observed significant editing (16.5%) in the guide pool editing the DHS+55 site with ABE101 (SEQ ID NOs: 1744 and 2021) (based on MG3-6_3-8_3-4, PAM NNRMWW), which correlates with our theoretical prediction of this ABE's targetability **(FIG. 33A)**. Deconvolution of the pooled guides identified two guides that edit both adenines in the GATA1 site with an average A→G editing efficiency across both A's of 44.81% and 24.02% **(FIG. 33C)**. DHS+55 is a critical enhancer binding site in the BCL11A gene. Disruption of this site with an ABE to increase in γ-globin level is presumed a safer alternative to Cas9-mediated indels.

[0477] *hPAH*

[0478] Phenylalanine hydroxylase (PAH) deficiency results in intolerance to the dietary intake of the essential amino acid phenylalanine and produces a spectrum of disorders. The risk of adverse outcomes varies based on the degree of PAH deficiency. Over 500 mutations have been reported in the coding sequence as well as in the intervening sequence of the PAH gene (Regier and Greene, 2000). The current best therapeutic to treat PAH deficiency is an oral medication, Sapropterin, that serves as a cofactor of the PAH protein and can improve the activity of some mutant forms of PAH. However, it is not effective against the most commonly reported SNVs in the PAH gene - 1222C>T (p.R408W), 1066-11G>A, 1315+1G>A **(FIG. 34A)**.

[0479] We reasoned that the chimeric ABEs disclosed herein can offer a permanent and effective alternative to sapropterin by directly correcting the causative SNVs. To test the chimeric ABEs on these SNVs, a new cell line was created that harbors these various mutations and conducted pooled ABE screening pipeline on these engineered K562 cells. The chimeric ABEs were tested with 10 guides tiled across three of the SNVs in this gene.

[0480] While the chimeric ABEs did not demonstrate any appreciable editing at 1066-11G>A or 1315+1G>A sites (**FIG. 34C** and **FIG. 34D**), ABE101 (SEQ ID NOs: 1744 and 2021) corrected the 1222C>T (p.R408W) SNV with a pooled 15% efficiency **(FIG. 34B)** in the high-throughput pooled screening. At 1222C>T (p.R408W) SNV, the potential for bystander editing outcome due to editing of neighboring as in the editing window of the base editors was observed.

[0481] The data presented in this example highlight the strength of the PID-swappable chimeric ABE toolbox disclosed herein, especially when deployed in conjunction with high-throughput guide screening. Our suite of novel ABEs boasts more targetability across therapeutically-

relevant genes and theoretical sites in the human genome when compared to the well-characterized SpCas9 or SaCas9 ABEs. This is corroborated by the findings here as many chimeric ABEs showed efficient editing with multiple guides within the same gene. With broadened targetability and high efficiency, the chimeric ABEs disclosed herein have the potential to install precise base edits at many therapeutically relevant loci that were typically inaccessible to previously reported base editors.

[0482] References

[0483] Regier DS, Greene CL. Phenylalanine Hydroxylase Deficiency. 2000 Jan 10 [Updated 2017 Jan 5]. In: Adam MP, Feldman J, Mirzaa GM, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2024.

EQUIVALENTS

[0484] The disclosure may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The foregoing embodiments are therefore to be considered in all respects illustrative rather than limiting on the disclosure described herein. Scope of the disclosure is thus indicated by the appended claims rather than by the foregoing description, and all changes that come within the meaning and range of equivalency of the claims are intended to be embraced therein.

CLAIMS

WHAT IS CLAIMED IS:

1. An engineered base editing system comprising:

(a) a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415; and

(b) an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

2. The engineered base editing system of claim 1, wherein the base editor comprises a sequence having at least 95% sequence identity to any one of SEQ ID NOs: 1654-1703, and 2021-2023.

3. The engineered base editing system of claim 1, wherein the base editor comprises a sequence having 100% sequence identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023.

4. An engineered base editing system comprising:

(a) a base editor encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757, and

(b) an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

5. The engineered base editing system of claim 4, wherein the base editor is encoded by a nucleic acid sequence having at least 90% identity to any one of SEQ ID NOs: 1727-1757.

6. The engineered base editing system of claim 4, wherein the base editor is encoded by a nucleic acid sequence having 100% identity to any one of SEQ ID NOs: 1727-1757.

7. The engineered base editing system of any one of claims 1-6, wherein the base editor comprises a deaminase.

8. The engineered base editing system of claim 7, wherein the deaminase binds non-covalently to the endonuclease.

9. The engineered base editing system of claim 7, wherein the deaminase is covalently linked to the endonuclease.

10. The engineered base editing system of claim 7, wherein the deaminase is fused to the endonuclease.

11. The engineered base editing system of any one of claims 1-10, wherein the engineered guide polynucleotide is a single guide nucleic acid.

12. The engineered base editing system of any one of claims 1-11, wherein the engineered guide polynucleotide is a dual guide nucleic acid.

13. The engineered base editing system of any one of claims 1-12, wherein the engineered guide polynucleotide is RNA.

14. The engineered base editing system of any one of claims 1-13, wherein the endonuclease binds non-covalently to the engineered guide polynucleotide.

15. The engineered base editing system of any one of claims 1-13, wherein the endonuclease is covalently linked to the engineered guide polynucleotide.

16. The engineered base editing system of any one of claims 1-15, wherein the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019.

17. The engineered base editing system of any one of claims 1-16, wherein the engineered guide polynucleotide comprises a sequence having at least 90% sequence identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019.

18. The engineered base editing system of any one of claims 1-17, wherein the engineered guide polynucleotide comprises a sequence having 100% sequence identity to any one of SEQ ID NOs: 88-96, 488-489, 679-680, 876, 917-931, 963-967, 1099-1105, 1187-1195, 1416-1418, 1427-1428, 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019.

19. The engineered base editing system of any one of claims 1-15, wherein the engineered guide polynucleotide comprises a sequence having at least 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1431-1454, 1479-1483, 1489-1490, 1705-1710, and 1758-2019.

20. The engineered base editing system of claim 19, wherein the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1431-1454, 1704, and 2010-2019.

21. The engineered base editing system of claim 19, wherein the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1890-1976.

22. The engineered base editing system of claim 19, wherein the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1977-2009.

23. The engineered base editing system of claim 19, wherein the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1479-1483 and 1758-1889.

24. The engineered base editing system of any one of claims 1-23, wherein the base editor comprises a nickase domain.

25. The engineered base editing system of claim 24, wherein the nickase comprises an aspartate to alanine mutation at residue 9 relative to SEQ ID NO: 70, residue 13 relative to SEQ ID NOs: 71, 72, or 74, residue 12 relative to SEQ ID NO: 73, residue 17 relative to SEQ ID NO: 75, residue 23 relative to SEQ ID NO: 76, or residue 10 relative to SEQ ID NO: 597, or any combination thereof.

26. The engineered base editing system of any one of claims 1-25, wherein the base editor further comprises a uracil DNA glycosylase inhibitor sequence.

27. The engineered base editing system of any one of claims 1-26, wherein the base editor further comprises a FAM72A sequence.

28. The engineered base editing system of claim 27, wherein the FAM72A sequence has at least 80% identity to SEQ ID NO: 1121.

29. A nucleic acid encoding the engineered base editing system of any one of claims 1-28.

30. A vector comprising the nucleic acid of claim 29.

31. The vector of claim 30, wherein the vector is a plasmid, a minicircle, a CELiD, an adeno-associated virus (AAV) derived virion, a lentivirus, or an adenovirus.

32. A cell comprising the engineered base editing system of any one of claims 1-28, the nucleic acid of claim 29., or the vector of any one of claims 30-31.

33. The cell of claim 32, wherein the cell is a eukaryotic cell.

34. The cell of claim 32, wherein the cell is a mammalian cell.

35. The cell of claim 32, wherein the cell is an immortalized cell.

36. The cell of claim 32, wherein the cell is an insect cell.

37. The cell of claim 32, wherein the cell is a yeast cell.

38. The cell of claim 32, wherein the cell is a plant cell.

39. The cell of claim 32, wherein the cell is a fungal cell.

40. The cell of claim 32, wherein the cell is a prokaryotic cell.

41. The cell of claim 32, wherein the cell is an A549, HEK-293, HEK-293T, BHK, CHO, HeLa, MRC5, Sf9, Cos-1, Cos-7, Vero, BSC 1, BSC 40, BMT 10, WI38, HeLa, Saos, C2C12, L cell, HT1080, HepG2, Huh7, K562, primary cell, or a derivative thereof.

42. The cell of claim 32, wherein the cell is an engineered cell.

43. The cell of claim 32, wherein the cell is a stable cell.

44. A method for modifying a target nucleic acid sequence, comprising: contacting the target nucleic acid sequence using the engineered base editing system of any one of claims 1-29.

45. The method of claim 44, wherein modifying the target nucleic acid sequence comprises converting an adenine to a guanine in the target nucleic acid sequence.

46. The method of claim 44, wherein modifying the target nucleic acid sequence comprises converting a cytosine to a uracil in the target nucleic acid sequence.

47. The method of any one of claims 44-45, wherein the target nucleic acid sequence comprises deoxyribonucleic acid (DNA).

48. The method of any one of claims 44-47, wherein the target nucleic acid sequence comprises ribonucleic acid (RNA).

49. The method of any one of claims 44-48, wherein the target nucleic acid sequence comprises genomic DNA, viral DNA, viral RNA, or bacterial DNA.

50. The method of any one of claims 44-49, wherein the target nucleic acid sequence is modified *in vitro*.

51. The method of any one of claims 44-49, wherein the target nucleic acid sequence is modified *in vivo*.

52. The method of any one of claims 44-49, wherein the target nucleic acid sequence is modified *ex vivo*.

53. The method of any one of claims 44-52, wherein the target nucleic acid sequence is modified within a cell.

54. The method of claim 53, wherein the cell is a prokaryotic cell, a bacterial cell, a eukaryotic cell, a fungal cell, a plant cell, an animal cell, a mammalian cell, a rodent cell, a primate cell, a human cell, or a primary cell.

55. A method of modifying a nucleic acid encoding ANGPTL3 comprising contacting the nucleic acid sequence encoding ANGPTL3 with an engineered base editing system, said base editing system comprising:

a) a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415, or encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757; and

b) an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

56. The method of claim 55, wherein the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1479-1483 and 1758-1889.

57. A method of modifying a nucleic acid encoding APOA1 comprising contacting the nucleic acid sequence encoding APOA1 with an engineered base editing system, said base editing system comprising:

a) a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415, or encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757; and;

b) an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

58. The method of claim 57, wherein the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1431-1454, 1704, and 2010-2019.

59. A method of modifying a nucleic acid encoding BCL11A comprising contacting the nucleic acid sequence encoding BCL11A with an engineered base editing system, said base editing system comprising:

a) a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415, or encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757; and

b) an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

60. The method of claim 59, wherein the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1890-1976.

61. A method of modifying a nucleic acid encoding PAH comprising contacting the nucleic acid sequence encoding PAH with an engineered base editing system, said base editing system comprising:

a) a base editor comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1654-1703 and 2021-2023, wherein the sequence does not comprise any one of the sequences selected from SEQ ID NO: 1128-1160 and 1363-1415, or encoded by a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 1727-1757; and

b) an engineered guide polynucleotide which forms a complex with an endonuclease of the base editor and comprises a spacer sequence that hybridizes to a target nucleic acid sequence.

62. The method of claim 61, wherein the engineered guide polynucleotide comprises a sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1977-2009.

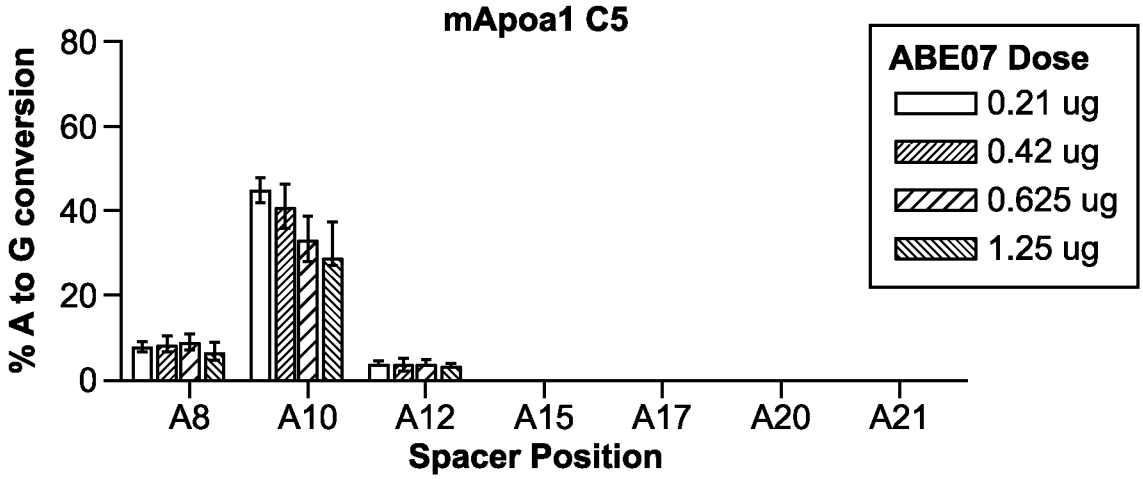

FIG. 1A

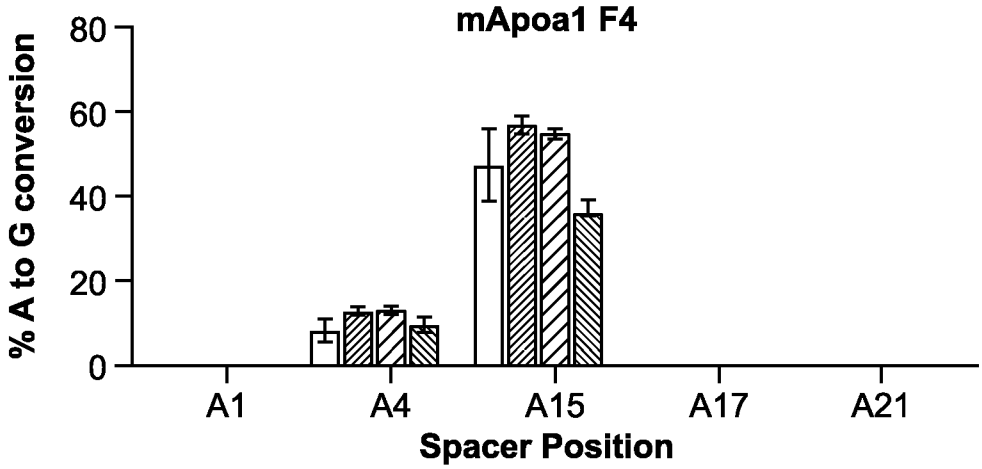

FIG. 1B

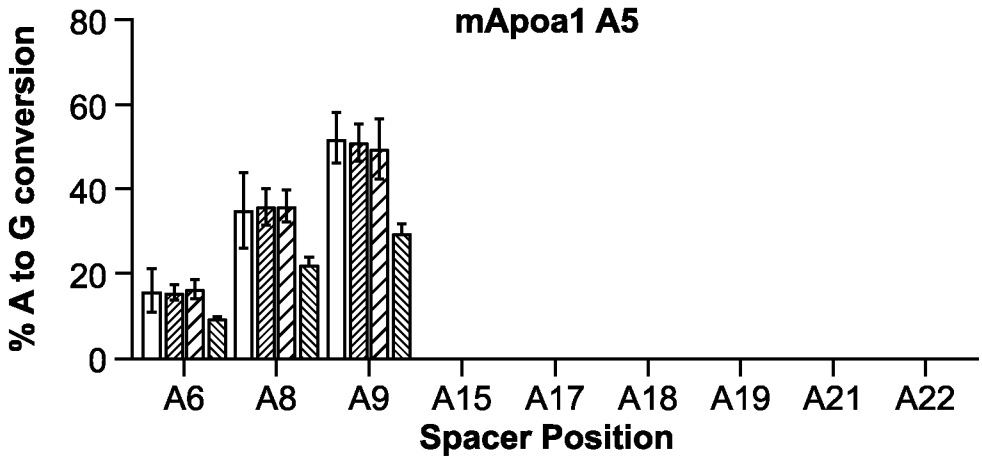

FIG. 1C

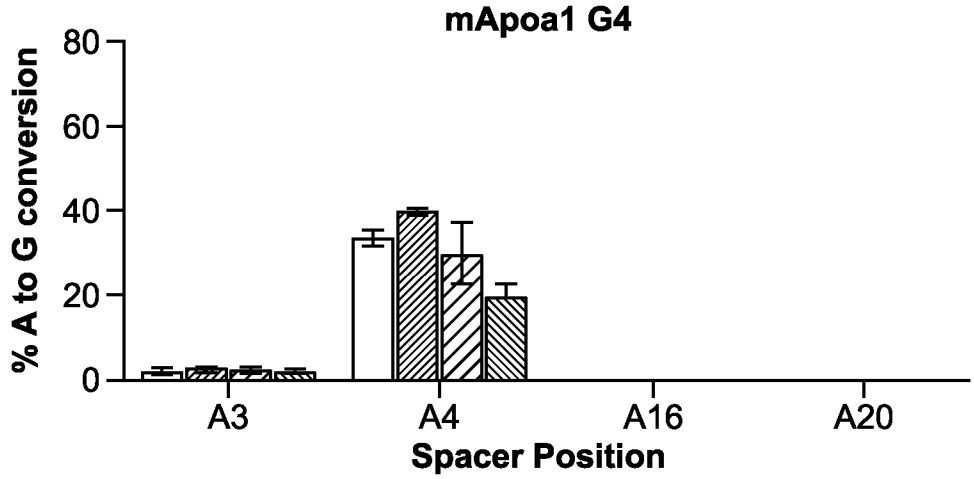

FIG. 1D

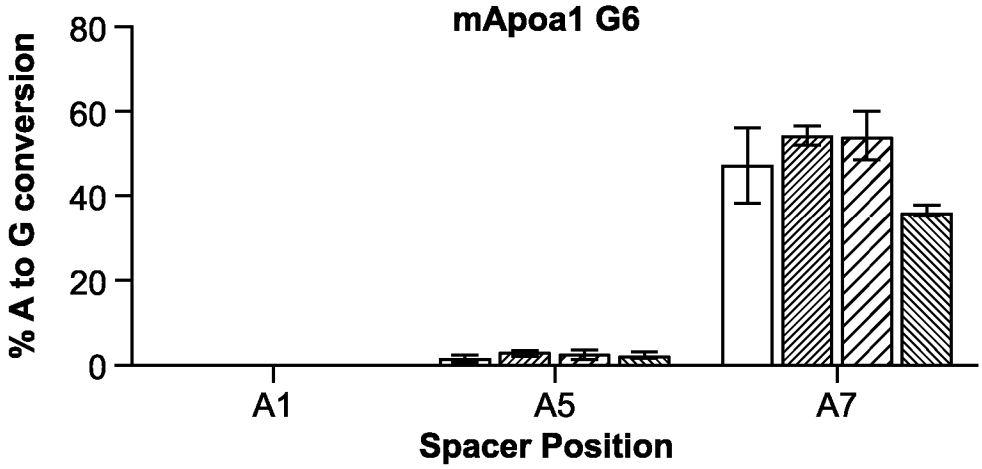

FIG. 1E

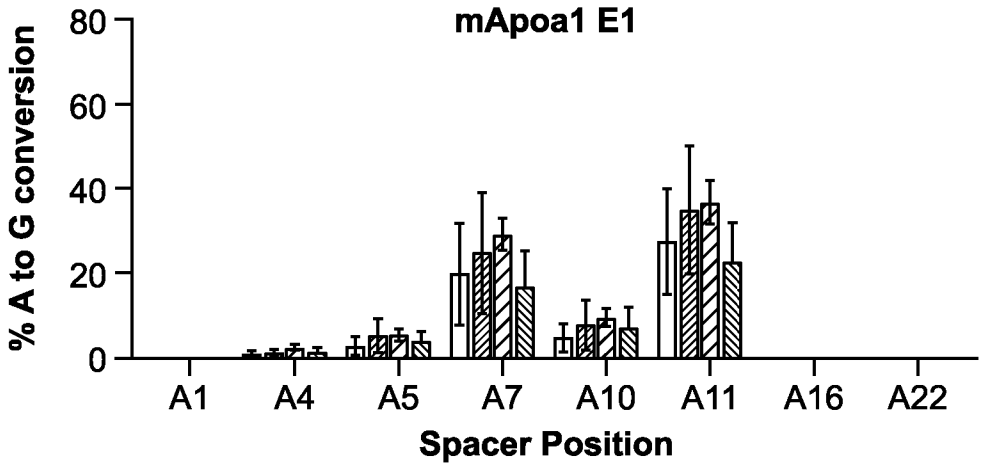

FIG. 1F

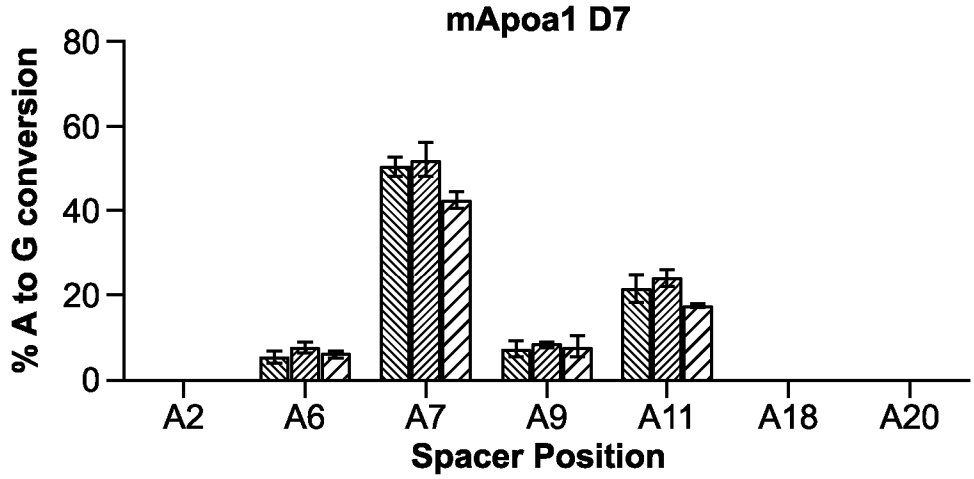

FIG. 1G

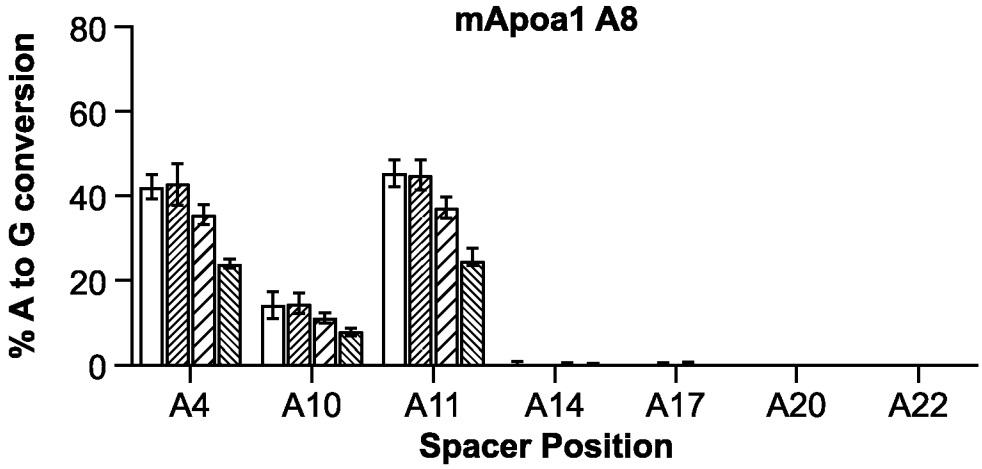

FIG. 1H

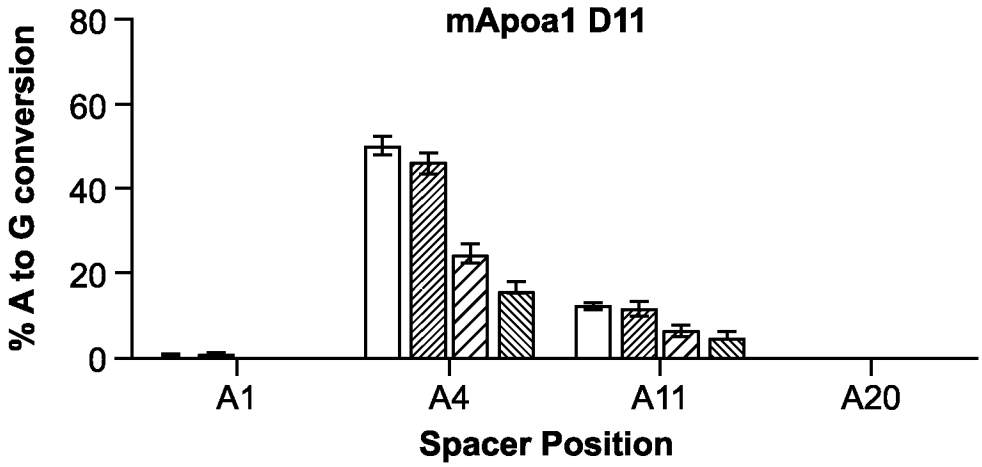

FIG. 1I

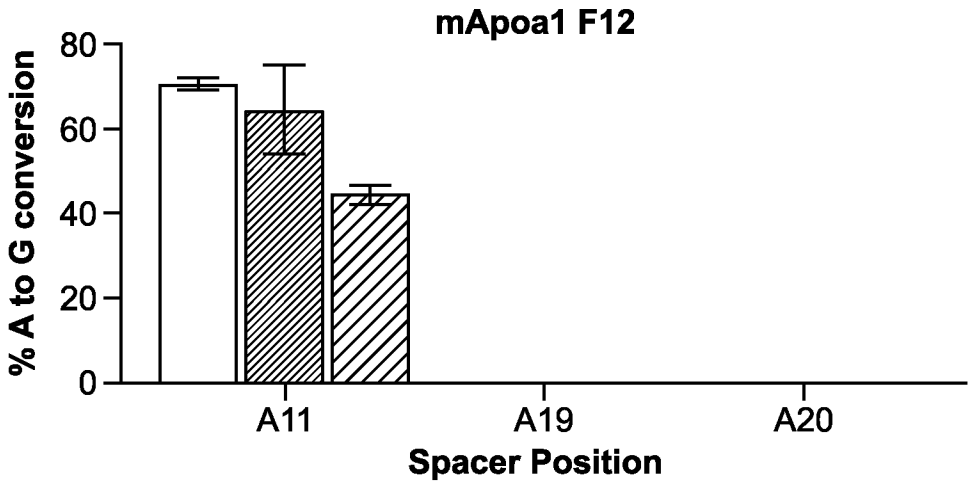
mApoa1 F12
% A to G conversion
80
60
40
20
0
A11
A19
A20
Spacer Position

FIG. 1J

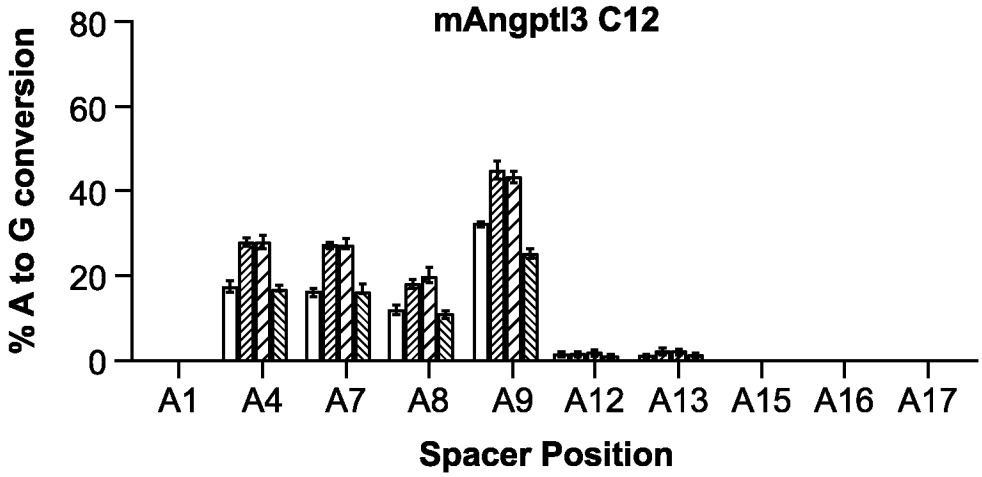
mAngptl3 C12
% A to G conversion
80
60
40
20
0
A1
A4
A7
A8
A9
A12
A13
A15
A16
A17
Spacer Position

FIG. 1K

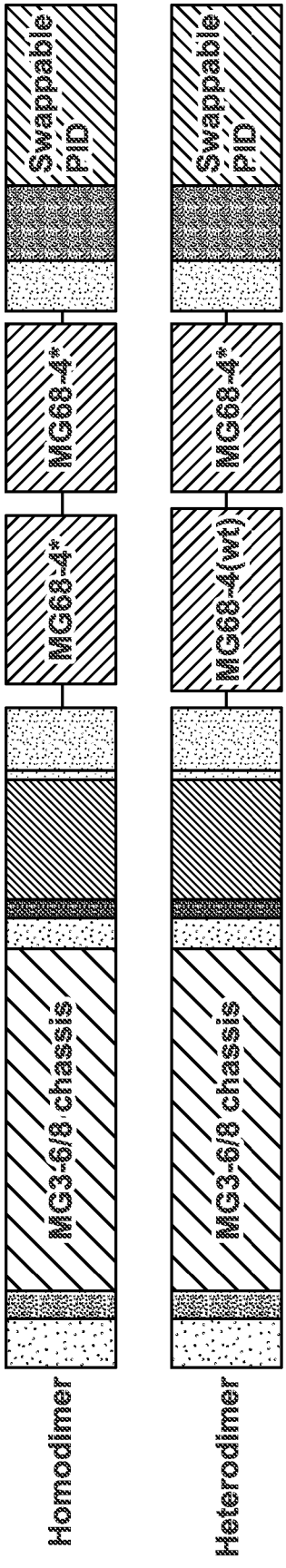
Swappable PID
MG68-4*
MG68-4*
MG3-6/8 chassis
Homodimer
Swappable PID
MG68-4*
MG68-4(wt)
MG3-6/8 chassis
Heterodimer

FIG. 2

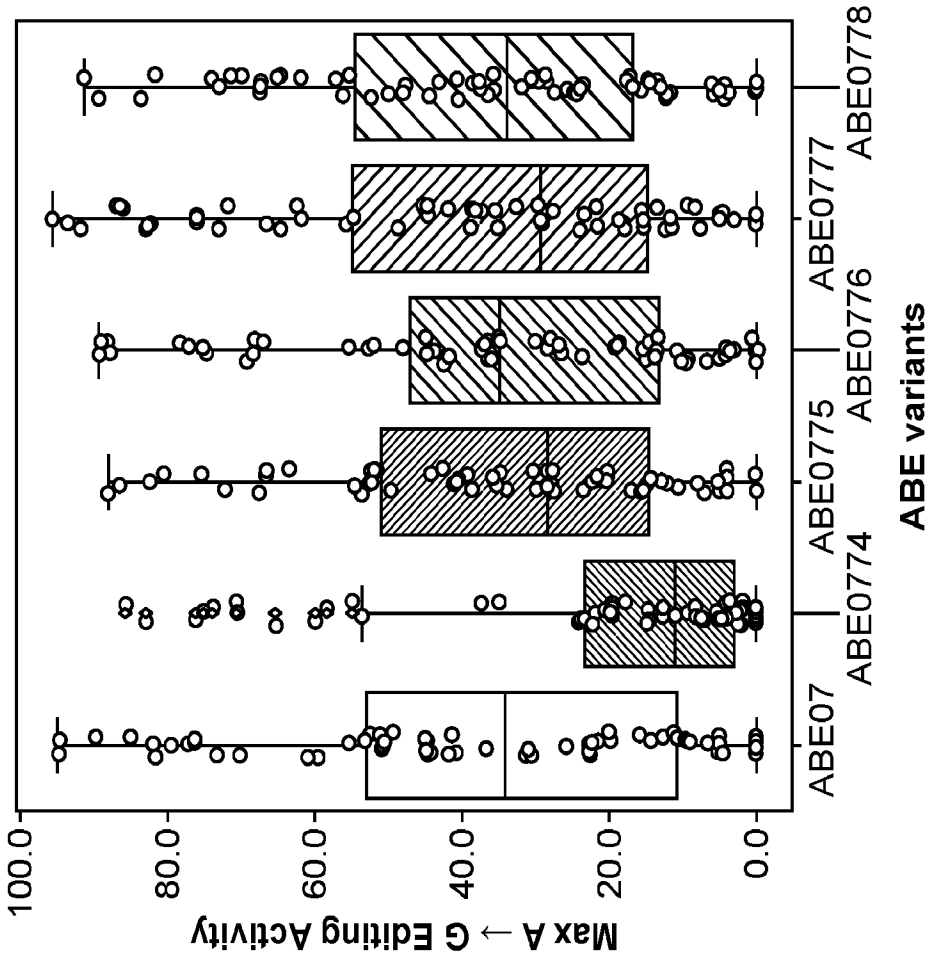
Max A → G Editing Activity
100.0
80.0
60.0
40.0
20.0
0.0
ABE07
ABE0774
ABE0775
ABE0776
ABE0777
ABE0778
ABE variants

FIG. 3B

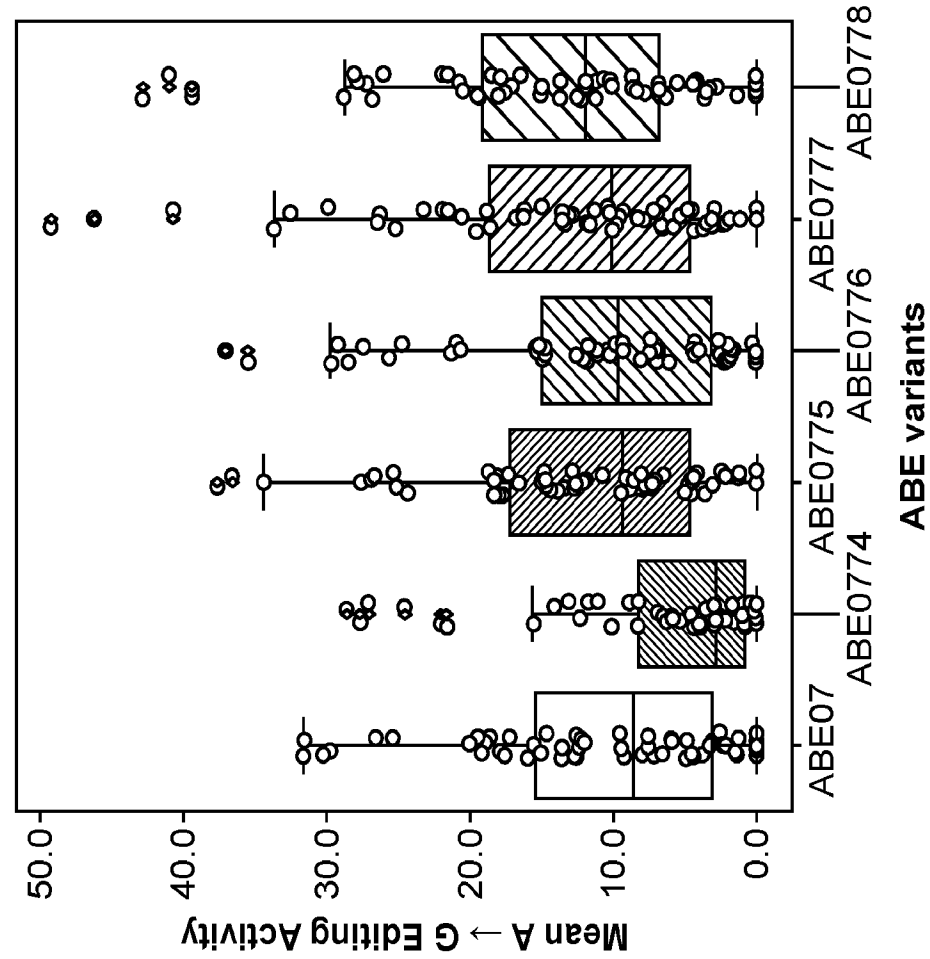
Mean A → G Editing Activity
50.0
40.0
30.0
20.0
10.0
0.0
ABE07
ABE0774
ABE0775
ABE0776
ABE0777
ABE0778
ABE variants

FIG. 3A

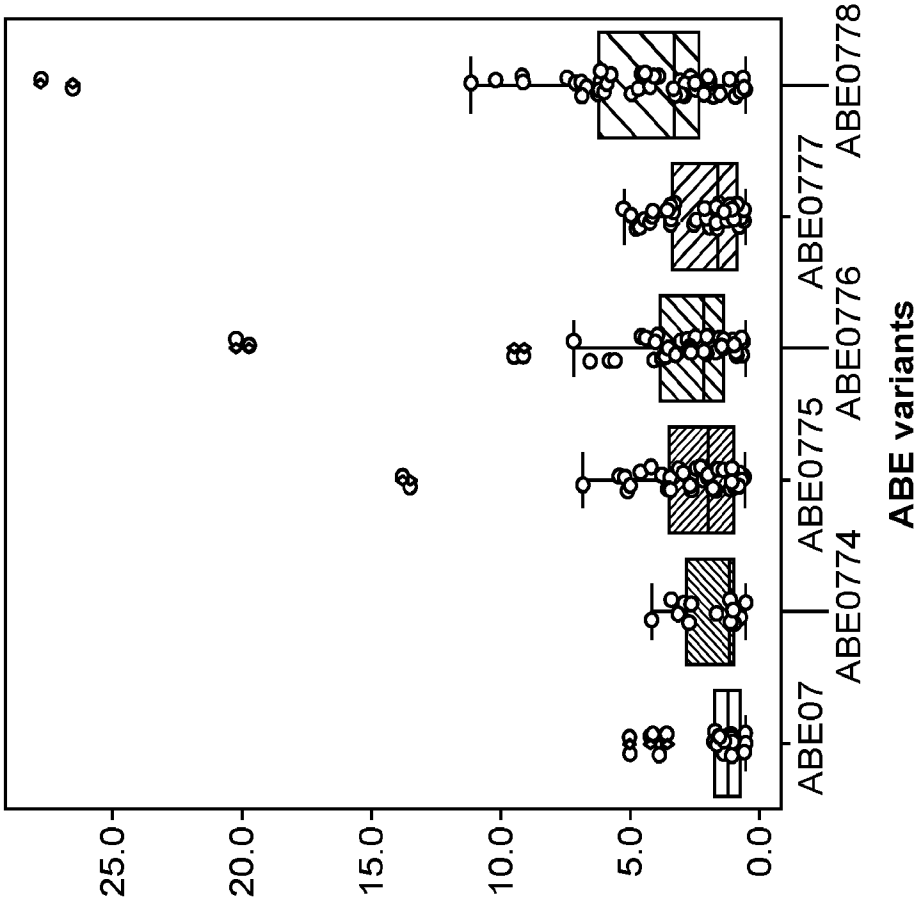
Max C → G Editing Activity
25.0
20.0
15.0
10.0
5.0
0.0
ABE07
ABE0774
ABE0775
ABE0776
ABE0777
ABE0778
ABE variants

FIG. 4B

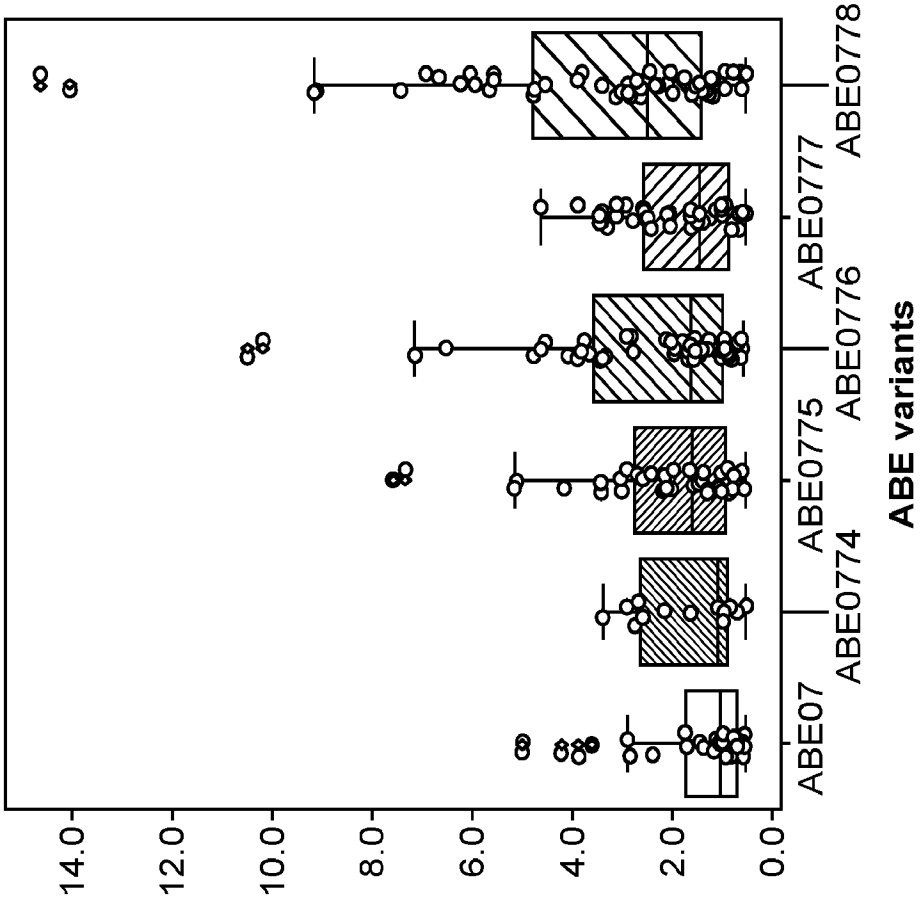
Mean C → G Editing Activity
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
ABE07
ABE0774
ABE0775
ABE0776
ABE0777
ABE0778
ABE variants

FIG. 4A

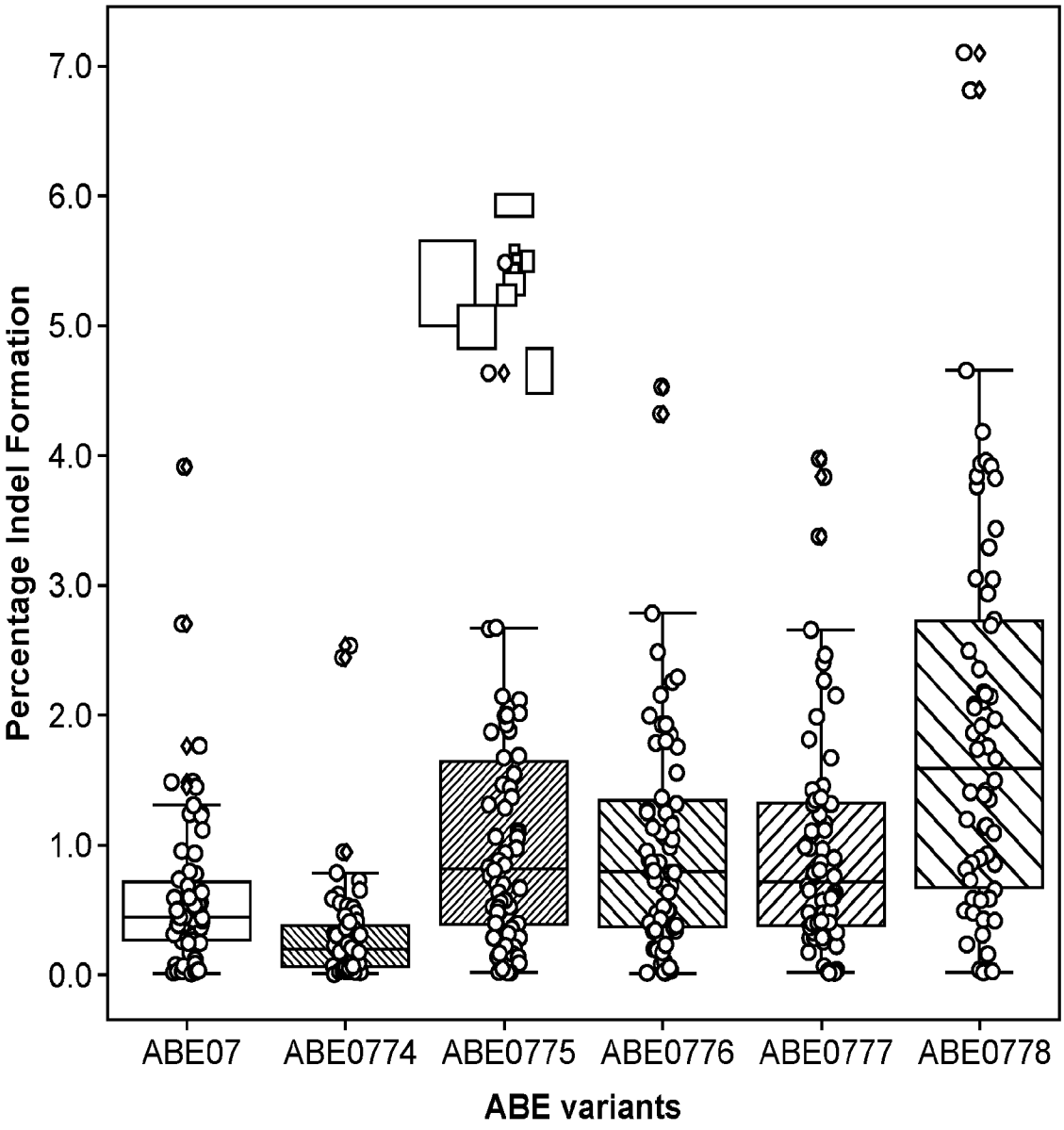
Percentage Indel Formation
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
ABE07
ABE0774
ABE0775
ABE0776
ABE0777
ABE0778
ABE variants

FIG. 5

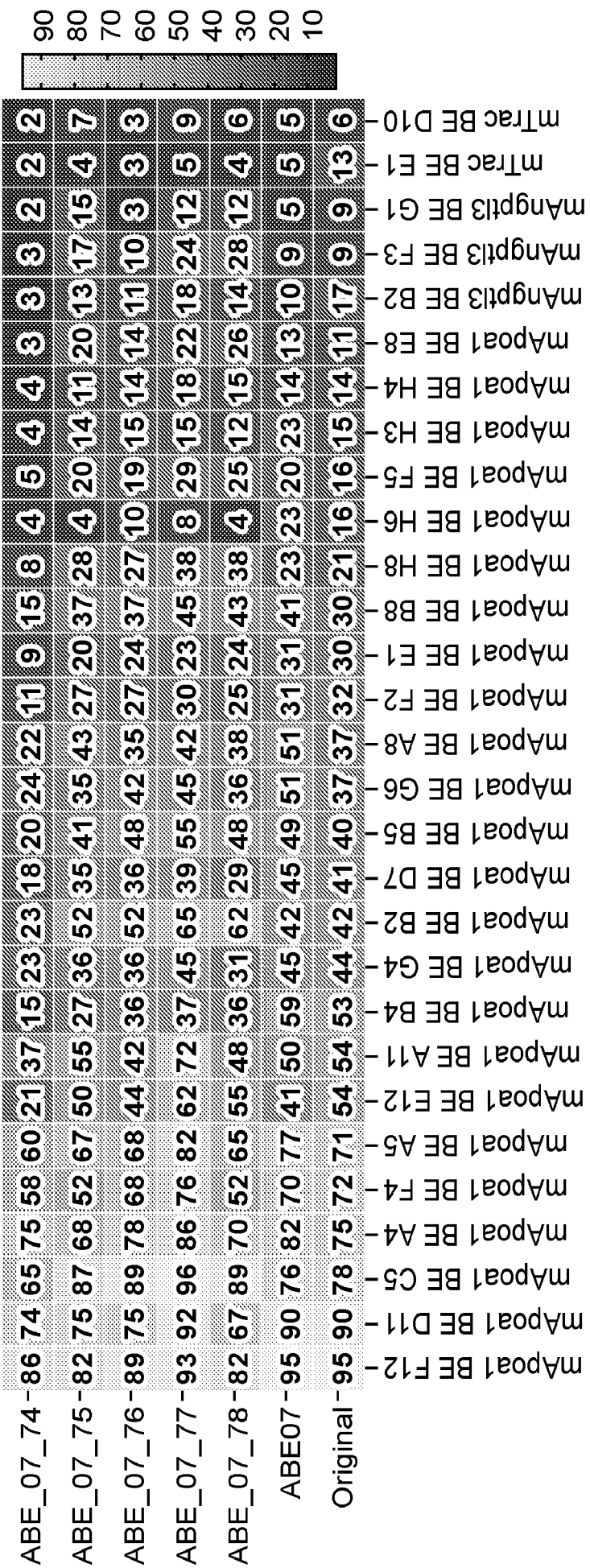
90
80
70
60
50
40
30
20
10
ABE_07_74
ABE_07_75
ABE_07_76
ABE_07_77
ABE_07_78
ABE07
Original
mApoa1 BE F12
mApoa1 BE D11
mApoa1 BE C5
mApoa1 BE A4
mApoa1 BE F4
mApoa1 BE A5
mApoa1 BE E12
mApoa1 BE A11
mApoa1 BE B4
mApoa1 BE G4
mApoa1 BE B2
mApoa1 BE D7
mApoa1 BE B5
mApoa1 BE G6
mApoa1 BE A8
mApoa1 BE F2
mApoa1 BE E1
mApoa1 BE B8
mApoa1 BE H8
mApoa1 BE H6
mApoa1 BE F5
mApoa1 BE H3
mApoa1 BE H4
mApoa1 BE E8
mAngptl3 BE B2
mAngptl3 BE F3
mAngptl3 BE G1
mTrac BE E1
mTrac BE D10

FIG. 6

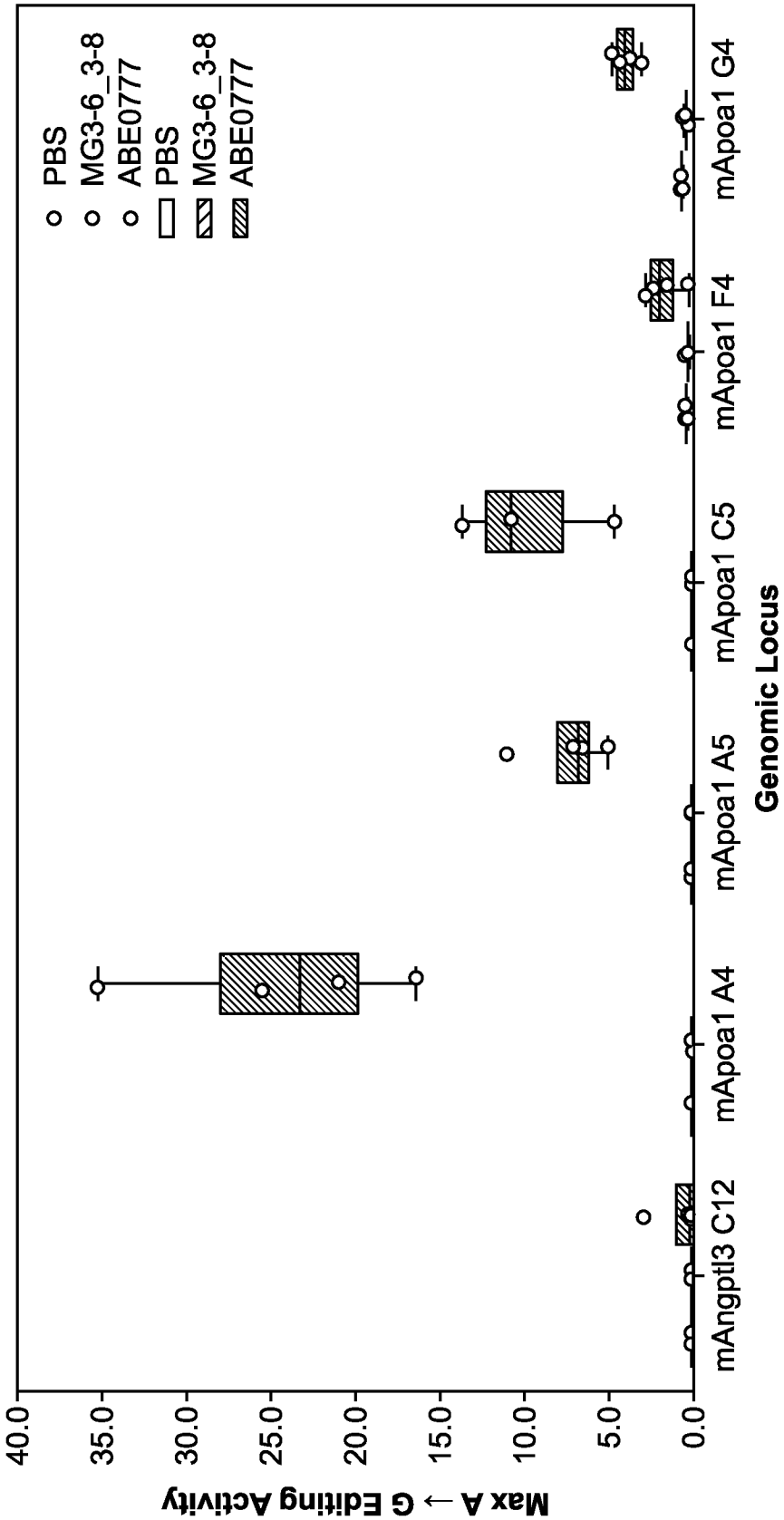
PBS
MG3-6_3-8
ABE0777
PBS
MG3-6_3-8
ABE0777
Max A → G Editing Activity
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
mAngptl3 C12
mApoa1 A4
mApoa1 A5
mApoa1 C5
mApoa1 F4
mApoa1 G4
Genomic Locus

FIG. 7

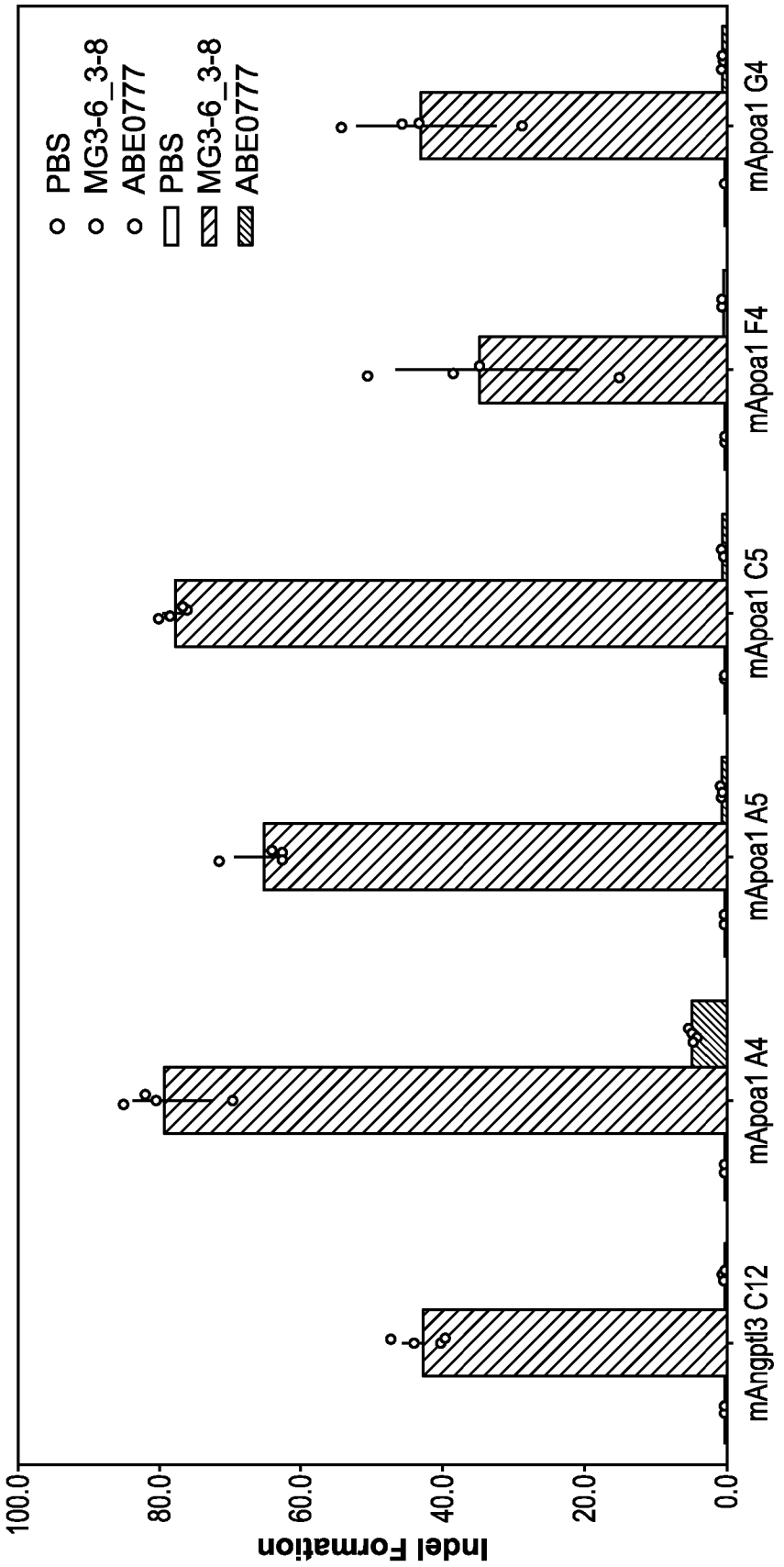
PBS
MG3-6_3-8
ABE0777
PBS
MG3-6_3-8
ABE0777
Indel Formation
100.0
80.0
60.0
40.0
20.0
0.0
mAngptl3 C12
mApoa1 A4
mApoa1 A5
mApoa1 C5
mApoa1 F4
mApoa1 G4
Genomic Locus

FIG. 8

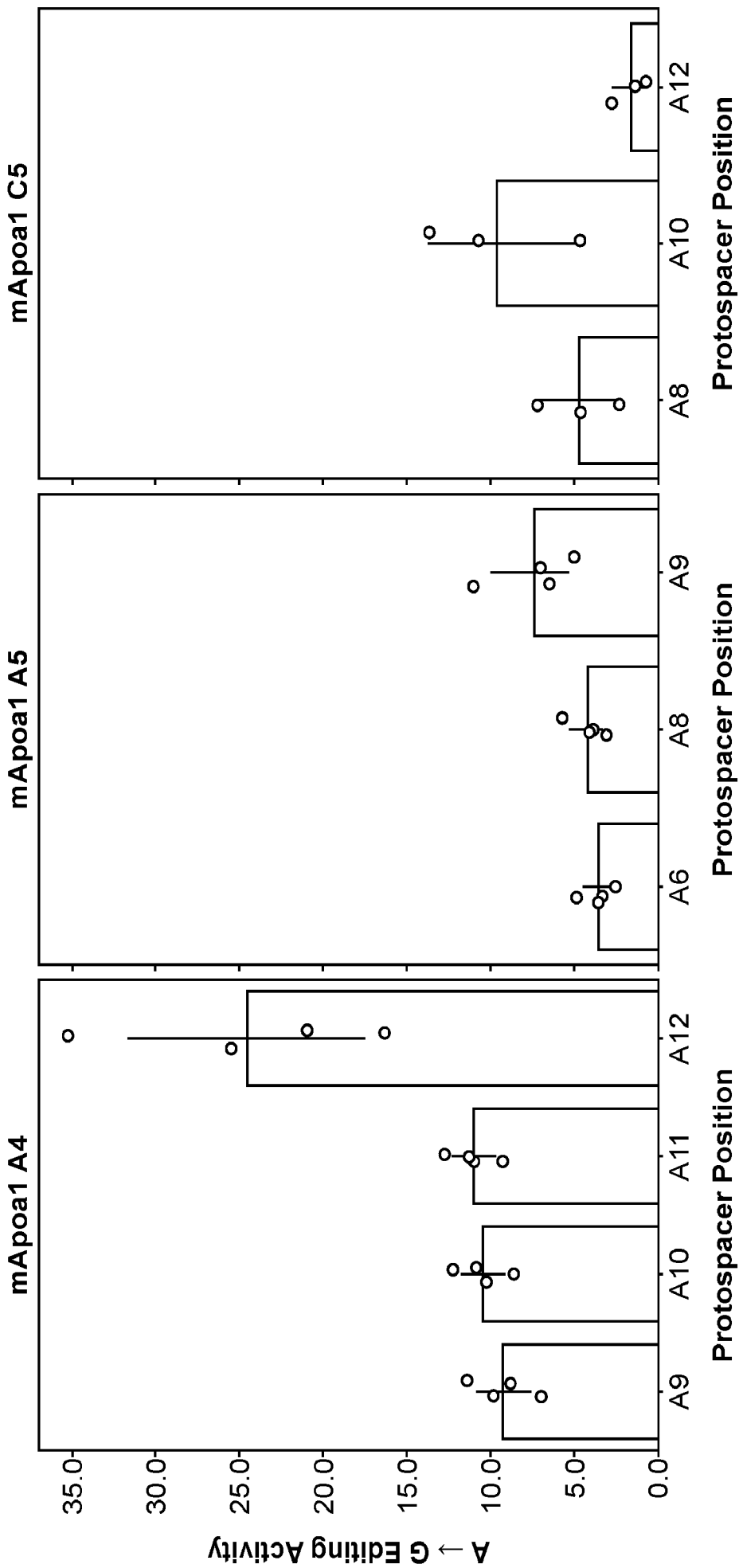
mApoa1 A4
mApoa1 A5
mApoa1 C5
A → G Editing Activity
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
A9
A10
A11
A12
A6
A8
A9
A8
A10
A12
Protospacer Position
Protospacer Position
Protospacer Position

FIG. 9

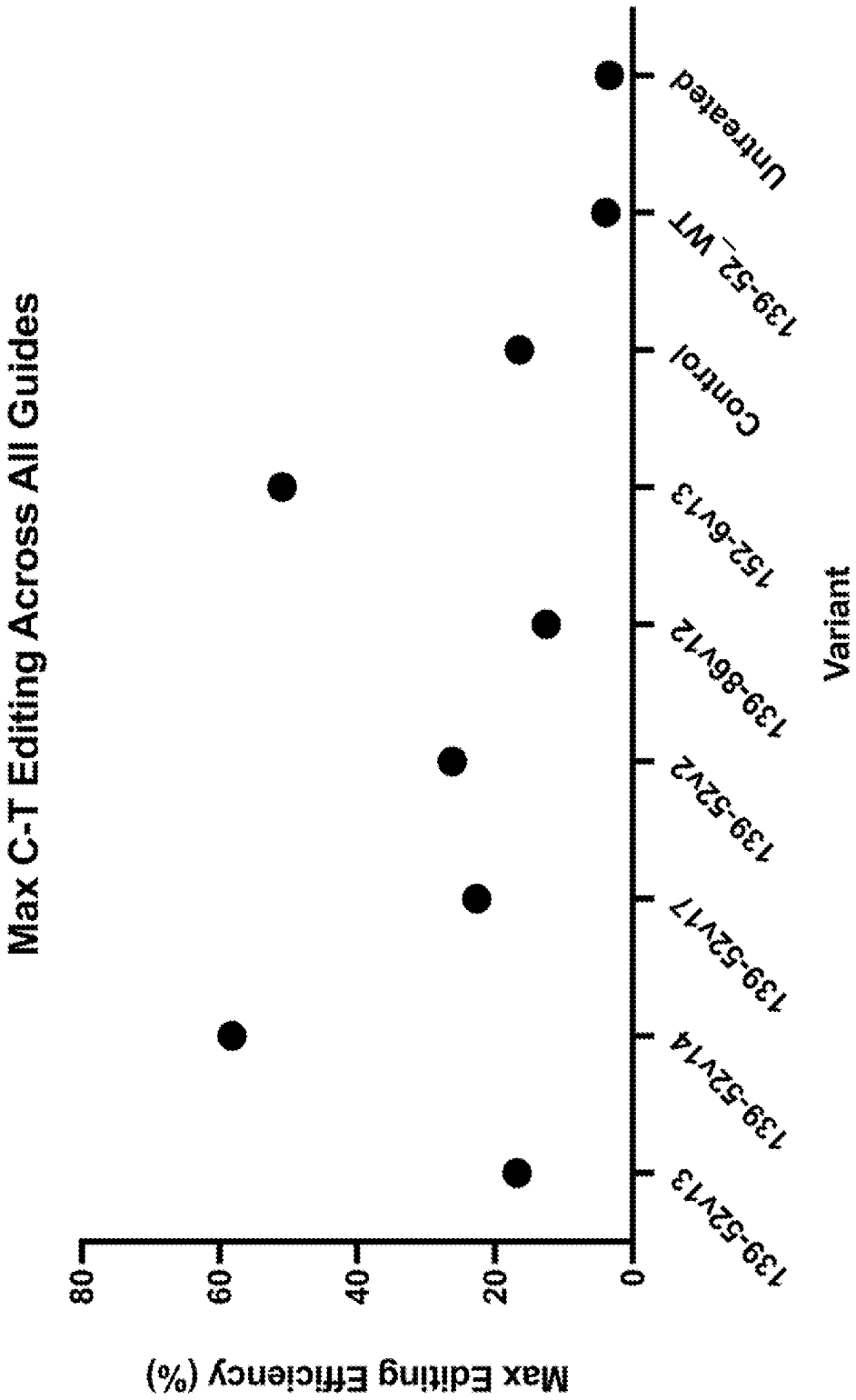
Max C-T Editing Across All Guides
Max Editing Efficiency (%)
0
20
40
60
80
Variant
139-52V13
139-52V14
139-52V17
139-52V2
139-86V12
152-6V13
Control
139-52_WT
Untreated

FIG. 10

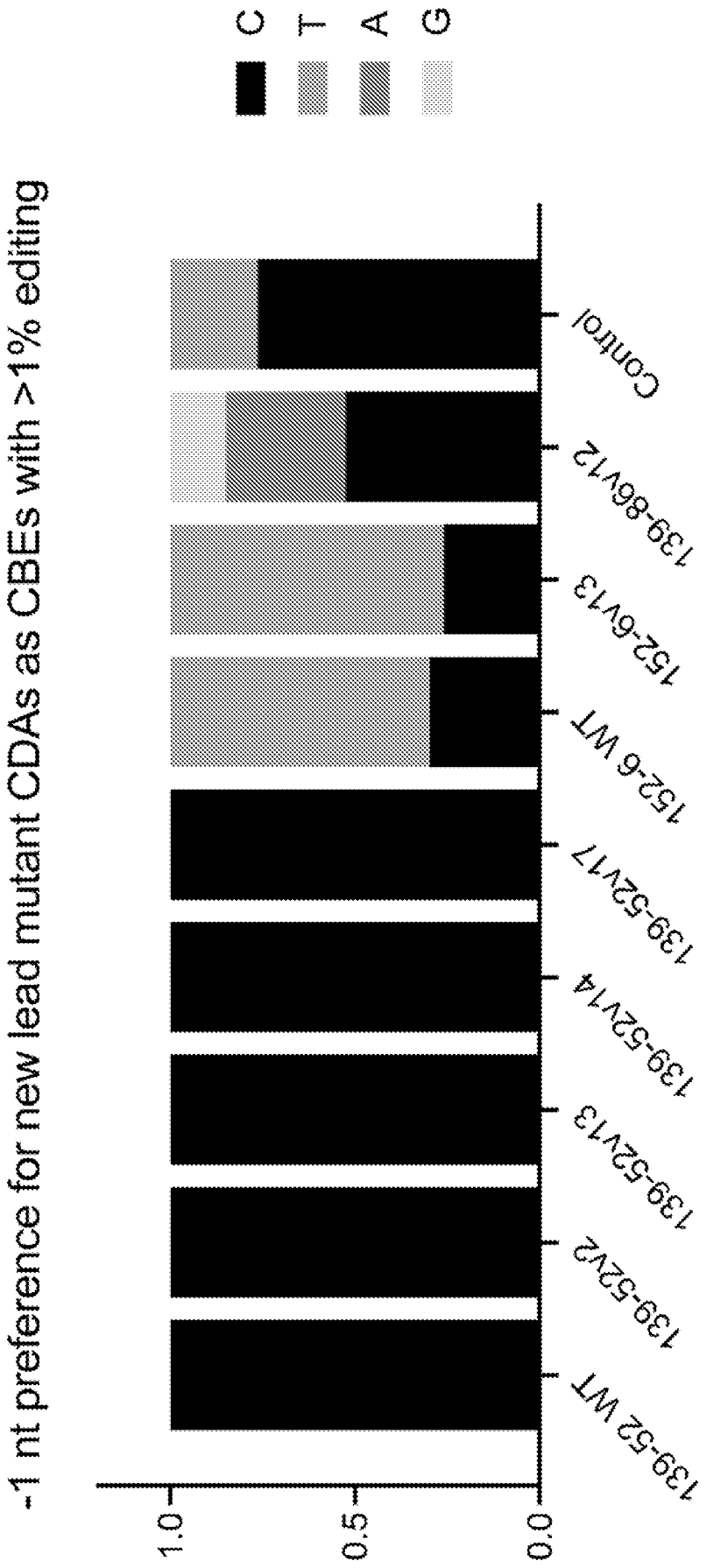
C
T
A
G
-1 nt preference for new lead mutant CDAs as CBEs with >1% editing
1.0
0.5
0.0
139-52 WT
139-52V2
139-52V13
139-52V14
139-52V17
152-6 WT
152-6V13
139-86V12
Control

FIG. 11

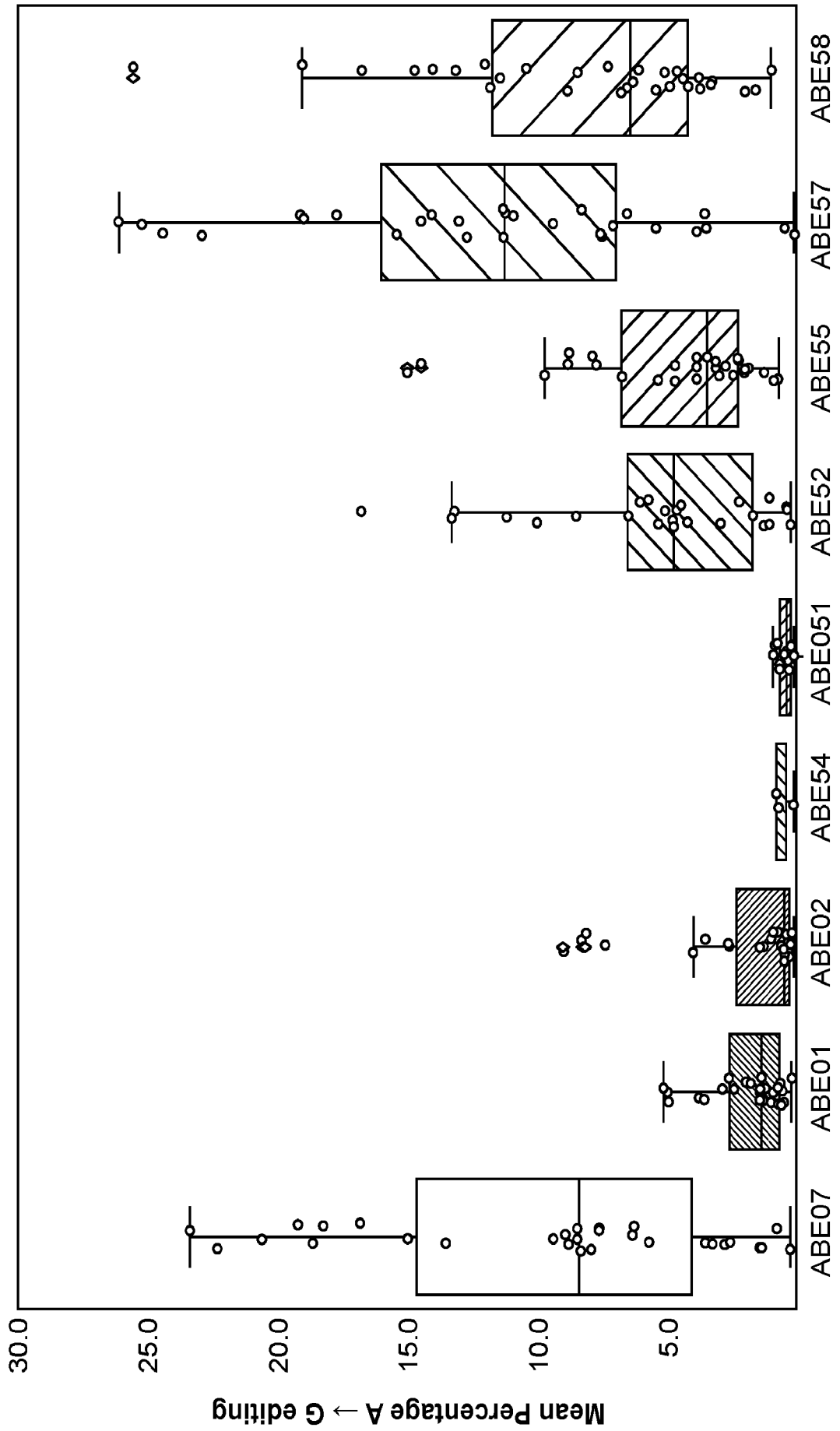
Mean Percentage A → G editing
30.0
25.0
20.0
15.0
10.0
5.0
ABE07
ABE01
ABE02
ABE54
ABE051
ABE52
ABE55
ABE57
ABE58
Variant

FIG. 12A

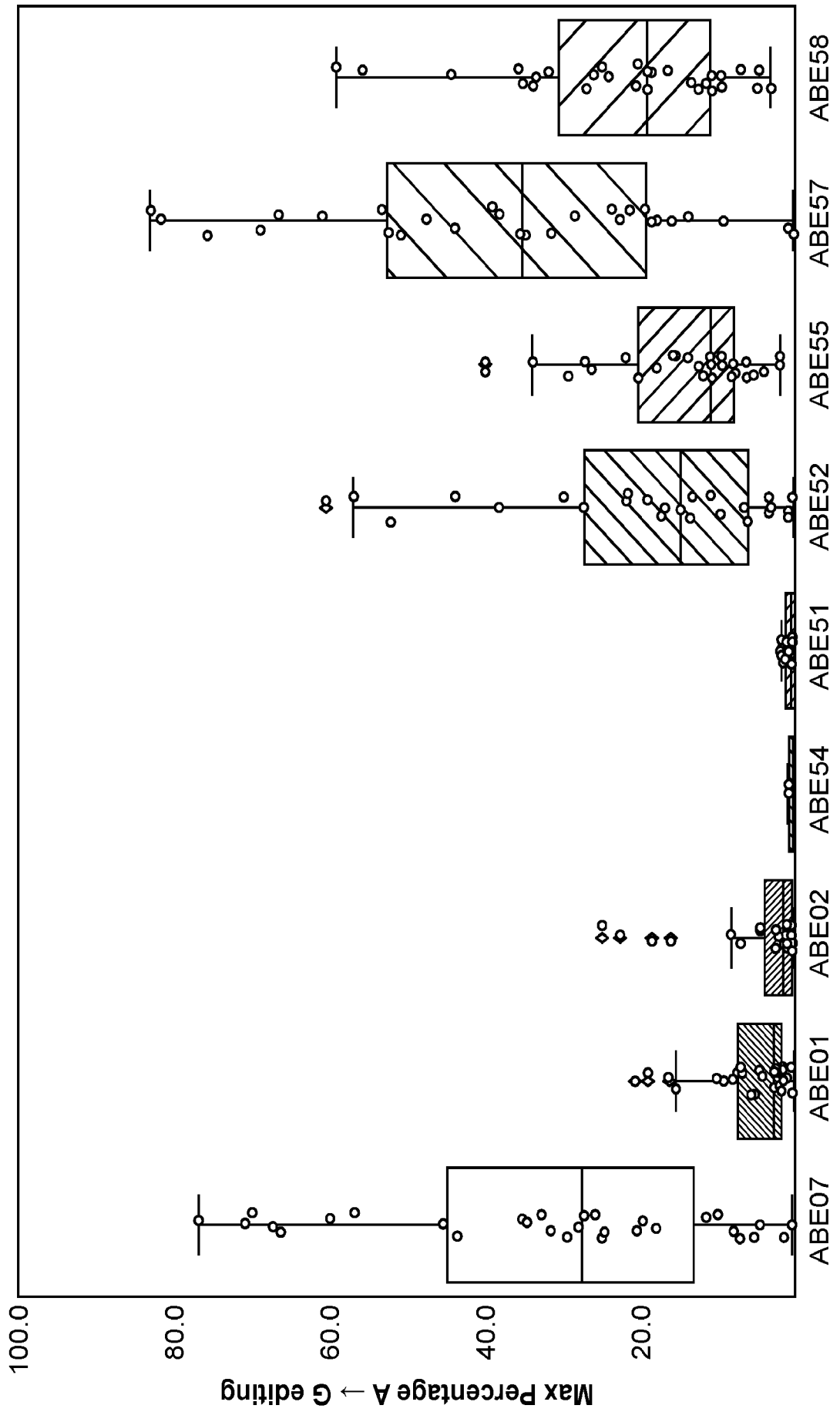
Max Percentage A → G editing
100.0
80.0
60.0
40.0
20.0
ABE07
ABE01
ABE02
ABE54
ABE51
ABE52
ABE55
ABE57
ABE58
Variant

FIG. 12B

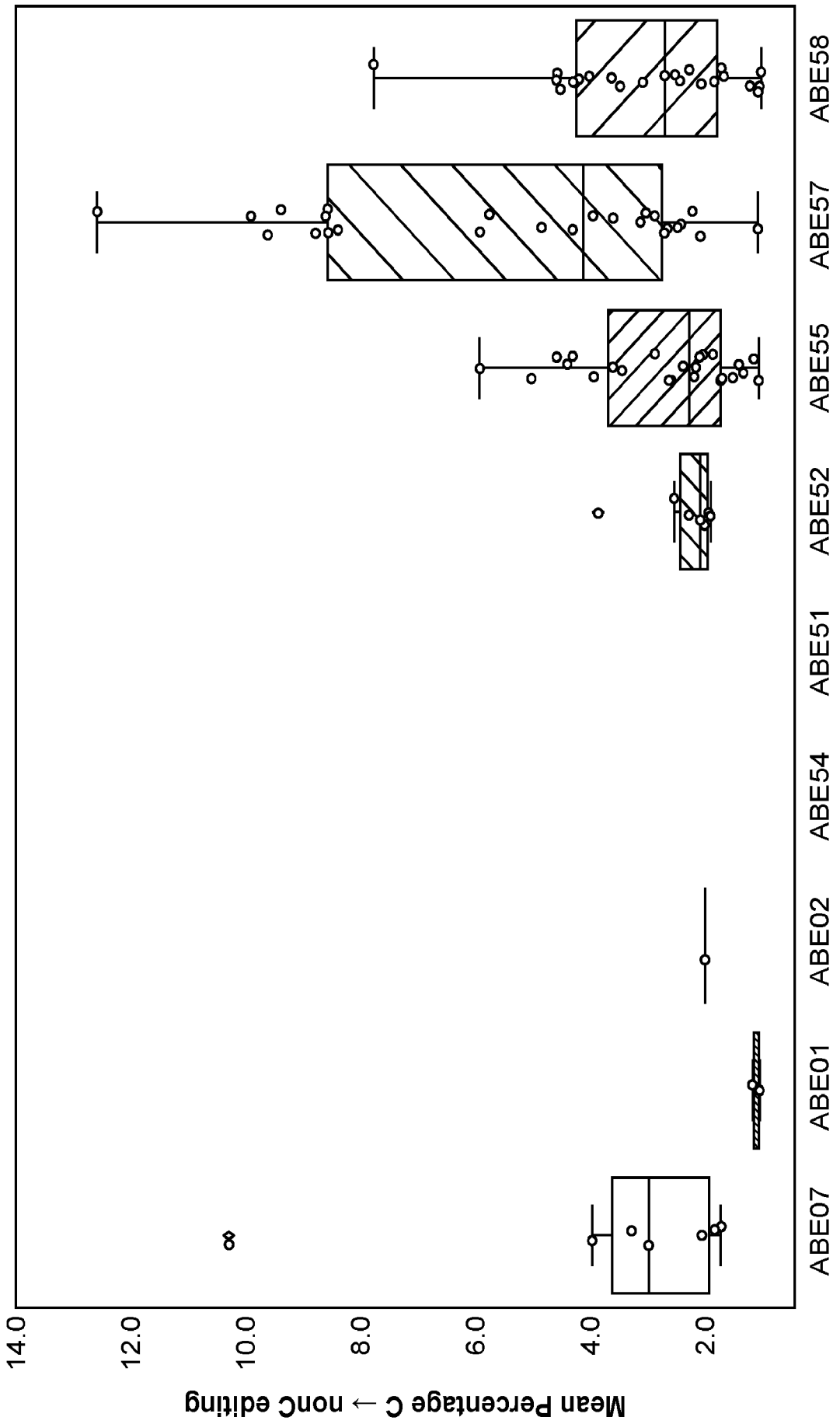
Mean Percentage C → nonC editing
14.0
12.0
10.0
8.0
6.0
4.0
2.0
ABE07
ABE01
ABE02
ABE54
ABE51
ABE52
ABE55
ABE57
ABE58
Variant

FIG. 13A

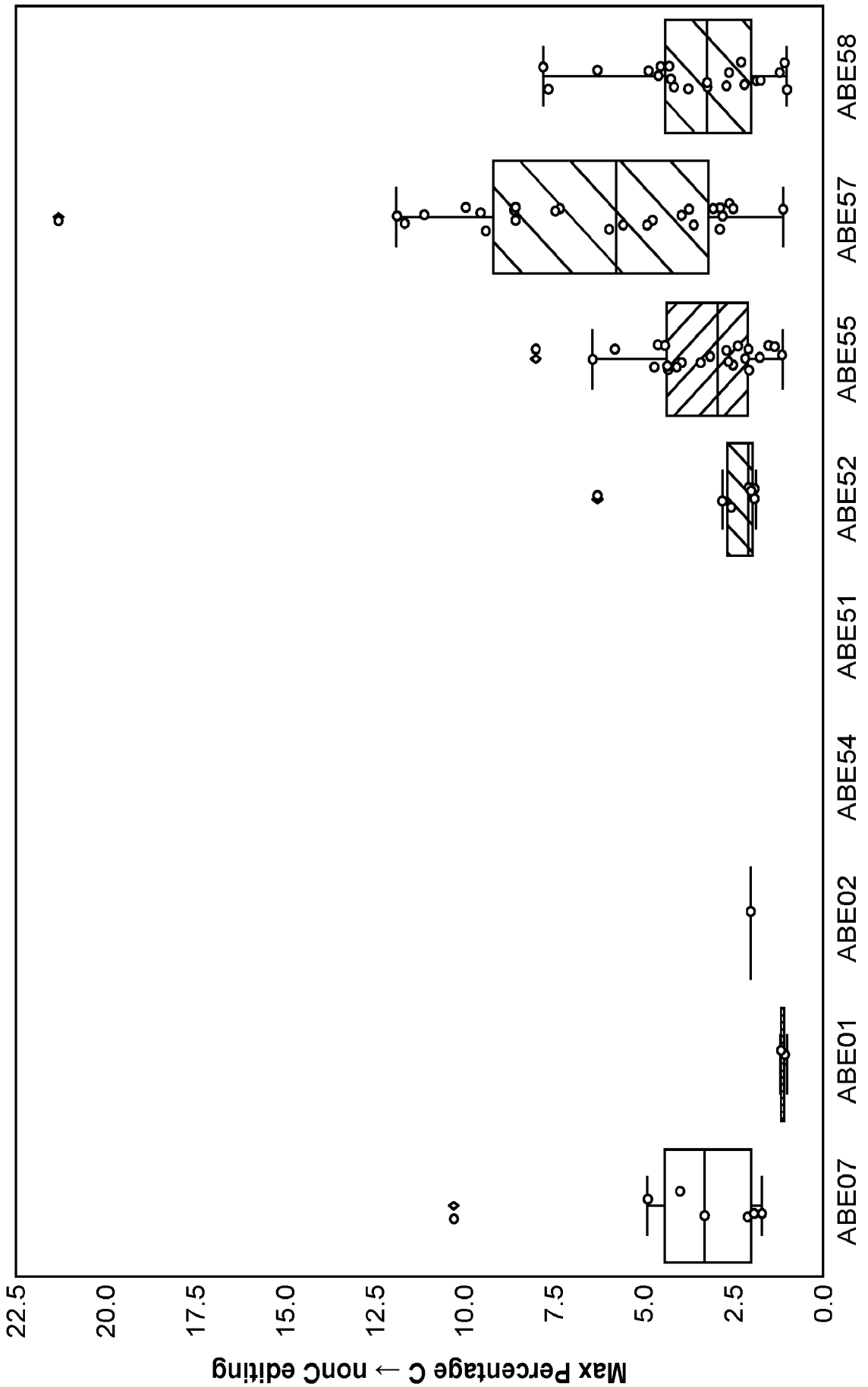
Max Percentage C → nonC editing
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
ABE07
ABE01
ABE02
ABE54
ABE51
ABE52
ABE55
ABE57
ABE58
Variant

FIG. 13B

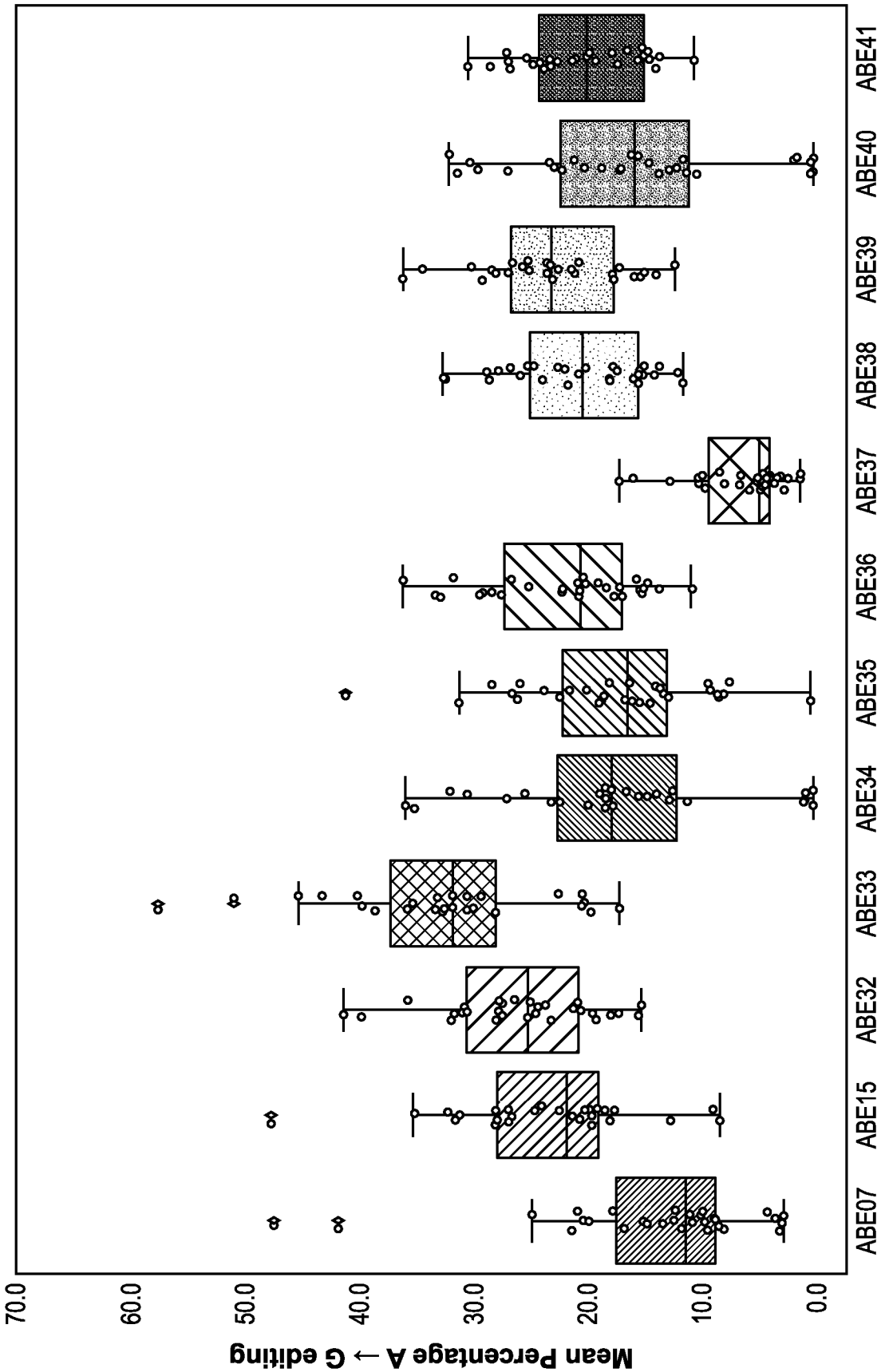
Mean Percentage A → G editing
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
ABE07
ABE15
ABE32
ABE33
ABE34
ABE35
ABE36
ABE37
ABE38
ABE39
ABE40
ABE41
Variant

FIG. 14A

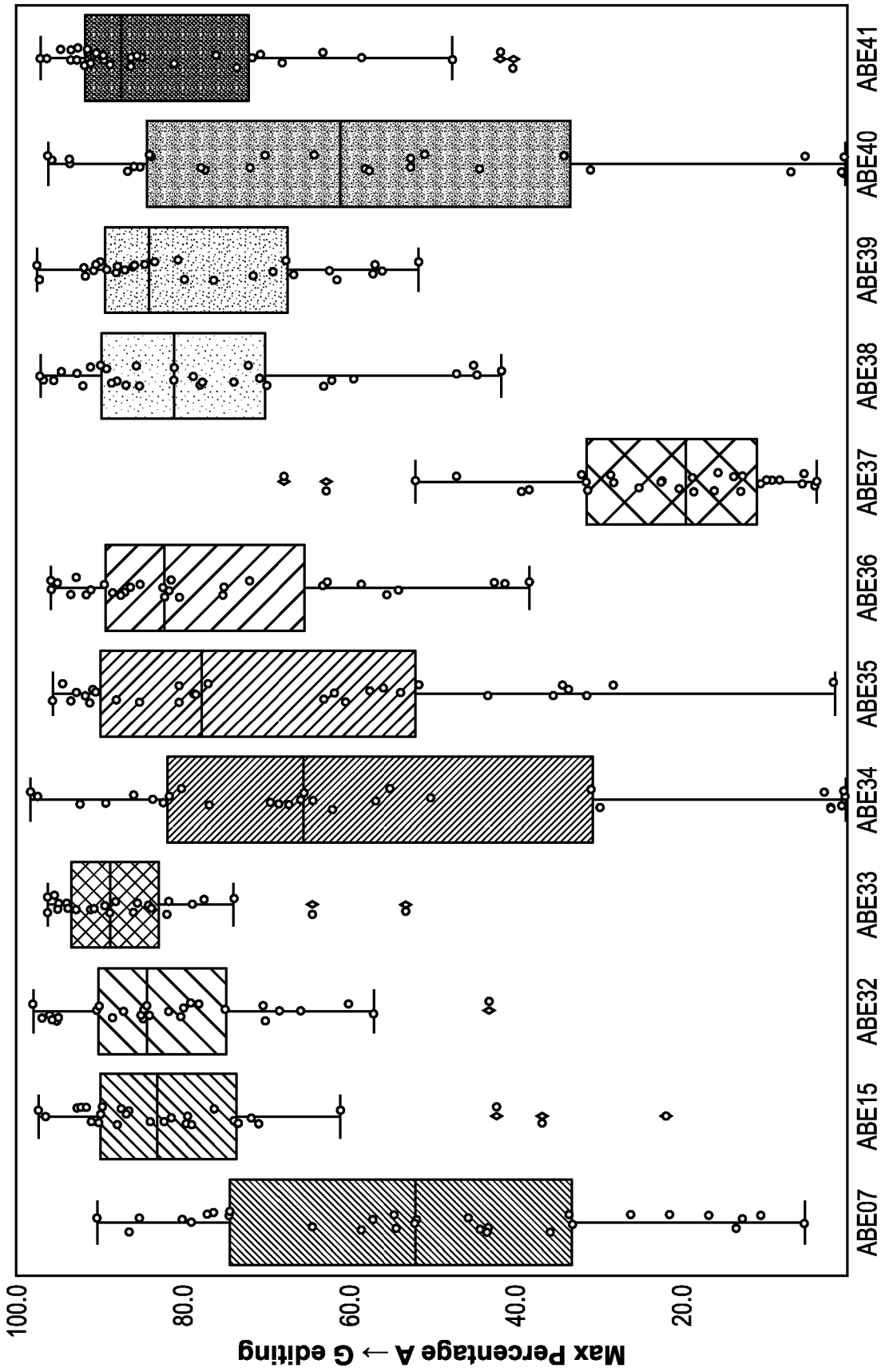
Max Percentage A → G editing
100.0
80.0
60.0
40.0
20.0
Variant
ABE07
ABE15
ABE32
ABE33
ABE34
ABE35
ABE36
ABE37
ABE38
ABE39
ABE40
ABE41

FIG. 14B

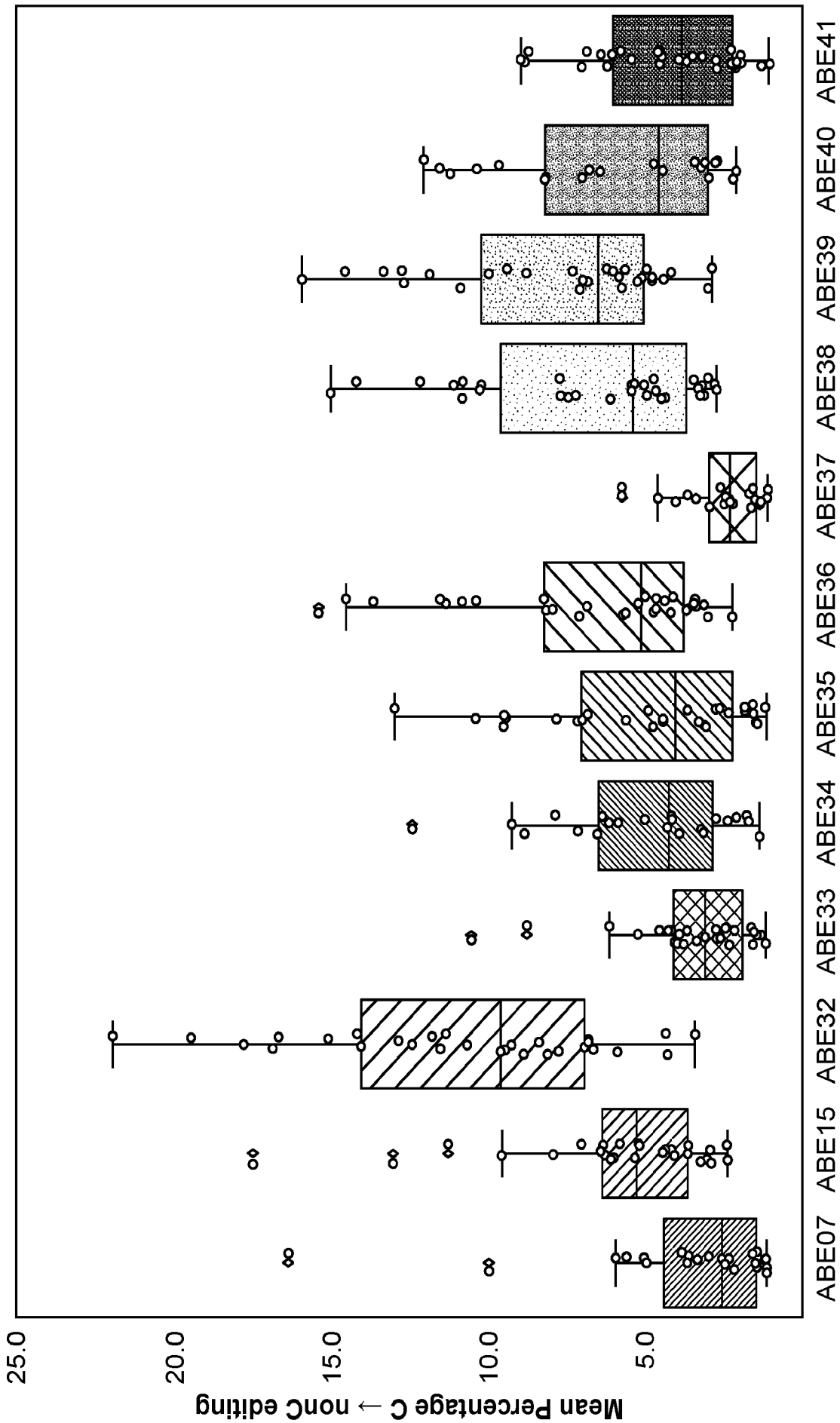
Mean Percentage C → nonC editing
25.0
20.0
15.0
10.0
5.0
ABE07
ABE15
ABE32
ABE33
ABE34
ABE35
ABE36
ABE37
ABE38
ABE39
ABE40
ABE41
Variant

FIG. 15A

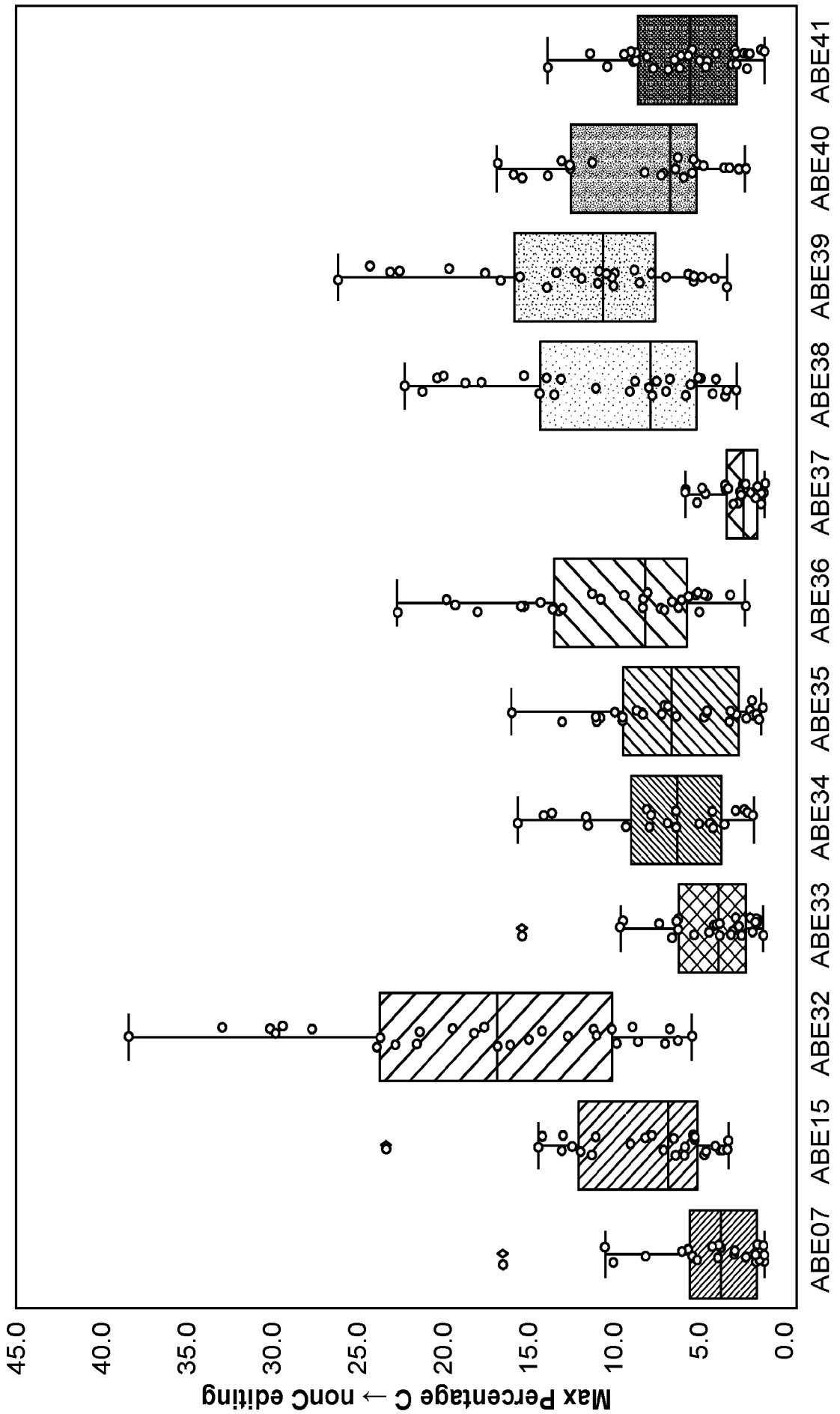
Max Percentage C → nonC editing
45.0
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
ABE07
ABE15
ABE32
ABE33
ABE34
ABE35
ABE36
ABE37
ABE38
ABE39
ABE40
ABE41
Variant

FIG. 15B

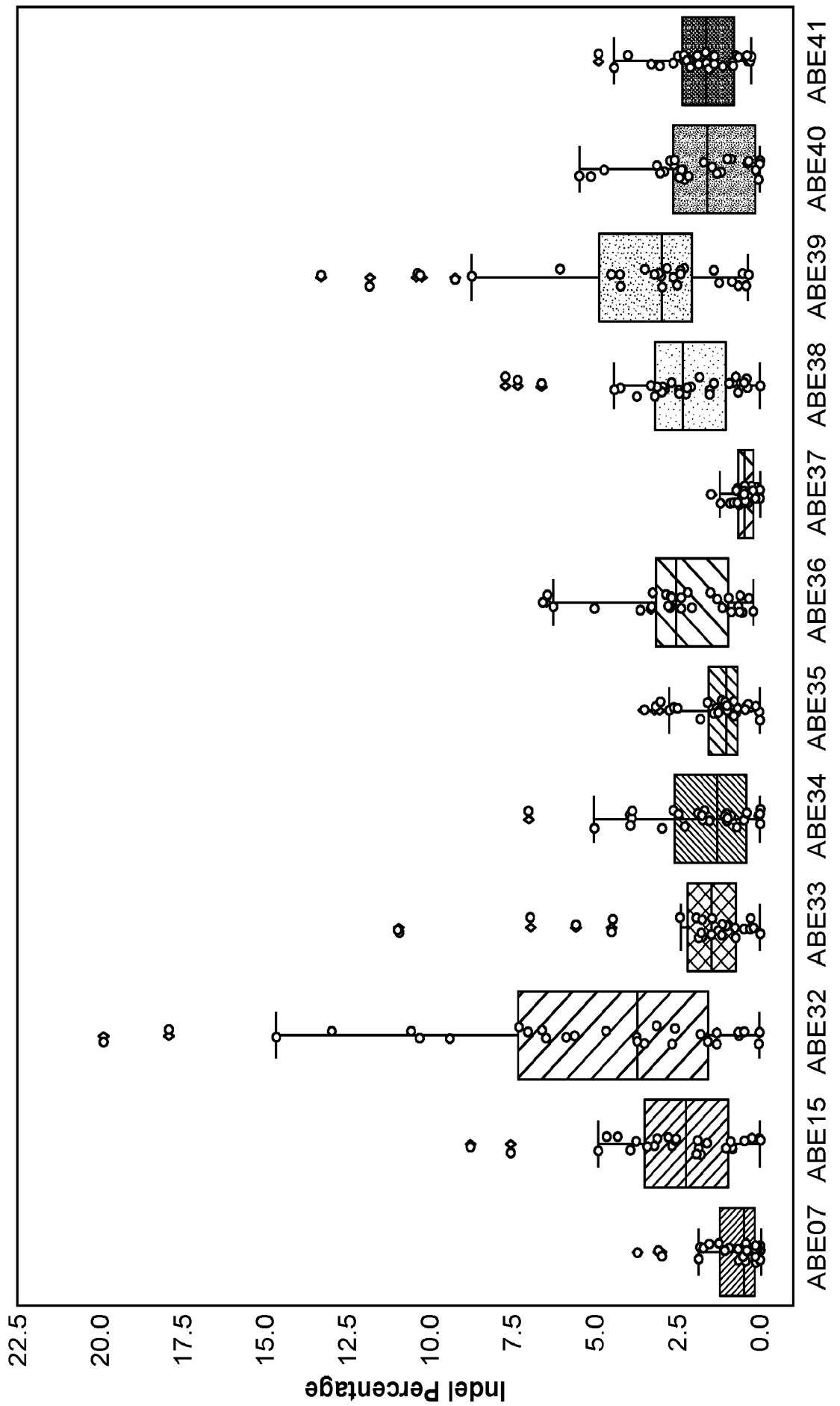
Indel Percentage
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
ABE07
ABE15
ABE32
ABE33
ABE34
ABE35
ABE36
ABE37
ABE38
ABE39
ABE40
ABE41
Variant

FIG. 16

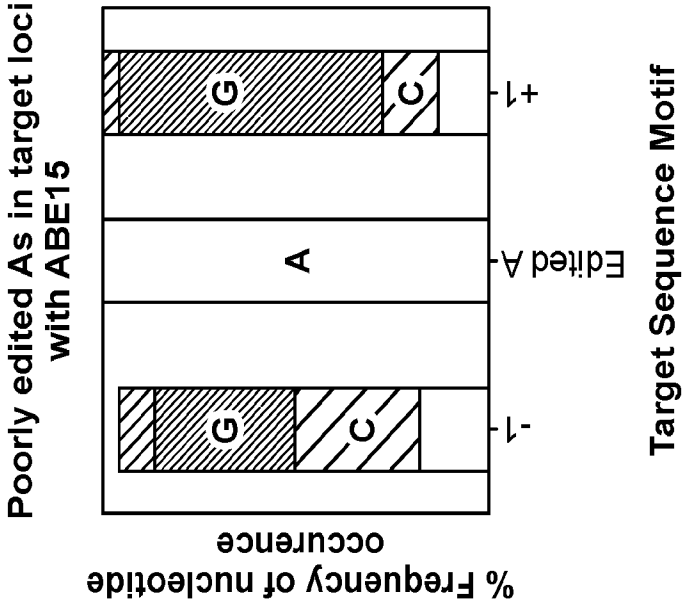

FIG. 17C

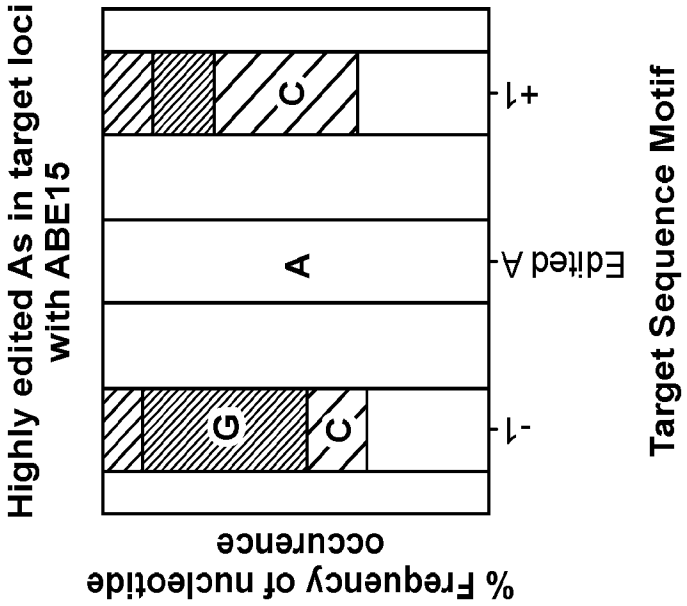

FIG. 17B

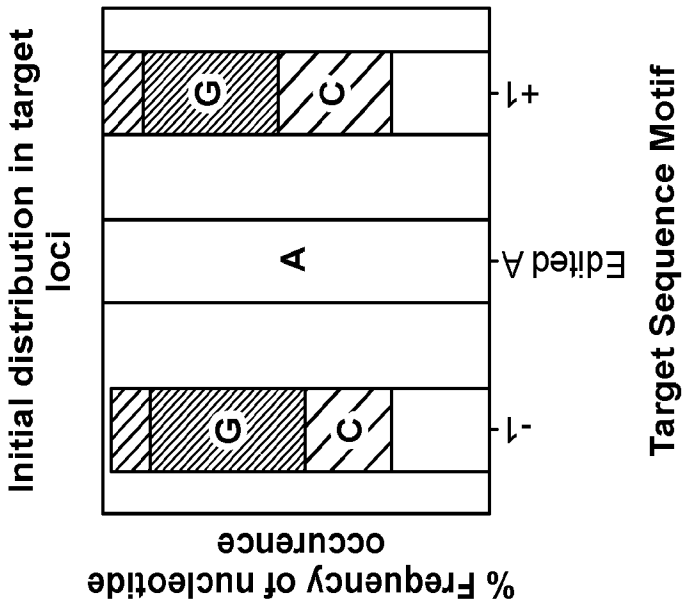

FIG. 17A

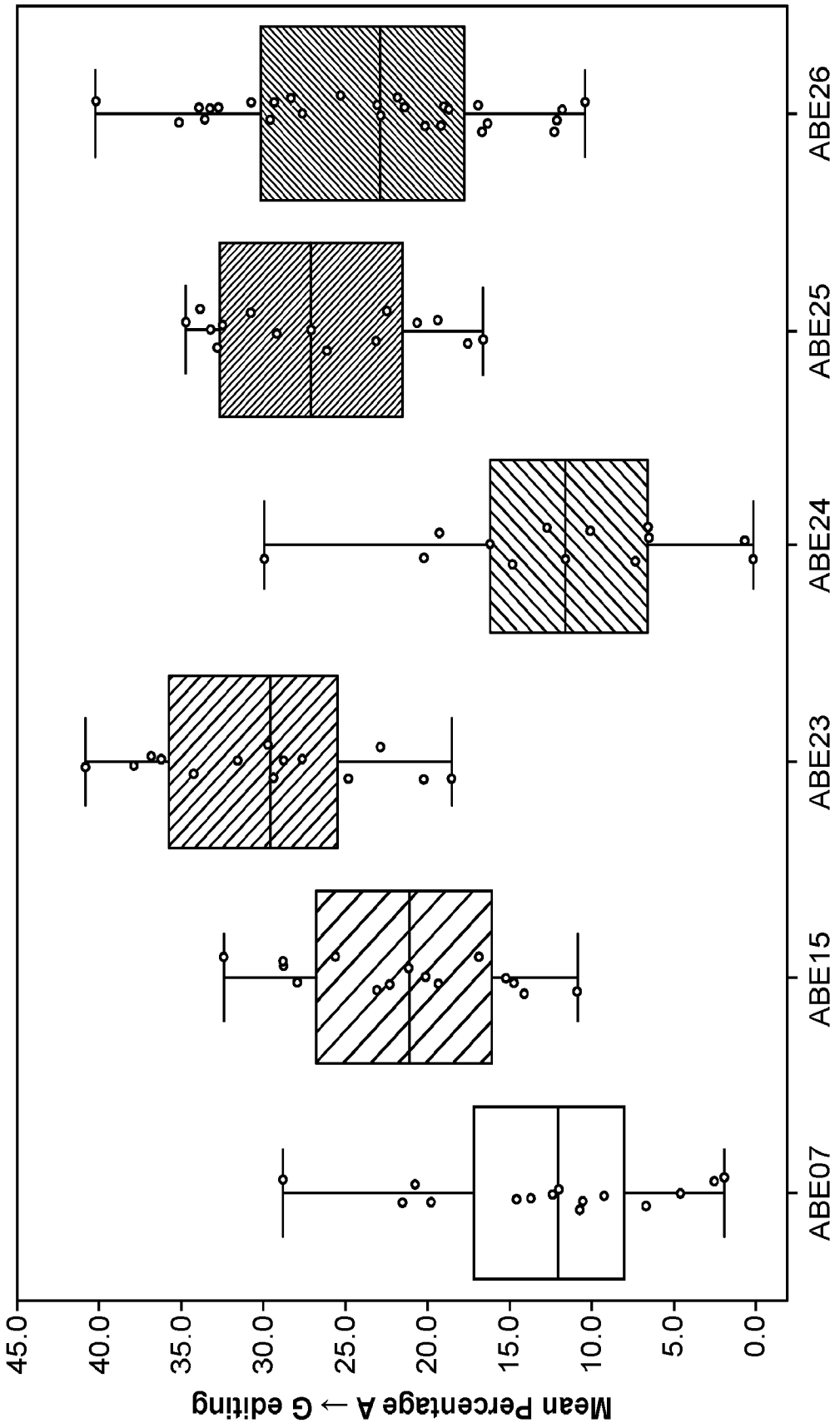
Mean Percentage A → G editing
45.0
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
ABE07
ABE15
ABE23
ABE24
ABE25
ABE26
Variant

FIG. 18A

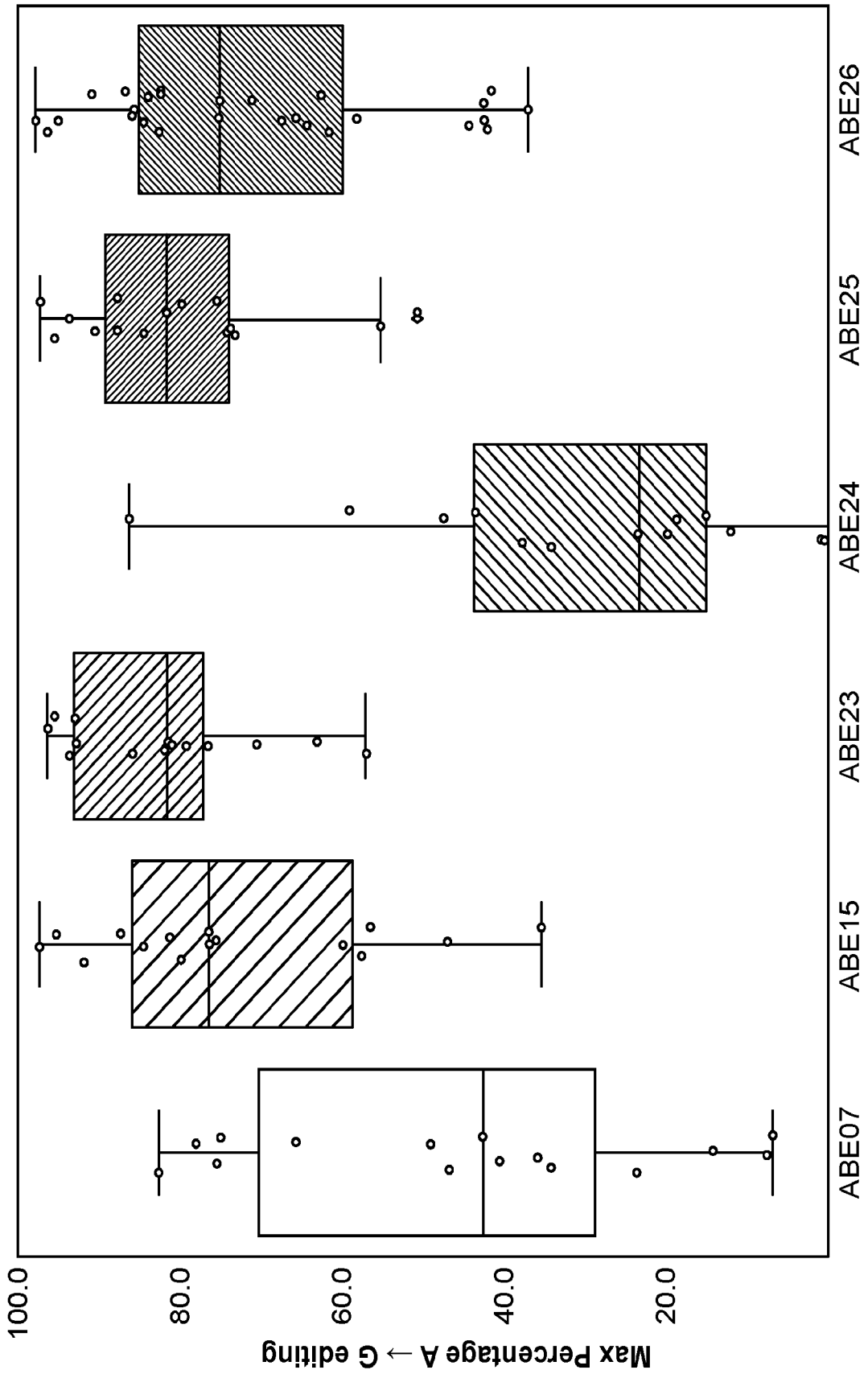
Max Percentage A → G editing
100.0
80.0
60.0
40.0
20.0
ABE07
ABE15
ABE23
ABE24
ABE25
ABE26
Variant

FIG. 18B

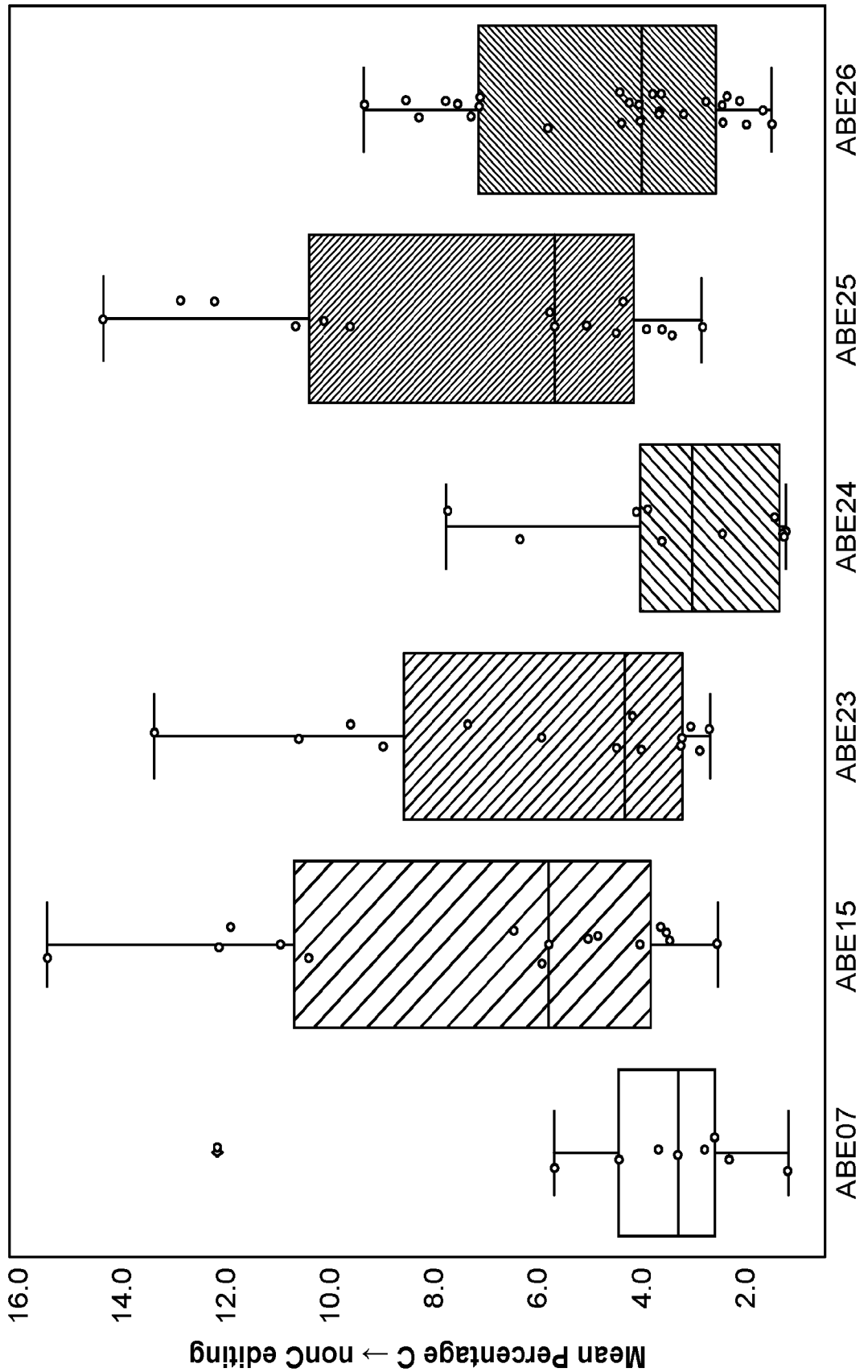
Mean Percentage C → nonC editing
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
ABE07
ABE15
ABE23
ABE24
ABE25
ABE26
Variant

FIG. 19A

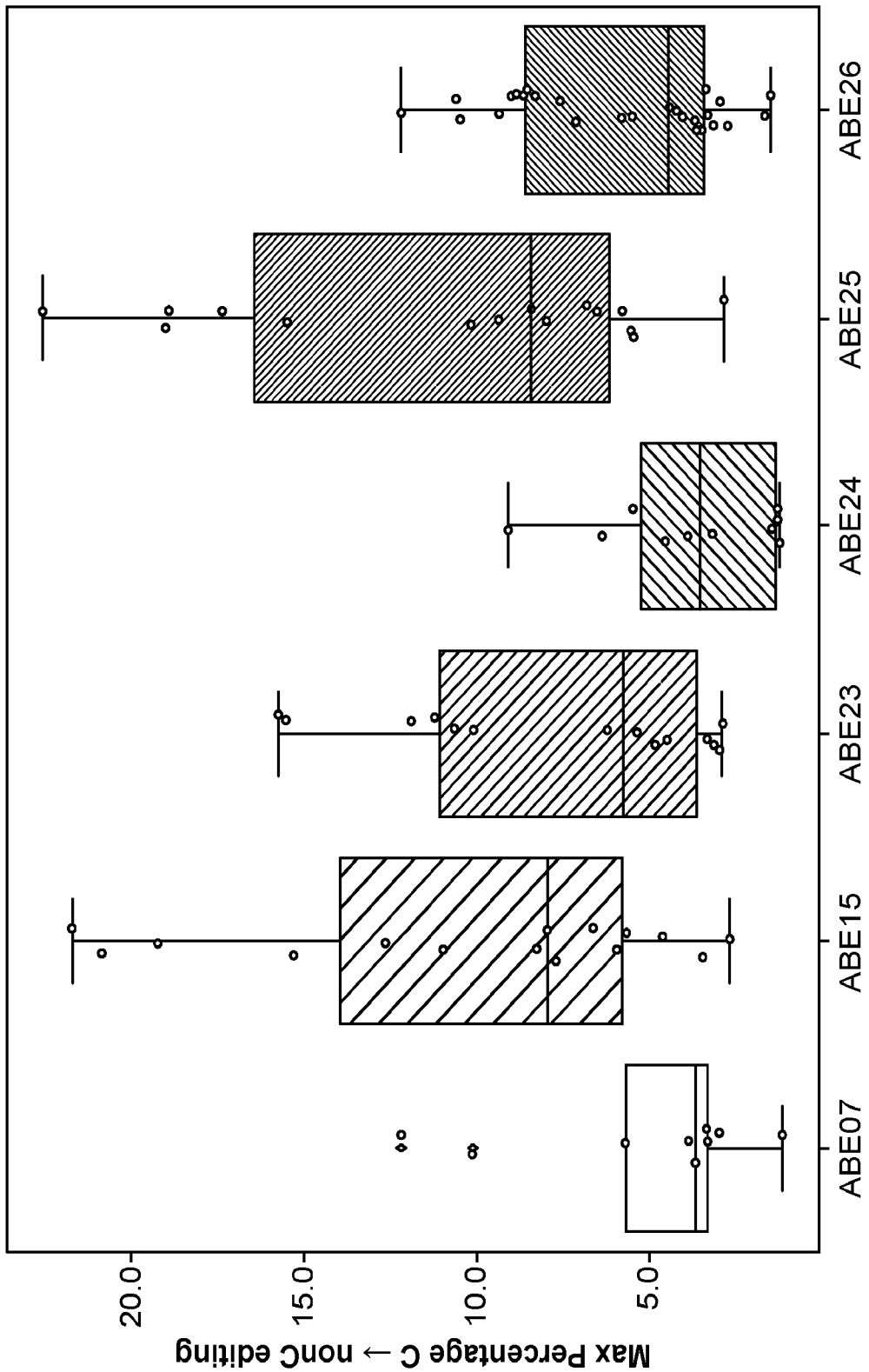
Max Percentage C → nonC editing
20.0
15.0
10.0
5.0
ABE07
ABE15
ABE23
ABE24
ABE25
ABE26
Variant

FIG. 19B

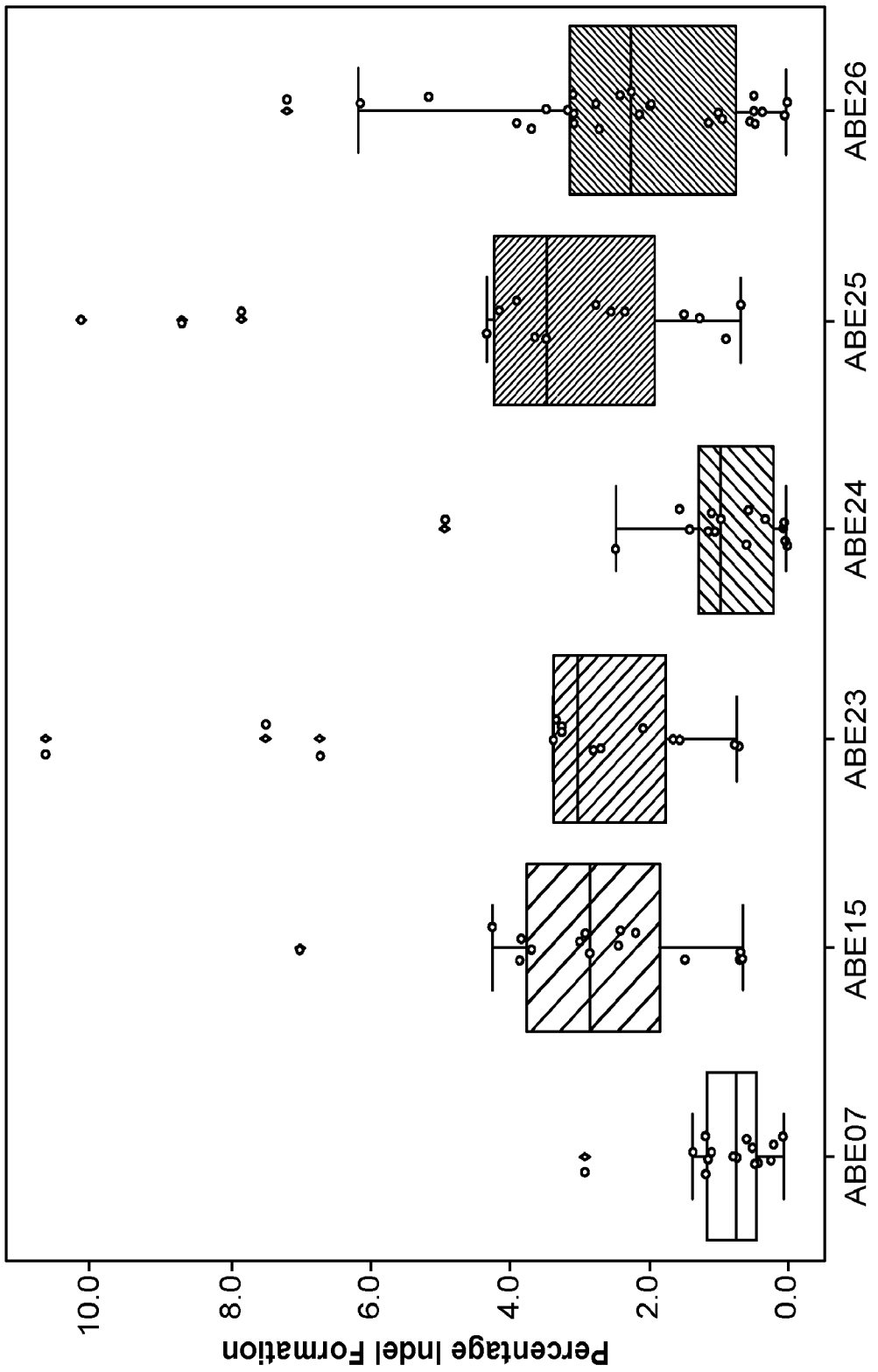
Percentage Indel Formation
10.0
8.0
6.0
4.0
2.0
0.0
ABE07
ABE15
ABE23
ABE24
ABE25
ABE26
Variant

FIG. 20A

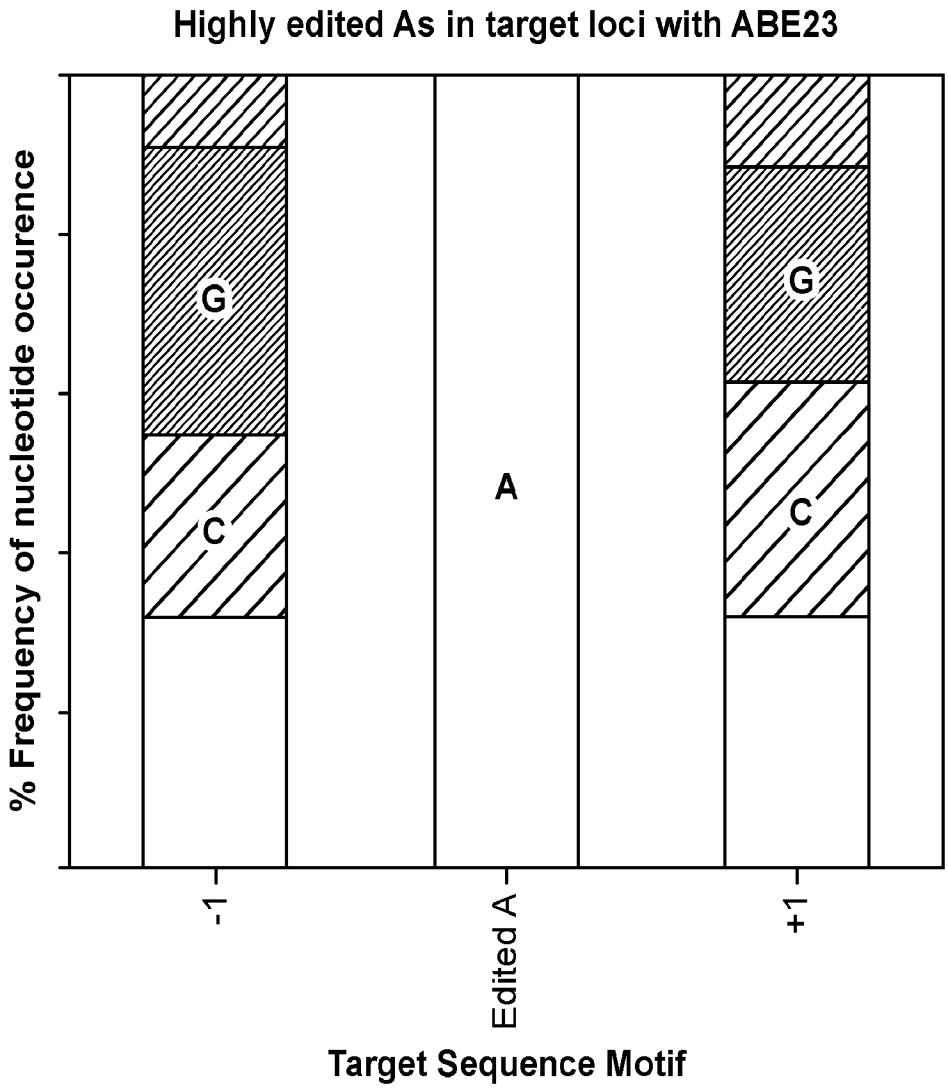
Highly edited As in target loci with ABE23
% Frequency of nucleotide occurence
G
C
A
G
C
-1
Edited A
+1
Target Sequence Motif

FIG. 20B

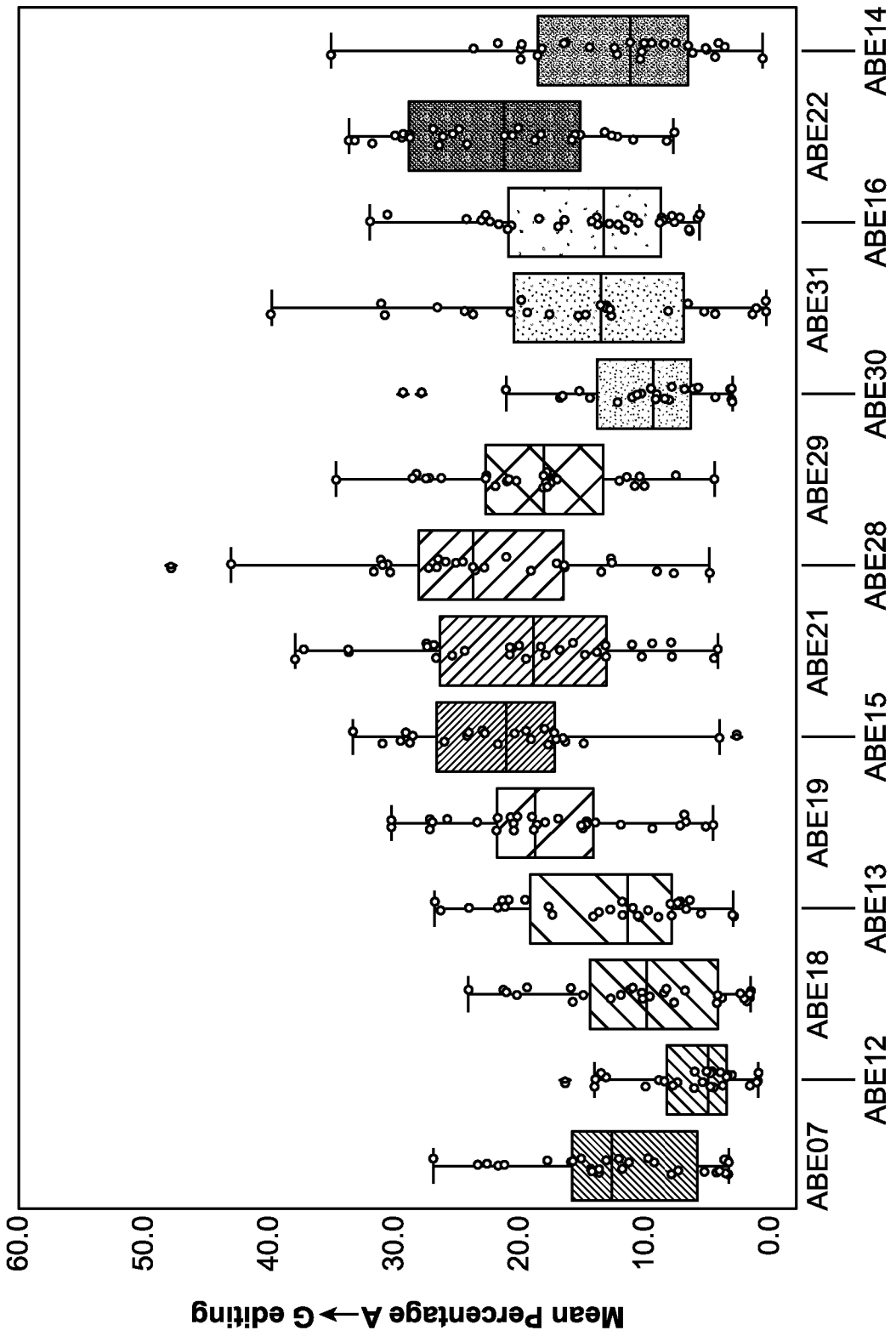
Mean Percentage A→G editing
60.0
50.0
40.0
30.0
20.0
10.0
0.0
ABE07
ABE12
ABE18
ABE13
ABE19
ABE15
ABE21
ABE28
ABE29
ABE30
ABE31
ABE16
ABE22
ABE14
Variant

FIG. 21A

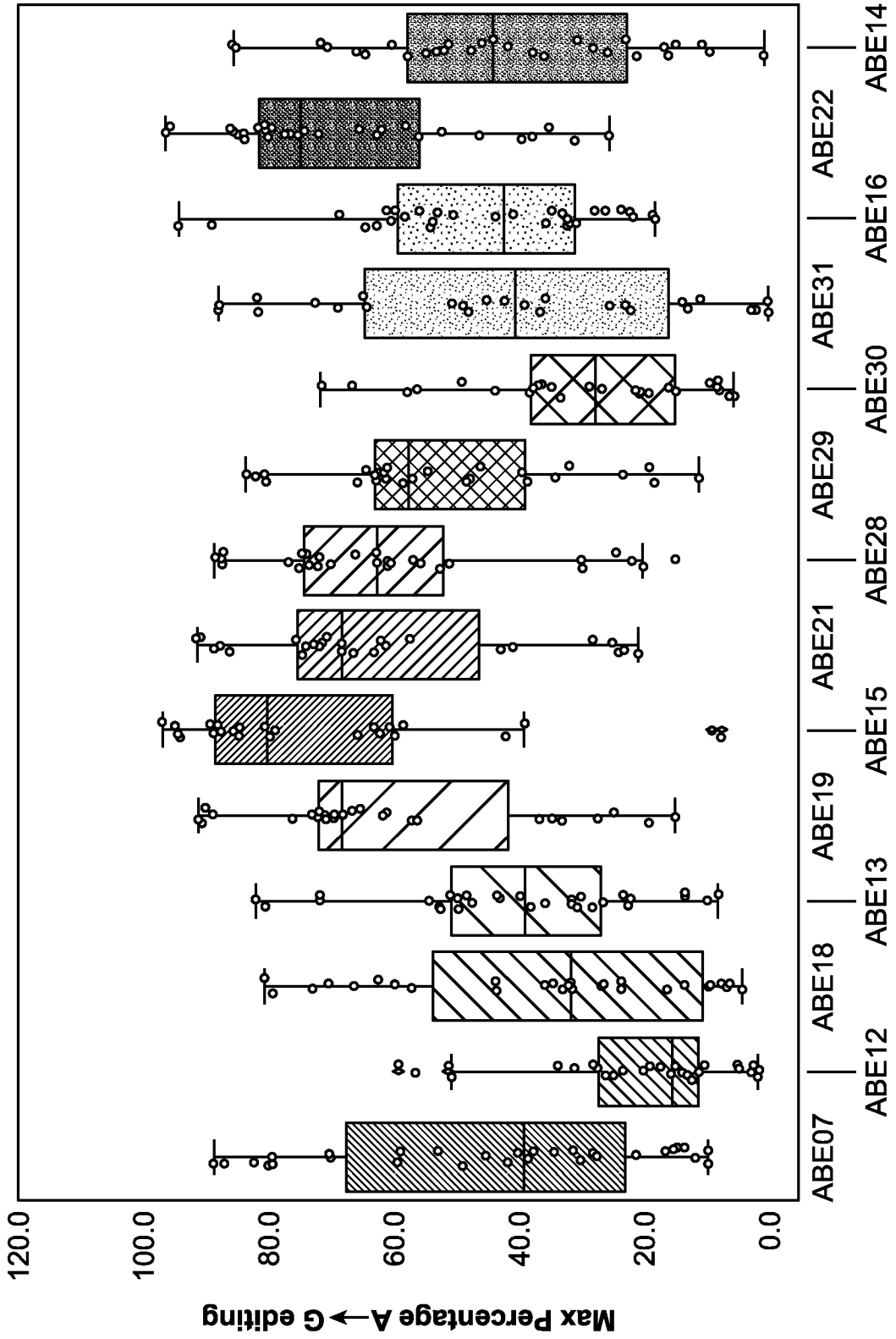
Max Percentage A→G editing
120.0
100.0
80.0
60.0
40.0
20.0
0.0
ABE07
ABE12
ABE18
ABE13
ABE19
ABE15
ABE21
ABE28
ABE29
ABE30
ABE31
ABE16
ABE22
ABE14
Variant

FIG. 21B

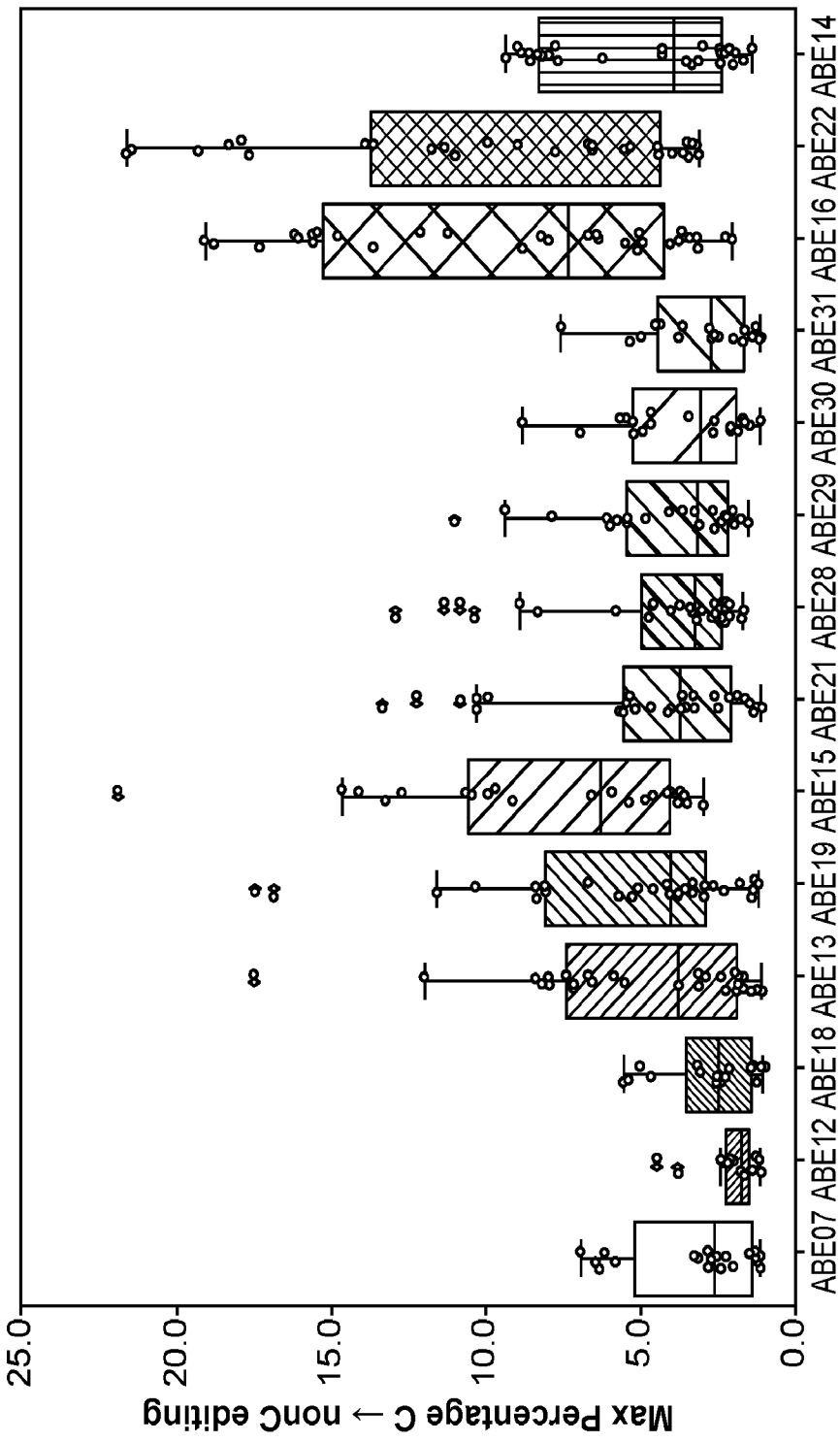
Max Percentage C → nonC editing
25.0
20.0
15.0
10.0
5.0
0.0
ABE07 ABE12 ABE18 ABE13 ABE19 ABE15 ABE21 ABE28 ABE29 ABE30 ABE31 ABE16 ABE22 ABE14
Variant

FIG. 22A

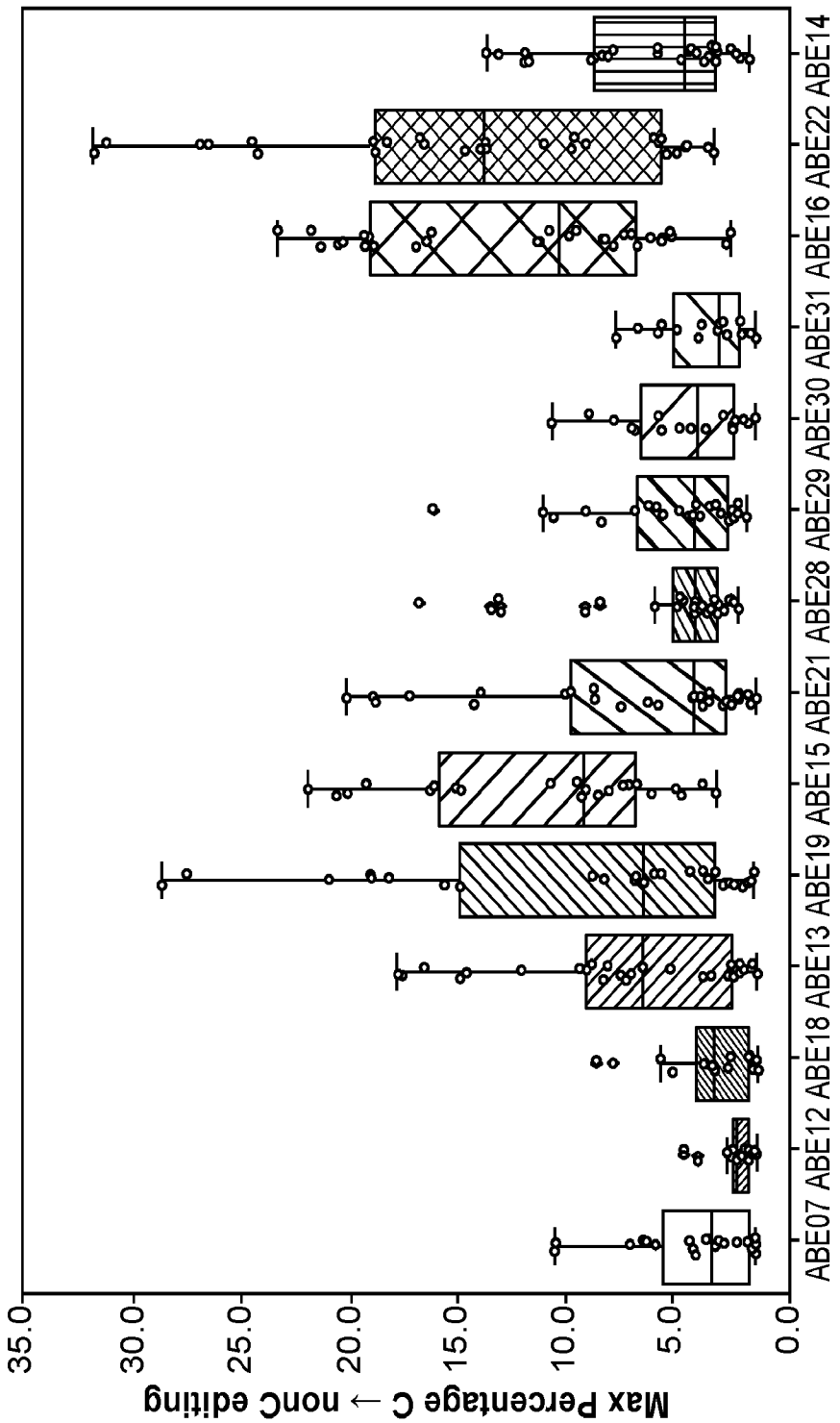
Max Percentage C → nonC editing
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
ABE07 ABE12 ABE18 ABE13 ABE19 ABE15 ABE21 ABE28 ABE29 ABE30 ABE31 ABE16 ABE22 ABE14
Variant

FIG. 22B

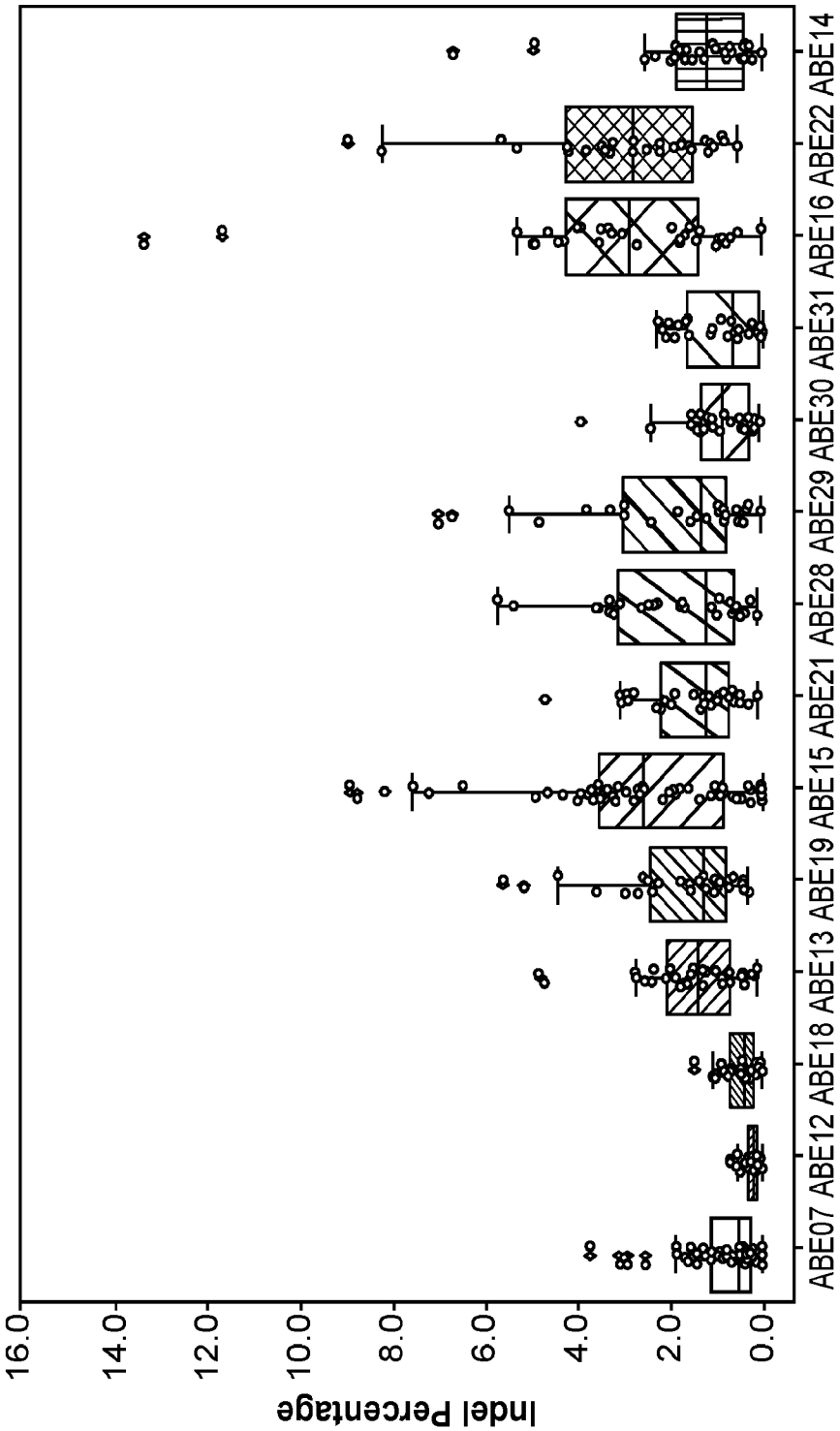
Indel Percentage
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
ABE07 ABE12 ABE18 ABE13 ABE19 ABE15 ABE21 ABE28 ABE29 ABE30 ABE31 ABE16 ABE22 ABE14
Variant

FIG. 23

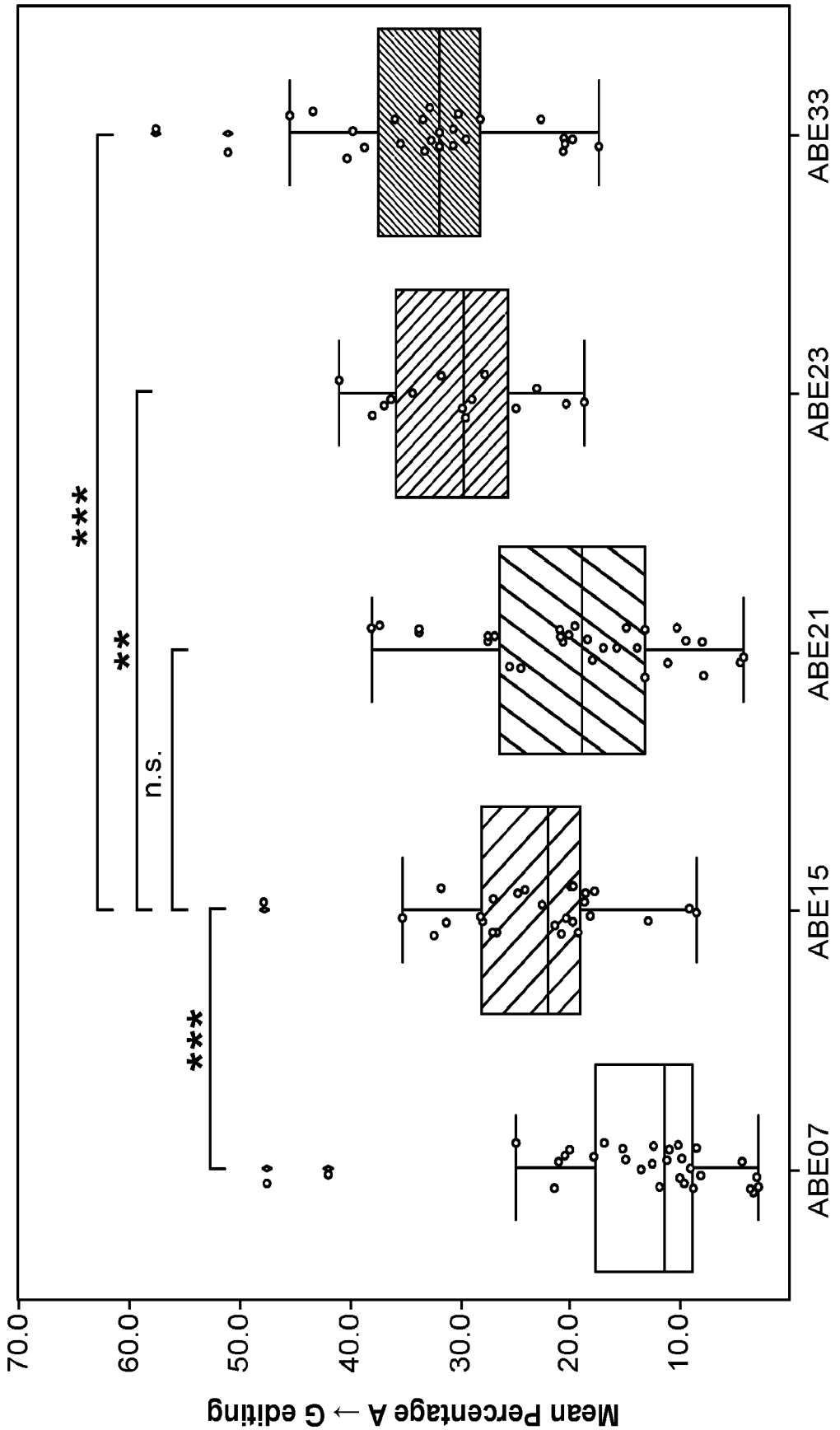
Mean Percentage A → G editing
70.0
60.0
50.0
40.0
30.0
20.0
10.0
***
n.s.
**
***
ABE07
ABE15
ABE21
ABE23
ABE33
Variant

FIG. 24A

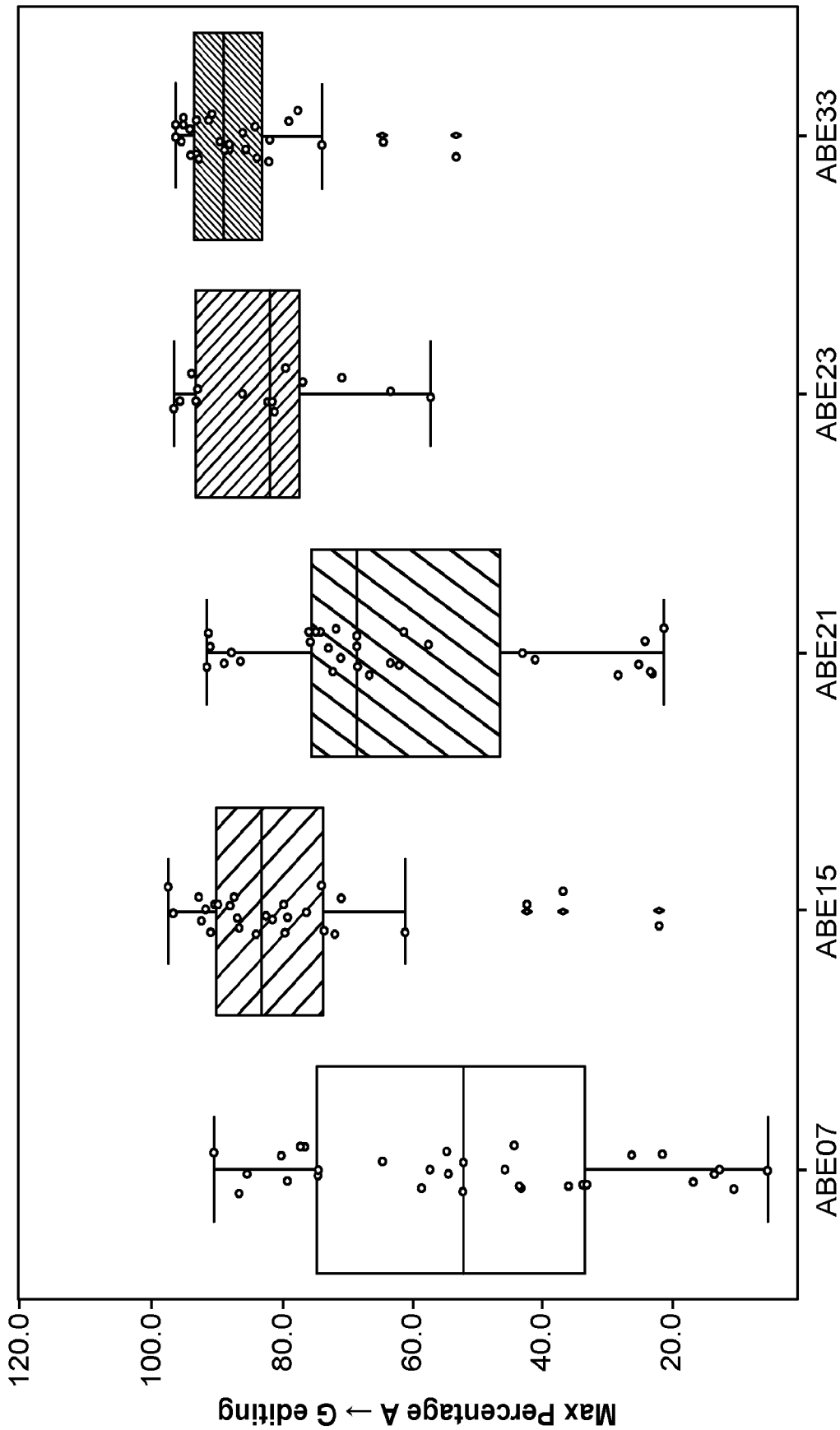
Max Percentage A → G editing
120.0
100.0
80.0
60.0
40.0
20.0
ABE07
ABE15
ABE21
ABE23
ABE33
Variant

FIG. 24B

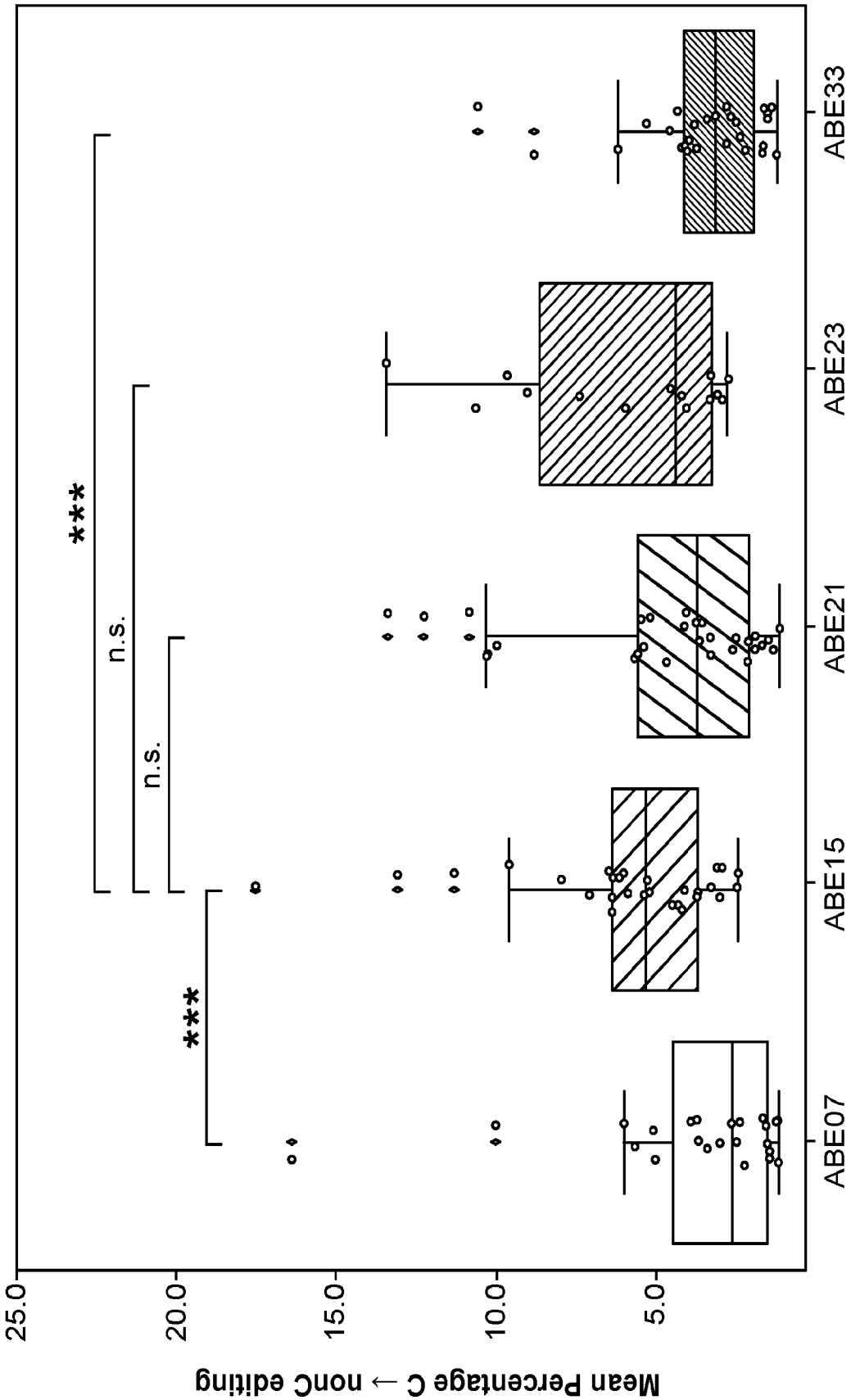
Mean Percentage C → nonC editing
25.0
20.0
15.0
10.0
5.0
ABE07
ABE15
ABE21
ABE23
ABE33
Variant
****
n.s.
n.s.
***

FIG. 25A

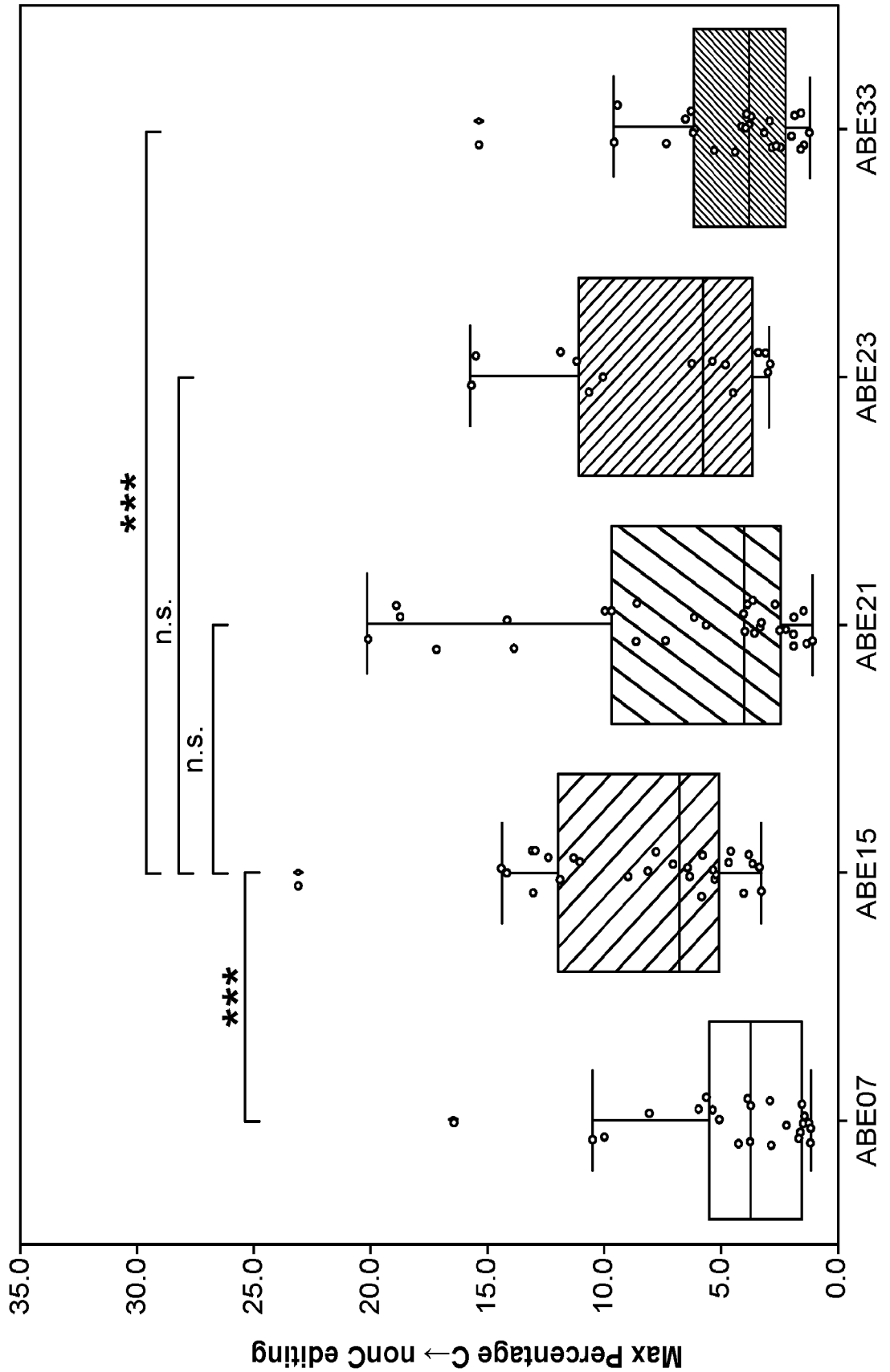
Max Percentage C→ nonC editing
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
****
n.s.
n.s.
***
ABE07
ABE15
ABE21
ABE23
ABE33
Variant

FIG. 25B

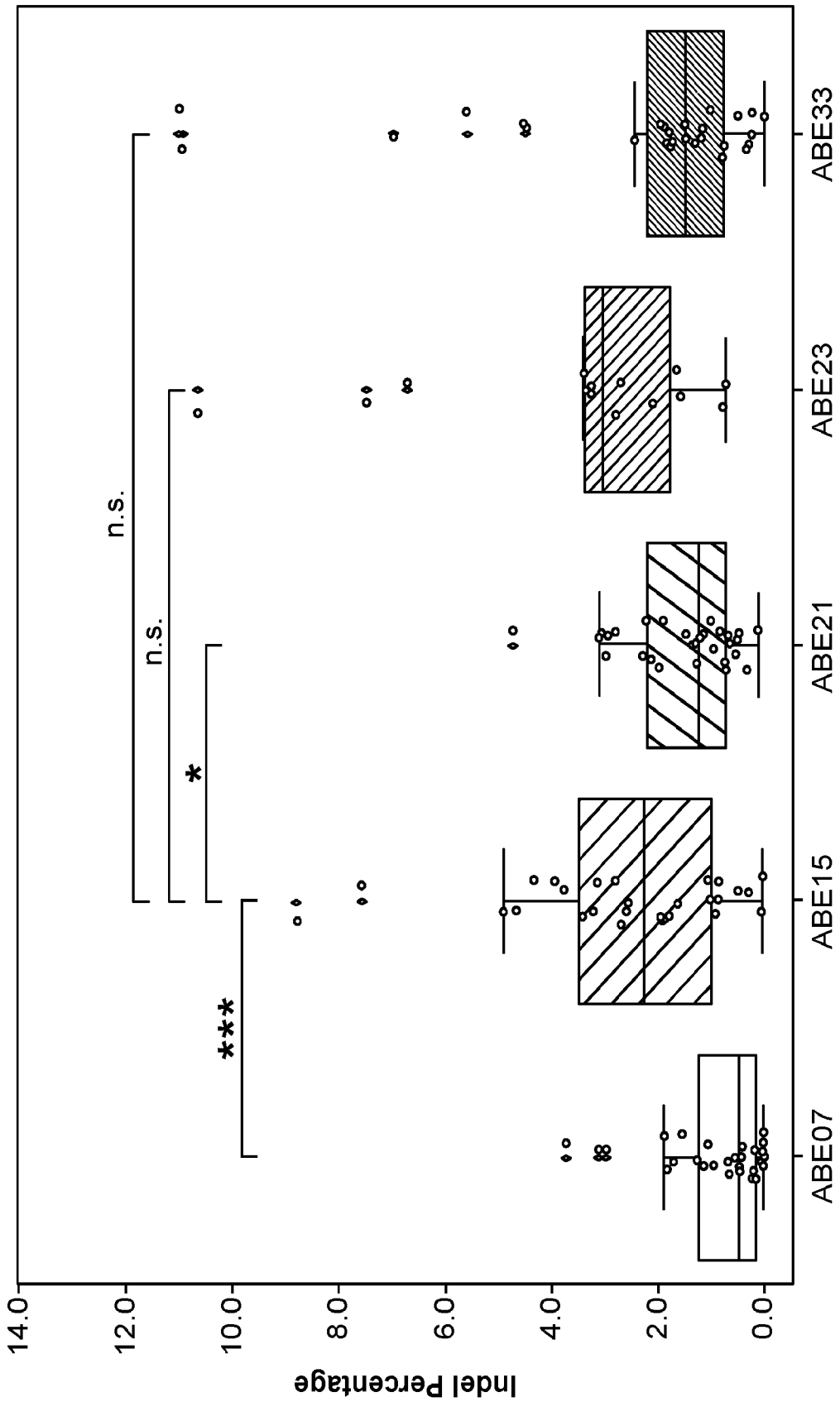
Indel Percentage
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
n.s.
n.s.
*
***
ABE07
ABE15
ABE21
ABE23
ABE33
Variant

FIG. 26

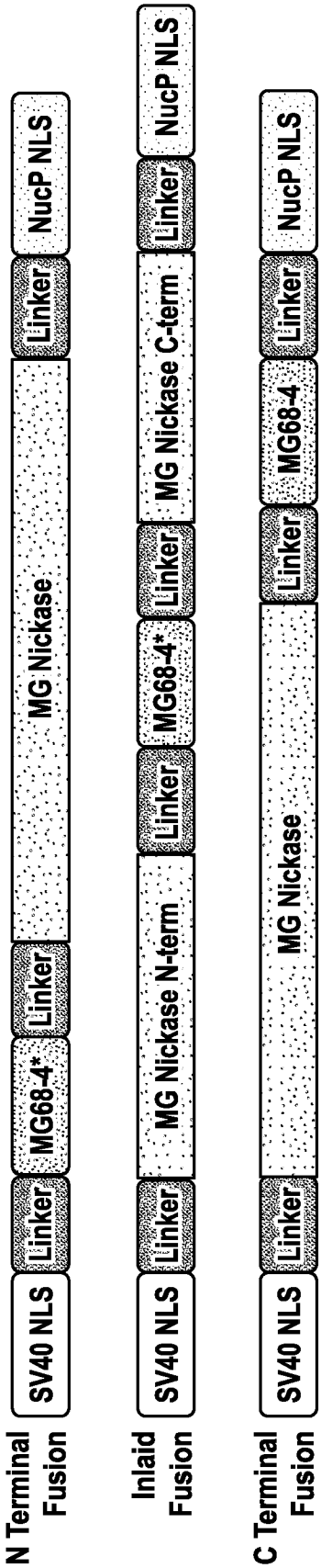
N Terminal Fusion
SV40 NLS
Linker
MG68-4*
Linker
MG Nickase
Linker
NucP NLS
Inlaid Fusion
SV40 NLS
Linker
MG Nickase N-term
Linker
MG68-4*
Linker
MG Nickase C-term
Linker
NucP NLS
C Terminal Fusion
SV40 NLS
Linker
MG Nickase
Linker
MG68-4
Linker
NucP NLS

FIG. 27A

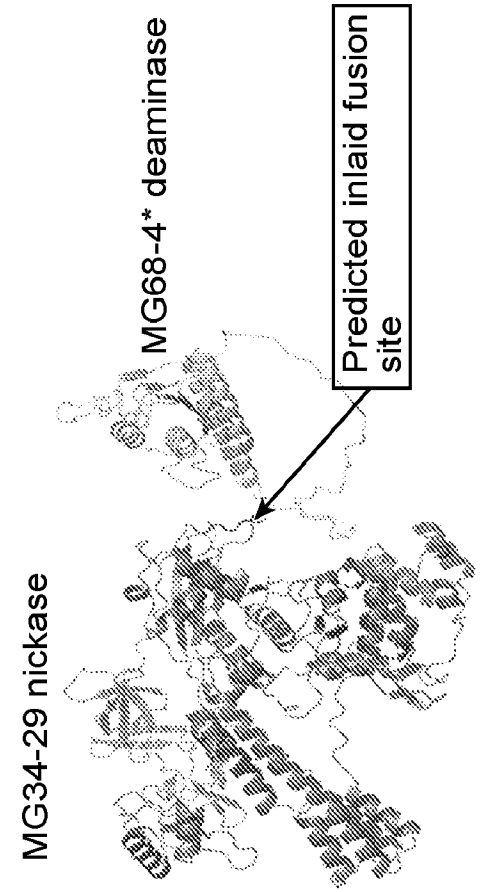

FIG. 27C

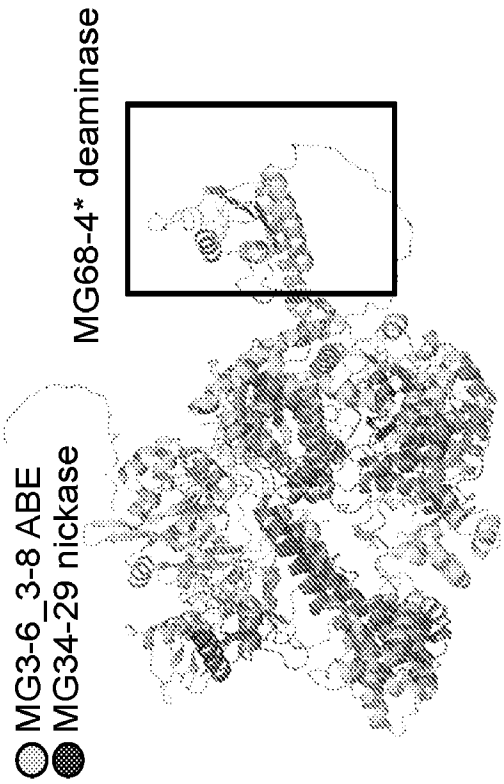

FIG. 27B

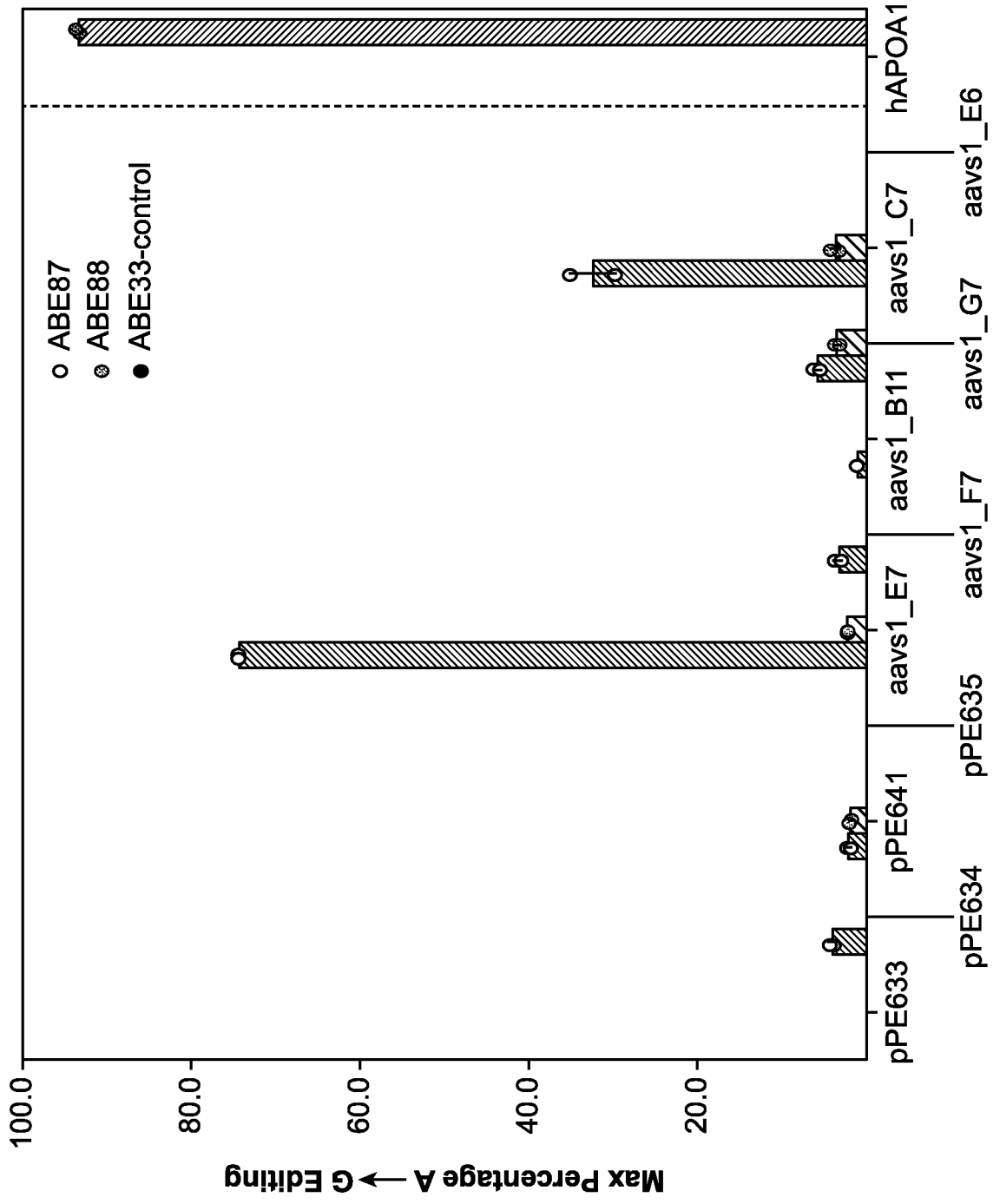
ABE87
ABE88
ABE33-control
Max Percentage A → G Editing
100.0
80.0
60.0
40.0
20.0
pPE633
pPE634
pPE641
pPE635
aavs1_E7
aavs1_F7
aavs1_B11
aavs1_G7
aavs1_C7
aavs1_E6
hAPOA1
Target Site

FIG. 28A

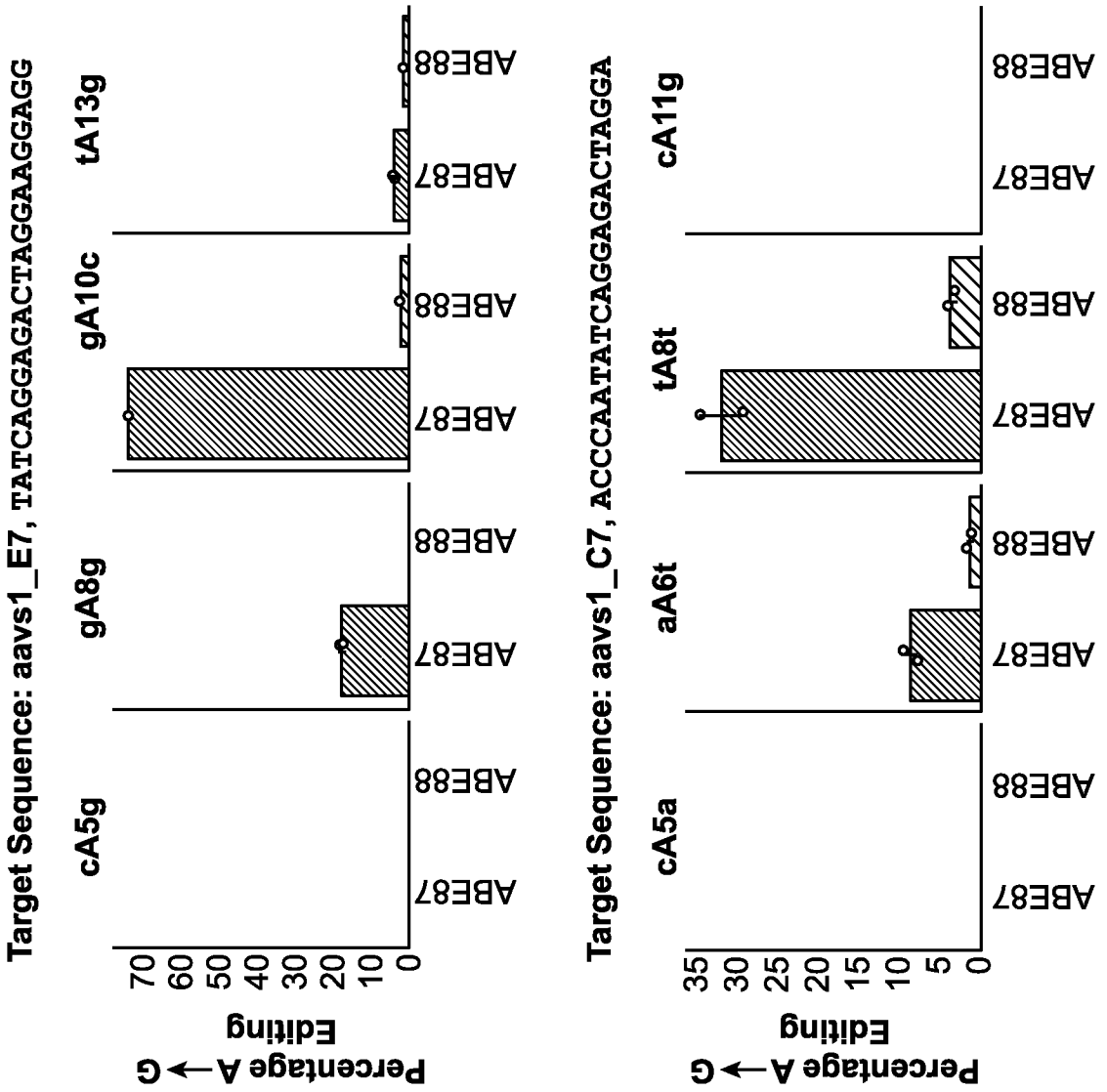
Target Sequence: aavs1_E7, TATCAGGAGACTAGGAAGGAGG
cA5g
gA8g
gA10c
tA13g
ABE87
ABE88
Percentage A→G Editing
0
10
20
30
40
50
60
70
Target Sequence: aavs1_C7, ACCCAATATCAGGAGACTAGGA
cA5a
aA6t
tA8t
cA11g
0
5
10
15
20
25
30
35

FIG. 28B

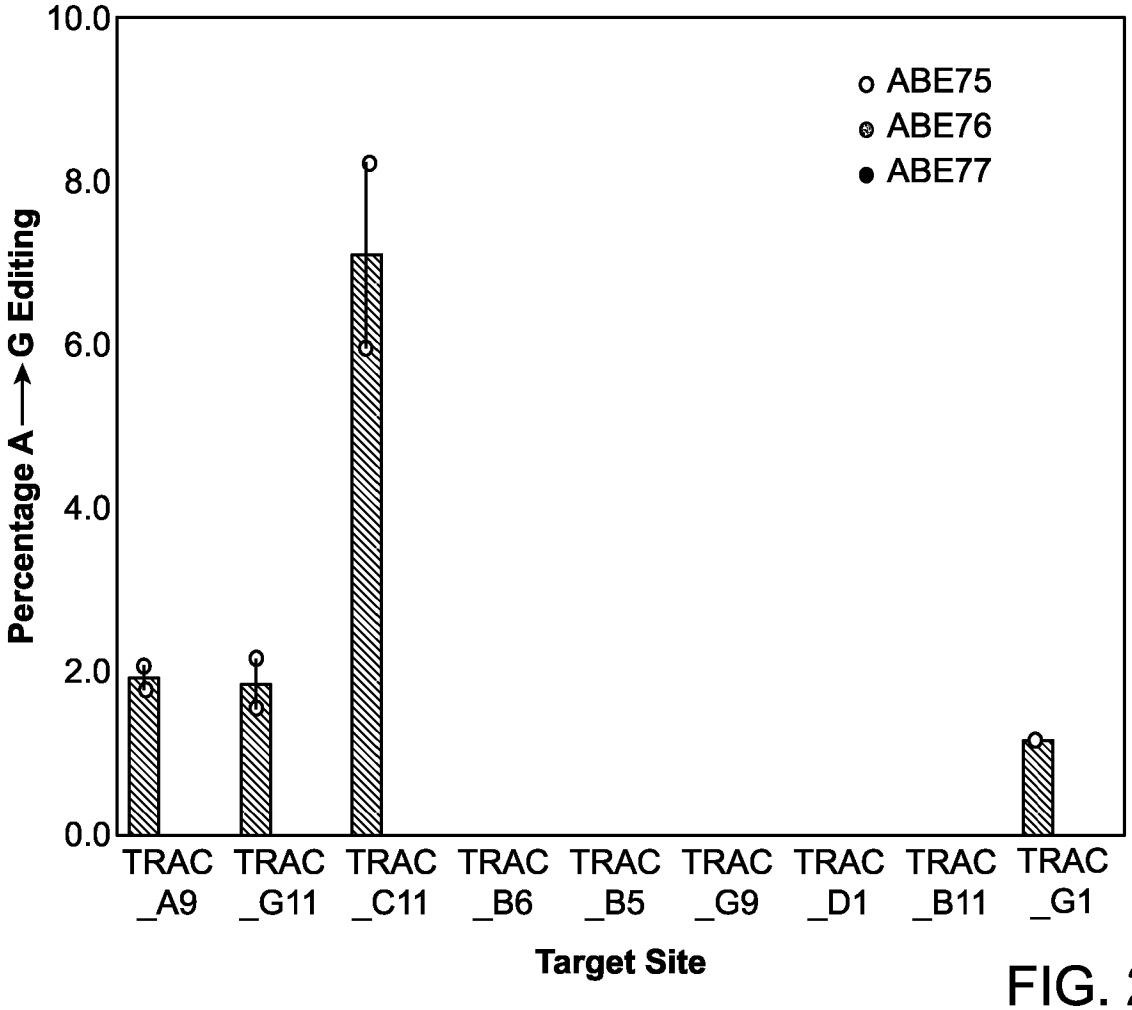

FIG. 28C

Target Sequence: TRAC_C11, CAGCCGCAGCGTCATGAGCAGATT

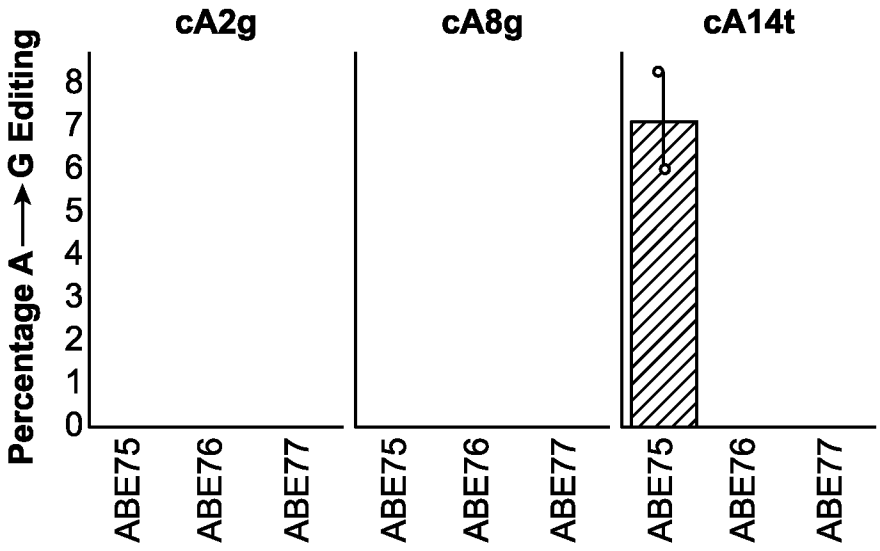

FIG. 28D

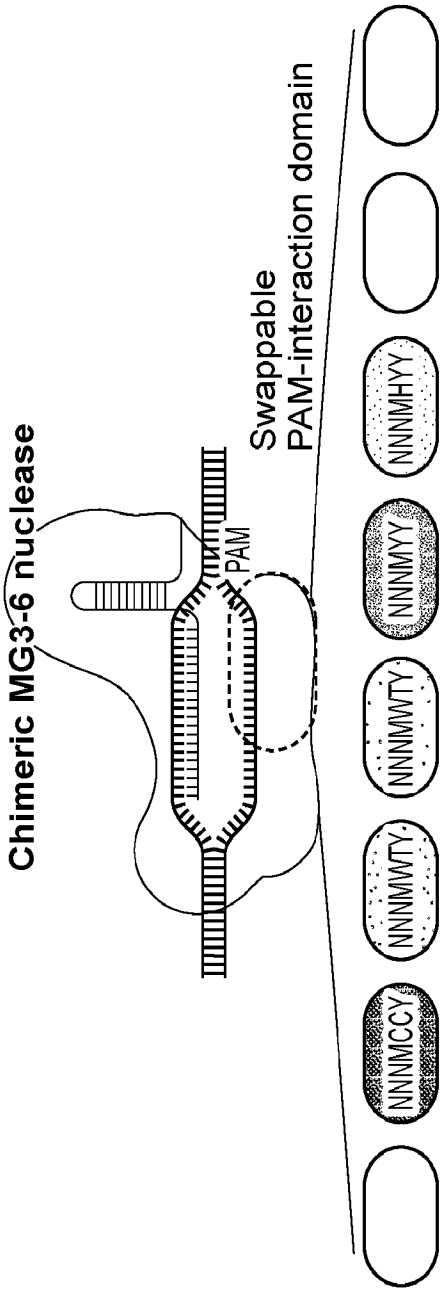

FIG. 29B

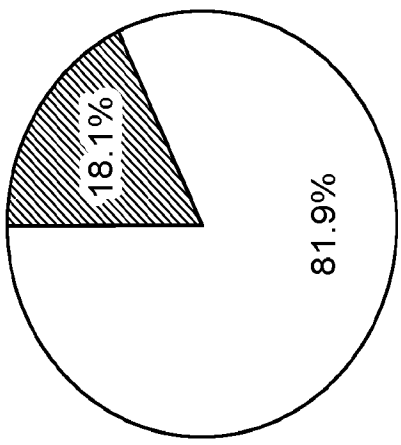

FIG. 29A

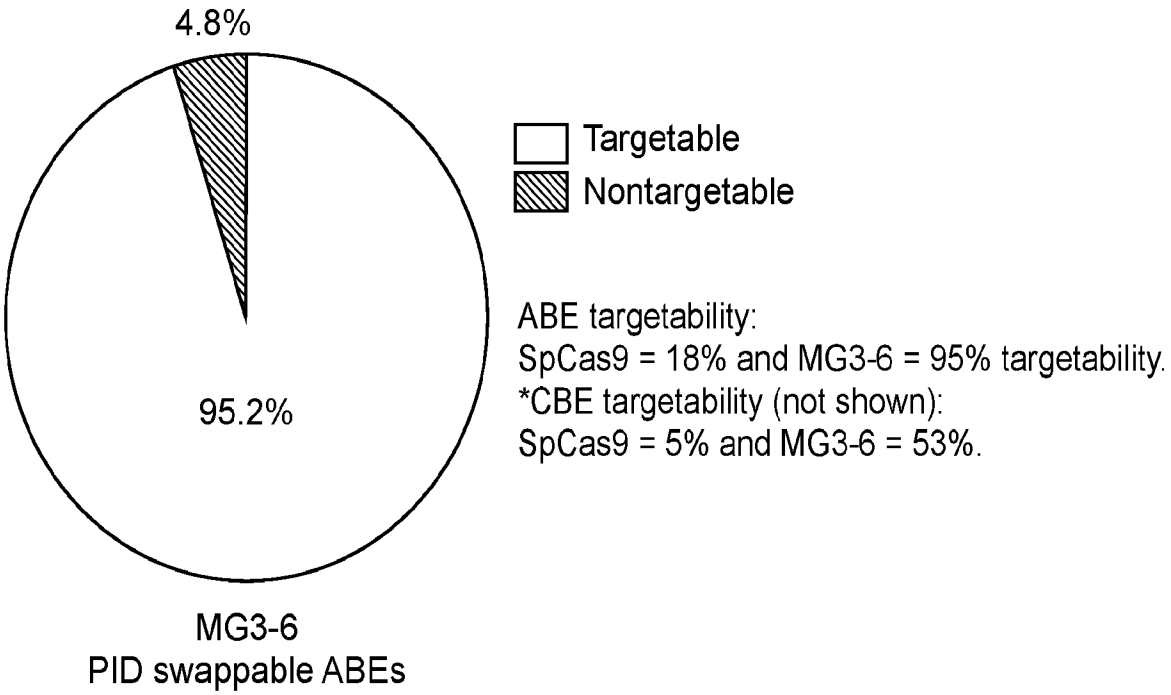
4.8%
95.2%
Targetable
Nontargetable
ABE targetability:
SpCas9 = 18% and MG3-6 = 95% targetability.
*CBE targetability (not shown):
SpCas9 = 5% and MG3-6 = 53%.
MG3-6
PID swappable ABEs

FIG. 29C

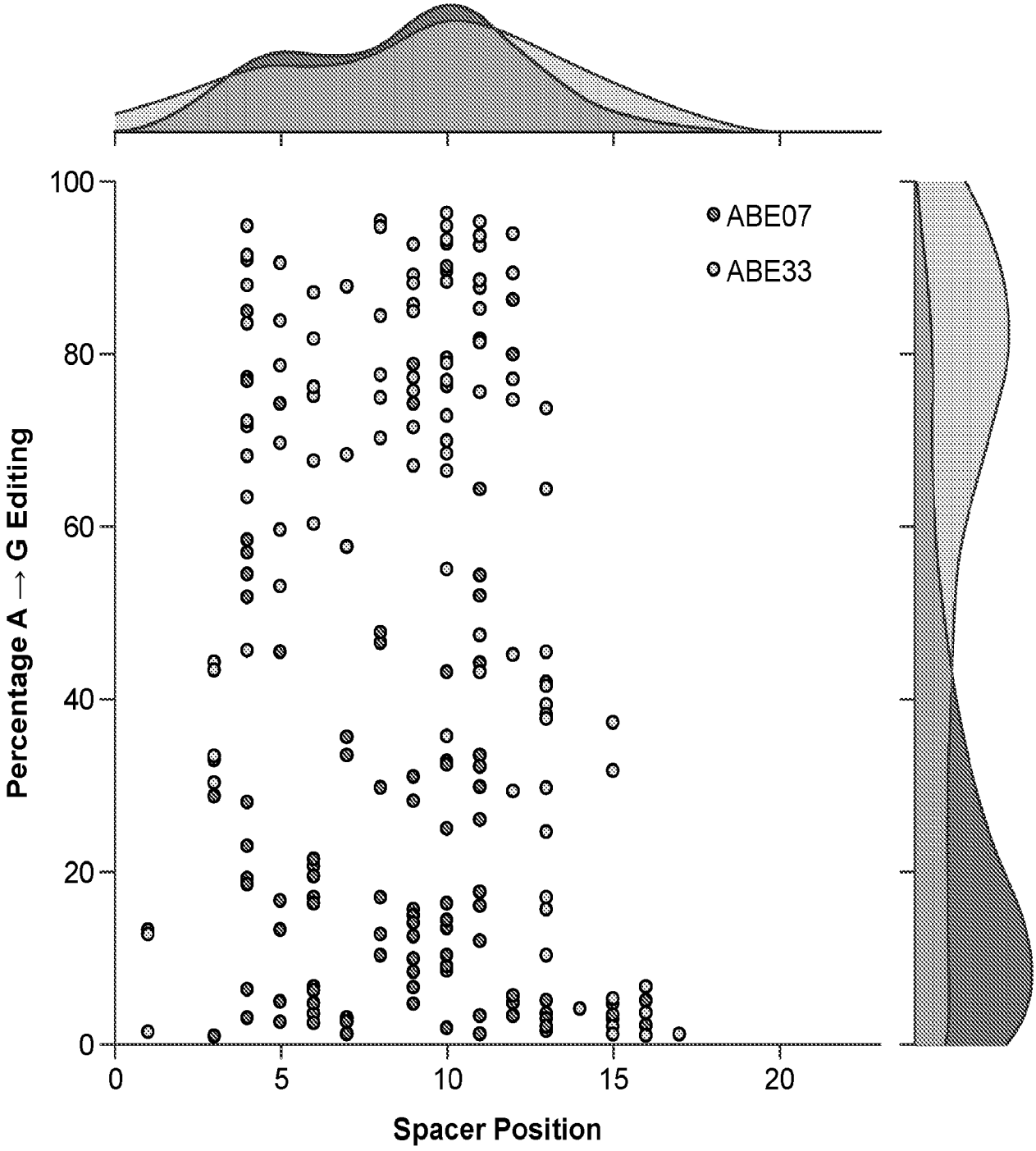
ABE07
ABE33
Percentage A → G Editing
Spacer Position
0
20
40
60
80
100
0
5
10
15
20

FIG. 30A

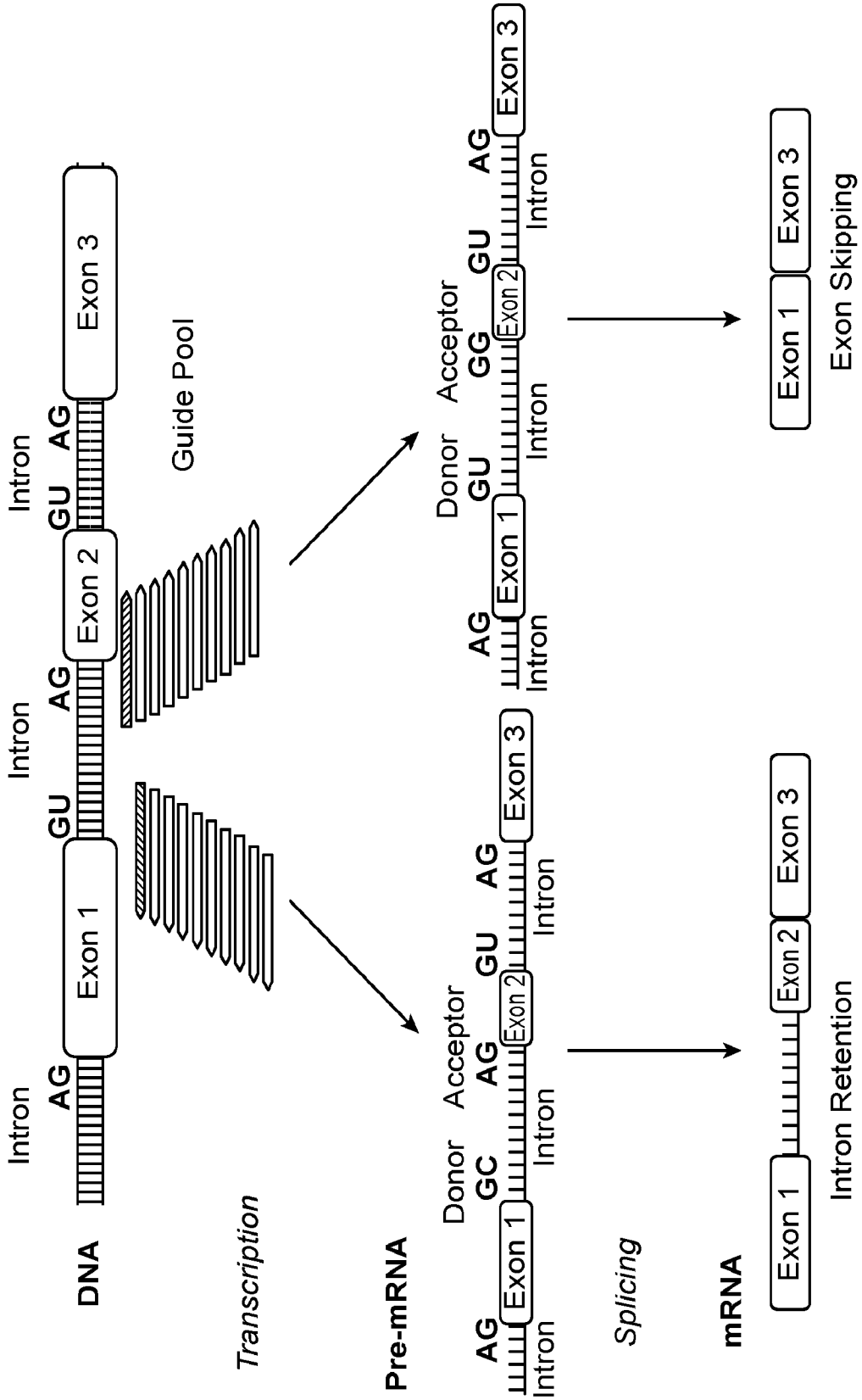
DNA
Intron
AG
Exon 1
GU
Intron
AG
Exon 2
GU
Intron
AG
Exon 3
Transcription
Guide Pool
Pre-mRNA
Donor
GC
Acceptor
AG
Exon 1
Intron
AG
Exon 2
GU
Intron
AG
Exon 3
Splicing
mRNA
Exon 1
Exon 2
Exon 3
Intron Retention
Donor
GU
Acceptor
AG
Exon 1
Intron
AG
Exon 2
GG
GU
Intron
AG
Exon 3
Exon 1
Exon 3
Exon Skipping

FIG. 30B

Pooled Guide Screen

12 ABE Chimeras

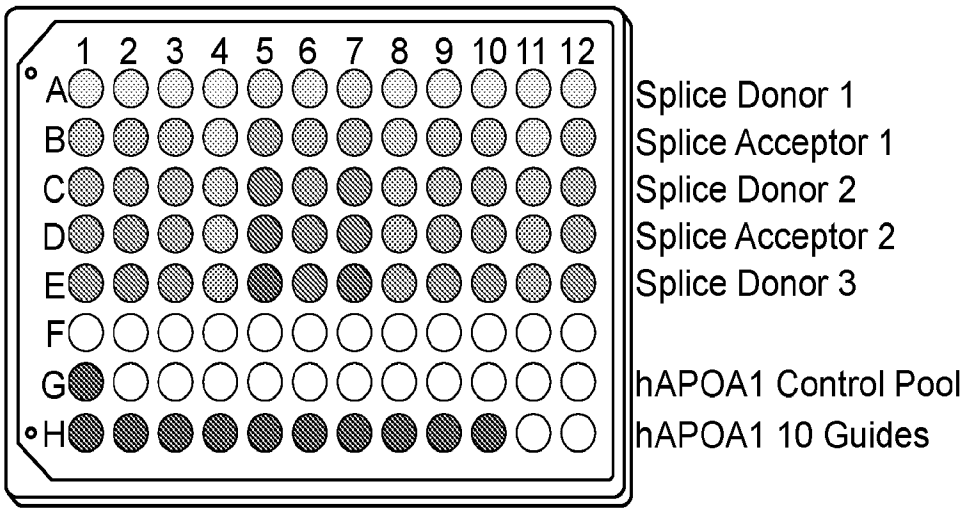

FIG. 30C

Deconvolution Screen

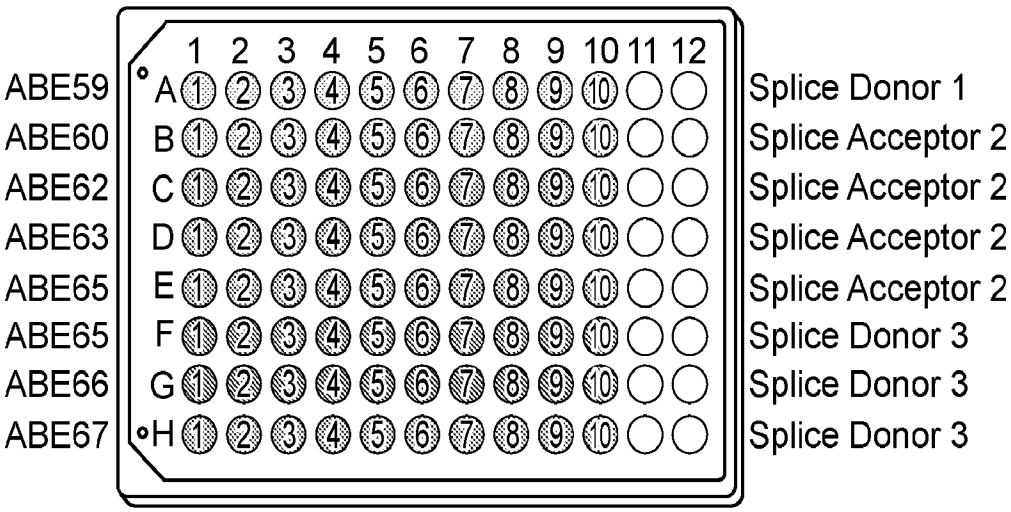

FIG. 30D

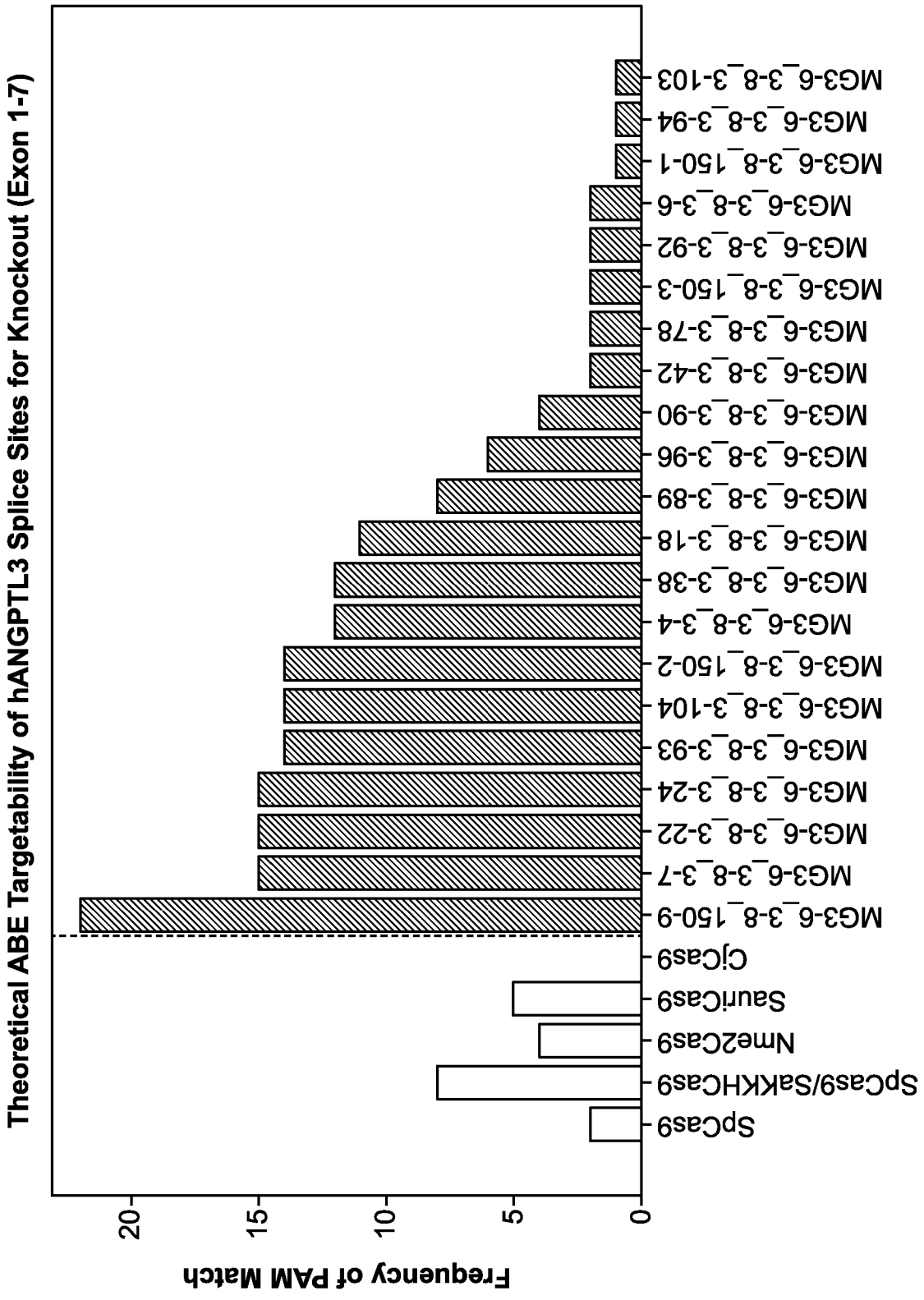
Theoretical ABE Targetability of hANGPTL3 Splice Sites for Knockout (Exon 1-7)
Frequency of PAM Match
0
5
10
15
20
SpCas9
SpCas9/SaKKHCas9
Nme2Cas9
SauriCas9
CjCas9
MG3-6_3-8_150-9
MG3-6_3-8_3-7
MG3-6_3-8_3-22
MG3-6_3-8_3-24
MG3-6_3-8_3-93
MG3-6_3-8_3-104
MG3-6_3-8_150-2
MG3-6_3-8_3-4
MG3-6_3-8_3-38
MG3-6_3-8_3-18
MG3-6_3-8_3-89
MG3-6_3-8_3-96
MG3-6_3-8_3-90
MG3-6_3-8_3-42
MG3-6_3-8_3-78
MG3-6_3-8_150-3
MG3-6_3-8_3-92
MG3-6_3-8_3-6
MG3-6_3-8_150-1
MG3-6_3-8_3-94
MG3-6_3-8_3-103
MG3-6 Chimeras

FIG. 31A

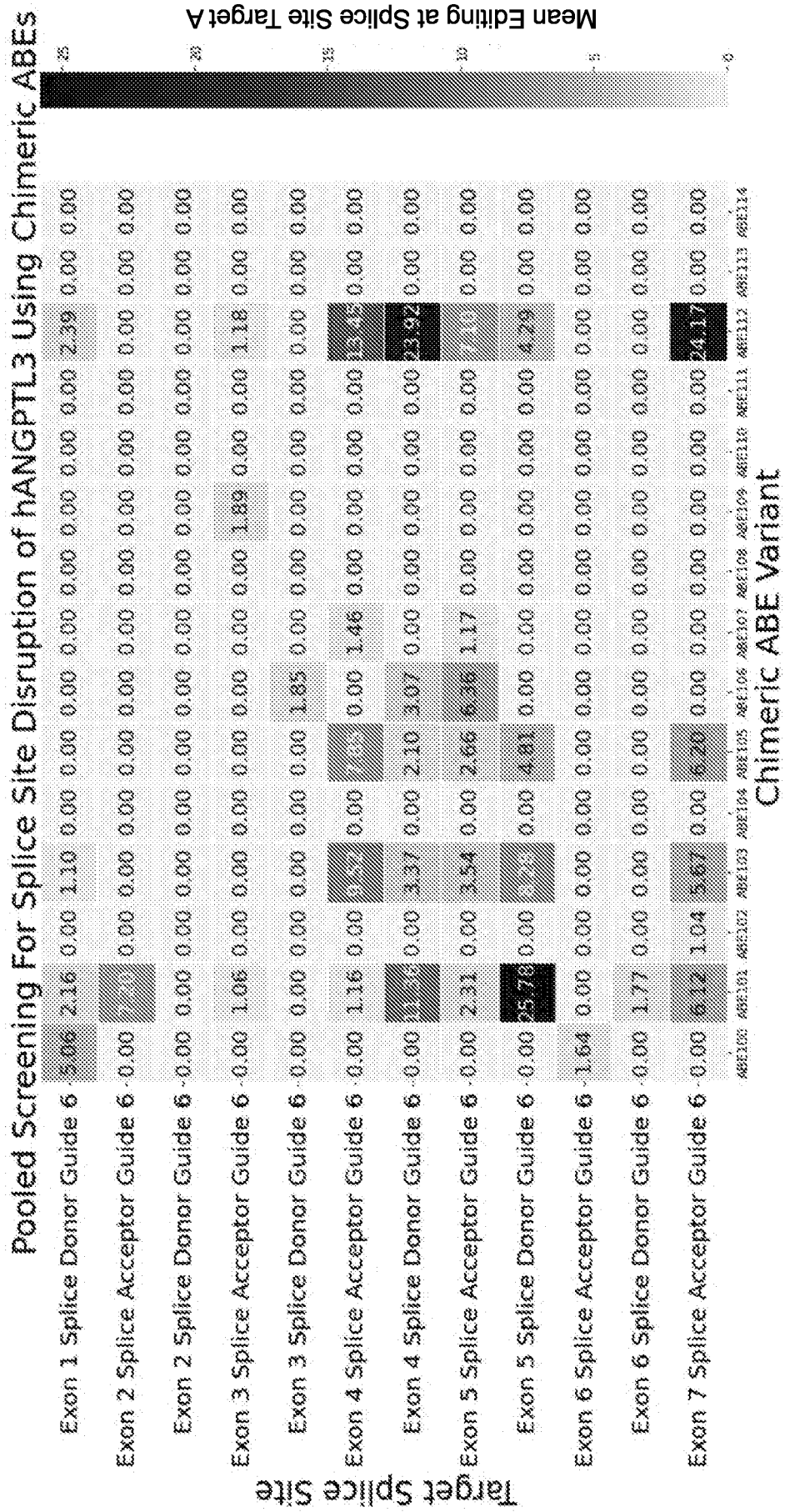
Pooled Screening For Splice Site Disruption of hANGPTL3 Using Chimeric ABEs
Mean Editing at Splice Site Target A
25
20
15
10
5
0
Target Splice Site
Exon 1 Splice Donor Guide 6
Exon 2 Splice Acceptor Guide 6
Exon 2 Splice Donor Guide 6
Exon 3 Splice Acceptor Guide 6
Exon 3 Splice Donor Guide 6
Exon 4 Splice Acceptor Guide 6
Exon 4 Splice Donor Guide 6
Exon 5 Splice Acceptor Guide 6
Exon 5 Splice Donor Guide 6
Exon 6 Splice Acceptor Guide 6
Exon 6 Splice Donor Guide 6
Exon 7 Splice Acceptor Guide 6
3.06 2.16 0.00 1.10 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 2.39 0.00 0.00
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
0.00 1.06 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.89 0.00 0.00 0.00 1.18 0.00 0.00 0.00
0.00 0.00 0.00 0.00 0.00 0.00 1.85 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
0.00 1.16 0.00 0.00 0.00 0.00 0.00 1.46 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
0.00 0.00 3.37 0.00 2.10 3.07 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
0.00 2.31 0.00 3.54 0.00 2.66 6.36 1.17 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
0.00 35.78 0.00 0.00 4.81 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 4.29 0.00 0.00 0.00
1.64 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
0.00 1.77 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
0.00 6.12 1.04 5.67 0.00 6.20 0.00 0.00 0.00 0.00 0.00 0.00 0.00 24.17 0.00 0.00 0.00
ABE100 ABE101 ABE102 ABE103 ABE104 ABE105 ABE106 ABE107 ABE108 ABE109 ABE110 ABE111 ABE112 ABE113 ABE114
Chimeric ABE Variant

FIG. 31B

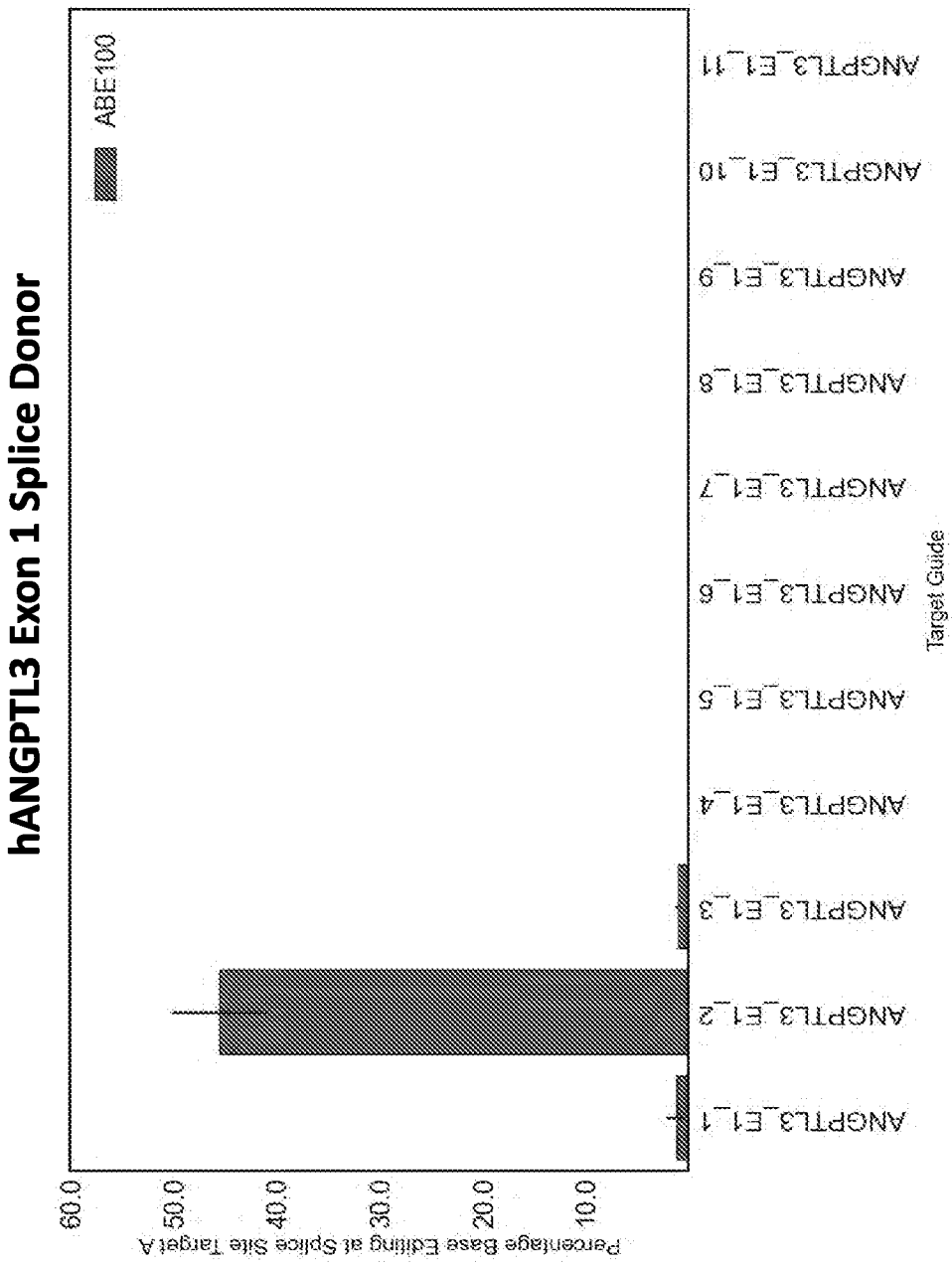
hANGPTL3 Exon 1 Splice Donor
ABE100
Percentage Base Editing at Splice Site Target A
60.0
50.0
40.0
30.0
20.0
10.0
ANGPTL3_E1_1
ANGPTL3_E1_2
ANGPTL3_E1_3
ANGPTL3_E1_4
ANGPTL3_E1_5
ANGPTL3_E1_6
ANGPTL3_E1_7
ANGPTL3_E1_8
ANGPTL3_E1_9
ANGPTL3_E1_10
ANGPTL3_E1_11
Target Guide

FIG. 32A

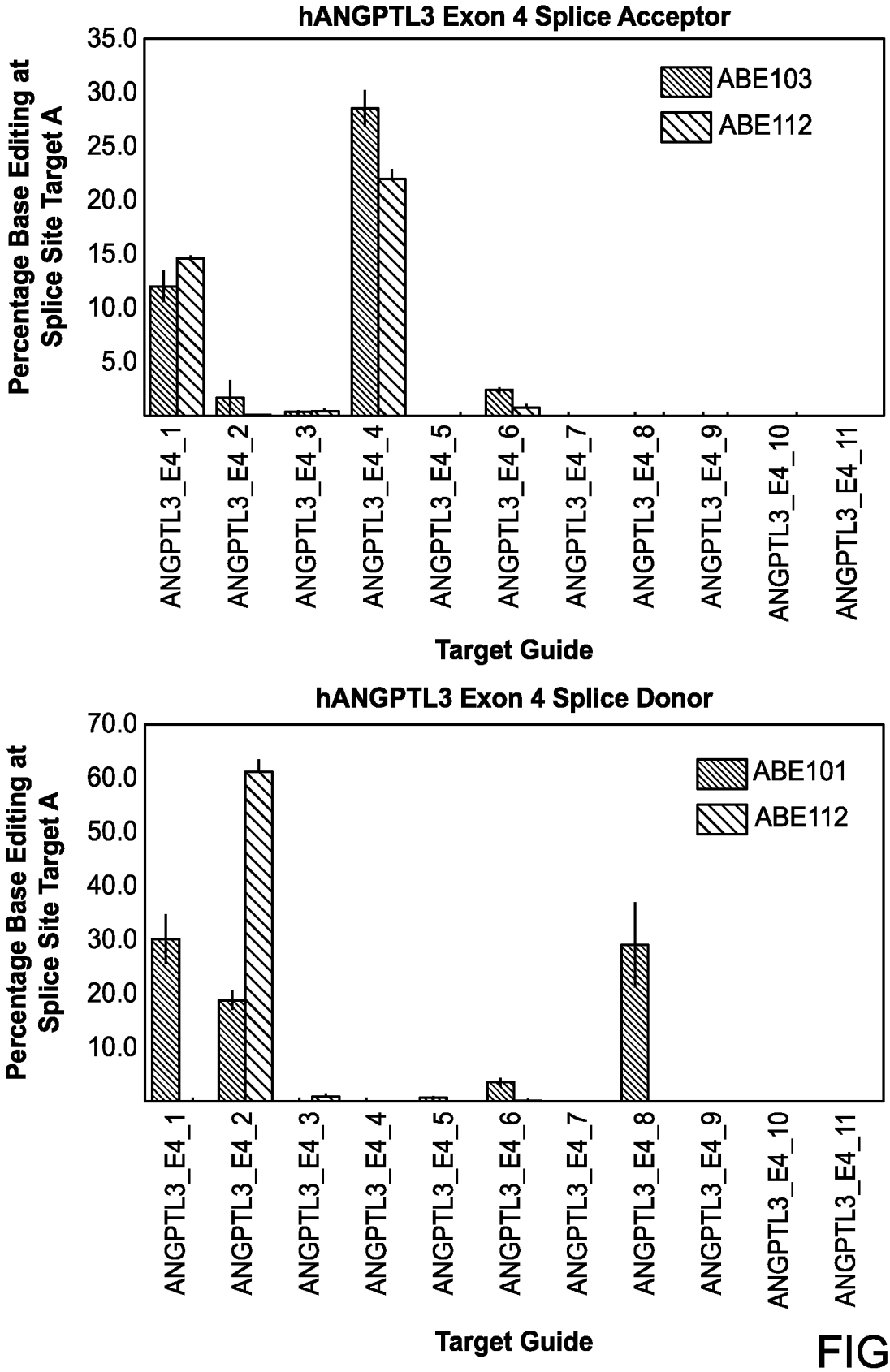
hANGPTL3 Exon 4 Splice Acceptor
Percentage Base Editing at Splice Site Target A
35.0
30.0
25.0
20.0
15.0
10.0
5.0
ABE103
ABE112
ANGPTL3_E4_1
ANGPTL3_E4_2
ANGPTL3_E4_3
ANGPTL3_E4_4
ANGPTL3_E4_5
ANGPTL3_E4_6
ANGPTL3_E4_7
ANGPTL3_E4_8
ANGPTL3_E4_9
ANGPTL3_E4_10
ANGPTL3_E4_11
Target Guide
hANGPTL3 Exon 4 Splice Donor
Percentage Base Editing at Splice Site Target A
70.0
60.0
50.0
40.0
30.0
20.0
10.0
ABE101
ABE112
ANGPTL3_E4_1
ANGPTL3_E4_2
ANGPTL3_E4_3
ANGPTL3_E4_4
ANGPTL3_E4_5
ANGPTL3_E4_6
ANGPTL3_E4_7
ANGPTL3_E4_8
ANGPTL3_E4_9
ANGPTL3_E4_10
ANGPTL3_E4_11
Target Guide

FIG. 32B

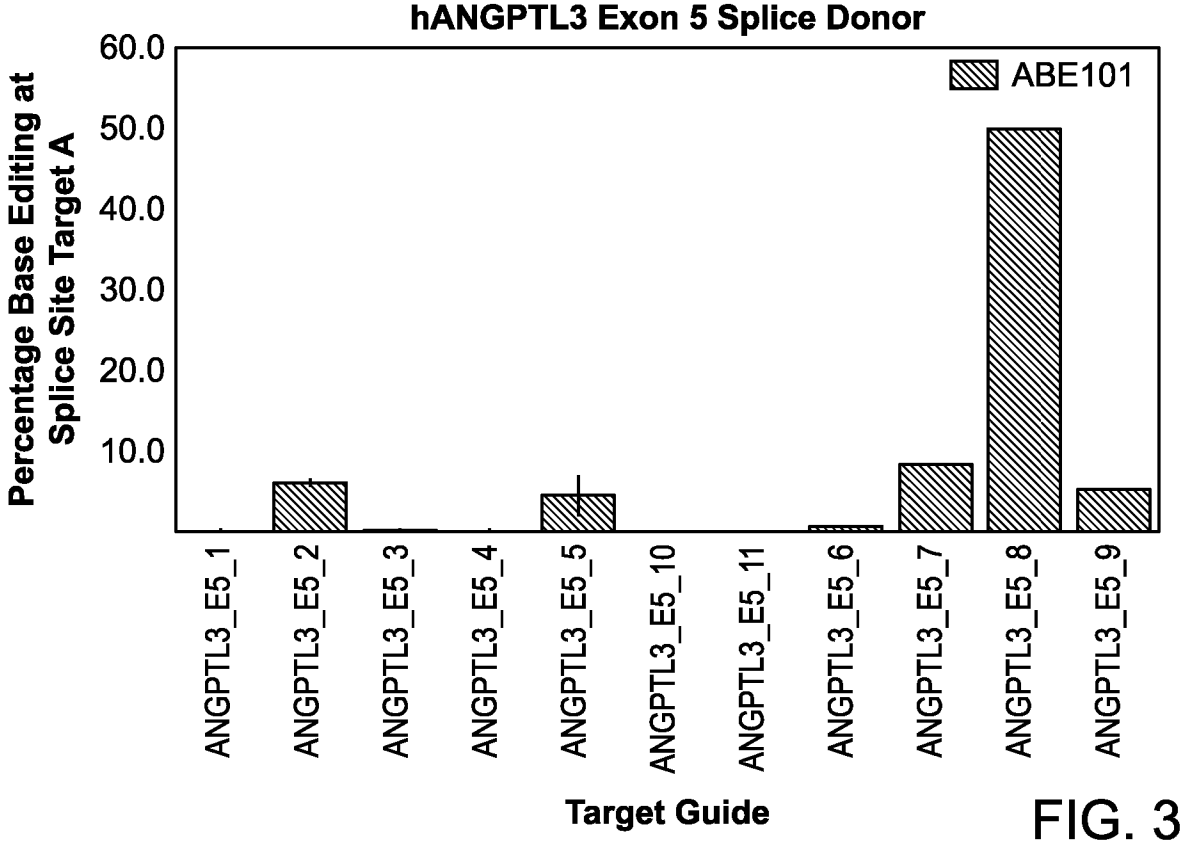

FIG. 32C

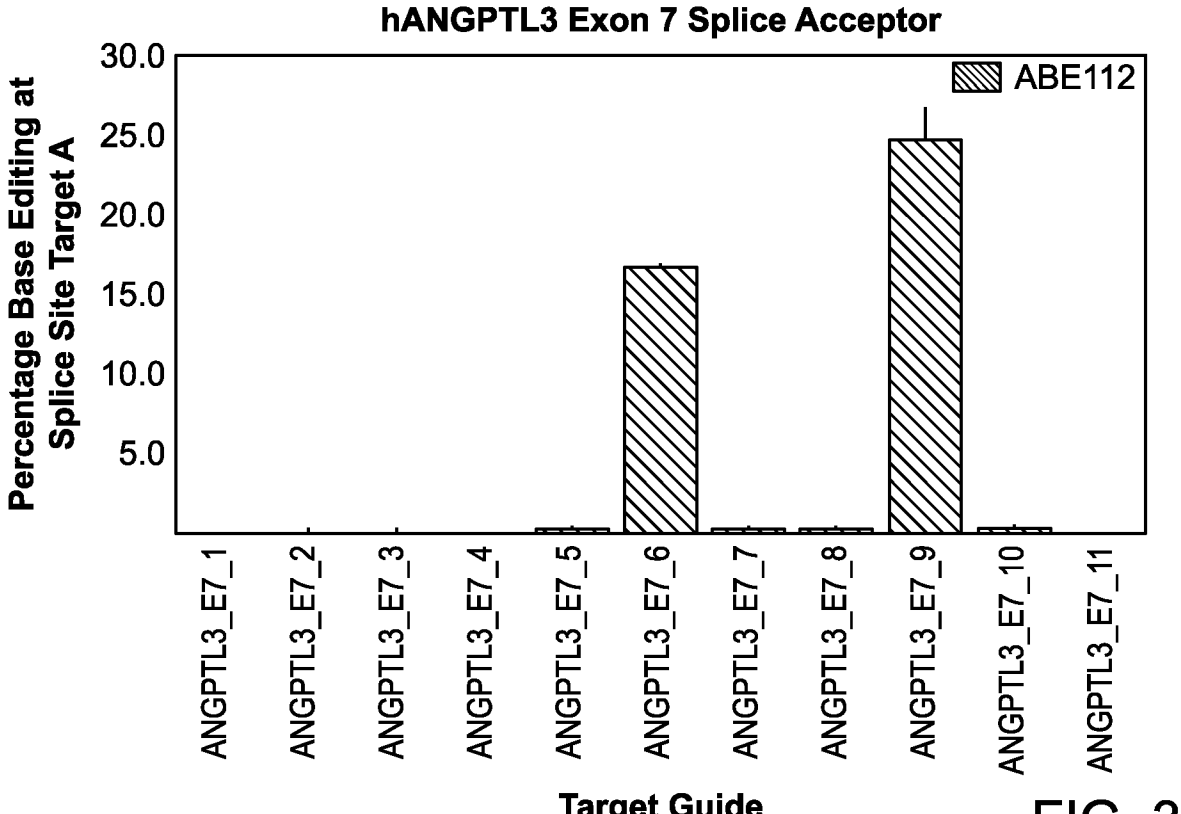

FIG. 32D

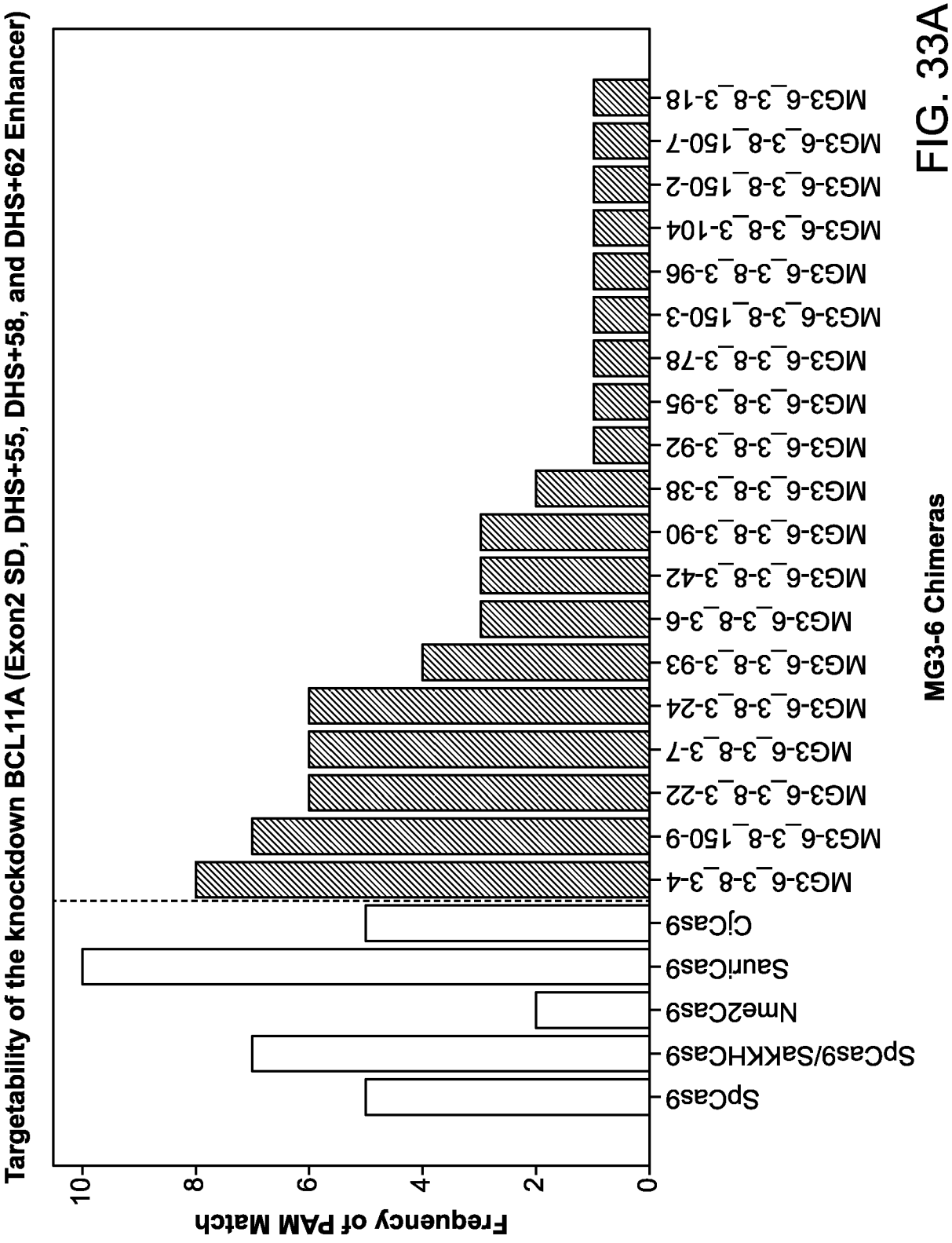
Targetability of the knockdown BCL11A (Exon2 SD, DHS+55, DHS+58, and DHS+62 Enhancer)
Frequency of PAM Match
0
2
4
6
8
10
SpCas9
SpCas9/SaKKHCas9
Nme2Cas9
SauriCas9
CjCas9
MG3-6_3-8_3-4
MG3-6_3-8_150-9
MG3-6_3-8_3-22
MG3-6_3-8_3-7
MG3-6_3-8_3-24
MG3-6_3-8_3-93
MG3-6_3-8_3-6
MG3-6_3-8_3-42
MG3-6_3-8_3-90
MG3-6_3-8_3-38
MG3-6_3-8_3-92
MG3-6_3-8_3-95
MG3-6_3-8_3-78
MG3-6_3-8_150-3
MG3-6_3-8_3-96
MG3-6_3-8_3-104
MG3-6_3-8_150-2
MG3-6_3-8_150-7
MG3-6_3-8_3-18
MG3-6 Chimeras

FIG. 33A

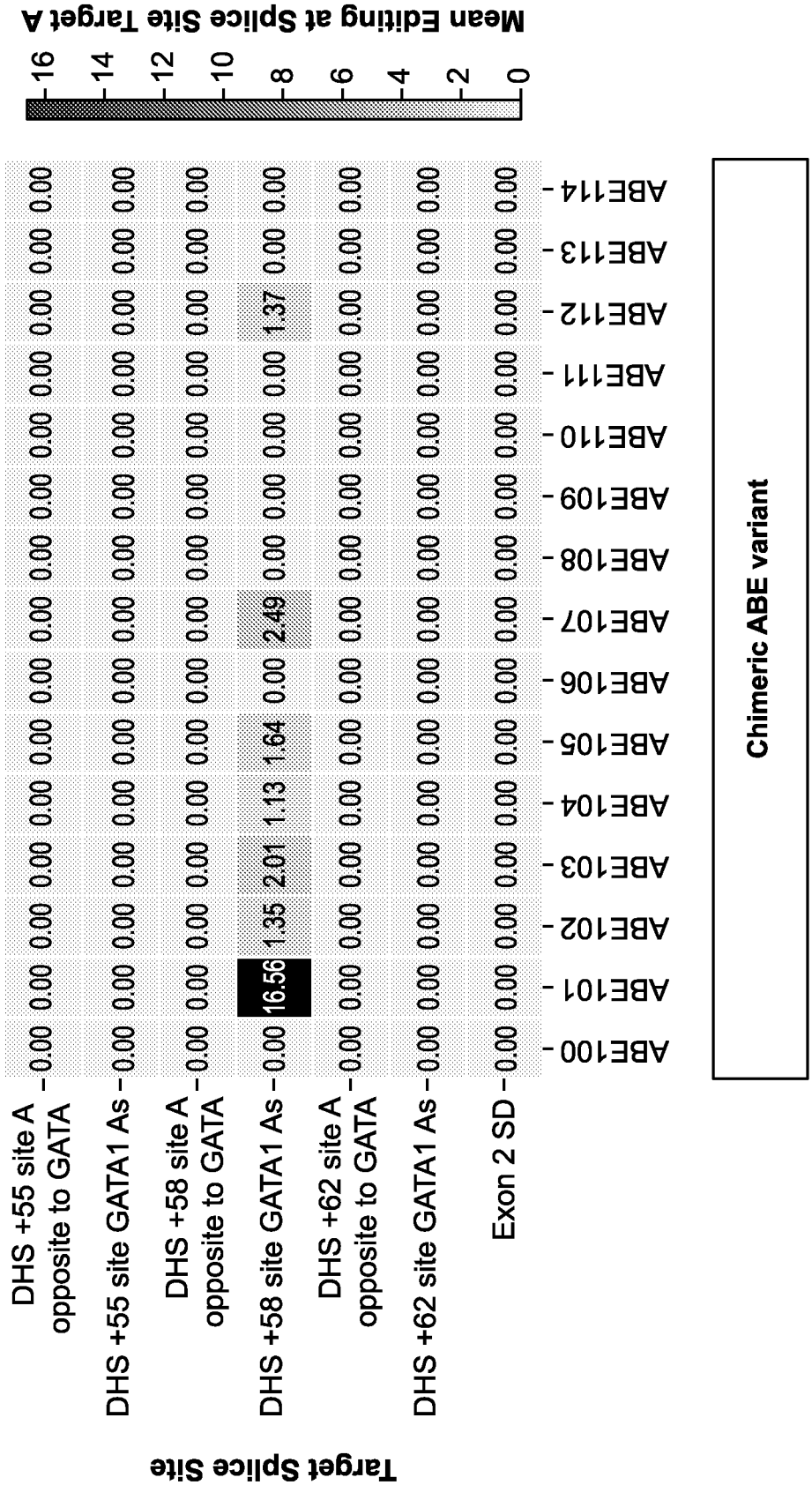
Mean Editing at Splice Site Target A
16
14
12
10
8
6
4
2
0
Target Splice Site
DHS +55 site A opposite to GATA
DHS +55 site GATA1 As
DHS +58 site A opposite to GATA
DHS +58 site GATA1 As
DHS +62 site A opposite to GATA
DHS +62 site GATA1 As
Exon 2 SD
ABE100
ABE101
ABE102
ABE103
ABE104
ABE105
ABE106
ABE107
ABE108
ABE109
ABE110
ABE111
ABE112
ABE113
ABE114
Chimeric ABE variant

FIG. 33B

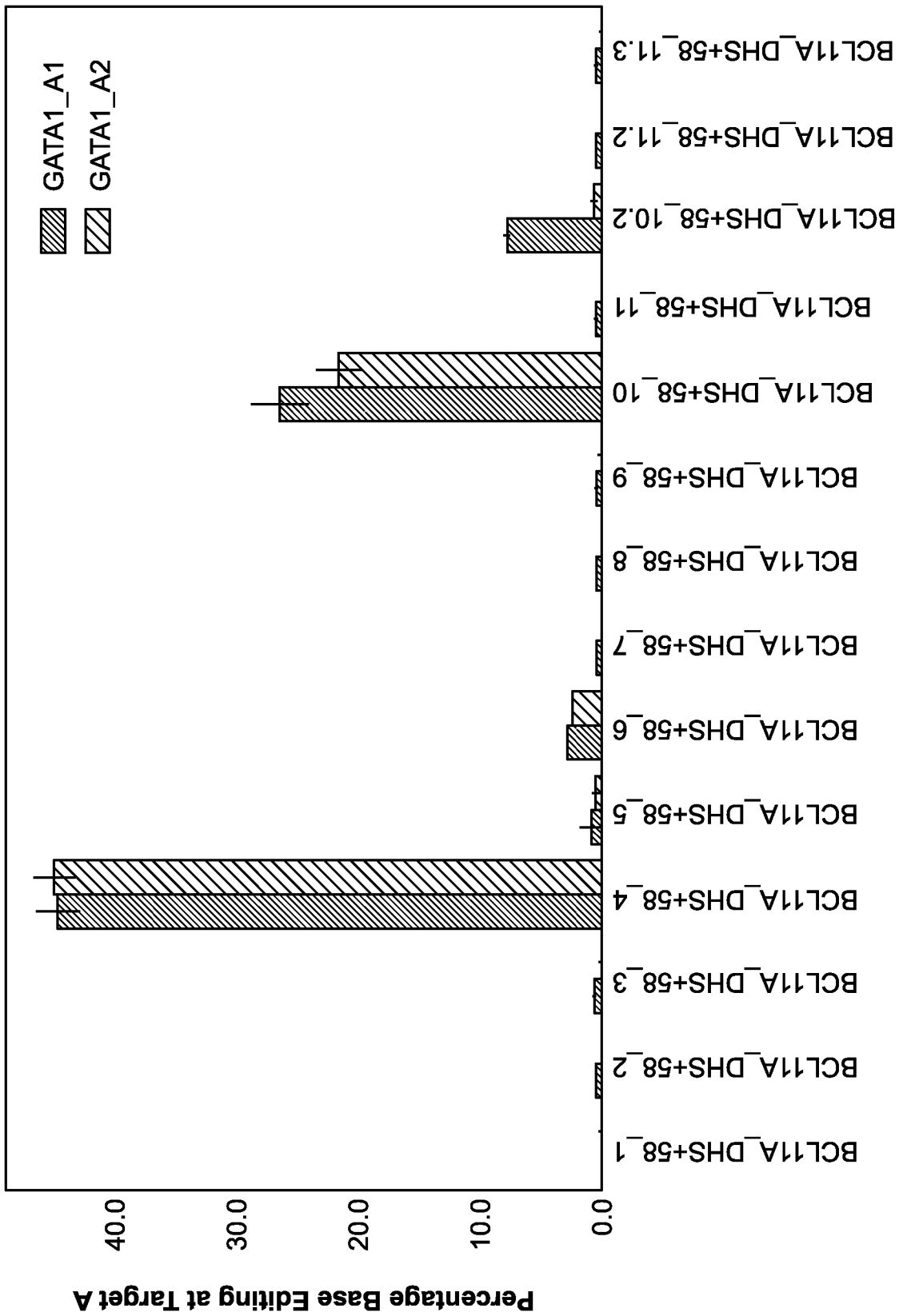
GATA1_A1
GATA1_A2
Percentage Base Editing at Target A
0.0
10.0
20.0
30.0
40.0
Target Guide
BCL11A_DHS+58_1
BCL11A_DHS+58_2
BCL11A_DHS+58_3
BCL11A_DHS+58_4
BCL11A_DHS+58_5
BCL11A_DHS+58_6
BCL11A_DHS+58_7
BCL11A_DHS+58_8
BCL11A_DHS+58_9
BCL11A_DHS+58_10
BCL11A_DHS+58_11
BCL11A_DHS+58_10.2
BCL11A_DHS+58_11.2
BCL11A_DHS+58_11.3

FIG. 33C

Table 2. Common Pathogenic Changes in *PAH* and Their Responsiveness to Sapropterin

cDNA	Protein	Cases in PAHdb	Responsive to Sapropterin
c.1222C>T	p.Arg408Trp	6.7%	<10%
c.1066-11G>A (IVS10-11G>A)		5.3%	<10%
c.194T>C	p.Ile65Thr	4.1%	89%
c.782G>A	p.Arg261Gln	3.6%	78%
c.842C>T	p.Pro281Leu	2.9%	None [Leuders et al 2014, biopku.org]
c.1315+1G>A (IVS12+1G>A)		2.8%	12.5% [biopku.org] None [Leuders et al 2014]
c.473G>A	p.Arg158Gln	2.7%	<10%

Data obtained from: PAHdb accessed 5/8/2016 (biopku.org); and Leuders et al [2014]. All changes with >2.5% frequency in the PAHdb database were included. In database searches, homozygosity was assumed for calculations; however, this is a rare finding in consanguineous individuals. It is recommended that all affected individuals be tested for personal responsiveness. Genetic changes shown affect >2.5% of the database population. See biopku.org for the most up-to-date information and additional references.

FIG. 34A

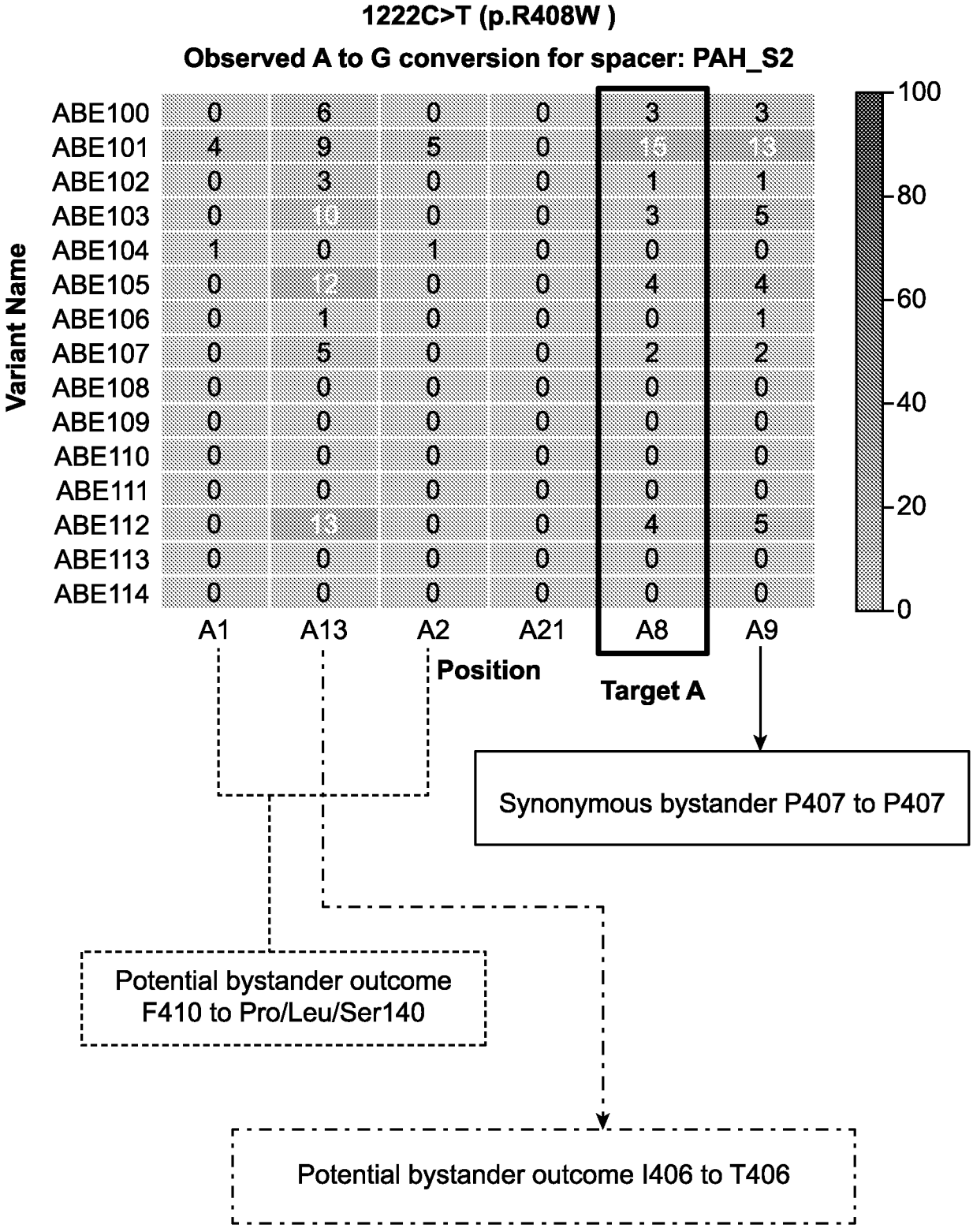
1222C>T (p.R408W)
Observed A to G conversion for spacer: PAH_S2
Variant Name
ABE100 0 6 0 0 3 3
ABE101 4 9 5 0 15 13
ABE102 0 3 0 0 1 1
ABE103 0 10 0 0 3 5
ABE104 1 0 1 0 0 0
ABE105 0 12 0 0 4 4
ABE106 0 1 0 0 0 1
ABE107 0 5 0 0 2 2
ABE108 0 0 0 0 0 0
ABE109 0 0 0 0 0 0
ABE110 0 0 0 0 0 0
ABE111 0 0 0 0 0 0
ABE112 0 13 0 0 4 5
ABE113 0 0 0 0 0 0
ABE114 0 0 0 0 0 0
A1 A13 A2 A21 A8 A9
Position
Target A
100
80
60
40
20
0
Synonymous bystander P407 to P407
Potential bystander outcome
F410 to Pro/Leu/Ser140
Potential bystander outcome I406 to T406

FIG. 34B

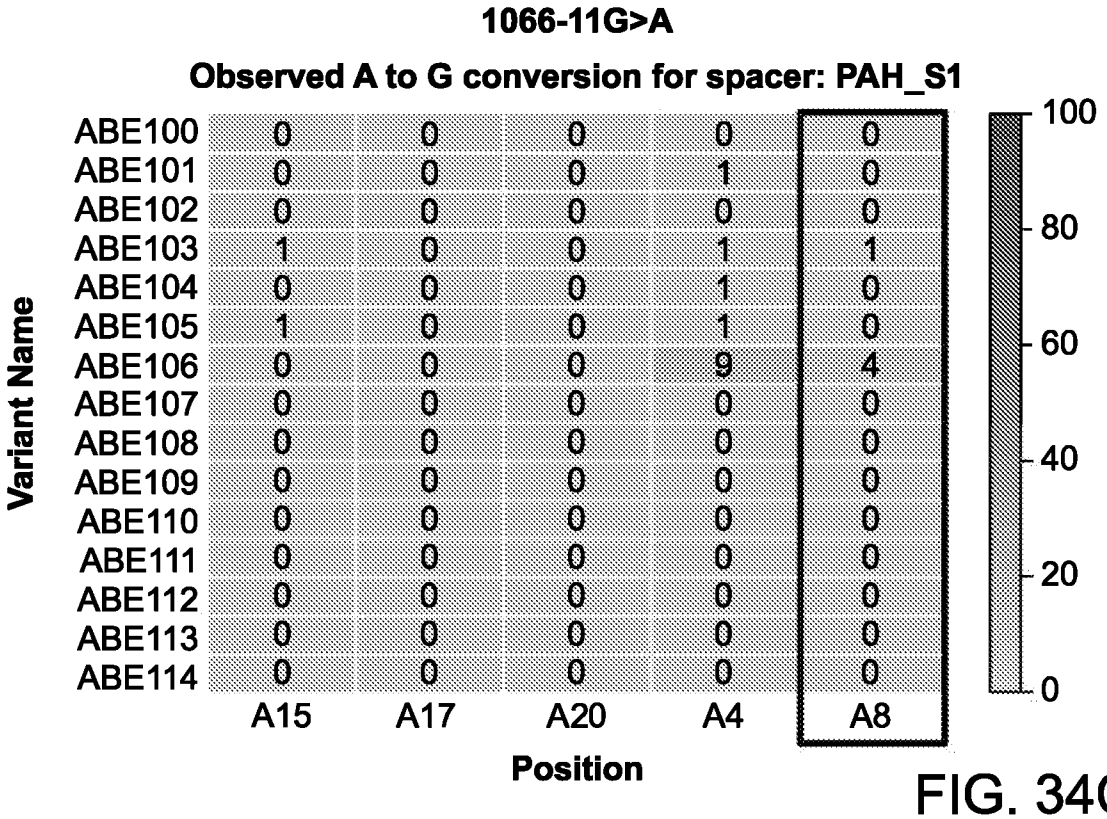

FIG. 34C

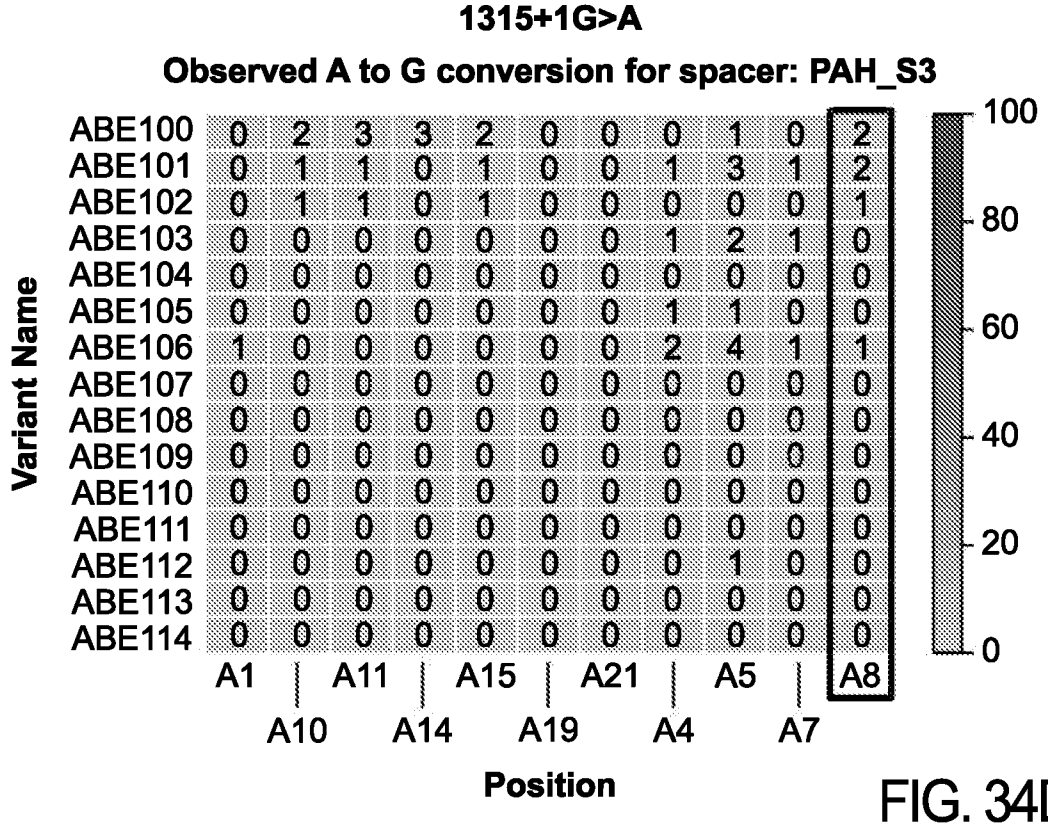

FIG. 34D