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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 March 2024 (28.03.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/064910 A1

(51) International Patent Classification:
C12N 15/113 (2010.01) *C12N 9/10* (2006.01)
C07K 19/00 (2006.01)

(21) International Application Number:
PCT/US2023/074931

(22) International Filing Date:
22 September 2023 (22.09.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/409,607 23 September 2022 (23.09.2022) US
63/502,328 15 May 2023 (15.05.2023) US
63/516,063 27 July 2023 (27.07.2023) US
63/581,229 07 September 2023 (07.09.2023) US

(71) Applicant: CHROMA MEDICINE, INC. [US/US]; 201 Brookline Ave., Suite 1101, Boston, Massachusetts 02215 (US).

(72) Inventors: JAFFE, Aron Brandon; 201 Brookline Ave., Suite 1101, Boston, Massachusetts 02215 (US). **ABUBUCKER, Noorussahar**; 201 Brookline Ave., Suite 1101, Boston, Massachusetts 02215 (US). **ANGLERO-RODRIGUEZ, Yesseinia**; 201 Brookline Ave., Suite 1101, Boston, Massachusetts 02215 (US). **MYER, Vic**; 201 Brookline Ave., Suite 1101, Boston, Massachusetts 02215 (US). **LOMBARDO, Angelo Leone**; 201 Brookline Ave., Suite 1101, Boston, Massachusetts 02215 (US). **CAPPELLUTI, Martino Alfredo**; 201 Brookline Ave., Suite 1101, Boston, Massachusetts 02215 (US).

(74) Agent: CYNN, Esther; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, California 94304 (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,

(54) Title: COMPOSITIONS AND METHODS FOR EPIGENETIC REGULATION OF HBV GENE EXPRESSION

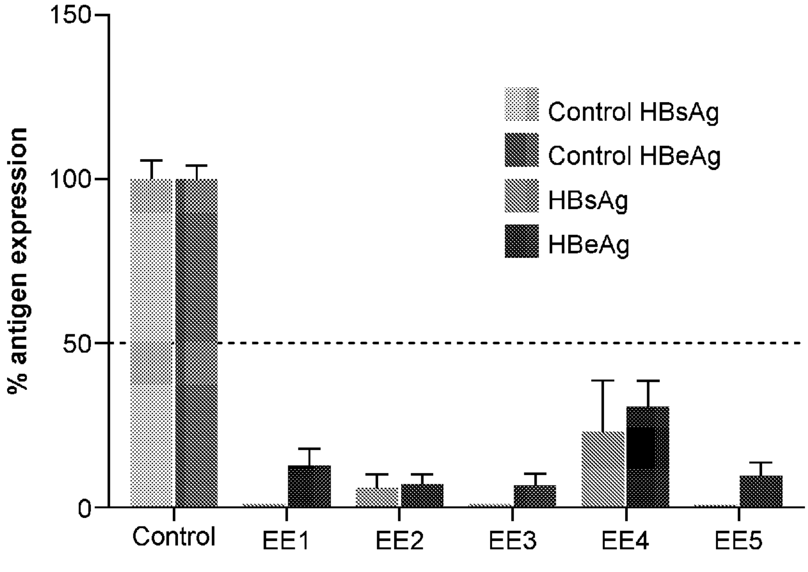


FIGURE 14D

(57) Abstract: This invention relates to compositions, methods, strategies, and treatment modalities related to the epigenetic modification of hepatitis B virus (HBV) genes.

WO 2024/064910 A1

[Continued on next page]

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).

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

COMPOSITIONS AND METHODS FOR EPIGENETIC REGULATION OF HBV GENE EXPRESSION

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 63/409,607, filed September 23, 2022, U.S. Provisional Application No. 63/502,328, filed May 15, 2023, U.S. Provisional Application No. 63/516,063, filed July 27, 2023, and U.S. Provisional Application No. 63/581,229, filed September 7, 2023, each of which is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

[0002] Despite available treatments, chronic hepatitis B (CHB) remains a high unmet medical need, with more than 250 million carriers of hepatitis B virus (HBV) worldwide and approximately 800,000 annual deaths due to HBV-related liver disease. Current approved CHB therapies elicit a functional cure rate (defined as durable HBsAg loss and undetectable serum HBV after completing a course of treatment) of less than 20%. Accordingly, there is a need for improved clinical modalities targeting HBV.

SUMMARY OF THE INVENTION

[0003] Some aspects of the present disclosure provide systems, compositions, strategies, and methods for the epigenetic modification of HBV, including HBV in host cells and organisms.

[0004] Some aspects of this disclosure provide methods of modifying an epigenetic state of a hepatitis B virus (HBV) gene or genome, comprising contacting the HBV gene or genome with an epigenetic editing system, wherein the epigenetic editing system comprises a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or one or more nucleic acid molecules encoding thereof, wherein the first DNA binding domain binds a first target region of the HBV gene or genome, and wherein the contacting results in a reduction of: number of HBV viral episomes, replication of the HBV gene or genome, and/or expression of a protein product encoded by the HBV gene or genome, wherein said reduction is at least about 20% compared to contacting the HBV gene or genome with a suitable control, and/or wherein said reduction of the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome is at least about 20% compared to the the number, replication, and/or expression in the subject before administering. Some aspects of this disclosure provide methods of treating an HBV infection in a subject comprising administering an epigenetic editing system to the subject, wherein the epigenetic editing system comprises a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or one or more nucleic acid molecules encoding thereof, wherein the first DNA binding domain binds a first target region of a HBV gene or genome, and wherein the contacting results in a reduction of: number of HBV viral episomes, replication of the HBV gene or genome, and/or expression of a protein product encoded by the HBV gene or genome, wherein said reduction is at least about 20% compared to administering a suitable control, and/or wherein said reduction of the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome is at least about 20% compared to the the number, replication, and/or expression in the subject before administering. Some aspects of this disclosure provide methods of modulating expression of an HBV gene or genome comprising contacting the HBV gene or genome with an epigenetic editing system, wherein the epigenetic editing system comprises a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or one or more nucleic acid molecules encoding thereof, wherein the first DNA binding domain binds a first target region of the HBV gene or genome, and wherein the contacting results in a reduction of expression of a gene product encoded by the HBV gene or genome, optionally, wherein the gene product is a nucleic acid or a protein, wherein said reduction is at least about 20% compared to contacting the HBV genome with a suitable control, and/or wherein said reduction of gene product encoded by the HBV gene or genome is at least about 20% compared to the expression in the subject before administering. Some aspects of this disclosure provide methods of inhibiting viral replication in a cell infected with an HBV comprising administering an epigenetic editing system, wherein the epigenetic editing system comprises a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or one or more nucleic acid molecules encoding thereof, wherein the first DNA binding domain binds a first target region of a HBV gene or genome, and wherein the epigenetic editing system targets a target region of the HBV gene or genome, and wherein the contacting results in a reduction of number of HBV viral episomes or replication of the HBV gene or genome, wherein said reduction is at least about 20% compared to administering a suitable control, and/or wherein said reduction of the number of HBV viral episomes or replication of the HBV gene or genome is at least about 20% compared to the the number and/or replication in the subject before administering. Some aspects of this disclosure provide methods comprising administering an epigenetic editing system to a subject in need thereof, wherein the epigenetic editing system comprises a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or one or more nucleic acid molecules encoding thereof, wherein the first DNA binding domain binds a first target region of a HBV gene or genome, and wherein the contacting results in a reduction of: number of HBV viral episomes, replication of the HBV gene or genome, or expression of a protein product encoded by the HBV gene or genome, wherein said reduction is at least about 20% compared to administering a suitable control , and/or wherein said reduction of the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome is at least about 20% compared to the the number, replication, and/or expression in the subject before administering. In some embodiments, the HBV genome is a covalently closed circular DNA (cccDNA) or an HBV integrated DNA. In some embodiments, the HBV genome comprises HBV genotype A, HBV genotype B, HBV genotype C, HBV genotype D, HBV genotype E, HBV genotype F, HBV genotype G or HBV genotype H. In some embodiments, the HBV genome comprises a sequence with at least 80% identity to an HBV genome sequence provided herein. In some embodiments, the first target region is located in a region of the HBV genome within nucleotide 0-303, 1000-2448 or 2802-3182 of an HBV genome provided herein In some embodiments, the first target region of the HBV genome is located in a CpG island. In some embodiments, the first target region of the HBV genome is located in a promotor. In some embodiments, the first target region of the HBV genome is located in a section of the HBV genome that encodes a transcript selected from the group consisting of a pgRNA, a precure mRNA, a preS mRNA, a S mRNA, and a X mRNA. In some embodiments, the first DNA binding domain comprises a CRISPR-Cas protein. In some embodiments, the epigenetic editing system further comprises a first guide RNA (gRNA) that comprises a region complementary to a strand of the first target region. In some embodiments, the gRNA comprises a sequence selected from a gRNA provided and/or disclosed herein, e.g., in Table 14 and/or 15. In some embodiments, the first DNA binding domain comprises a zinc-finger protein. In some embodiments, the zinc-finger protein comprises a zinc-finger motif with a sequence selected from any zinc finger or zinc finger motif provided herein, e.g., in Table 1. In some embodiments, the zinc-finger protein comprises a sequence of any of the zinc finger epigenetic repressors provided herein. In some embodiments, the transcriptional repressor domain comprises ZIM3 In some embodiments, the first DNMT domain is a DNMT3A domain or a DNMT3L domain. In some embodiments, the first DNMT domain comprises a sequence of a DNMT domain provided herein. In some embodiments, the epigenetic editing system further comprises a second DNMT domain or a nucleic acid encoding thereof. In some embodiments, the second DNMT domain is a DNMT3A domain or a DNMT3L domain. In some embodiments, the second DNMT domain comprises a sequence of a DNMT domain provided herein. In some embodiments, the epigenetic editing system comprises a fusion protein or a nucleic acid encoding thereof, and wherein the fusion protein comprises the first DNA binding domain, the first DNMT domain, the repressor domain and the second DNMT domain. In some embodiments, the fusion protein further comprises a nuclear localization sequence (NLS). In some embodiments, the fusion protein comprises a sequence of a fusion protein provided herein. In some embodiments, the epigenetic editing system further comprises a second DNA binding domain or a nucleic acid encoding thereof, wherein the second DNA binding domain binds a second target region of the HBV genome. In some embodiments, the second target region is located in a region of the HBV genome within nucleotide 0-303, 1000-2448 or 2802-3182. In some embodiments, the second target region of the HBV genome is located in a CpG island. In some embodiments, the second target region of the HBV genome is located in a promotor. In some embodiments, the second target region of the HBV genome is located in a section of the HBV genome that encodes a transcript selected from the group consisting of a pgRNA, a precure mRNA, a preS mRNA, a S mRNA, and a X mRNA. In some embodiments, the second DNA binding domain comprises a CRISPR-Cas protein. In some embodiments, the epigenetic editing system further comprises a second gRNA that comprises a region complementary to a strand of the second target region. In some embodiments, the gRNA comprises a sequence selected from a gRNA sequence provided herein, e.g., a sequence provided and/or disclosed in Table 14 and/or 15. In some embodiments, the second DNA binding domain comprises a zinc-finger protein. In some embodiments, the zinc-finger protein comprises a zinc-finger motif with a sequence selected from a zinc finger motif sequence provided herein, e.g., a zinc finger motif provided in Table 1. In some embodiments, the zinc-finger protein comprises a sequence of a zinc finger motif provided in Table 1. In some embodiments, the epigenetic editing system comprises a first fusion protein or a first nucleic acid encoding thereof and a second fusion protein or a second nucleic acid encoding thereof, wherein the first fusion protein comprises the first DNA binding domain and the first DNMT domain, and wherein the second fusion protein comprises the second DNA binding domain and the transcriptional repressor domain. In some embodiments, the first fusion protein comprises a sequence of a fusion protein provided herein. In some embodiments, the second fusion protein comprises a sequence of a fusion protein provided herein. In some embodiments, the epigenetic editing system further comprises a third DNA binding domain or a nucleic acid encoding thereof, wherein the third DNA binding domain binds to a third target region of the HBV genome. In some embodiments, the third target region is located in a region of the HBV genome within nucleotide 0-303, 1000-2448 or 2802-3182. In some embodiments, the third target region of the HBV genome is located in a CpG island. In some embodiments, the third target region of the HBV genome is located in a promotor. In some embodiments, the third target region of the HBV genome is located in a section of the HBV genome that encodes a transcript selected from the group consisting of a pgRNA, a precure mRNA, a preS mRNA, a S mRNA, and a X mRNA. In some embodiments, the third DNA binding domain comprises a CRISPR-Cas protein. In some embodiments, the epigenetic editing system further comprises a third gRNA that comprises a region complementary to a strand of the third target region. In some embodiments, the third gRNA comprises a sequence selected from a gRNA sequence provided herein, e.g., of a gRNA sequence provided and/or disclosed in Table 14 and/or 15. In some embodiments, the third DNA binding domain comprises a zinc-finger protein. In some embodiments, the zinc-finger protein comprises a zinc-finger motif with a sequence selected from a zinc finger motif provided herein. In some embodiments, the zinc-finger protein comprises a sequence of a zinc finger motif provided in Table 1. In some embodiments, the epigenetic editing system further comprises a second DNMT domain or a nucleic acid encoding thereof. In some embodiments, the second DNMT domain is a DNMT3A domain or a DNMT3L domain. In some embodiments, the epigenetic editing system comprises a third fusion protein or a nucleic acid encoding thereof, wherein the third fusion protein comprises the third DNA binding domain and the second DNMT domain. In some embodiments, the third fusion protein comprises a sequence of a fusion protein provided herein. In some embodiments, the epigenetic editing system comprises a nucleic acid sequence provided in Table 20. In some embodiments, the reduction of the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome is at least about 20% compared to the the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome measured or observed before contacting the HBV genome with the epigentic editing system, or before administering the epigenetic editing system to the subject. In some embodiments, the reduction of the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome is at least about 25%, at least about 50%, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 99%, at least about 99.5%, at least about 99.8%, at least about 99.9%, at least about 99.95%, at least about 99.99%, or more than 99.99%, compared to the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome measured or observed before contacting the HBV genome with the epigentic editing system, or before administering the epigenetic editing system to the subject.

[0005] Some aspects of this disclosure provide epigenetic editing systems comprising: a fusion protein or a nucleic acid encoding the fusion protein, wherein the fusion protein comprises: (a) a DNA-binding domain that binds a target region of a HBV gene or genome, (b) a first DNA methyltransferase (DNMT) domain, and (c) a transcriptional repressor domain. In some embodiments, the epigenetic editing system is capable of reducing a number of the HBV viral episome, replication of the HBV, or expression of a gene product encoded by the HBV gene or genome, wherein said reduction is at least about 20% compared to contacting the HBV gene or genome with a suitable control. In some embodiments, the HBV genome is a covalently closed circular DNA (cccDNA) or an HBV integrated DNA. In some embodiments, the HBV genome comprises HBV genotype A, HBV genotype B, HBV genotype C, HBV genotype D, HBV genotype E, HBV genotype F, HBV genotype G or HBV genotype H. In some embodiments, the HBV genome comprises a sequence with at least 80% identity to an HBV genome sequence provided herein. In some embodiments, the target region is located in a region of the HBV genome within nucleotide 0-303, 1000-2448 or 2802-3182 of an HBV genome sequence provided herein. In some embodiments, the target region of the HBV genome is located in a CpG island. In some embodiments, the target region of the HBV genome is located in a promotor. In some embodiments, the target region of the HBV genome is located in a section of the HBV genome that encodes a transcript selected from the group consisting of a pgRNA, a precure mRNA, a preS mRNA, a S mRNA, and a X mRNA. In some embodiments, the DNA binding domain comprises a CRISPR-Cas protein. In some embodiments, the epigenetic editing system further comprises a gRNA that comprises a region complementary to a strand of the target region. In some embodiments, the gRNA comprises a sequence selected from a gRNA sequence provided herein, e.g., in Table 14 and/or 15. In some embodiments, the DNA binding domain comprises a zinc-finger protein In some embodiments, the zinc-finger protein comprises a zinc-finger motif with a sequence selected from a zinc finger motif provided herein. In some embodiments, the zinc-finger protein comprises a sequence of a zinc finger motif provided in Table 1. In some embodiments, the transcriptional repressor domain comprises a sequence of a transcriptional repressor provided herein. In some embodiments, the first DNMT domain is a DNMT3A domain or a DNMT3L domain. In some embodiments, the DNMT domain comprises a sequence of a DNMT domain provided herein. In some embodiments, the fusion protein further comprises a second DNMT domain. In some embodiments, the second DNMT domain is a DNMT3A domain or a DNMT3L domain. In some embodiments, the fusion protein further comprises a nuclear localization sequence (NLS). In some embodiments, the fusion protein comprises a sequence of a fusion protein provided herein. Some aspects of the present disclosure provide epigenetic editing systems comprising: a first fusion protein or a nucleic acid encoding the first fusion protein, wherein the first fusion protein comprises a first DNA binding domain and a first DNMT domain, wherein the first DNA binding domain binds a first target region of a HBV genome, and a second fusion protein or a nucleic acid encoding the second fusion protein, wherein the second fusion protein comprises a second DNA binding domain and a transcriptional repressor domain, wherein the second DNA binding domain binds a second target region of the HBV genome. In some embodiments, the epigenetic editing system is capable of reducing a number of the HBV viral episome, replication of the HBV, or expression of a gene product encoded by the HBV genome, wherein said reduction is at least about 20% compared to contacting the HBV genome with a suitable control. In some embodiments, the HBV genome is a covalently closed circular DNA (cccDNA) or an HBV integrated DNA. In some embodiments, the HBV genome comprises HBV genotype A, HBV genotype B, HBV genotype C, HBV genotype D, HBV genotype E, HBV genotype F, HBV genotype G or HBV genotype H In some embodiments, the HBV genome comprises a sequence with at least 80% identity to an HBV genome provided herein. In some embodiments, the epigenetic editing system further comprises a third fusion protein or a nucleic acid encoding the third fusion protein, wherein the third fusion protein comprises a third DNA binding domain and a second DNMT domain, wherein the third DNA binding domain binds a third target region of the HBV genome. In some embodiments, the first target region, the second target region or the third target region is located in a region of the HBV genome within nucleotide 0-303, 1000-2448 or 2802-3182 of an HBV genome provided herein In some embodiments, the first target region, the second target region or the third target region of the HBV genome is located in a CpG island In some embodiments, the first target region, the second target region or the third target region of the HBV genome is located in a promotor In some embodiments, the first target region, the second target region or the third target region of the HBV genome is located in a section of the HBV genome that encodes a transcript selected from the group consisting of a pgRNA, a precure mRNA, a preS mRNA, a S mRNA, and a X mRNA In some embodiments, the first DNA binding domain, the second DNA binding domain or the third DNA binding domain comprises a CRISPR-Cas protein. In some embodiments, the epigenetic editing system further comprises a first gRNA that comprises a region complementary to a strand of the first target region, a second gRNA that comprises a region complementary to a strand of the second target region or a third RNA that comprises a region complementary to a strand of the third target region. In some embodiments, the first gRNA comprises a sequence selected from a gRNA sequence provided herein, e.g., provided and/or disclosed in Table 14 and/or 15, the second gRNA comprises a sequence selected from a gRNA sequence provided herein, e.g., provided and/or disclosed in Table 14 and/or 15, and/or the third gRNA comprises a sequence selected from a gRNA sequence provided and/or disclosed herein, e.g., provided and/or disclosed in Table 14 and/or 15. In some embodiments, the first DNA binding domain, the second DNA binding domain or the third DNA binding domain comprises a zinc-finger protein In some embodiments, the zinc-finger protein comprises a zinc-finger motif with a sequence selected from a zinc finger motif provided herein In some embodiments, the zinc-finger protein comprises a sequence of a zinc finger motif provided in Table 1. In some embodiments, the transcriptional repressor domain comprises ZIM3. In some embodiments, the first DNMT domain is a DNMT3A domain or a DNMT3L domain. In some embodiments, the first DNMT domain comprises a sequence of a DNMT provided herein. In some embodiments, the second DNMT domain is a DNMT3A domain or a DNMT3L domain. In some embodiments, the second DNMT domain comprises a sequence of a DNMT domain provided herein. In some embodiments, the first fusion protein comprises a sequence of a fusion protein provided herein. In some embodiments, the second fusion protein comprises a sequence of a fusion protein provided herein. In some embodiments, the third fusion protein comprises a sequence of a fusion protein provided herein. In some embodiments of any of the previous methods, the epigenetic editing system comprises a nucleic acid sequence provided in Table 20. In some embodiments, the reduction of the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome is at least about 20% compared to the the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome measured or observed before contacting the HBV genome with the epigentic editing system, or before administering the epigenetic editing system to the subject. In some embodiments, the reduction of the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome is at least about 25%, at least about 50%, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 99%, at least about 99.5%, at least about 99.8%, at least about 99.9%, at least about 99.95%, at least about 99.99%, or more than 99.99%, compared to the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome measured or observed before contacting the HBV genome with the epigentic editing system, or before administering the epigenetic editing system to the subject.

[0006] Some aspects of the present disclosure provide a method of treating an HDV infection in a subject comprising administering an epigenetic editing system to the subject, wherein the epigenetic editing system comprises a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or one or more nucleic acid molecules encoding thereof, wherein the first DNA binding domain binds a first target region of a HBV gene or genome, and wherein the contacting results in a reduction of: number of HDV viral episomes, replication of the HDV gene or genome, or expression of a protein product encoded by the HDV gene or genome, wherein said reduction is at least about 20% compared to administering a suitable control. Some aspects of the present disclosure provide a method of inhibiting viral replication in a cell infected with an HDV comprising administering an epigenetic editing system, wherein the epigenetic editing system comprises a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or one or more nucleic acid molecules encoding thereof, wherein the first DNA binding domain binds a first target region of a HBV gene or genome, and wherein the epigenetic editing system targets a target region of the HBV gene or genome, and wherein the contacting results in a reduction of number of HDV viral episomes or replication of the HDV gene or genome, wherein said reduction is at least about 20% compared to administering a suitable control. In some embodiments, the first DNA binding domain comprises a CRISPR-Cas protein. In some embodiments, the epigenetic editing system further comprises a first guide RNA (gRNA) that comprises a region complementary to a strand of the first target region. In some embodiments, the gRNA comprises a sequence selected from a gRNA provided herein, e.g., in Table 14 and/or 15. In some embodiments, the first DNA binding domain comprises a zinc-finger protein. In some embodiments, the zinc-finger protein comprises a zinc-finger motif with a sequence selected from any zinc finger or zinc finger motif provided herein, e.g., in Table 1 or Table 20. In some embodiments, the zinc-finger protein comprises a sequence of any of the zinc finger epigenetic repressors provided herein. In some embodiments, the transcriptional repressor domain comprises ZIM3. In some embodiments, the first DNMT domain is a DNMT3A domain or a DNMT3L domain. In some embodiments, the first DNMT domain comprises a sequence of a DNMT domain provided herein. In some embodiments, the epigenetic editing system further comprises a second DNMT domain or a nucleic acid encoding thereof. In some embodiments, the second DNMT domain is a DNMT3A domain or a DNMT3L domain. In some embodiments, the second DNMT domain comprises a sequence of a DNMT domain provided herein. In some embodiments, the epigenetic editing system comprises a fusion protein or a nucleic acid encoding thereof, and wherein the fusion protein comprises the first DNA binding domain, the first DNMT domain, the repressor domain and the second DNMT domain. In some embodiments, the fusion protein further comprises a nuclear localization sequence (NLS). In some embodiments, the fusion protein comprises a sequence of a fusion protein provided herein. In some embodiments, the first DNA binding domain binds a target region of a HBV gene or genome encoding or controlling expression of an S-antigen. In some embodiments, the epigenetic editing system comprises a nucleic acid sequence provided in Table 20. In some embodiments, the reduction of the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome is at least about 20% compared to the the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome measured or observed before contacting the HBV genome with the epigentic editing system, or before administering the epigenetic editing system to the subject. In some embodiments, the reduction of the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome is at least about 25%, at least about 50%, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 99%, at least about 99.5%, at least about 99.8%, at least about 99.9%, at least about 99.95%, at least about 99.99%, or more than 99.99%, compared to the number of HBV viral episomes, of replication of the HBV gene or genome, or of expression of a protein product encoded by the HBV gene or genome measured or observed before contacting the HBV genome with the epigentic editing system, or before administering the epigenetic editing system to the subject.

[0007] Other features, objectives, and advantages of the invention are apparent in the detailed description that follows. It should be understood, however, that the detailed description, while indicating embodiments and embodiments of the invention, is given by way of illustration only, not limitation. Various changes and modifications within the scope of the invention will become apparent to those skilled in the art from the detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] **Figure 1** is a diagram illustrating an exemplary structure of a circular HBV genome. HBV genes and CpG islands are indicated. Exemplary target sites for CRISPR-based epigenetic repressors (red arrows) as well as for zinc-finger-based epigenetic repressors (green arrows) are identified.

[0009] **Figure 2** is a heat map showing conservation of guide RNA target domains across different HBV genotypes.

[0010] **Figure 3** is a bar graph illustrating the geographical distribution of different HBV genotypes.

[0011] **Figure 4A** is a diagram describing the experimental timeline for testing different CRISPR-based epigenetic repressors in HepAD38 cells, which express HPV in a doxycycline-inducible manner. **Figure 4B** is a diagram showing the repression of HBV by various CRISPR-based epigenetic repressors (#1.1-3.2). Controls: UT: untransfected control; GFP: transfection control without repressor; HBV-KO: CRISPR nuclease mediated knockout; sgRNA scramble: CRISPR-based repressor with sgRNA not targeting HBV; B2M: CRISPR-based repressor with sgRNA targeting B2M.

[0012] **Figure 5A** is a diagram describing the experimental timeline for testing different CRISPR-based epigenetic repressors in a HepG2-NTCP infection model (see, e.g., Methods Mol Biol. 2017;1540:1-14). **Figure 5B** is a diagram showing the expression of HBe antigen (via ELISA) at different times after treatment of HBV-infected Hep2G-NTCT cells with different doses of CRISPR-based epigenetic repressors (ETRs), or with different doses of Cas9 nuclease targeting HBV (Cas9), plotted normalized to the expression value of HBe antigen measured for a negative control (empty).

[0013] **Figure 6** is a diagram describing the experimental timeline for a guide RNA screen testing different CRISPR-based epigenetic repressor systems in a HepG2-NTCP infection model with ELISA readout for HBe and HBs antigens at day 6.

[0014] **Figure 7** is a diagram showing QC results from different LNP batches used in the guide screen.

[0015] **Figure 8** is a bar graph showing the expression of HBe and HBs for an exemplary CRISPR-based epigenetic repressor (#3.2), calculated as the percentage of the expression of the respective antigen measured for a non-targeting control.

[0016] **Figure 9** is a diagram showing HBe expression values measured in the guide RNA screen for different guides (calculated as a percentage of the expression of HBe measured for a non-targeting control). Each guide/repressor combination is represented by a dot. A 50% repression cutoff is shown as a horizontal line. The position of the respective guide RNA within the HBV genome (shown at the bottom of the graph) is mapped on the X-axis. The position and the measured modulation of HBe expression for exemplary guide RNA #3.2 is indicated by red lines.

[0017] **Figure 10** is a diagram showing HBs expression values measured in the guide RNA screen for different guides (calculated as a percentage of the expression of HBs measured for a non-targeting control). Each guide/repressor combination is represented by a dot. A 50%

repression cutoff is shown as a horizontal line. The position of the respective guide RNA within the HBV genome (shown at the bottom of the graph) is mapped on the X-axis. The position and the measured modulation of HBs expression for exemplary guide RNA #3.2 is indicated by red lines.

[0018] **Figure 11** is a diagram showing a correlation between HBs and HBe expression for the guides tested. The graph on the right shows HBe and HBs repression efficiencies for 25 exemplary guides.

[0019] **Figure 12A** is a diagram describing the experimental timeline for a guide RNA assay testing CRISPR-off single construct epigenetic editor in combination with individual exemplary gRNAs in a HepG2-NTCP infection model with ELISA readout for HBe and HBs antigens at day 6; and **Figure 12B** is a graph summarizing the percentage reduction in HBV antigens at day 6 relative to non-targeting control.

[0020] **Figure 13A** is a diagram describing the experimental timeline for a guide RNA assay testing CRISPR-off single construct epigenetic editor in combination with individual exemplary gRNAs in a PLC/PRF/5 cell model with ELISA readout for HBs antigen at day 4; and **Figure 13B** is a graph summarizing the percentage reduction in HBs antigen at day 4 relative to non-targeting control.

[0021] **Figure 14A** is a diagram describing the experimental timeline for a guide RNA assay testing CRISPR-off single construct epigenetic editor in combination with individual exemplary gRNAs in a PXB cell model with ELISA readout for HBe and HBs antigens at day 6; and **Figure 14B** is a graph summarizing the percentage reduction in HBV antigens at day 6 relative to non-targeting control. **Figure 14C** is a diagram describing the experimental timeline for a guide RNA assay testing CRISPR-off single construct epigenetic editor in combination with individual exemplary gRNAs in a PXB cell model with ELISA readout for HBe and HBs antigens at day 12. **Figure 14D** is a graph summarizing the percentage reduction in HBV antigens at day 12 relative to non-targeting control. Bars represent mean±SEM; N=5. EE1= PLA002 and gRNA#007, EE2=PLA002 and gRNA#008, EE3=PLA002 and gRNA#009, EE4=PLA002 and gRNA#015, and EE5=PLA002 and gRNA#011.

[0022] **Figure 15A** is a diagram describing the experimental timeline for a zinc finger assay testing ZF-off single construct epigenetic editor that contains individual exemplary zinc finger motif in a HepG2-NTCP infection model with ELISA readout for HBe and HBs antigens at day 6; and **Figure 15B** is a graph summarizing the percentage reduction in HBV antigens at day 6 relative to non-targeting control. "N" denotes non-targeting control, "P" denotes the positive control, and the individual numbers on the x-axis denote exemplary constructs tested in the experiment, for instance, "1" represents "mRNA0001" construct, and "20" represents "mRNA0020" construct.

[0023] **Figure 16A** is a graph summarizing the results of top ten ZF-off constructs from **Figure 15B**. **Figure 16B** is a diagram showing HBsAg (top) and HBeAg (middle) expression values measured in the ZF-off screen (calculated as a percentage of the expression of HBsAg or HBeAg--top and middle, respectively--measured for a non-targeting control). Each ZF-off construct is represented by a dot. 50% and 60% repression cutoffs are shown as horizontal lines. The position of the respective guide RNA within the HBV genome (bottom) is mapped on the X-axis.

[0024] **Figure 17** is an experimental timeline for testing dose response (top) and two graphs showing dose response of % HbsAg (bottom left) and % HbeAg (bottom right) in HepG2-NTCP cells upon administration of ZF fusion proteins. The mRNA corresponding to the ZF motif for each fusion protein is indicated.

[0025] **Figure 18** is an experimental timeline for testing durable silencing of HBsAg (top) and a graph showing the durability of HBsAg silencing by ZF fusion proteins (bottom). The mRNA corresponding to the ZF motif for each fusion protein is indicated.

[0026] **Figure 19** is an experimental timeline for testing HBsAg silencing in a PLC/PRF/5 *in vitro* model (top) and a graph showing % HBsAg relative to control on Day 14 after administration of ZF fusion proteins. The mRNA corresponding to the ZF motif for each fusion protein is indicated. Information about the % match to target for each construct is also indicated.

[0027] **Figure 20A** is a volcano plot showing differentially expressed (DE) genes for an exemplary ZF specificity assay. DE genes are shown with dots. **Figure 20B** is a volcano plot showing DE for CRISPR-off and gRNA epigenetic editors. Points represent genes with their change in expression (x-axis) and statistical significance of that change (y-axis). EE1= PLA002 and gRNA#007, EE2=PLA002 and gRNA#008, EE3=PLA002 and gRNA#009, EE4=PLA002 and gRNA#015, and EE5=PLA002 and gRNA#011. Also shown are results for low specificity and host target gene controls. **Figures 20C-20D** are scatter plots showing methylation levels between treatment (y-axis) and control (x-axis) for 935,000 CpG sites in the human genome. Lines represent thresholds for changes in methylation considered significant (absolute [methylation difference]>= 0.2). DMRs are noted on each figure. Results for a host target (PCSK9, next-to-final panel) as well as a low specificity control (final panel) are also shown. **Figure 20C** shows the results versus effector only. **Figure 20D** shows the results versus no treatment. EE1= PLA002 and gRNA#007, EE2=PLA002 and gRNA#008, EE3=PLA002 and gRNA#009, EE4=PLA002 and gRNA#015, EE5=PLA002 and gRNA#011, EE6=PLA002 and gRNA#003, and EE7=PLA002 and gRNA#016.

[0028] **Figure 21** is an illustration of an experimental schematic for an *in vivo* study of multiplexing ZF fusion protein effectors.

DETAILED DESCRIPTION OF THE INVENTION

[0029] The present disclosure provides epigenetic editors, and strategies and methods of using such epigenetic editors, for regulating expression of HBV. By altering expression of HBV, and in particular, by repressing expression of HBV, e.g., of a gene comprised in the HBV genome or a gene product encoded by the HBV genome, the compositions and methods described herein are useful to suppress viral function in infected cells, e.g., in the context of treating an HBV infection in a human subject, or in the context of treating CHB.

[0030] The structure and biology of HBV as well as HBV-associated diseases have been reported (see, for example, Yuen, MF., Chen, DS., Dusheiko, G. et al. Hepatitis B virus infection. Nat Rev Dis Primers 4, 18035 (2018), incorporated herein by reference in its entirety).

[0031] Exemplary HBV sequences can be found at various NCBI database entries, e.g., representative sequences can be found under accession numbers NC_00397 and U95551, which are incorporated herein by reference in their entirety, and the sequences of which are provided elsewhere herein.

[0032] A number of treatment options for HBV has been reported, but there remains a need for effective treatment of HBV infections. Genetic editing approaches targeting HBV genomes for cutting of genomic DNA are associated with a risk of off-target cutting and genomic translocations. The present epigenetic editors and related methods of use have several advantages compared to other genome engineering methods, including increased efficiency, decreased risk of translocation, and durable silencing of HBV.

[0033] Hepatitis D virus (HDV) is the smallest pathogen known to infect humans. HDV infection is only found in patients infected with HBV, as HDV relies on HBV functions for most of its functions, including viral packaging, infectivity, transmission, and inhibition of host immunity. About 5% of patients with HBV infection also have an HDV infection. HDV uses HBV S-antigen (HBsAg) as a capsid protein, and HDV infection is therefore dependent on HBV S-antigen production. Decreasing HBV S-antigen expression also reduces HDV infectivity. The structure and biology of HDV has been reported (see, for example, Asselah and Rizzetto, Hepatitis D Virus Infection, The New England Journal of Medicine (389;1; July 6, 2023), incorporated herein by reference in its entirety). In some embodiments of the present disclosure, HDV infection is addressed through methods targeting an HBV gene or genome.

[0034] In some embodiments, an epigenetic editor as described herein may comprise one or more fusion proteins, wherein each fusion protein comprises a DNA-binding domain linked to one or more effector domains for epigenetic modification. In certain embodiments, where the DNA-binding domain is a polynucleotide guided DNA-binding domain, the epigenetic editor may further comprise one or more guide polynucleotides. DNA-binding domains, effector domains, and guide polynucleotides of an epigenetic editor as described herein may be selected, e.g., from those described below, in any functional combination.

[0035] The epigenetic editors described herein may be expressed in a host cell transiently, or may be integrated in a genome of the host cell; such cells and their progeny are also contemplated by the present disclosure. Both transiently expressed and integrated epigenetic editors or components thereof can effect stable epigenetic modifications. For example, after introducing to a host cell an epigenetic editor described herein, the target gene in the host cell may be stably or permanently repressed or silenced. For example, in some embodiments provided herein, a transiently expressed epigenetic editor comprising a DNMT3A domain, a DNMT3L domain, and a KRAB domain effects stable epigenetic modifications. For example, in some embodiments provided herein, a constitutively expressed epigenetic editor comprising DNMT3A and a DNMT3L domain effects stable epigenetic modifications. In some embodiments, expression of the target gene is reduced or silenced for at least 1 week, at least 2 weeks, at least 3 weeks, at least 4 weeks, at least 5 weeks, at least 6 weeks, at least 7 weeks, at least 2 months, at least 3 months, at least 4 months, at least 5 months, at least 6 months, at least 1 year, at least 2 years, or for the entire lifetime of the cell or the subject carrying the cell, as compared to the level of expression in the absence of the epigenetic editor. The epigenetic modification may be inherited by the progeny of the host cells into which the epigenetic editor was introduced.

[0036] The present epigenetic editors may be introduced to a patient in need thereof (e.g., a human patient), e.g., into the patient's hepatocytes, biliary epithelial cells (cholangiocytes), stellate cells, Kupffer cells, and liver sinusoidal endothelial cells.

I. DNA-Binding Domains

[0037] An epigenetic editor described herein may comprise one or more DNA-binding domains that direct the effector domain(s) of the epigenetic editor to target sequences within an HBV genome. A DNA-binding domain as described herein may be, e.g., a polynucleotide guided DNA-binding domain, a zinc finger protein (ZFP) domain, a transcription activator like effector (TALE) domain, a meganuclease DNA-binding domain, and the like. Examples of DNA-binding domains can be found in U.S. Patent 11,162,114, which is incorporated by refence herein in its entirety.

[0038] In some embodiments, a DNA-binding domain described herein is encoded by its native coding sequence. In other embodiments, the DNA-binding domain is encoded by a nucleotide sequence that has been codon-optimized for optimal expression in human cells.

A. Polynucleotide Guided DNA-Binding Domains

[0039] In some embodiments, a DNA-binding domain herein may be a protein domain directed by a guide nucleic acid sequence (e.g., a guide RNA sequence) to a target site in an HBV genome. In certain embodiments, the protein domain may be derived from a CRISPR-associated nuclease, such as a Class I or II CRISPR-associated nuclease. In some embodiments, the protein domain may be derived from a Cas nuclease such as a Type II, Type IIA, Type IIB, Type IIC, Type V, or Type VI Cas nuclease. In certain embodiments, the protein domain may be derived from a Class II Cas nuclease selected from Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Cas14a, Cas14b, Cas14c, CasX, CasY, CasPhi, C2c4, C2c8, C2c9, C2c10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csf1, Csf2, CsO, Csf4, and homologues and modified versions thereof. “Derived from” is used to mean that the protein domain comprises the full polypeptide sequence of the parent protein, or comprises a variant thereof (e.g., with amino acid residue deletions, insertions, and/or substitutions). The variant retains the desired function of the parent protein (e.g., the ability to form a complex with the guide nucleic acid sequence and the target DNA).

[0040] In some embodiments, the CRISPR-associated protein domain may be a Cas9 domain described herein. Cas9 may, for example, refer to a polypeptide with at least about 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity and/or sequence similarity to a wildtype Cas9 polypeptide described herein. In some embodiments, said wildtype polypeptide is Cas9 from *Streptococcus pyogenes* (NCBI Ref. No. NC_002737.2 (SEQ ID NO: 1)) and/or UniProt Ref. No. Q99ZW2 (SEQ ID NO: 2). In some embodiments, said wildtype polypeptide is Cas9 from *Staphylococcus aureus* (SEQ ID NO: 3). In some embodiments, the CRISPR-associated protein domain is a Cpf1 domain or protein, or a polypeptide with at least about 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity and/or sequence similarity to a wildtype Cpf1 polypeptide described herein (e.g., Cpf1 from *Franscisella novicida* (UniProt Ref. No. U2UMQ6 or SEQ ID NO: 4). In certain embodiments, the CRISPR-associated protein domain may be a modified form of the wildtype protein comprising one or more amino acid residue changes such as a deletion, an insertion, or a substitution; a fusion or chimera; or any combination thereof.

[0041] Cas9 sequences and structures of variant Cas9 orthologs have been described for various organisms. Exemplary organisms from which a Cas9 domain herein can be derived include, but are not limited to, *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus sp.*, *Staphylococcus aureus*, *Listeria innocua*, *Lactobacillus gasseri*, *Francisella novicida*, *Wolinella succinogenes*, *Sutterella wadsworthensis*, *Gamma proteobacterium*, *Neisseria meningitidis*, *Campylobacter jejuni*, *Pasteurella multocida*, *Fibrobacter succinogene*, *Rhodospirillum rubrum*, *Nocardiopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycoides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Lactobacillus buchneri*, *Treponema denticola*, *Microscilla marina*, *Burkholderiales bacterium*, *Polar omonas naphthalenivorans*, *Polar omonas sp.*, *Crocosphaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor becscii*, *Candidatus Desulforudis*, *Clostridium botulinum*, *Clostridium difficile*, *Finegoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionium*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira sp.*, *Lyngbya sp.*, *Microcoleus chthonoplastes*, *Oscillator ia sp.*, *Petrotoga mobilis*, *Thermosipho africanus*, *Streptococcus pasteurianus*, *Neisseria cinerea*, *Campylobacter lari*, *Parvibaculum lavamentivorans*, *Coryne bacterium diphtheria*, and *Acaryochloris marina*. Cas9 sequences also include those from the organisms and loci disclosed in Chylinski et al., *RNA Biol.* (2013) 10(5):726-37.

[0042] In some embodiments, the Cas9 domain is from *Streptococcus pyogenes*. In some embodiments, the Cas9 domain is from *Staphylococcus aureus*.

[0043] Other Cas domains are also contemplated for use in the epigenetic editors herein. These include, for example, those from CasX (Cas12E) (e.g., SEQ ID NO: 5), CasY (Cas12d) (e.g., SEQ ID NO: 6), Casφ (CasPhi) (e.g., SEQ ID NO: 7), Cas12f1 (Cas14a) (e.g., SEQ ID

NO: 8), Cas12f2 (Cas14b) (e.g., SEQ ID NO: 9), Cas12f3 (Cas14c) (e.g., SEQ ID NO: 10), and C2c8 (e.g., SEQ ID NO: 11).

[0044] For epigenetic editing, the nuclease-derived protein domain (e.g., a Cas9 or Cpf1 domain) may have reduced or no nuclease activity through mutations such that the protein domain does not cleave DNA or has reduced DNA-cleaving activity while retaining the ability to complex with the guide nucleic acid sequence (e.g., guide RNA) and the target DNA. For example, the nuclease activity may be reduced by at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% compared to the wildtype domain. In some embodiments, a CRISPR-associated protein domain described herein is catalytically inactive ("dead"). Examples of such domains include, for example, dCas9 ("dead" Cas9), dCpf1, ddCpf1, dCasPhi, ddCas12a, dLbCpf1, and dFnCpf1. A dCas9 protein domain, for example, may comprise one, two, or more mutations as compared to wildtype Cas9 that abrogate its nuclease activity. The DNA cleavage domain of Cas9 is known to include two subdomains: the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A (in RuvC1) and H840A (in HNH) completely inactivate the nuclease activity of SpCas9. SaCas9, similarly, may be inactivated by the mutations D10A and N580A. In some embodiments, the dCas9 comprises at least one mutation in the HNH subdomain and/or the RuvC1 subdomain that reduces or abrogates nuclease activity. In some embodiments, the dCas9 only comprises a RuvC1 subdomain, or only comprises an HNH subdomain. It is to be understood that any mutation that inactivates the RuvC1 and/or the HNH domain may be included in a dCas9 herein, e.g., insertion, deletion, or single or multiple amino acid substitution in the RuvC1 domain and/or the HNH domain.

[0045] In some embodiments, a dCas9 protein herein comprises a mutation at position(s) corresponding to position D10 (e.g., D10A), H840 (e.g., H840A), or both, of a wildtype SpCas9 sequence as numbered in the sequence provided at UniProt Accession No. Q99ZW2 (SEQ ID NO: 2). In particular embodiments, the dCas9 comprises the amino acid sequence of dSpCas9 (D10A and H840A) (SEQ ID NO: 12).

[0046] In some embodiments, a dCas9 protein as described herein comprises a mutation at position(s) corresponding to position D10 (e.g., D10A), N580 (e.g., N580A), or both, of a wildtype SaCas9 sequence (e.g., SEQ ID NO: 9). In particular embodiments, the dCas9 comprises the amino acid sequence of dSaCas9 (D10A and N580A) (SEQ ID NO.: 13).

[0047] Additional suitable mutations that inactivate Cas9 will be apparent to those of skill in the art based on this disclosure and knowledge in the field and are within the scope of this disclosure. Such mutations may include, but are not limited to, D839A, N863A, and/or K603R in SpCas9. The present disclosure contemplates any mutations that reduce or abrogate the nuclease activity of any Cas9 described herein (e.g., mutations corresponding to any of the Cas9 mutations described herein).

[0048] A dCpf1 protein domain may comprise one, two, or more mutations as compared to wildtype Cpf1 that reduce or abrogate its nuclease activity. The Cpf1 protein has a RuvC-like endonuclease domain that is similar to the RuvC domain of Cas9, but does not have an HNH endonuclease domain, and the N-terminal of Cpf1 does not have the alpha-helical recognition lobe of Cas9. In some embodiments, the dCpf1 comprises one or more mutations corresponding to position D917A, E1006A, or D1255A as numbered in the sequence of the *Francisella novicida* Cpf1 protein (FnCpf1; SEQ ID NO: 4). In certain embodiments, the dCpf1 protein comprises mutations corresponding to D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, or D917A/ E1006A/D1255A, or corresponding mutation(s) in any of the Cpf1 amino acid sequences described herein. In some embodiments, the dCpf1 comprises a D917A mutation. In particular embodiments, the dCpf1 comprises the amino acid sequence of dFnCpf1 (SEQ ID NO: 14).

[0049] Further nuclease inactive CRISPR-associated protein domains contemplated herein include those from, for example, dNmeCas9 (e.g., SEQ ID NO: 15), dCjCas9 (e.g., SEQ ID NO: 16), dSt1Cas9 (e.g., SEQ ID NO: 17), dSt3Cas9 (e.g., SEQ ID NO: 18), dLbCpf1 (e.g., SEQ ID NO: 19), dAsCpf1 (e.g., SEQ ID NO: 20), denAsCpf1 (e.g., SEQ ID NO: 21), dHFAsCpf1 (e.g., SEQ ID NO: 22), dRVRAsCpf1 (e.g., SEQ ID NO: 23), dRRAsCpf1 (e.g., SEQ ID NO: 24), dCasX (e.g., SEQ ID NO: 25), and dCasPhi (e.g., SEQ ID NO: 26).

[0050] In some embodiments, a Cas9 domain described herein may be a high fidelity Cas9 domain, e.g., comprising one or more mutations that decrease electrostatic interactions between the Cas9 domain and the sugar-phosphate backbone of DNA to confer increased target binding specificity. In certain embodiments, the high fidelity Cas9 domain may be nuclease inactive as described herein.

[0051] A CRISPR-associated protein domain described herein may recognize a protospacer adjacent motif (PAM) sequence in a target gene. A “PAM” sequence is typically a 2 to 6 bp DNA sequence immediately following the sequence targeted by the CRISPR-associated protein domain. The PAM sequence is required for CRISPR protein binding and cleavage but is not part of the target sequence. The CRISPR-associated protein domain may either recognize a naturally occurring or canonical PAM sequence or may have altered PAM specificity. CRISPR-associated protein domains that bind to non-canonical PAM sequences have been described in the art. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver et al., *Nature* (2015) 523(7561):481-5 and Kleinstiver et al., *Nat Biotechnol.* (2015) 33:1293-8. Such Cas9 domains may include, for example, those from "VRER" SpCas9, "EQR" SpCas9, "VQR" SpCas9, "SpG Cas9," "SpRYCas9," and "KKH" SaCas9. Nuclease inactive versions of these Cas9 domains are also contemplated, such as nuclease inactive VRER SpCas9 (e.g., SEQ ID NO: 27), nuclease inactive EQR SpCas9 (e.g., SEQ ID NO: 28), nuclease inactive VQR SpCas9 (e.g., SEQ ID NO: 29), nuclease inactive SpG Cas9 (e.g., SEQ ID NO: 30), nuclease inactive SpRY Cas9 (e.g., SEQ ID NO: 31), and nuclease inactive KKH SaCas9 (e.g., SEQ ID NO: 32). Another example is the Cas9 of *Francisella novicida* engineered to recognize 5'-YG-3' (where "Y" is a pyrimidine).

[0052] Additional suitable CRISPR-associated proteins, orthologs, and variants, including nuclease inactive variants and sequences, will be apparent to those of skill in the art based on this disclosure.

[0053] Guide RNAs that can be used in conjunction with the CRISPR-associated protein domains herein are further described in Section II below.

B. Zinc Finger Protein Domains

[0054] In some embodiments, the DNA-binding domain of an epigenetic editor described herein comprises a zinc finger protein (ZFP) domain (or "ZF domain" as used herein). ZFPs are proteins having at least one zinc finger, and bind to DNA in a sequence-specific manner. A "zinc finger" (ZF) or "zinc finger motif" (ZF motif) refers to a polypeptide domain comprising a beta-beta-alpha (ββα)-protein fold stabilized by a zinc ion. A ZF binds from two to four base pairs of nucleotides, typically three or four base pairs (contiguous or noncontiguous). Each ZF typically comprises approximately 30 amino acids. ZFP domains may contain multiple ZFs that make tandem contacts with their target nucleic acid sequence. A tandem array of ZFs may be engineered to generate artificial ZFPs that bind desired nucleic acid targets. ZFPs may be rationally designed by using databases comprising triplet (or quadruplet) nucleotide sequences and individual ZF amino acid sequences, in which each triplet or quadruplet nucleotide sequence is associated with one or more amino acid sequences of ZFs that bind the particular triplet or quadruplet sequence. See, e.g., U.S. Patents 6,453,242, 6,534,261, and 8,772,453.

[0055] ZFPs are widespread in eukaryotic cells, and may belong to, e.g., C2H2 class, CCHC class, PHD class, or RING class. An exemplary motif characterizing one class of these proteins (C2H2 class) is -Cys-$(X)_{2\text{-}4}$-Cys-$(X)_{12}$-His-$(X)_{3\text{-}5}$-His- (SEQ ID NO:1091), where X is any independently chosen amino acid. In some embodiments, a ZFP domain herein may comprise a ZF array comprising sequential C2H2-ZFs each contacting three or more sequential nucleotides. Additional architectures, e.g. as described in Paschon et al., *Nat. Commun.* 10, 1133 (2019), are also possible.

[0056] A ZFP domain of an epigenetic editor described herein may include 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more ZFs. The ZFP domain may include an array of two-finger or three-finger units, e.g., 3, 4, 5, 6, 7, 8, 9 or 10 or more units, wherein each unit binds a subsite in the target sequence. In some embodiments, a ZFP domain comprising at least three ZFs recognizes a target DNA sequence of 9 or 10 nucleotides. In some embodiments, a ZFP domain comprising at least four ZFs recognizes a target DNA sequence of 12 to 14 nucleotides. In some embodiments, a ZFP domain comprising at least six ZFs recognizes a target DNA sequence of 18 to 21 nucleotides.

[0057] In some embodiments, ZFs in a ZFP domain described herein are connected via peptide linkers. The peptide linkers may be, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more amino acids in length. In some embodiments, a linker comprises 5 or more amino acids. In some embodiments, a linker comprises 7-17 amino acids. The linker may be flexible or rigid.

[0058] In some embodiments a zinc finger array may have the sequence:

> SRPGERPFQCRICMRNFSXXXXXXXH*XX*THTGEKPFQCRICMRNFSXXXXXXXH*XX*TH[**linker**]FQCRICMRNFSXXXXXXXH*XX*THTGEKPFQCRICMRNFSXXXXXXXH*XX*TH[**linker**]PFQCRICMRNFSXXXXXXXH*XX*THTGEKPFQCRICMRNFSXXXXXXXH*XX*THLRGS (SEQ ID NO: 1084),

or a sequence at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical thereto, where "XXXXXXX" represents the amino acids of the ZF recognition helix, which confers DNA-binding specificity upon the zinc finger; each X may be independently chosen. In the above sequence, "*XX*" in italics may be TR, LR or LK, and "[linker]" represents a linker sequence. In some embodiments, the linker sequence is TGSQKP (SEQ ID NO: 1085); this linker may be used when sub-sites targeted by the ZFs are adjacent. In some embodiments, the linker sequence is TGGGGSQKP (SEQ ID NO: 1086); this linker may be used when there is a base between the sub-sites targeted by the zinc fingers. The two indicated linkers may be the same or different.

[0059] ZFP domains herein may contain arrays of two or more adjacent ZFs that are directly adjacent to one another (e.g., separated by a short (canonical) linker sequence), or are separated by longer, flexible or structured polypeptide sequences. In some embodiments, directly adjacent fingers bind to contiguous nucleic acid sequences, i.e., to adjacent trinucleotides/triplets. In some embodiments, adjacent fingers cross-bind between each other's respective target triplets, which may help to strengthen or enhance the recognition of the target sequence, and leads to the binding of overlapping sequences. In some embodiments, distant ZFs within the ZFP domain may recognize (or bind to) non-contiguous nucleotide sequences.

[0060] The amino acid sequences of the ZF DNA-recognition helices of exemplary ZFP domains herein, and their HBV target sequences, are shown below in **Table 1**.

Table 1. Zinc finger transcriptional repressors for silencing HBV. ZF sequences of exemplary ZFP domains are presented. SEQ ID Nos for target sequences and ZF can be found in Table 20 sequence listing.

ZFP	SEQ ID	Target Sequence	Start	End	Strd	F1	F2	F3	F4	F5	F6
ZFP894	33	GATGAGGCATAGCAGCAG	415	432	-	KKFNLLQ	RQDNLNS	RSHNLKL	QSTTLKR	RNTNLTR	IKHNLAR
ZFP895	34	GATGAGGCATAGCAGCAG	415	432	-	KKFNLLQ	RKDYLIS	RSHNLKL	QSTTLKR	RQDNLGR	VVNNLNR
ZFP896	35	GATGAGGCATAGCAGCAG	415	432	-	KKFNLLQ	RKDYLIS	RSHNLRL	QSTTLKR	RQDNLGR	VVNNLNR
ZFP899	36	GATGATTAGGCAGAGGTG	1828	1845	-	RRHILDR	RQDNLGR	QSTTLKR	RRDGLAG	VHHNLVR	ISHNLAR
ZFP900	37	GATGATTAGGCAGAGGTG	1828	1845	-	RREVLEN	RRDNLNR	QSTTLKR	RRDGLAG	VHHNLVR	ISHNLAR
ZFP901	38	GATGATTAGGCAGAGGTG	1828	1845	-	RRAVLDR	RQDNLGR	QSTTLKR	RRDGLAG	VHHNLVR	ISHNLAR
ZFP902	39	GGATTCAGCGCCGACGGG	1433	1450	-	RQEHLVR	EGGNLMR	SDRRDLD	SFQSYLE	RPNHLAI	QSPHLKR
ZFP903	40	GGATTCAGCGCCGACGGG	1433	1450	-	RREHLVR	DPSNLQR	SDRRDLD	SFQSYLE	RPNHLAI	QSPHLKR
ZFP904	41	GGATTCAGCGCCGACGGG	1433	1450	-	RREHLVR	DMGNLGR	SDRRDLD	SFQSYLE	RPNHLAI	QSPHLKR
ZFP907	42	GGCAGTAGTCGGAACAGGG	90	108	-	KKDHLHR	QKEILTR	QSAHLKR	ETGSLRR	QSHSLKS	ESGHLKR
ZFP908	43	GGCAGTAGTCGGAACAGGG	90	108	-	KKDHLHR	QKEILTR	QSAHLKR	DRTPLNR	QSHSLKS	ESGHLKR
ZFP909	44	GGCAGTAGTCGGAACAGGG	90	108	-	KTDHLAR	QKEILTR	QSAHLKR	ETGSLRR	QKHHLVT	ENSKLRR
ZFP912	45	GTAAACTGAGCCAGGAGAA	664	682	-	QAGNLVR	QNSHLRR	DLSTLRR	QNEHLKV	GGTALRM	QRSSLVR
ZFP913	46	GTAAACTGAGCCAGGAGAA	664	682	-	QRGNLQR	QTTHLSR	DGSTLRR	QKTHLAV	GGTALRM	QRSSLVR
ZFP914	47	GTAAACTGAGCCAGGAGAA	664	682	-	QRGNLQR	QTTHLSR	DLSTLRR	QNEHLKV	GGSALSM	QRSSLVR

ZFP930	48	ACGGTGGTCTCCATGCGAC	1605	1623	-	DRGNLTR	QARSLRA	EKASLIK	DHSSLKR	RRFILSR	RNDSLKC
ZFP931	49	ACGGTGGTCTCCATGCGAC	1605	1623	-	DRGNLTR	QARSLRA	DKSSLRK	DHSSLKR	RNFILQR	RNDTLII
ZFP932	50	ACGGTGGTCTCCATGCGAC	1605	1623	-	DRGNLTR	QARSLRA	CNGSLKK	DHSSLKR	RNFILQR	RNDTLII
ZFP933	51	GCTGGATGTGTCTGCGGCG	372	393	+	RTDTLAR	RTDSLPR	DHSSLKR	QPHGLAH	QSAHLKR	VGNSLSR
ZFP934	52	GCTGGATGTGTCTGCGGCG	372	393	+	RTDTLAR	RTDSLPR	DHSSLKR	QPHGLRH	QSAHLKR	VGNSLSR
ZFP935	53	GCTGGATGTGTCTGCGGCG	372	393	+	RTDTLAR	RLDMLAR	DHSSLKR	QPHGLST	QQAHLVR	VHESLKR
ZFP938	54	GTCTGCGAGGCGAGGGAG	2381	2398	-	RADNLGR	RNTHLSY	RGDGLRR	RRDNLNR	RARNLTL	DPSSLKR
ZFP939	55	GTCTGCGAGGCGAGGGAG	2381	2398	-	RADNLGR	RNTHLSY	RKLGLLR	RQDNLGR	RARNLTL	DPSSLKR
ZFP940	56	GTCTGCGAGGCGAGGGAG	2381	2398	-	RADNLGR	RNTHLSY	RKLGLLR	RQDNLGR	RRRNLQL	DHSSLKR
ZFP943	57	GTTGCCGGGCAACGGGGTA	1146	1164	-	QQSSLLR	RREHLVR	GLTALRT	ERAKLIR	AKRDLDR	VNSSLTR
ZFP944	58	GTTGCCGGGCAACGGGGTA	1146	1164	-	QQSSLLR	RREHLVR	GLTALRT	ERAKLIR	LRKDLVR	VRHSLTR
ZFP945	59	GTTGCCGGGCAACGGGGTA	1146	1164	-	QASALSR	RREHLVR	GLTALRT	ERAKLIR	AKRDLDR	VNSSLTR
ZFP951	60	CGAGAAAGTGAAAGCCTGC	1085	1103	-	RGRNLEM	DSSVLRR	QNANLKR	QKHHLAV	QRSNLAR	QKVHLEA
ZFP952	61	CGAGAAAGTGAAAGCCTGC	1085	1103	-	RRRNLDV	DSSVLRR	QNANLKR	QKHHLAV	QRSNLAR	QKVHLEA
ZFP953	62	CGAGAAAGTGAAAGCCTGC	1085	1103	-	RGRNLAI	DSSVLRR	LKSNLHR	LKQHLVV	LKTNLAR	QKCHLKA
ZFP956	63	GAGGCTTGAACAGTAGGAC	1856	1874	-	DGSNLRR	RIDNLDG	QRRYLVE	QQTNLAR	QRSDLTR	RGDNLNR
ZFP957	64	GAGGCTTGAACAGTAGGAC	1856	1874	-	DPSNLQR	RRDNLPK	TTFNLRV	QTQNLTR	HKETLNR	REDNLGR
ZFP958	65	GAGGCTTGAACAGTAGGAC	1856	1874	-	DPSNLQR	RRDNLPK	QRRYLVE	QQTNLAR	QRSDLTR	RGDNLNR
ZFP961	66	GAGGTTGGGGACTGCGAA	312	329	-	QQTNLTR	ANRTLVH	EEANLRR	RGEHLTR	TNSSLTR	RIDNLIR
ZFP962	67	GAGGTTGGGGACTGCGAA	312	329	-	QQTNLTR	ANRTLVH	EEANLRR	RREHLVR	MTSSLRR	RQDNLGR
ZFP963	68	GAGGTTGGGGACTGCGAA	312	329	-	QQTNLTR	ANRTLVH	EEANLRR	RGEHLTR	MTSSLRR	RQDNLGR
ZFP964	69	GATGATGTGGTATTGGGG	742	762	+	RATHLTR	RADVLKG	QRSSLVR	RKDALHV	VHHNLVR	ISHNLAR

ZFP965	70	GATGATGTGGTATTGGGG	742	762	+	RATHLTR	RADVLKG	QSSSLVR	RKERLAT	VRHNLTR	ISHNLAR
ZFP966	71	GATGATGTGGTATTGGGG	742	762	+	KKDHLHR	RKESLTV	QSSSLVR	RKERLAT	VHHNLVR	ISHNLAR
ZFP969	72	GATGATGTGGTATTGGGGG	742	763	+	RVDHLHR	RREHLSG	QSSSLVR	RKERLAT	VAHNLTR	ISHNLAR
ZFP970	73	GATGATGTGGTATTGGGGG	742	763	+	RKHHLGR	RREHLTI	QSSSLVR	RKERLAT	VAHNLTR	ISHNLAR
ZFP971	74	GATGATGTGGTATTGGGGG	742	763	+	RVDHLHR	RSDHLSL	QSSSLVR	RKERLAT	VAHNLTR	ISHNLAR
ZFP984	75	GCAGTAGTCGGAACAGGG	90	107	-	KTDHLAR	QKEILTR	QSAHLKR	ETGSLRR	QSSSLVR	QTNTLGR
ZFP985	76	GCAGTAGTCGGAACAGGG	90	107	-	KKDHLHR	QKEILTR	QSAHLKR	ETGSLRR	QSSSLVR	QGGTLRR
ZFP986	77	GCAGTAGTCGGAACAGGG	90	107	-	KKDHLHR	QKEILTR	QSAHLKR	DPTSLNR	QSSSLVR	QTNTLGR
ZFP989	78	GCATAGCAGCAGGATGAA	409	426	-	QQTNLTR	VGGNLAR	KRYNLYQ	RQDNLNT	RSHNLKL	QSTTLKR
ZFP990	79	GCATAGCAGCAGGATGAA	409	426	-	QQTNLTR	VGGNLSR	KRYNLYQ	RQDNLNT	RSHNLRL	QSTTLKR
ZFP991	80	GCATAGCAGCAGGATGAA	409	426	-	QQTNLTR	VGGNLSR	KKFNLLQ	RRDNLKS	RSHNLKL	QSTTLKR
ZFP994	81	GGCGTTCACGGTGGTCTCC	1612	1630	-	DKSSLRK	DHSSLKR	RNFILQR	RNDTLII	TSTLLKR	LKEHLTR
ZFP995	82	GGCGTTCACGGTGGTCTCC	1612	1630	-	CNGSLKK	DHSSLKR	RNFILAR	RQDILVV	HKSSLTR	ESGHLKR
ZFP996	83	GGCGTTCACGGTGGTCTCC	1612	1630	-	CNGSLKK	DHSSLKR	RNFILAR	RQDILVV	TSTLLKR	LKEHLTR
ZFP999	84	GTTGGTGAGTGATTGGAG	327	344	-	TNNNLAR	RTDSLTL	QREHLTT	RRDNLNR	RRQKLTI	HKSSLTR
ZFP1000	85	GTTGGTGAGTGATTGGAG	327	344	-	TNNNLAR	RTDSLTL	QREHLTT	RGDNLKR	RRQKLTI	HKSSLTR
ZFP1001	86	GTTGGTGAGTGATTGGAG	327	344	-	TNNNLAR	RTDSLTL	QREHLNG	RGDNLAR	RRQKLTI	HKSSLTR
ZFP1005	87	GGAGGTTGGGGACTGCGAA	312	330	-	QQTNLTR	ANRTLVH	DPANLRR	RQEHLVR	MKHHLGR	QNSHLRR
ZFP1006	88	GGAGGTTGGGGACTGCGAA	312	330	-	QQTNLTR	ANRTLVH	EEANLRR	RREHLVR	MKHHLGR	QNSHLRR
ZFP1007	89	GGAGGTTGGGGACTGCGAA	312	330	-	QQTNLTR	ANRTLVH	DPANLRR	RQEHLVR	LKQHLVR	QGGHLAR
ZFP1008	90	GGATGATGTGGTATTGGGG	741	762	+	RNTHLAR	RADVLKG	QRSSLVR	RKDALHV	QNEHLKV	QNSHLRR
ZFP1009	91	GGATGATGTGGTATTGGGG	741	762	+	RNTHLAR	RADVLKG	QSSSLVR	RKERLAT	QKTHLAV	QGGHLKR

ZFP1010	92	GGATGATGTGGTATTGGGG	741	762	+	RNTHLAR	RADVLKG	QSSSLVR	RKERLAT	QKTHLAV	QNSHLRR
ZFP1013	93	GGATGTGTCTGCGGCGTT	375	395	+	HKSSLTR	ESGHLKR	RRRNLTL	DRSSLKR	QPHSLAV	QKPHLSR
ZFP1014	94	GGATGTGTCTGCGGCGTT	375	395	+	HKSSLTR	EGGHLKR	RRRNLQL	DHSSLKR	RRQHLQY	QSAHLKR
ZFP1015	95	GGATGTGTCTGCGGCGTT	375	395	+	HKSSLTR	EGGHLKR	RRRNLTL	DRSSLKR	RRQHLQY	QSAHLKR
ZFP1018	96	GGGGGTTGCGTCAGCAAAC	1184	1202	-	GHTALRN	QSGTLHR	DHSSLKR	AMRSLMG	RRSRLVR	RGEHLTR
ZFP1019	97	GGGGGTTGCGTCAGCAAAC	1184	1202	-	GHTALRN	QSTTLKR	DHSSLKR	QQRSLVG	EAHHLSR	RTEHLAR
ZFP1020	98	GGGGGTTGCGTCAGCAAAC	1184	1202	-	GHTALRN	QSTTLKR	DHSSLKR	AMRSLMG	RQSRLQR	RREHLVR
ZFP1023	99	GTTGTTAGACGACGAGGCA	2342	2363	+	QGETLKR	RADNLRR	DKANLTR	DQGNLIR	HRHVLIN	TNSSLTR
ZFP1024	100	GTTGTTAGACGACGAGGCA	2342	2363	+	QGETLKR	RADNLRR	DSSNLRR	DQGNLIR	HKSSLTR	IRTSLKR
ZFP1025	101	GTTGTTAGACGACGAGGCA	2342	2363	+	QGETLKR	RADNLRR	EQGNLLR	DGGNLGR	HRHVLIN	TNSSLTR

[0061] In some embodiments, the ZFP domain of the present epigenetic editor binds to a target sequence provided herein. In further embodiments, the ZFP domain comprises, in order, the F1-F6 amino acid sequences of any one of the zinc finger proteins as shown in **Table 1** and **Table 20**. The F1-F6 amino acid sequences may be placed within the ZF framework sequence of SEQ ID NO: 1084, or within any other ZF framework known in the art.

C. TALEs

[0062] In some embodiments, the DNA-binding domain of an epigenetic editor described herein comprises a transcription activator-like effector (TALE) domain. The DNA-binding domain of a TALE comprises a highly conserved sequence of about 33-34 amino acids, with a repeat variable di-residue (RVD) at positions 12 and 13 that is central to the recognition of specific nucleotides. TALEs can be engineered to bind practically any desired DNA sequence. Methods for programming TALEs are known in the art. For example, such methods are described in Carroll et al., *Genet Soc Amer.* (2011) 188(4):773-82; Miller et al., *Nat Biotechnol.* (2007) 25(7):778-85; Christian et al., *Genetics* (2008) 186(2):757-61; Li et al., *Nucl Acids Res.* (2010) 39(1):359-72; and Moscou et al., *Science* (2009) 326(5959):1501.

D. Other DNA-Binding Domains

[0063] Other DNA-binding domains are contemplated for the epigenetic editors described herein. In some embodiments, the DNA-binding domain comprises an argonaute protein domain, e.g., from *Natronobacterium gregoryi* (NgAgo). NgAgo is a ssDNA-guided endonuclease that is guided to its target site by 5' phosphorylated ssDNA (gDNA), where it produces double-strand breaks. In contrast to Cas9, the NgAgo-gDNA system does not require a protospacer-adjacent motif (PAM). Thus, using a nuclease inactive NgAgo (dNgAgo) can greatly expand the bases that may be targeted. The characterization and use of NgAgo have been described, e.g., in Gao et al., *Nat Biotechnol.* (2016) 34(7):768-73; Swarts et al., *Nature* (2014) 507(7491):258-61; and Swarts et al., *Nucl Acids Res.* (2015) 43(10):5120-9.

[0064] In some embodiments, the DNA-binding domain comprises an inactivated nuclease, for example, an inactivated meganuclease. Additional non-limiting examples of DNA-binding domains include tetracycline-controlled repressor (tetR) DNA-binding domains, leucine zippers, helix-loop-helix (HLH) domains, helix-turn-helix domains, β-sheet motifs, steroid receptor motifs, bZIP domains homeodomains, and AT-hooks.

II. Guide Polynucleotides

[0065] Epigenetic editors described herein that comprise a polynucleotide guided DNA-binding domain may also include a guide polynucleotide that is capable of forming a complex with the DNA-binding domain. The guide polynucleotide may comprise RNA, DNA, or a mixture of both. For example, where the polynucleotide guided DNA-binding domain is a CRISPR-associated protein domain, the guide polynucleotide may be a guide RNA (gRNA). A “guide RNA” or “gRNA” refers to a nucleic acid that is able to hybridize to a target sequence and direct binding of the CRISPR-Cas complex to the target sequence. Methods of using guide polynucleotide sequences with programmable DNA-binding proteins (e.g., CRISPR-associated protein domains) for site-specific DNA targeting (e.g., to modify a genome) are known in the art.

[0066] A guide polynucleotide sequence (e.g., a gRNA sequence) may comprises two parts: 1) a nucleotide sequence comprising a “targeting sequence” that is complementary to a target nucleic acid sequence (“target sequence”), e.g., to a nucleic acid sequence comprised in a genomic target site; and 2) a nucleotide sequence that binds a polynucleotide guided DNA-binding domain (e.g., a CRISPR-Cas protein domain). The nucleotide sequence in 1) may comprise a targeting sequence that is 100% complementary to a genomic nucleic acid sequence, e.g., a nucleic acid sequence comprised in a genomic target site, and thus may hybridize to the target nucleic acid sequence. The nucleotide sequence in 1) may be referred to as, e.g., a crispr RNA, or crRNA. The nucleotide sequence in 2) may be referred to as a scaffold sequence of a guide nucleic acid, e.g., a tracrRNA, or an activating region of a guide nucleic acid, and may comprise a stem-loop structure. Parts 1) and 2) as described above may be fused to form one single guide (e.g., a single guide RNA, or sgRNA), or may be on two separate nucleic acid molecules. In some embodiments, a guide polynucleotide comprises parts 1) and 2) connected by a linker. In some embodiments, a guide polynucleotide comprises parts 1) and 2) connected by a non-nucleic acid linker, for example, a peptide linker or a chemical linker.

[0067] Part 2 (the scaffold sequence) of a guide polynucleotide as described herein may be, for example, as described in Jinek et al., *Science* (2012) 337:816-21; U.S. Patent Publication 2016/0208288; or U.S. Patent Publication 2016/0200779. Variants of part 2) are also contemplated by the present disclosure. For example, the tetraloop and stem loop of a gRNA scaffold (tracrRNA) sequence may be modified to include RNA aptamers, which can be bound by specific protein domains. In some embodiments, such modified gRNAs can be used to facilitate the recruitment of repressive or activating domains fused to the protein-interacting RNA aptamers.

[0068] A gRNA as provided herein typically comprises a targeting domain and a binding domain. The targeting domain (also termed “targeting sequence”) may comprise a nucleic acid sequence that binds to a target site, e.g., to a genomic nucleic acid molecule within a cell. The target site may be a double-stranded DNA sequence comprising a PAM sequence as well as the target sequence, which is located on the same strand as, and directly adjacent to, the PAM sequence. The targeting domain of the gRNA may comprise an RNA sequence that corresponds to the target sequence, i.e., it resembles the sequence of the target domain, sometimes with one or more mismatches, but typically comprising an RNA sequence instead of a DNA sequence. The targeting domain of the gRNA thus may base pair (in full or partial complementarity) with the sequence of the double-stranded target site that is complementary to the target sequence, and thus with the strand complementary to the strand that comprises the PAM sequence. It will be understood that the targeting domain of the gRNA typically does not include a sequence that resembles the PAM sequence. It will further be understood that the location of the PAM may be 5’ or 3’ of the target sequence, depending on the nuclease employed. For example, the PAM is typically 3’ of the target sequence for Cas9 nucleases, and 5’ of the target sequence for Casl2a nucleases. For an illustration of the location of the PAM and the mechanism of gRNA binding to a target site, see, e.g., Figure 1 of Vanegas et al., *Fungal Biol Biotechnol.* (2019) 6:6, which is incorporated by reference herein. For additional illustration and description of the mechanism of gRNA targeting of an RNA-guided nuclease to a target site, see Fu et al., *Nat Biotechnol* (2014) 32(3):279-84 and Sternberg et al., *Nature* (2014) 507(7490):62-7, each incorporated herein by reference.

[0069] In some embodiments, the targeting domain sequence comprises between 17 and 30 nucleotides and corresponds fully to the target sequence (i.e., without any mismatch nucleotides). In some embodiments, however, the targeting domain sequence may comprise one or more, but typically not more than 4, mismatches, e.g., 1, 2, 3, or 4 mismatches. As the targeting domain is part of gRNA, which is an RNA molecule, it will typically comprise ribonucleotides, while the DNA targeting domain will comprise deoxyribonucleotides.

[0070] An exemplary illustration of a Cas9 target site, comprising a 22 nucleotide target domain, and an NGG PAM sequence, as well as of a gRNA comprising a targeting domain that fully corresponds to the target sequence (and thus base pairs with full complementarity with the DNA strand complementary to the strand comprising the target sequence and PAM) is provided below:

```
[                 target domain (DNA)         ][ PAM  ]
5'-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-G-G-3' (DNA)
3'-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-C-C-5' (DNA)
   | | | | | | | | | | | | | | | | | | | | | |
5'-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-[ gRNA scaffold]-3' (RNA)
[            targeting domain ( RNA)          ][  binding domain  ]
```

[0071] An exemplary illustration of a Casl2a target site, comprising a 22 nucleotide target domain, and a TTN PAM sequence, as well as of a gRNA comprising a targeting domain that fully corresponds to the target sequence (and thus base pairs with full complementarity with the DNA strand complementary to the strand comprising the target sequence and PAM) is provided below:

```
          [  PAM  ][            target domain ( DNA)             ]
          5'-T-T-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-3' (DNA)
          3'-A-A-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-5' (DNA)
                   | | | | | | | | | | | | | | | | | | | | | |
5'-[gRNA scaffold]-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-N-3' (RNA)
[ binding domain  ][             targeting domain  ( RNA)         ]
```

[0072] While not wishing to be bound by theory, at least in some embodiments, it is believed that the length and complementarity of the targeting domain with the target sequence contributes to specificity of the interaction of the gRNA/Cas9 molecule complex with a target nucleic acid. In some embodiments, the targeting domain of a gRNA provided herein is 5 to 50 nucleotides in length. In some embodiments, the targeting domain is 15 to 25 nucleotides in length. In some embodiments, the targeting domain is 18 to 22 nucleotides in length. In some embodiments, the targeting domain is 19-21 nucleotides in length. In some embodiments, the targeting domain is 15 nucleotides in length. In some embodiments, the targeting domain is 16 nucleotides in length. In some embodiments, the targeting domain is 17 nucleotides in length. In some embodiments, the targeting domain is 18 nucleotides in length. In some embodiments, the targeting domain is 19 nucleotides in length. In some embodiments, the targeting domain is 20 nucleotides in length. In some embodiments, the targeting domain is 21 nucleotides in length. In some embodiments, the targeting domain is 22 nucleotides in length. In some embodiments, the targeting domain is 23 nucleotides in length. In some embodiments, the targeting domain is 24 nucleotides in length. In some embodiments, the targeting domain is 25 nucleotides in length. In certain embodiments, the targeting domain fully corresponds, without mismatch, to a target sequence provided herein, or a part thereof. In some embodiments, the targeting domain of a gRNA provided herein comprises 1 mismatch relative to a target sequence provided herein. In some embodiments, the targetindg domain comprises 2 mismatches relative to the target sequence. In some embodiments, the target domain comprises 3 mismatches relative to the target sequence.

[0073] Methods for designing, selecting, and validating gRNAs are described herein and known in the art. Software tools can be used to optimize the gRNAs corresponding to a target DNA sequence, e.g., to minimize total off-target activity across the genome. For example, DNA sequence searching algorithms can be used to identify a target sequence in crRNAs of a gRNA for use with Cas9. Exemplary gRNA design tools include the ones described in Bae et al., *Bioinformatics* (2014) 30:1473-5.

[0074] Guide polynucleotides (e.g., gRNAs) described herein may be of various lengths. In some embodiments, the length of the spacer or targeting sequence depends on the CRISPR-associated protein component of the epigenetic editor system used. For example, Cas proteins from different bacterial species have varying optimal targeting sequence lengths. Accordingly, the spacer sequence may comprise, e.g., 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more than 50 nucleotides in length. In some embodiments, the spacer comprises 10-24, 11-20, 11-16, 18-24, 19-21, or 20 nucleotides in length. In some embodiments, a guide polynucleotide (e.g., gRNA) is from 15-100 (e.g., 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50) nucleotides in length and comprises a spacer sequence of at least 10 (e.g., 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50) contiguous nucleotides complementary to the target sequence. In some embodiments, a guide polynucleotide described herein may be truncated, e.g., by 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50 or more nucleotides.

[0075] In certain embodiments, the 3’ end of the *HBV* target sequence is immediately adjacent to a PAM sequence (e.g., a canonical PAM sequence such as NGG for SpCas9). The degree of complementarity between the targeting sequence of the guide polynucleotide (e.g., the spacer sequence of a gRNA) and the target sequence may be at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%. In particular embodiments, the targeting and the target sequence may be 100% complementary. In other embodiments, the targeting sequence and the target sequence may contain, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 mismatches.

[0076] A guide polynucleotide (e.g., gRNA) may be modified with, for example, chemical alterations and synthetic modifications. A modified gRNA, for instance, can include an alteration or replacement of one or both of the non-linking phosphate oxygens and/or of one or more of the linking phosphate oxygens in the phosphodiester backbone linkage, an alteration of the ribose sugar (e.g., of the 2’ hydroxyl on the ribose sugar), an alteration of the phosphate moiety, modification or replacement of a naturally occurring nucleobase, modification or replacement of the ribose-phosphate backbone, modification of the 3’ end and/or 5’ end of the oligonucleotide, replacement of a terminal phosphate group or conjugation of a moiety, cap, or linker, or any combination thereof.

[0077] In some embodiments, one or more ribose groups of the gRNA may be modified. Examples of chemical modifications to the ribose group include, but are not limited to, 2’-O-methyl (2’-OMe), 2’-fluoro (2’-F), 2’-deoxy, 2’-O-(2-methoxyethyl) (2’-MOE), 2’-NH2, 2’-O-allyl, 2’-O-ethylamine, 2’-O-cyanoethyl, 2’-O-acetalester, or a bicyclic nucleotide such as locked nucleic acid (LNA), 2’-(5-constrained ethyl (S-cEt)), constrained MOE, or 2’-0,4’-C-aminomethylene bridged nucleic acid (2’,4’-BNANC). 2’-O-methyl modification and/or 2’-fluoro modification may increase binding affinity and/or nuclease stability of the gRNA oligonucleotides.

[0078] In some embodiments, one or more phosphate groups of the gRNA may be chemically modified. Examples of chemical modifications to a phosphate group include, but are not limited to, a phosphorothioate (PS), phosphonoacetate (PACE), thiophosphonoacetate (thioPACE), amide, triazole, phosphonate, and phosphotriester modification. In some embodiments, a guide polynucleotide described herein may comprise one, two, three, or more PS linkages at or near the 5’ end and/or the 3’ end; the PS linkages may be contiguous or noncontiguous.

[0079] In some embodiments, the gRNA herein comprises a mixture of ribonucleotides and deoxyribonucleotides and/or one or more PS linkages.

[0080] In some embodiments, one or more nucleobases of the gRNA may be chemically modified. Examples of chemically modified nucleobases include, but are not limited to, 2-thiouridine, 4-thiouridine, N6-methyladenosine, pseudouridine, 2,6-diaminopurine, inosine, thymidine, 5-methylcytosine, 5-substituted pyrimidine, isoguanine, isocytosine, and nucleobases with halogenated aromatic groups. Chemical modifications can be made in the spacer region, the tracr RNA region, the stem loop, or any combination thereof.

[0081] **Table 2** below lists exemplary target sequences for epigenetic modification of HBV, as well as the coordinates of the start and end positions of the targeted site on the HBV genome.

Table 2. Targeting Domain Sequences of Exemplary gRNAs Targeting HBV. The following target sites were identified as suitable for targeting with an epigenetic repressor:

SEQ IDs	Target domain sequence	Start	End	Strand
333	CCTGCTGGTGGCTCCAGTTC	57	77	+
334	CTGAACTGGAGCCACCAGCA	59	79	-
335	CCTGAACTGGAGCCACCAGC	60	80	-
336	CCTCGAGAAGATTGACGATA	115	135	-
337	TCGTCAATCTTCTCGAGGAT	117	137	+
338	CGTCAATCTTCTCGAGGATT	118	138	+
339	GTCAATCTTCTCGAGGATTG	119	139	+
340	AACATGGAGAACATCACATC	153	173	+
341	AACATCACATCAGGATTCCT	162	182	+
342	CTAGACTCTGCGGTATTGTG	233	253	-
343	TACCGCAGAGTCTAGACTCG	238	258	+
344	CGCAGAGTCTAGACTCGTGG	241	261	+
345	CACCACGAGTCTAGACTCTG	243	263	-
346	TGGACTTCTCTCAATTTTCT	261	281	+
347	GGACTTCTCTCAATTTTCTA	262	282	+
348	GACTTCTCTCAATTTTCTAG	263	283	+
349	ACTTCTCTCAATTTTCTAGG	264	284	+
350	CGAATTTTGGCCAAGACACA	295	315	-
351	AGGTTGGGGACTGCGAATTT	309	328	-
352	GGCATAGCAGCAGGATGAAG	408	427	-
353	AGAAGATGAGGCATAGCAGC	417	436	-
354	GCTATGCCTCATCTTCTTGT	420	439	+
355	GAAGAACCAACAAGAAGATG	429	448	-
356	CATCTTCTTGTTGGTTCTTC	429	448	+
357	CCCGTTTGTCCTCTAATTCC	469	488	+
358	CCTGGAATTAGAGGACAAAC	472	491	-
359	TCCTGGAATTAGAGGACAAA	473	492	-

360	TACTAGTGCCATTTGTTCAG	680	699	+
361	CCATTTGTTCAGTGGTTCGT	688	707	+
362	CATTTGTTCAGTGGTTCGTA	689	708	+
363	CCTACGAACCACTGAACAAA	691	710	-
364	TTTCAGTTATATGGATGATG	731	750	+
365	CAAAAGAAAATTGGTAACAG	799	818	-
366	TACCAATTTTCTTTTGTCTT	803	822	+
367	ACCAATTTTCTTTTGTCTTT	804	823	+
368	ACCCAAAGACAAAAGAAAAT	808	827	-
369	TGACATACTTTCCAATCAAT	975	994	-
370	CACTTTCTCGCCAACTTACA	1093	1113	+
371	CACAGAAAGGCCTTGTAAGT	1106	1126	-
372	TGAACCTTTACCCCGTTGCC	1137	1157	+
373	GGGCAACGGGGTAAAGGTTC	1138	1158	-
374	TTTACCCCGTTGCCCGGCAA	1143	1163	+
375	GTTGCCGGGCAACGGGGTAA	1144	1164	-
376	CCCGTTGCCCGGCAACGGCC	1148	1168	+
377	CTGGCCGTTGCCGGGCAACG	1150	1170	-
378	CCTGGCCGTTGCCGGGCAAC	1151	1171	-
379	ACCTGGCCGTTGCCGGGCAA	1152	1172	-
380	GCACAGACCTGGCCGTTGCC	1158	1178	-
381	GGCACAGACCTGGCCGTTGC	1159	1179	-
382	GCAAACACTTGGCACAGACC	1169	1189	-
383	GGGTTGCGTCAGCAAACACT	1180	1200	-
384	TTTGCTGACGCAACCCCCAC	1184	1204	+
385	CTGACGCAACCCCCACTGGC	1188	1208	+
386	TGACGCAACCCCCACTGGCT	1189	1209	+
387	GACGCAACCCCCACTGGCTG	1190	1210	+
388	AACCCCCACTGGCTGGGGCT	1195	1215	+
389	TCCTCTGCCGATCCATACTG	1255	1275	+
390	TCCGCAGTATGGATCGGCAG	1259	1279	-
391	AGGAGTTCCGCAGTATGGAT	1265	1285	-
392	CGGCTAGGAGTTCCGCAGTA	1270	1290	-
393	TGCGAGCAAAACAAGCGGCT	1285	1305	-
394	CCGCTTGTTTTGCTCGCAGC	1287	1307	+
395	CCTGCTGCGAGCAAAACAAG	1290	1310	-
396	TGTTTTGCTCGCAGCAGGTC	1292	1312	+
397	GCAGCACAGCCTAGCAGCCA	1376	1396	-
398	TGCTAGGCTGTGCTGCCAAC	1380	1400	+
399	GCTGCCAACTGGATCCTGCG	1391	1411	+
400	CTGCCAACTGGATCCTGCGC	1392	1412	+
401	CGTCCCGCGCAGGATCCAGT	1398	1418	-

402	AAACAAAGGACGTCCCGCGC	1408	1428	-
403	GTCCTTTGTTTACGTCCCGT	1417	1437	+
404	CGCCGACGGGACGTAAACAA	1422	1442	-
405	TGCCGTTCCGACCGACCACG	1504	1523	+
406	AGGTGCGCCCCGTGGTCGGT	1513	1533	-
407	AGAGAGGTGCGCCCCGTGGT	1517	1537	-
408	GTAAAGAGAGGTGCGCCCCG	1521	1541	-
409	GGGGCGCACCTCTCTTTACG	1522	1542	+
410	CGGGGAGTCCGCGTAAAGAG	1533	1553	-
411	CAGATGAGAAGGCACAGACG	1551	1571	-
412	GTCTGTGCCTTCTCATCTGC	1552	1572	+
413	GGCAGATGAGAAGGCACAGA	1553	1573	-
414	GCAGATGAGAAGGCACAGAC	1553	1572	-
415	ACACGGTCCGGCAGATGAGA	1562	1582	-
416	GAAGCGAAGTGCACACGGTC	1574	1594	-
417	GAGGTGAAGCGAAGTGCACA	1579	1599	-
418	CTTCACCTCTGCACGTCGCA	1590	1610	+
419	GGTCTCCATGCGACGTGCAG	1598	1618	-
420	TGCCCAAGGTCTTACATAAG	1640	1660	+
421	GTCCTCTTATGTAAGACCTT	1645	1665	-
422	AGTCCTCTTATGTAAGACCT	1646	1666	-
423	GTCTTACATAAGAGGACTCT	1648	1668	+
424	AATGTCAACGACCGACCTTG	1680	1700	+
425	TTTGAAGTATGCCTCAAGGT	1694	1714	-
426	AGTCTTTGAAGTATGCCTCA	1698	1718	-
427	AAGACTGTTTGTTTAAAGAC	1712	1732	+
428	AGACTGTTTGTTTAAAGACT	1713	1733	+
429	CTGTTTGTTTAAAGACTGGG	1716	1736	+
430	GTTTAAAGACTGGGAGGAGT	1722	1742	+
431	TCTTTGTACTAGGAGGCTGT	1766	1786	+
432	AGGAGGCTGTAGGCATAAAT	1776	1796	+
433	GTGAAAAAGTTGCATGGTGC	1810	1830	-
434	GCAGAGGTGAAAAAGTTGCA	1816	1836	-
435	AACAAGAGATGATTAGGCAG	1832	1852	-
436	GACATGAACAAGAGATGATT	1838	1858	-
437	AGCTTGGAGGCTTGAACAGT	1860	1880	-
438	CAAGCCTCCAAGCTGTGCCT	1866	1886	+
439	AAGCCTCCAAGCTGTGCCTT	1867	1887	+
440	CCTCCAAGCTGTGCCTTGGG	1871	1890	+
441	CCACCCAAGGCACAGCTTGG	1873	1893	-
442	AGCTGTGCCTTGGGTGGCTT	1876	1896	+
443	AAGCCACCCAAGGCACAGCT	1876	1896	-

444	GCTGTGCCTTGGGTGGCTTT	1877	1897	+
445	CTGTGCCTTGGGTGGCTTTG	1878	1898	+
446	TAGCTCCAAATTCTTTATAA	1916	1936	-
447	GTAGCTCCAAATTCTTTATA	1917	1937	-
448	TAAAGAATTTGGAGCTACTG	1919	1939	+
449	ATGACTCTAGCTACCTGGGT	2097	2117	+
450	CACATTTCTTGTCTCACTTT	2211	2231	+
451	TAGTTTCCGGAAGTGTTGAT	2321	2341	-
452	CGTCTAACAACAGTAGTTTC	2334	2354	-
453	ACTACTGTTGTTAGACGACG	2337	2357	+
454	CTGTTGTTAGACGACGAGGC	2341	2361	+
455	CGAGGGAGTTCTTCTTCTAG	2368	2388	-
456	GCGAGGGAGTTCTTCTTCTA	2369	2389	-
457	GGCGAGGGAGTTCTTCTTCT	2370	2390	-
458	CTCCCTCGCCTCGCAGACGA	2380	2400	+
459	GACCTTCGTCTGCGAGGCGA	2385	2405	-
460	AGACCTTCGTCTGCGAGGCG	2386	2406	-
461	GATTGAGACCTTCGTCTGCG	2391	2411	-
462	GATTGAGATCTTCTGCGACG	2415	2435	-
463	GTCGCAGAAGATCTCAATCT	2416	2436	+
464	TCGCAGAAGATCTCAATCTC	2417	2437	+
465	ATATGGTGACCCACAAAATG	2807	2827	-
466	TTTGTGGGTCACCATATTCT	2810	2830	+
467	TTGTGGGTCACCATATTCTT	2811	2831	+
468	GCTGGATCCAACTGGTGGTC	2894	2914	-
469	CACCCCAAAAGGCCTCCGTG	3026	3046	-
470	CCTTTTGGGGTGGAGCCCTC	3034	3054	+
471	CCTGAGGGCTCCACCCCAAA	3037	3057	-
472	GGGGTGGAGCCCTCAGGCTC	3040	3060	+
473	GGGTGGAGCCCTCAGGCTCA	3041	3061	+
474	CGATTGGTGGAGGCAGGAGG	3092	3112	-
475	CTCATCCTCAGGCCATGCAG	3159	3179	+
102	GATGAGGCATAGCAGCAG	415	432	-
103	GATGATTAGGCAGAGGTG	1828	1845	-
104	GGATTCAGCGCCGACGGG	1433	1450	-
105	GGCAGTAGTCGGAACAGGG	90	108	-
106	GTAAACTGAGCCAGGAGAA	664	682	-
107	ACGGTGGTCTCCATGCGAC	1605	1623	-
108	GCTGGATGTGTCTGCGGCG	372	393	+
109	GTCTGCGAGGCGAGGGAG	2381	2398	-
110	GTTGCCGGGCAACGGGGTA	1146	1164	-
111	CGAGAAAGTGAAAGCCTGC	1085	1103	-

112	GAGGCTTGAACAGTAGGAC	1856	1874	-
113	GAGGTTGGGGACTGCGAA	312	329	-
114	GATGATGTGGTATTGGGG	742	762	+
115	GATGATGTGGTATTGGGGG	742	763	+
116	GCAGTAGTCGGAACAGGG	90	107	-
117	GCATAGCAGCAGGATGAA	409	426	-
118	GGCGTTCACGGTGGTCTCC	1612	1630	-
119	GTTGGTGAGTGATTGGAG	327	344	-
120	GGAGGTTGGGGACTGCGAA	312	330	-
121	GGATGATGTGGTATTGGGG	741	762	+
122	GGATGTGTCTGCGGCGTT	375	395	+
123	GGGGGTTGCGTCAGCAAAC	1184	1202	-
124	GTTGTTAGACGACGAGGCA	2342	2363	+

[0082] Target domains identified above that are adjacent to a PAM sequence, e.g., an *S. pyogenes* Cas9 PAM sequence, can be targeted by a CRISPR-based epigenetic repressor, e.g., an epigenetic repressor comprising a dCas9 DNA-binding domain. For example, target sites 1-143 are suitable for dCas9-based epigenetic repressor targeting.

[0083] A suitable gRNA for targeting any of the target domain sequences would, in some embodiments, comprise a target domain sequence that is the RNA-equivalent sequence of the provided DNA sequence of the targeting domain sequence (i.e., an RNA nucleotide of that sequence instead of the provided DNA nucleotide, with uracil instead of thymine), and a suitable tracr RNA sequence.

[0084] Any tracr sequence known in the art is contemplated for a gRNA described herein. In some embodiments, a gRNA described herein has a tracr sequence shown in **Table 3** below, or a tracr sequence at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to the tracr sequence shown below (SEQ: SEQ ID NO).

Table 3. Exemplary TRACR Sequences

SEQ	Sequence (5' to 3')
1087	GUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
1088	GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
1089	GUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
1090	GUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU

[0085] In some embodiments, the gRNA herein is provided to the cell directly (e.g., through an RNP complex together with the CRISPR-associated protein domain). In some embodiments, the gRNA is provided to the cell through an expression vector (e.g., a plasmid vector or a viral vector) introduced into the cell, where the cell then expresses the gRNA from the expression vector. Methods of introducing gRNAs and expression vectors into cells are well known in the art.

III. Effector Domains

[0086] Epigenetic editors described herein include one or more effector protein domains (also “epigenetic effector domains,” or “effector domains,” as used herein) that effect epigenetic modification of a target gene. An epigenetic editor with one or more effector domains may modulate expression of a target gene without altering its nucleobase sequence. In some embodiments, an effector domain described herein may provide repression or silencing of expression of HBV or an HBV gene, e.g., by repressing transcription or by modifying or remodeling HBV chromatin. Such effector domains are also referred to herein as “repression domains,” “repressor domains,” “epigenetic repressor domains,” or “epigenetic repression domains.” Non-limiting examples of chemical modifications that may be mediated by effector domains include methylation, demethylation, acetylation, deacetylation, phosphorylation, SUMOylation and/or ubiquitination of DNA or histone residues.

[0087] In some embodiments, an effector domain of an epigenetic editor described herein may make histone tail modifications, e.g., by adding or removing active marks on histone tails.

[0088] In some embodiments, an effector domain of an epigenetic editor described herein may comprise or recruit a transcription-related protein, e.g., a transcription repressor. The transcription-related protein may be endogenous or exogenous.

[0089] In some embodiments, an effector domain of an epigenetic editor described herein may, for example, comprise a protein that directly or indirectly blocks access of a transcription factor to the gene of interest harboring the target sequence.

[0090] An effector domain may be a full-length protein or a fragment thereof that retains the epigenetic effector function (a “functional domain”). Functional domains that are capable of modulating (e.g., repressing) gene expression can be derived from a larger protein. For example, functional domains that can reduce target gene expression may be identified based on sequences of repressor proteins. Amino acid sequences of gene expression-modulating proteins may be obtained from available genome browsers, such as the UCSD genome browser or Ensembl genome browser. Protein annotation databases such as UniProt or Pfam can be used to identify functional domains within the full protein sequence. As a starting point, the largest sequence, encompassing all regions identified by different databases, may be tested for gene expression modulation activity. Various truncations then may be tested to identify the minimal functional unit.

[0091] Variants of effector domains described herein are also contemplated by the present disclosure. A variant may, for example, refer to a polypeptide with at least about 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity and/or sequence similarity to a wildtype effector domain described herein. In particular embodiments, the variant retains at least about 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% of the epigenetic effector function of the wildtype effector domain.

[0092] In some embodiments, an epigenetic editor described herein may comprise 1 effector domain, 2 effector domains, 3 effector domains, 4 effector domains, 5 effector domains, 6 effector domains, 7 effector domains, 8 effector domains, 9 effector domains, 10 effector domains, or more. In certain embodiments, the epigenetic editor comprises one or more fusion proteins (e.g., one, two, or three fusion proteins), each with one or more effector domains (e.g., one, two, or three effector domains) linked to a DNA-binding domain. In some embodiments, the effector domains may induce a combination of epigenetic modifications, e.g., transcription repression and DNA methylation, DNA methylation and histone deacetylation, DNA methylation and histone demethylation, DNA methylation and histone methylation, DNA methylation and histone phosphorylation, DNA methylation and histone ubiquitylation, DNA methylation, and histone SUMOylation.

[0093] In certain embodiments, an effector domain described herein (e.g., DNMT3A and/or DNMT3L) is encoded by a nucleotide sequence as found in the native genome (e.g., human or murine) for that effector domain. In other embodiments, an effector domain described herein is encoded by a nucleotide sequence that has been codon-optimized for optimal expression in human cells.

[0094] Effector domains described herein may include, for example, transcriptional repressors, DNA methyltransferases, and/or histone modifiers, as further detailed below.

A. Transcriptional Repressors

[0095] In some embodiments, an epigenetic effector domain described herein mediates repression of a target gene's expression (e.g., transcription). The effector domain may comprise, e.g., a Krüppel-associated box (KRAB) repression domain, a Repressor Element Silencing Transcription Factor (REST) repression domain, a KRAB-associated protein 1 (KAP1) domain, a MAD domain, a FKHR (forkhead in rhabdosarcoma gene) repressor domain, an EGR-1 (early growth response gene product-1) repressor domain, an ets2 repressor factor repressor domain (ERD), a MAD smSIN3 interaction domain (SID), a WRPW motif of the hairy-related basic helix-loop-helix (bHLH) repressor proteins, an HP1 alpha chromo-shadow repression domain, an HP1 beta repression domain, or any combination thereof. The effector domain may recruit one or more protein domains that repress expression of the target gene, e.g., through a scaffold protein. In some embodiments, the effector domain may recruit or interact with a scaffold protein domain that recruits a PRMT protein, a HDAC protein, a SETDB1 protein, or a NuRD protein domain.

[0096] In some embodiments, the effector domain comprises a functional domain derived from a zinc finger repressor protein, such as a KRAB domain. KRAB domains are found in approximately 400 human ZFP-based transcription factors. Descriptions of KRAB domains may be found, for example, in Ecco et al., *Development* (2017) 144(15):2719-29 and Lambert et al., *Cell* (2018) 172:650-65.

[0097] In certain embodiments, the effector domain comprises a repression domain (e.g., KRAB) derived from KOX1/ZNF10, KOX8/ZNF708, ZNF43, ZNF184, ZNF91, HPF4, HTF10, or HTF34. In some embodiments, the effector domain comprises a repression domain (e.g., KRAB) derived from ZIM3, ZNF436, ZNF257, ZNF675, ZNF490, ZNF320, ZNF331, ZNF816, ZNF680, ZNF41, ZNF189, ZNF528, ZNF543, ZNF554, ZNF140, ZNF610, ZNF264, ZNF350, ZNF8, ZNF582, ZNF30, ZNF324, ZNF98, ZNF669, ZNF677, ZNF596, ZNF214, ZNF37, ZNF34, ZNF250, ZNF547, ZNF273, ZNF354, ZFP82, ZNF224, ZNF33, ZNF45, ZNF175, ZNF595, ZNF184, ZNF419, ZFP28-1, ZFP28-2, ZNF18, ZNF213, ZNF394, ZFP1, ZFP14, ZNF416, ZNF557, ZNF566, ZNF729, ZIM2, ZNF254, ZNF764, ZNF785, or any combination thereof. For example, the repression domain may be a KRAB domain derived from KOX1, ZIM3, ZFP28, or ZN627. In particular embodiments, the repression domain is a ZIM3 KRAB domain. In further embodiments, the effector domain is derived from a human protein, e.g., a human ZIM3, a human KOX1, a human ZFP28, or a human ZN627.

[0098] Exemplary effector domains that may reduce or silence target gene expression are provided in **Table 4** below (SEQ: SEQ ID NO, see **Table 20** for sequences of exemplary effector domains). Further examples of repressors and transcriptional repressor domains can be found, e.g., in PCT Patent Publication WO 2021/226077 and Tycko et al., *Cell* (2020) 183(7):2020-35, each of which is incorporated herein by reference in its entirety.

Table 4. Exemplary Effector Domains Suitable for Silencing Gene Expression

Protein	SEQ
ZIM3	495
ZNF436	496
ZNF257	497
ZNF675	498
ZNF490	499
ZNF320	500
ZNF331	501
ZNF816	502
ZNF680	503
ZNF41	504
ZNF189	505
ZNF528	506
ZNF543	507
ZNF554	508
ZNF140	509
ZNF610	510
ZNF264	511
ZNF350	512
ZNF8	513
ZNF582	514
ZNF30	515
ZNF324	516
ZNF98	517
ZNF669	518
ZNF677	519
ZNF596	520
ZNF214	521
ZNF37A	522
ZNF34	523
ZNF250	524
ZNF547	525
ZNF273	526
ZNF354A	527
ZFP82	528
ZNF224	529
ZNF33A	530

Protein	SEQ
ZN440_HUMAN	761
ZN583_HUMAN	762
ZN441_HUMAN	763
ZNF43_HUMAN	764
CBX5_HUMAN	765
ZN589_HUMAN	766
ZNF10_HUMAN	767
ZN563_HUMAN	768
ZN561_HUMAN	769
ZN136_HUMAN	770
ZN630_HUMAN	771
ZN527_HUMAN	772
ZN333_HUMAN	773
Z324B_HUMAN	774
ZN786_HUMAN	775
ZN709_HUMAN	776
ZN792_HUMAN	777
ZN599_HUMAN	778
ZN613_HUMAN	779
ZF69B_HUMAN	780
ZN799_HUMAN	781
ZN569_HUMAN	782
ZN564_HUMAN	783
ZN546_HUMAN	784
ZFP92_HUMAN	785
YAF2_HUMAN	786
ZN723_HUMAN	787
ZNF34_HUMAN	788
ZN439_HUMAN	789
ZFP57_HUMAN	790
ZNF19_HUMAN	791
ZN404_HUMAN	792
ZN274_HUMAN	793
CBX3_HUMAN	794
ZNF30_HUMAN	795
ZN250_HUMAN	796

Protein	SEQ
ZNF45	531
ZNF175	532
ZNF595	533
ZNF184	534
ZNF419	535
ZFP28-1	536
ZFP28-2	537
ZNF18	538
ZNF213	539
ZNF394	540
ZFP1	541
ZFP14	542
ZNF416	543
ZNF557	544
ZNF566	545
ZNF729	546
ZIM2	547
ZNF254	548
ZNF764	549
ZNF785	550
ZNF10 (KOX1)	551
CBX5 (chromoshadow domain)	552
RYBP (YAF2_RYBP component of PRC1)	553
YAF2 (YAF2_RYBP component of PRC1)	554
MGA (component of PRC1.6)	555
CBX1 (chromoshadow)	556
SCMH1 (SAM_1/SPM)	557
MPP8 (Chromodomain)	558
SUMO3 (Rad60-SLD)	559
HERC2 (Cyt-b5)	560
BIN1 (SH3_9)	561
PCGF2 (RING finger protein domain)	562
TOX (HMG box)	563
FOXA1 (HNF3A C-terminal domain)	564

Protein	SEQ
ZN570_HUMAN	797
ZN675_HUMAN	798
ZN695_HUMAN	799
ZN548_HUMAN	800
ZN132_HUMAN	801
ZN738_HUMAN	802
ZN420_HUMAN	803
ZN626_HUMAN	804
ZN559_HUMAN	805
ZN460_HUMAN	806
ZN268_HUMAN	807
ZN304_HUMAN	808
ZIM2_HUMAN	809
ZN605_HUMAN	810
ZN844_HUMAN	811
SUMO5_HUMAN	812
ZN101_HUMAN	813
ZN783_HUMAN	814
ZN417_HUMAN	815
ZN182_HUMAN	816
ZN823_HUMAN	817
ZN177_HUMAN	818
ZN197_HUMAN	819
ZN717_HUMAN	820
ZN669_HUMAN	821
ZN256_HUMAN	822
ZN251_HUMAN	823
CBX4_HUMAN	824
PCGF2_HUMAN	825
CDY2_HUMAN	826
CDYL2_HUMAN	827
HERC2_HUMAN	828
ZN562_HUMAN	829
ZN461_HUMAN	830

Protein	SEQ
FOXA2 (HNF3B C-terminal domain)	565
IRF2BP1 (IRF-2BP1_2 N-terminal domain)	566
IRF2BP2 (IRF-2BP1_2 N-terminal domain)	567
IRF2BPL IRF-2BP1_2 N-terminal domain	568
HOXA13 (homeodomain)	569
HOXB13 (homeodomain)	570
HOXC13 (homeodomain)	571
HOXA11 (homeodomain)	572
HOXC11 (homeodomain)	573
HOXC10 (homeodomain)	574
HOXA10 (homeodomain)	575
HOXB9 (homeodomain)	576
HOXA9 (homeodomain)	577
ZFP28_HUMAN	578
ZN334_HUMAN	579
ZN568_HUMAN	580
ZN37A_HUMAN	581
ZN181_HUMAN	582
ZN510_HUMAN	583
ZN862_HUMAN	584
ZN140_HUMAN	585
ZN208_HUMAN	586
ZN248_HUMAN	587
ZN571_HUMAN	588
ZN699_HUMAN	589
ZN726_HUMAN	590
ZIK1_HUMAN	591
ZNF2_HUMAN	592
Z705F_HUMAN	593
ZNF14_HUMAN	594
ZN471_HUMAN	595
ZN624_HUMAN	596
ZNF84_HUMAN	597
ZNF7_HUMAN	598
ZN891_HUMAN	599

Protein	SEQ
Z324A_HUMAN	831
ZN766_HUMAN	832
ID2_HUMAN	833
TOX_HUMAN	834
ZN274_HUMAN	835
SCMH1_HUMAN	836
ZN214_HUMAN	837
CBX7_HUMAN	838
ID1_HUMAN	839
CREM_HUMAN	840
SCX_HUMAN	841
ASCL1_HUMAN	842
ZN764_HUMAN	843
SCML2_HUMAN	844
TWST1_HUMAN	845
CREB1_HUMAN	846
TERF1_HUMAN	847
ID3_HUMAN	848
CBX8_HUMAN	849
CBX4_HUMAN	850
GSX1_HUMAN	851
NKX22_HUMAN	852
ATF1_HUMAN	853
TWST2_HUMAN	854
ZNF17_HUMAN	855
TOX3_HUMAN	856
TOX4_HUMAN	857
ZMYM3_HUMAN	858
I2BP1_HUMAN	859
RHXF1_HUMAN	860
SSX2_HUMAN	861
I2BPL_HUMAN	862
ZN680_HUMAN	863
CBX1_HUMAN	864
TRI68_HUMAN	865

Protein	SEQ
ZN337_HUMAN	600
Z705G_HUMAN	601
ZN529_HUMAN	602
ZN729_HUMAN	603
ZN419_HUMAN	604
Z705A_HUMAN	605
ZNF45_HUMAN	606
ZN302_HUMAN	607
ZN486_HUMAN	608
ZN621_HUMAN	609
ZN688_HUMAN	610
ZN33A_HUMAN	611
ZN554_HUMAN	612
ZN878_HUMAN	613
ZN772_HUMAN	614
ZN224_HUMAN	615
ZN184_HUMAN	616
ZN544_HUMAN	617
ZNF57_HUMAN	618
ZN283_HUMAN	619
ZN549_HUMAN	620
ZN211_HUMAN	621
ZN615_HUMAN	622
ZN253_HUMAN	623
ZN226_HUMAN	624
ZN730_HUMAN	625
Z585A_HUMAN	626
ZN732_HUMAN	627
ZN681_HUMAN	628
ZN667_HUMAN	629
ZN649_HUMAN	630
ZN470_HUMAN	631
ZN484_HUMAN	632
ZN431_HUMAN	633
ZN382_HUMAN	634
ZN254_HUMAN	635
ZN124_HUMAN	636

Protein	SEQ
HXA13_HUMAN	866
PHC3_HUMAN	867
TCF24_HUMAN	868
CBX3_HUMAN	869
HXB13_HUMAN	870
HEY1_HUMAN	871
PHC2_HUMAN	872
ZNF81_HUMAN	873
FIGLA_HUMAN	874
SAM11_HUMAN	875
KMT2B_HUMAN	876
HEY2_HUMAN	877
JDP2_HUMAN	878
HXC13_HUMAN	879
ASCL4_HUMAN	880
HHEX_HUMAN	881
HERC2_HUMAN	882
GSX2_HUMAN	883
BIN1_HUMAN	884
ETV7_HUMAN	885
ASCL3_HUMAN	886
PHC1_HUMAN	887
OTP_HUMAN	888
I2BP2_HUMAN	889
VGLL2_HUMAN	890
HXA11_HUMAN	891
PDLI4_HUMAN	892
ASCL2_HUMAN	893
CDX4_HUMAN	894
ZN860_HUMAN	895
LMBL4_HUMAN	896
PDIP3_HUMAN	897
NKX25_HUMAN	898
CEBPB_HUMAN	899
ISL1_HUMAN	900
CDX2_HUMAN	901
PROP1_HUMAN	902

Protein	SEQ
ZN607_HUMAN	637
ZN317_HUMAN	638
ZN620_HUMAN	639
ZN141_HUMAN	640
ZN584_HUMAN	641
ZN540_HUMAN	642
ZN75D_HUMAN	643
ZN555_HUMAN	644
ZN658_HUMAN	645
ZN684_HUMAN	646
RBAK_HUMAN	647
ZN829_HUMAN	648
ZN582_HUMAN	649
ZN112_HUMAN	650
ZN716_HUMAN	651
HKR1_HUMAN	652
ZN350_HUMAN	653
ZN480_HUMAN	654
ZN416_HUMAN	655
ZNF92_HUMAN	656
ZN100_HUMAN	657
ZN736_HUMAN	658
ZNF74_HUMAN	659
CBX1_HUMAN	660
ZN443_HUMAN	661
ZN195_HUMAN	662
ZN530_HUMAN	663
ZN782_HUMAN	664
ZN791_HUMAN	665
ZN331_HUMAN	666
Z354C_HUMAN	667
ZN157_HUMAN	668
ZN727_HUMAN	669
ZN550_HUMAN	670
ZN793_HUMAN	671
ZN235_HUMAN	672
ZNF8_HUMAN	673

Protein	SEQ
SIN3B_HUMAN	903
SMBT1_HUMAN	904
HXC11_HUMAN	905
HXC10_HUMAN	906
PRS6A_HUMAN	907
VSX1_HUMAN	908
NKX23_HUMAN	909
MTG16_HUMAN	910
HMX3_HUMAN	911
HMX1_HUMAN	912
KIF22_HUMAN	913
CSTF2_HUMAN	914
CEBPE_HUMAN	915
DLX2_HUMAN	916
ZMYM3_HUMAN	917
PPARG_HUMAN	918
PRIC1_HUMAN	919
UNC4_HUMAN	920
BARX2_HUMAN	921
ALX3_HUMAN	922
TCF15_HUMAN	923
TERA_HUMAN	924
VSX2_HUMAN	925
HXD12_HUMAN	926
CDX1_HUMAN	927
TCF23_HUMAN	928
ALX1_HUMAN	929
HXA10_HUMAN	930
RX_HUMAN	931
CXXC5_HUMAN	932
SCML1_HUMAN	933
NFIL3_HUMAN	934
DLX6_HUMAN	935
MTG8_HUMAN	936
CBX8_HUMAN	937
CEBPD_HUMAN	938
SEC13_HUMAN	939

Protein	SEQ
ZN724_HUMAN	674
ZN573_HUMAN	675
ZN577_HUMAN	676
ZN789_HUMAN	677
ZN718_HUMAN	678
ZN300_HUMAN	679
ZN383_HUMAN	680
ZN429_HUMAN	681
ZN677_HUMAN	682
ZN850_HUMAN	683
ZN454_HUMAN	684
ZN257_HUMAN	685
ZN264_HUMAN	686
ZFP82_HUMAN	687
ZFP14_HUMAN	688
ZN485_HUMAN	689
ZN737_HUMAN	690
ZNF44_HUMAN	691
ZN596_HUMAN	692
ZN565_HUMAN	693
ZN543_HUMAN	694
ZFP69_HUMAN	695
SUMO1_HUMAN	696
ZNF12_HUMAN	697
ZN169_HUMAN	698
ZN433_HUMAN	699
SUMO3_HUMAN	700
ZNF98_HUMAN	701
ZN175_HUMAN	702
ZN347_HUMAN	703
ZNF25_HUMAN	704
ZN519_HUMAN	705
Z585B_HUMAN	706
ZIM3_HUMAN	707
ZN517_HUMAN	708
ZN846_HUMAN	709
ZN230_HUMAN	710

Protein	SEQ
FIP1_HUMAN	940
ALX4_HUMAN	941
LHX3_HUMAN	942
PRIC2_HUMAN	943
MAGI3_HUMAN	944
NELL1_HUMAN	945
PRRX1_HUMAN	946
MTG8R_HUMAN	947
RAX2_HUMAN	948
DLX3_HUMAN	949
DLX1_HUMAN	950
NKX26_HUMAN	951
NAB1_HUMAN	952
SAMD7_HUMAN	953
PITX3_HUMAN	954
WDR5_HUMAN	955
MEOX2_HUMAN	956
NAB2_HUMAN	957
DHX8_HUMAN	958
FOXA2_HUMAN	959
CBX6_HUMAN	960
EMX2_HUMAN	961
CPSF6_HUMAN	962
HXC12_HUMAN	963
KDM4B_HUMAN	964
LMBL3_HUMAN	965
PHX2A_HUMAN	966
EMX1_HUMAN	967
NC2B_HUMAN	968
DLX4_HUMAN	969
SRY_HUMAN	970
ZN777_HUMAN	971
NELL1_HUMAN	972
ZN398_HUMAN	973
GATA3_HUMAN	974
BSH_HUMAN	975
SF3B4_HUMAN	976

Protein	SEQ
ZNF66_HUMAN	711
ZFP1_HUMAN	712
ZN713_HUMAN	713
ZN816_HUMAN	714
ZN426_HUMAN	715
ZN674_HUMAN	716
ZN627_HUMAN	717
ZNF20_HUMAN	718
Z587B_HUMAN	719
ZN316_HUMAN	720
ZN233_HUMAN	721
ZN611_HUMAN	722
ZN556_HUMAN	723
ZN234_HUMAN	724
ZN560_HUMAN	725
ZNF77_HUMAN	726
ZN682_HUMAN	727
ZN614_HUMAN	728
ZN785_HUMAN	729
ZN445_HUMAN	730
ZFP30_HUMAN	731
ZN225_HUMAN	732
ZN551_HUMAN	733
ZN610_HUMAN	734
ZN528_HUMAN	735
ZN284_HUMAN	736
ZN418_HUMAN	737
MPP8_HUMAN	738
ZN490_HUMAN	739
ZN805_HUMAN	740
Z780B_HUMAN	741
ZN763_HUMAN	742
ZN285_HUMAN	743
ZNF85_HUMAN	744
ZN223_HUMAN	745
ZNF90_HUMAN	746
ZN557_HUMAN	747

Protein	SEQ
TEAD1_HUMAN	977
TEAD3_HUMAN	978
RGAP1_HUMAN	979
PHF1_HUMAN	980
FOXA1_HUMAN	981
GATA2_HUMAN	982
FOXO3_HUMAN	983
ZN212_HUMAN	984
IRX4_HUMAN	985
ZBED6_HUMAN	986
LHX4_HUMAN	987
SIN3A_HUMAN	988
RBBP7_HUMAN	989
NKX61_HUMAN	990
TRI68_HUMAN	991
R51A1_HUMAN	992
MB3L1_HUMAN	993
DLX5_HUMAN	994
NOTC1_HUMAN	995
TERF2_HUMAN	996
ZN282_HUMAN	997
RGS12_HUMAN	998
ZN840_HUMAN	999
SPI2B_HUMAN	1000
PAX7_HUMAN	1001
NKX62_HUMAN	1002
ASXL2_HUMAN	1003
FOXO1_HUMAN	1004
GATA3_HUMAN	1005
GATA1_HUMAN	1006
ZMYM5_HUMAN	1007
ZN783_HUMAN	1008
SPI2B_HUMAN	1009
LRP1_HUMAN	1010
MIXL1_HUMAN	1011
SGT1_HUMAN	1012
LMCD1_HUMAN	1013

Protein	SEQ
ZN425_HUMAN	748
ZN229_HUMAN	749
ZN606_HUMAN	750
ZN155_HUMAN	751
ZN222_HUMAN	752
ZN442_HUMAN	753
ZNF91_HUMAN	754
ZN135_HUMAN	755
ZN778_HUMAN	756
RYBP_HUMAN	757
ZN534_HUMAN	758
ZN586_HUMAN	759
ZN567_HUMAN	760

Protein	SEQ
CEBPA_HUMAN	1014
GATA2_HUMAN	1015
SOX14_HUMAN	1016
WTIP_HUMAN	1017
PRP19_HUMAN	1018
CBX6_HUMAN	1019
NKX11_HUMAN	1020
RBBP4_HUMAN	1021
DMRT2_HUMAN	1022
SMCA2_HUMAN	1023
ZNF10_HUMAN	1024
EED_HUMAN	1025
RCOR1_HUMAN	1026

[0099] A functional analog of any one of the above-listed proteins, i.e., a molecule having the same or substantially the same biological function (e.g., retaining 70% or more, 80% or more, 90% or more, 95% or more, or 98% or more) of the protein's transcription factor function) is encompassed by the present disclosure. For example, the functional analog may be an isoform or a variant of the above-listed protein, e.g., containing a portion of the above protein with or without additional amino acid residues and/or containing mutations relative to the above protein. In some embodiments, the functional analog has a sequence identity that is at least 75, 80, 85, 90, 95, 98, or 99% to one of the sequences listed in **Table 4**. Homologs, orthologs, and mutants of the above-listed proteins are also contemplated.

[0100] In certain embodiments, an epigenetic editor described herein comprises a KRAB domain derived from KOX1, ZIM3, ZFP28, or ZN627, and/or an effector domain derived from KAP1, MECP2, HP1a, HP1b, CBX8, CDYL2, TOX, TOX3, TOX4, EED, EZH2, RBBP4, RCOR1, or SCML2, optionally wherein the parental protein is a human protein. In particular embodiments, an epigenetic editor described herein comprises a domain derived from KOX1, ZIM3, ZFP28, and/or ZN627, optionally wherein the parental protein is a human protein. In certain embodiments, the epigenetic editor may comprise a KRAB domain derived from KOX1 (ZNF10), e.g., a human KOX1. In certain embodiments, the epigenetic editor may comprise a KRAB domain derived from ZIM3 (ZNF657 or ZNF264), e.g., a human ZIM3. In certain embodiments, the epigenetic editor may comprise a KRAB domain derived from ZFP28, e.g., a human ZFP28. In certain embodiments, the epigenetic editor may comprise a KRAB domain derived from ZN627, e.g., a human ZN627. In certain embodiments, an epigenetic editor described herein may comprise a CDYL2, e.g., a human CDYL2, and/or a TOX domain (e.g., a human TOX domain) in combination with a KOX1 KRAB domain (e.g., a human KOX1 KRAB domain).

[0101] In certain embodiments, an epigenetic effector described herein comprises a repression domain derived from ZNF10 (SEQ ID NO: 1024). For example, the repression domain may comprise the sequence of SEQ ID NO: 1024, or a sequence at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 1024.

B. DNA Methyltransferases

[0102] In some embodiments, an effector domain of an epigenetic editor described herein alters target gene expression through DNA modification, such as methylation. Highly methylated areas of DNA tend to be less transcriptionally active than less methylated areas. DNA methylation occurs primarily at CpG sites (shorthand for "C-phosphate-G-" or "cytosine-phosphate-guanine" sites). Many mammalian genes have promoter regions near or including CpG islands (nucleic acid regions with a high frequency of CpG dinucleotides).

[0103] An effector domain described herein may be, e.g., a DNA methyltransferase (DNMT) or a catalytic domain thereof, or may be capable of recruiting a DNA methyltransferase. DNMTs encompass enzymes that catalyze the transfer of a methyl group to a DNA nucleotide, such as canonical cytosine-5 DNMTs that catalyze the addition of methyl groups to genomic DNA (e.g., DNMT1, DNMT3A, DNMT3B, and DNMT3C). This term also encompasses non-canonical family members that do not catalyze methylation themselves but that recruit (including activate) catalytically active DNMTs; a non-limiting example of such a DNMT is DNMT3L. *See, e.g.*, Lyko, *Nat Review* (2018) 19:81-92. Unless otherwise indicated, a DNMT domain may refer to a polypeptide domain derived from a catalytically active DNMT (e.g., DNMT1, DNMT3A, and DNMT3B) or from a catalytically inactive DNMT (e.g., DNMT3L). A DNMT may repress expression of the target gene through the recruitment of repressive regulatory proteins. In some embodiments, the methylation is at a CG (or CpG) dinucleotide sequence. In some embodiments, the methylation is at a CHG or CHH sequence, where H is any one of A, T, or C.In some embodiments, DNMTs in the epigenetic editors may include, e.g., DNMT1, DNMT3A, DNMT3B, and/or DNMT3C. In some embodiments, the DNMT is a mammalian (e.g., human or murine) DNMT. In particular embodiments, the DNMT is DNMT3A (e.g., human DNMT3A). In certain embodiments, an epigenetic editor described herein comprises a DNMT3A domain comprising SEQ ID NO: 1028, or a sequence at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ

ID NO: 1028. In certain embodiments, an epigenetic editor described herein comprises a DNMT3A domain comprising SEQ ID NO: 1029, or a sequence at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 1029. In some embodiments, the DNMT3A domain may have, e.g., a mutation at position H739 (such as H739A or H739E), R771 (such as R771L) and/or R836 (such as R836A or R836Q), or any combination thereof (numbering according to SEQ ID NO: 1028).

[0104] In some embodiments, an effector domain described herein may be a DNMT-like domain. As used herein a "DNMT-like domain" is a regulatory factor of DNA methyltransferase that may activate or recruit other DNMT domains, but does not itself possess methylation activity. In some embodiments, the DNMT-like domain is a mammalian (e.g., human or mouse) DNMT-like domain. In certain embodiments, the DNMT-like domain is DNMT3L, which may be, for example, human DNMT3L or mouse DNMT3L. In certain embodiments, an epigenetic editor described herein comprises a DNMT3L domain comprising SEQ ID NO: 1032, or a sequence at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 1032. In certain embodiments, an epigenetic editor herein comprises a DNMT3L domain comprising SEQ ID NO: 1033, or a sequence at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 1033. In certain embodiments, an epigenetic editor described herein comprises a DNMT3L domain comprising SEQ ID NO: 1034, or a sequence at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 1034. In certain embodiments, an epigenetic editor described herein comprises a DNMT3L domain comprising SEQ ID NO: 1035, or a sequence at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 1035. In some embodiments, the DNMT3L domain may have, e.g., a mutation corresponding to that at position D226 (such as D226V), Q268 (such as Q268K), or both (numbering according to SEQ ID NO: 1032).

[0105] In certain embodiments, an epigenetic editor herein may comprise comprising both DNMT and DNMT-like effector domains. For example, the epigenetic editor may comprise a DNMT3A-3L domain, wherein DNMT3A and DNMT3L may be covalently linked. In other embodiments, an epigenetic editor described herein may comprise an effector domain that comprises only a DNMT3A domain (e.g., human DNMT3A), or only a DNMT-like domain (e.g., DNMT3L, which may be human or mouse DNMT3L).

[0106] **Table 5** below provides exemplary methyltransferases from which an effector domain of an epigenetic editor described herein may be derived. See **Table 20** for sequences of these exemplary methyltransferases.

Table 5. Exemplary DNA Methyltransferase Sequences

Protein Name	**Species**	**Target**	**Protein Sequence**
DNMT1	Human	5mC	SEQ ID NO: 1027
DNMT3A	Human	5mC	SEQ ID NO: 1028
DNMT3A (catalytic domain)	Human	5mC	SEQ ID NO: 1029
DNMT3B	Human	5mC	SEQ ID NO: 1030
DNMT3C	Mouse	5mC	SEQ ID NO: 1031
DNMT3L	Human	5mC	SEQ ID NO: 1032
DNMT3L (catalytic domain)	Human	5mC	SEQ ID NO: 1033
DNMT3L	Mouse	5mC	SEQ ID NO: 1034
DNMT3L (catalytic domain)	Mouse	5mC	SEQ ID NO: 1035
TRDMT1 (DNMT2)	Human	tRNA 5mC	SEQ ID NO: 1036
M.MpeI	*Mycoplasma penetrans*	5mC	SEQ ID NO: 1037
M.SssI	*Spiroplasma monobiae*	5mC	SEQ ID NO: 1038
M.HpaII	*Haemophilus parainfluenzae*	5mC (CCGG)	SEQ ID NO: 1039
M.AluI	*Arthrobacter luteus*	5mC (AGCT)	SEQ ID NO: 1040
M.HaeIII	*Haemophilus aegyptius*	5mC (GGCC)	SEQ ID NO: 1041
M.HhaI	*Haemophilus haemolyticus*	5mC (GCGC)	SEQ ID NO: 1042
M.MspI	*Moraxella*	5mC (CCGG)	SEQ ID NO: 1043
Masc1	*Ascobolus*	5mC	SEQ ID NO: 1044
MET1	*Arabidopsis*	5mC	SEQ ID NO: 1045
Masc2	*Ascobolus*	5mC	SEQ ID NO: 1046
Dim-2	*Neurospora*	5mC	SEQ ID NO: 1047
dDnmt2	*Drosophila*	5mC	SEQ ID NO: 1048
Pmt1	*S. pombe*	5mC	SEQ ID NO: 1049
DRM1	*Arabidopsis*	5mC	SEQ ID NO: 1050
DRM2	*Arabidopsis*	5mC	SEQ ID NO: 1051
CMT1	*Arabidopsis*	5mC	SEQ ID NO: 1052
CMT2	*Arabidopsis*	5mC	SEQ ID NO: 1053
CMT3	*Arabidopsis*	5mC	SEQ ID NO: 1054
Rid	*Neurospora*	5mC	SEQ ID NO: 1055
hsdM gene	bacteria (*E. coli*, strain 12)	m6A	SEQ ID NO: 1056
hsdS gene	bacteria (*E. coli*, strain 12)	m6A	SEQ ID NO: 1057
M.TaqI	Bacteria (*Thermus aquaticus*)	m6A	SEQ ID NO: 1058

Protein Name	**Species**	**Target**	**Protein Sequence**
M.EcoDam	*E. coli*	m6A	SEQ ID NO: 1059
M.CcrMI	*Caulobacter crescentus*	m6A	SEQ ID NO: 1060
CamA	*Clostridioides difficile*	m6A	SEQ ID NO: 1061

[0107] A functional analog of any one of the above-listed proteins, i.e., a molecule having the same or substantially the same biological function (e.g., retaining 70% or more, 80% or more, 90% or more, 95% or more, or 98% or more) of the protein's DNA methylation function or recruiting function) is encompassed by the present disclosure. For example, the functional analog may be an isoform or a variant of the above-listed protein, e.g., containing a portion of the above protein with or without additional amino acid residues and/or containing mutations relative to the above protein. In some embodiments, the functional analog has a sequence identity that is at least 75, 80, 85, 90, 95, 98, or 99% to one of the sequences listed in **Table 5**. In some embodiments, the effector domain herein comprises only the functional domain (or functional analog thereof), e.g., the catalytical domain or recruiting domain, of the above-listed proteins.

[0108] As used herein, a DNMT domain (e.g., a DNMT3A domain or a DNMT3L domain) refers to a protein domain that is identical to the parental protein (e.g., a human or murine DNMT3A or DNMT3L) or a functional analog thereof (e.g., having a functional fragment, such as a catalytic fragment or recruiting fragment, of the parental protein; and/or having mutations that improve the activity of the DNMT protein).

[0109] An epigenetic editor herein may effect methylation at, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 or more CpG dinucleotide sequences in the target gene or chromosome. The CpG dinucleotide sequences may be located within or near the target gene in CpG islands, or may be located in a region that is not a CpG island. A CpG island generally refers to a nucleic acid sequence or chromosome region that comprises a high frequency of CpG dinucleotides. For example, a CpG island may comprise at least 50% GC content. The CpG island may have a high observed-to-expected CpG ratio, for example, an observed-to-expected CpG ratio of at least 60%. As used herein, an observed-to-expected CpG ratio is determined by Number of CpG * (sequence length) / (Number of C * Number of G). In some embodiments, the CpG island has an observed-to-expected CpG ratio of at least 60%, 70%, 80%, 90% or more. A CpG island may be a sequence or region of, e.g., at least 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, or 800 nucleotides. In some embodiments, only 1, or less than 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, or 50 CpG dinucleotides are methylated by the epigenetic editor.

[0110] In some embodiments, an epigenetic editor herein effects methylation at a hypomethylated nucleic acid sequence, i.e., a sequence that may lack methyl groups on the 5-methyl cytosine nucleotides (e.g., in CpG) as compared to a standard control. Hypomethylation may occur, for example, in aging cells or in cancer (e.g., early stages of neoplasia) relative to a younger cell or non-cancer cell, respectively.

[0111] In some embodiments, an epigenetic editor described herein induces methylation at a hypermethylated nucleic acid sequence.

[0112] In some embodiments, methylation may be introduced by the epigenetic editor at a site other than a CpG dinucleotide. For example, the target gene sequence may be methylated at the C nucleotide of CpA, CpT, or CpC sequences. In some embodiments, an epigenetic editor comprises a DNMT3A domain and effects methylation at CpG, CpA, CpT, CpC sequences, or any combination thereof. In some embodiments, an epigenetic editor comprises a DNMT3A domain that lacks a regulatory subdomain and only maintains a catalytic domain. In some embodiments, the epigenetic editor comprising a DNMT3A catalytic domain effects methylation exclusively at CpG sequences. In some embodiments, an epigenetic editor comprising a DNMT3A domain that comprises a mutation, e.g. a R836A or R836Q mutation (numbering according to SEQ ID NO: 1028), has higher methylation activity at CpA, CpC, and/or CpT sequences as compared to an epigenetic editor comprising a wildtype DNMT3A domain.

C. Histone Modifiers

[0113] In some embodiments, an effector domain of an epigenetic editor herein mediates histone modification. Histone modifications play a structural and biochemical role in gene transcription, such as by formation or disruption of the nucleosome structure that binds to the histone and prevents gene transcription. Histone modifications may include, for example, acetylation, deacetylation, methylation, phosphorylation, ubiquitination, SUMOylation and the like, e.g., at their N-terminal ends (“histone tails”). These modifications maintain or specifically convert chromatin structure, thereby controlling responses such as gene expression, DNA replication, DNA repair, and the like, which occur on chromosomal DNA. Post-translational modification of histones is an epigenetic regulatory mechanism and is considered essential for the genetic regulation of eukaryotic cells. Recent studies have revealed that chromatin remodeling factors such as SWI/SNF, RSC, NURF, NRD, and the like, which facilitate transcription factor access to DNA by modifying the nucleosome structure; histone acetyltransferases (HATs) that regulate the acetylation state of histones; and histone deacetylases (HDACs), act as important regulators.

[0114] In particular, the unstructured N-termini of histones may be modified by acetylation, deacetylation, methylation, ubiquitylation, phosphorylation, SUMOylation, ribosylation, citrullination O-GlcNAcylation, crotonylation, or any combination thereof. For example, histone acetyltransferases (HATs) utilize acetyl-CoA as a cofactor and catalyze the transfer of an acetyl group to the epsilon amino group of the lysine side chains. This neutralizes the lysine's positive charge and weakens the interactions between histones and DNA, thus opening the chromosomes for transcription factors to bind and initiate transcription. Acetylation of K14 and K9 lysines of histone H3 by histone acetyltransferase enzymes may be linked to transcriptional competence in humans. Lysine acetylation may directly or indirectly create binding sites for chromatin-modifying enzymes that regulate transcriptional activation. On the other hand, histone methylation of lysine 9 of histone H3 may be associated with heterochromatin, or transcriptionally silent chromatin.

[0115] In certain embodiments, an effector domain of an epigenetic editor described herein comprises a histone methyltransferase domain. The effector domain may comprise, for example, a DOT1L domain, a SET domain, a SUV39H1 domain, a G9a/EHMT2 protein domain, an EZH1 domain, an EZH2 domain, a SETDB1 domain, or any combination thereof. In particular embodiments, the effector domain comprises a histone-lysine-N-methyltransferase SETDB1 domain.

[0116] In some embodiments, the effector domain comprises a histone deacetylase protein domain. In certain embodiments, the effector domain comprises a HDAC family protein domain, for example, a HDAC1, HDAC3, HDAC5, HDAC7, or HDAC9 protein domain. In particular embodiments, the effector domain comprises a nucleosome remodeling and deacetylase complex (NURD), which removes acetyl groups from histones.

D. Other Effector Domains

[0117] In some embodiments, the effector domain comprises a tripartite motif containing protein (TRIM28, TIF1-beta, or KAP1). In certain embodiments, the effector domain comprises one or more KAP1 proteins. A KAP1 protein in an epigenetic editor herein may form a complex with one or more other effector domains of the epigenetic editor or one or more proteins involved in modulation of gene expression in a cellular environment. For example, KAP1 may be recruited by a KRAB domain of a transcriptional repressor. A KAP1 protein domain may interact with or recruit one or more protein complexes that reduces or silences gene expression. In some embodiments, KAP1 interacts with or recruits a histone deacetylase protein, a histone-lysine methyltransferase protein, a chromatin remodeling protein, and/or a heterochromatin protein. For example, a KAP1 protein domain may interact with or recruit a heterochromatin protein 1 (HP1) protein, a SETDB1 protein, an HDAC protein, and/or a NuRD protein complex component. In some embodiments, a KAP1 protein domain interacts with or recruits a ZFP90 protein (e.g., isoform 2 of ZFP90), and/or a FOXP3 protein. An exemplary KAP1 amino acid sequence is shown in SEQ ID NO: 1062.

[0118] In some embodiments, the effector domain comprises a protein domain that interacts with or is recruited by one or more DNA epigenetic marks. For example, the effector domain may comprise a methyl CpG binding protein 2 (MECP2) protein that interacts with methylated DNA nucleotides in the target gene (which may or may not be at a CpG island of the target gene). An MECP2 protein domain in an epigenetic editor described herein may induce condensed chromatin structure, thereby reducing or silencing expression of the target gene. In some embodiments, an MECP2 protein domain in an epigenetic editor described herein may interact with a histone deacetylase (e.g. HDAC), thereby repressing or silencing expression of the target gene. In some embodiments, an MECP2 protein domain in an epigenetic editor described herein may block access of a transcription factor or transcriptional activator to the target sequence, thereby repressing or silencing expression of the target gene. An exemplary MECP2 amino acid sequence is shown in SEQ ID NO: 1063.

[0119] Also contemplated as effector domains for the epigenetic editors described herein are, e.g., a chromoshadow domain, a ubiquitin-2 like Rad60 SUMO-like (Rad60-SLD/SUMO) domain, a chromatin organization modifier domain (Chromo) domain, a Yaf2/RYBP C-terminal binding motif domain (YAF2_RYBP), a CBX family C-terminal motif domain (CBX7_C), a zinc finger C3HC4 type (RING finger) domain (ZF-C3HC4_2), a cytochrome b5 domain (Cyt-b5), a helix-loop-helix domain (HLH), a helix-hairpin-helix motif domain (e.g., HHH_3), a high mobility group box domain (HMG-box), a basic leucine zipper domain (e.g., bZIP_1 or bZIP_2), a Myb_DNA-binding domain, a homeodomain, a MYM-type Zinc finger with FCS sequence domain (ZF-FCS), an interferon regulatory factor 2-binding protein zinc finger domain (IRF-2BP1_2), an SSX repression domain (SSXRD), a B-box-type zinc finger domain (ZF-B_box), a CXXC zinc finger domain (ZF-CXXC), a regulator of chromosome condensation 1 domain (RCC1), an SRC homology 3 domain (SH3_9), a sterile alpha motif domain (SAM_1), a sterile alpha motif domain (SAM_2), a sterile alpha motif/Pointed domain (SAM_PNT), a Vestigial/Tondu family domain (Vg_Tdu), a LIM domain, an RNA recognition motif domain (RRM_1), a paired amphipathic helix domain (PAH), a proteasomal ATPase OB C-terminal domain (Prot_ATP_ID_OB), a nervy homology 2 domain (NHR2), a hinge domain of cleavage stimulation factor subunit 2 (CSTF2_hinge), a PPAR gamma N-terminal region domain (PPARgamma_N), a CDC48 N-terminal domain (CDC48_2), a WD40 repeat domain (WD40), a Fip1 motif domain (Fip1), a PDZ domain (PDZ_6), a Von Willebrand factor type C domain (VWC), a NAB conserved region 1 domain (NCD1), an S1 RNA-binding domain (S1), an HNF3 C-terminal domain (HNF_C), a Tudor domain (Tudor_2), a histone-like transcription factor (CBF/NF-Y) and archaeal histone domain (CBFD_NFYB_HMF), a zinc finger protein domain (DUF3669), an EGF-like domain (cEGF), a GATA zinc finger domain (GATA), a TEA/ATTS domain (TEA), a phorbol esters/diacylglycerol binding domain (C1-1), polycomb-like MTF2 factor 2 domain (Mtf2_C), a transactivation domain of FOXO protein family (FOXO-TAD), a homeobox KN domain (Homeobox_KN), a BED zinc finger domain (ZF-BED), a zinc finger of C3HC4-type RING domain (ZF-C3HC4_4), a RAD51 interacting motif domain (RAD51_interact), a p55-binding region of a nethyl-CpG-binding domain protein MBD (MBDa), a Notch domain, a Raf-like Ras-binding domain (RBD), a Spin/Ssty family domain (Spin-Ssty), a PHD finger domain (PHD_3), a Low-density lipoprotein receptor domain class A (Ldl_recept_a), a CS domain, a DM DNA-binding domain, and a QLQ domain.

[0120] In some embodiments, the effector domain is a protein domain comprising a YAF2_RYBP domain or homeodomain or any combination thereof. In certain embodiments, the homeodomain of the YAF2_RYBP domain is a PRD domain, an NKL domain, a HOXL domain, or a LIM domain. In particular embodiments, the YAF2_RYBP domain may comprise a 32 amino acid Yaf2/RYBP C-terminal binding motif domain (32 aa RYBP).

[0121] In some embodiments, the effector domain comprises a protein domain selected from a group consisting of SUMO3 domain, Chromo domain from M phase phosphoprotein 8 (MPP8), chromoshadow domain from Chromobox 1 (CBX1), and SAM_1/SPM domain from Scm Polycomb Group Protein Homolog 1 (SCMH1).

[0122] In some embodiments, the effector domain comprises an HNF3 C-terminal domain (HNF_C). The HNF_C domain may be from FOXA1 or FOXA2. In certain embodiments, the HNF_C domain comprises an EH1 (engrailed homology 1) motif.

[0123] In some embodiments, the effector domain may comprise an interferon regulatory factor 2-binding protein zinc finger domain (IRF-2BP1_2), a Cyt-b5 domain from DNA repair factor HERC2 E3 ligase, a variant SH3 domain (SH3_9) from Bridging Integrator 1 (BIN1), an HMG-box domain from transcription factor TOX or ZF-C3HC4_2 RING finger domain from the polycomb component PCGF2, a Chromodomain-helicase-DNA binding protein 3 (CHD3) domain, or a ZNF783 domain.

IV. Epigenetic Editors

[0124] Provided herein are epigenetic editors, also referred to herein as epigenetic editing systems, that direct epigenetic modification(s) to a target sequence in a gene of interest, e.g., using one or more DNA-binding domains as described herein and one or more effector domains (e.g., epigenetic repression domains) as described herein, in any combination. The DNA-binding domain (in concert with a guide polynucleotide such as one described herein, where the DNA-binding domain is a polynucleotide guided DNA-binding domain) directs the effector domain to epigenetically modify the target sequence, resulting in gene repression or silencing that may be durable and inheritable across cell generations. In some aspects, the epigenetic editors described herein can repress or silence genes reversibly or irreversibly in cells.

[0125] In particular embodiments, an epigenetic editor described herein comprises one or more fusion proteins, each comprising (1) DNA-binding domain(s) and (2) effector domain(s). The effector domains may be on one or more fusion proteins comprised by the epigenetic editor. For example, a single fusion protein may comprise all of the effector domains with a DNA-binding domain. Alternatively, the effector domains or subsets thereof may be on separate fusion proteins, each with a DNA-binding domain (which may be the same or different). A fusion protein described herein may further comprise one or more linkers (e.g., peptide linkers), detectable tags, nuclear localization signals (NLSs), or any combination thereof. As used herein, a “fusion protein” refers to a chimeric protein in which two or more coding sequences (e.g., for DNA-binding domain(s) and/or effector domain(s)) are covalently or non-covalently joined, directly or indirectly.

[0126] In some embodiments, an epigenetic editor described herein comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, or more effector (e.g., repression) domains, which may be identical or different. In certain embodiments, two or more of said effector domains function synergistically. Combinations of effector domains may comprise DNA methylation domains, histone deacetylation domains, histone methylation domains, and/or scaffold domains that recruit any of the above. For example, an epigenetic editor described herein may comprise one or more transcriptional repressor domains (e.g., a KRAB domain such as KOX1, ZIM3, ZFP28, or ZN627 KRAB) in combination with one or more DNA methylation domains (e.g., a DNMT domain) and/or recruiter domain (e.g., a DNMT3L domain). Such an epigenetic editor may comprise, for instance, a KRAB domain, a DNMT3A domain, and a DNMT3L domain. An epigenetic editor can comprise a DNMT3A domain and a DNMT3L domain and preferably further comprise a KRAB domain. In some embodiments, the epigenetic editor further comprises an additional effector domain (e.g., a KAP1, MECP2, HP1b, CBX8, CDYL2, TOX,

TOX3, TOX4, EED, RBBP4, RCOR1, or SCML2 domain). In some embodiments, the additional effector domain is a CDYL2, TOX, TOX3, TOX4, or HP1a domain. For example, an epigenetic editor described herein may comprise a CDYL2 and/or a TOX domain in combination with a KRAB domain (e.g., a KOX1 KRAB domain).

A. Linkers

[0127] A fusion protein as described herein may comprise one or more linkers that connect components of the epigenetic editor. A linker may be a peptide or non-peptide linker.

[0128] In some embodiments, one or more linkers utilized in an epigenetic editor provided herein is a peptide linker, i.e., a linker comprising a peptide moiety. A peptide linker can be any length applicable to the epigenetic editor fusion proteins described herein. In some embodiments, the linker can comprise a peptide between 1 and 200 (e.g., between 1 and 80) amino acids. In some embodiments, the linker comprises from 1 to 5, 1 to 10, 1 to 20, 1 to 30, 1 to 40, 1 to 50, 1 to 60, 1 to 80, 1 to 100, 1 to 150, 1 to 200, 5 to 10, 5 to 20, 5 to 30, 5 to 40, 5 to 60, 5 to 80, 5 to 100, 5 to 150, 5 to 200, 10 to 20, 10 to 30, 10 to 40, 10 to 50, 10 to 60, 10 to 80, 10 to 100, 10 to 150, 10 to 200, 20 to 30, 20 to 40, 20 to 50, 20 to 60, 20 to 80, 20 to 100, 20 to 150, 20 to 200, 30 to 40, 30 to 50, 30 to 60, 30 to 80, 30 to 100, 30 to 150, 30 to 200, 40 to 50, 40 to 60, 40 to 80, 40 to 100, 40 to 150, 40 to 200, 50 to 60 50 to 80, 50 to 100, 50 to 150, 50 to 200, 60 to 80, 60 to 100, 60 to 150, 60 to 200, 80 to 100, 80 to 150, 80 to 200, 100 to 150, 100 to 200, or 150 to 200 amino acids in length. Longer or shorter linkers are also contemplated. In some embodiments, the peptide linker is 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 25, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 amino acids in length. For example, the peptide linker may be 4, 5, 16, 20, 24, 27, 32, 40, 64, 92, or 104 amino acids in length. The peptide linker may be a flexible or rigid linker. In particular embodiments, the peptide linker comprises the amino acid sequence of any one of SEQ ID NOs: 1064-1068 or a sequence at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical thereto.

[0129] In certain embodiments, the peptide linker is an XTEN linker. Such a linker may comprise part of the XTEN sequence (Schellenberger et al., *Nat Biotechnol* (2009) 27(1):1186-90), an unstructured hydrophilic polypeptide consisting only of residues G, S, P, T, E, and A. The term “XTEN” as used herein refers to a recombinant peptide or polypeptide lacking hydrophobic amino acid residues. XTEN linkers typically are unstructured and comprise a limited set of natural amino acids. Fusion of XTEN to proteins alters its hydrodynamic properties and reduces the rate of clearance and degradation of the fusion protein. These XTEN fusion proteins are produced using recombinant technology, without the need for chemical modifications, and degraded by natural pathways. The XTEN linker may be, for example, 5, 10, 16, 20, 26, or 80 amino acids in length. In some embodiments, the XTEN linker is 16 amino acids in length. In some embodiments, the XTEN linker is 80 amino acids in length. In certain embodiments, the XTEN linker may be XTEN10, XTEN16, XTEN20, or XTEN80. In certain embodiments, the XTEN linker may comprise the amino acid sequence of any one of SEQ ID NOs: 1069-1073 and 1092 or a sequence at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical thereto. In some embodiments, the XTEN linker may be XTEN10, XTEN16, XTEN20, or XTEN80.

[0130] In some embodiments, one or more linkers utilized in an epigenetic editor provided herein is a non-peptide linker. For example, the linker may be a carbon bond, a disulfide bond, or carbon-heteroatom bond. In certain embodiments, the linker is a carbon-nitrogen bond of an amide linkage. In certain embodiments, the linker is a cyclic or acyclic, substituted or unsubstituted, or branched or unbranched aliphatic or heteroaliphatic linker.

[0131] In some embodiments, one or more linkers utilized in an epigenetic editor provided herein is polymeric (e.g., polyethylene, polyethylene glycol, polyamide, polyester, etc.). The linker may comprise, for example, a monomer, dimer, or polymer of aminoalkanoic acid; an aminoalkanoic acid (e.g., glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, etc.); a monomer, dimer, or polymer of aminohexanoic acid (Ahx); or a polyethylene glycol moiety (PEG); or an aryl or heteroaryl moiety. In certain embodiments, the linker may be based on a carbocyclic moiety (e.g., cyclopentane or cyclohexane) or a phenyl ring. The linker may include functionalized moieties to facilitate attachment of a nucleophile (e.g., thiol, amino) from the peptide to the linker. Any electrophile may be used as part of the linker. Exemplary electrophiles include, but are not limited to, activated esters, activated amides, alkyl halides, aryl halides, acyl halides, and isothiocyanates.

[0132] Various linker lengths and flexibilities can be employed between any two components of an epigenetic editor (e.g., between an effector domain (e.g., a repressor domain) and a DNA-binding domain (e.g., a Cas9 domain), between a first effector domain and a second effector domain, etc.). The linkers may range from very flexible linkers, such as glycine/serine-rich linkers, to more rigid linkers, in order to achieve the optimal length for effector domain activity for the specific application. In some embodiments, the more flexible linkers are glycine/serine-rich linkers (GS-rich linkers), where more than 45% (e.g., more than 48, 50, 55, 60, 70, 80, or 90%) of the residues are glycine or serine residues. Non-limiting examples of the GS-rich linkers are (GGGGS)n (SEQ ID NO: 485), (G)n, and W linker (SEQ ID NO: 486). In some embodiments, the more rigid linkers are in the form of the form (EAAAK)n (SEQ ID NO:

487), (SGGS)n (SEQ ID NO: 488), and (XP)n (SEQ ID NO: 489). In the aforementioned formulae of flexible and rigid linkers, n may be any integer between 1 and 30. In some embodiments, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15. In some embodiments, the linker comprises a (GGS)n motif, wherein n is 1, 3, or 7 (SEQ ID NO: 490). In some embodiments, the linker comprises a (GGGGS)n motif, wherein n is 4 (SEQ ID NO: 491).

[0133] In some embodiments, a linker in an epigenetic editor described herein comprises a nuclear localization signal, for example, with the amino acid sequence of any one of SEQ ID NOs: 1074-1079. In some embodiments, a linker in an epigenetic editor described herein comprises an expression tag, e.g., a detectable tag such as a green fluorescence protein.

B. Nuclear Localization Signals

[0134] A fusion protein described herein may comprise one or more nuclear localization signals, and in certain embodiments, may comprise two or more nuclear localization signals. For example, the fusion protein may comprise 1, 2, 3, 4, or 5 nuclear localization signals. As used herein, a "nuclear localization signal" (NLS) is an amino acid sequence that directs proteins to the nucleus. In certain embodiments, the NLS may be an SV40 NLS. The fusion protein may comprise an NLS at its N-terminus, C-terminus, or both, and/or an NLS may be embedded in the middle of the fusion protein (e.g., at the N- or C- terminus of a DNA-binding domain or an effector domain). In certain embodiments, an NLS comprises the amino acid sequence of any one of SEQ ID NOs: 1074-1079, or a sequence at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the selected sequence. Additional NLSs are known in the art.

C. Tags

[0135] Epigenetic editors provided herein may comprise one or more additional sequences ("tags") for tracking, detection, and localization of the editors. In some embodiments, the epigenetic editor comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more detectable tags. Each of the detectable tags may be the same or different.

[0136] For example, an epigenetic editor fusion protein may comprise cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, poly-histidine tags (also referred to as histidine tags or His-tags), maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (e.g., Softag 1 or Softag 3), strep-tags, biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art. Sequences disclosed herein that are presented with tag sequences included are also contemplated without the presented tag sequences; similarly, sequences disclosed herein without tag sequences are also contemplated to include the addition of suitable sequences apparent to those of skill in the art.

D. Fusion Protein Configurations

[0137] A fusion protein of an epigenetic editor described herein may have its components structured in different configurations. For example, the DNA-binding domain may be at the C-terminus, the N-terminus, or in between two or more epigenetic effector domains or additional domains. In some embodiments, the DNA-binding domain is at the C-terminus of the epigenetic editor. In some embodiments, the DNA-binding domain is at the N-terminus of the epigenetic editor. In some embodiments, the DNA-binding domain is linked to one or more nuclear localization signals. In some embodiments, the DNA-binding domain is flanked by an epigenetic effector domain and/or an additional domain on both sides. In some embodiments, where “DBD” indicates DNA-binding domain and “ED” indicates effector domain, the epigenetic editor comprises the configuration of:

- N’]-[ED1]-[DBD]-[ED2]-[C’
- N’]-[ED1]-[DBD]-[ED2]-[ED3]-[C’
- N’]-[ED1]-[ED2]-[DBD]-[ED3]-[C’

or

- N’]-[ED1]-[ED2]-DBD]-[ED3]-[ED4]-[C’.

[0138] In some embodiments, an epigenetic editor comprises a DNA-binding domain (DBD), a DNA methyltransferase (DNMT) domain, and a transcriptional repressor (“repressor”) domain that represses or silences expression of a target gene. The DBD, DNMT, and transcriptional repressor domains may be any as described herein, in any combination. For example, an epigenetic editor can comprise a DBD, a DNMT3A domain, and a DNMT3L domain. An epigenetic editor can comprise a DBD, a DNMT3A domain, a DNMT3L domain, and preferably further comprise a KRAB domain. In some embodiments, the epigenetic editor comprises a fusion protein with the configuration of:

N’]-[DNA methyltransferase domain]-[DBD]-[repressor domain]-[C’
N’]-[repressor domain]-[DBD]-[DNA methyltransferase domain]-[C’
N’]-[DNA methyltransferase domain]-[repressor domain]- [DBD]-[C’
or
N’]-[repressor domain]-[DNA methyltransferase domain]- [DBD]-[C’.

[0139] In some embodiments, a connecting structure "]-["in any one of the epigenetic editor structures is a linker, e.g., a peptide linker; a detectable tag; a peptide bond; a nuclear localization signal; and/or a promoter or regulatory sequence. In an epigenetic editor structure, the multiple connecting structures "]-[" may be the same or may each be a different linker, tag, NLS, or peptide bond. In particular embodiments, the DNA methyltransferase domain comprises DNMT3A, DNMT3L, or both. In particular embodiments, the DBD is a catalytically inactive polynucleotide guided DNA-binding domain (e.g., a dCas9) or a ZFP domain. In particular embodiments, the repressor domain is a KRAB domain.

[0140] In some embodiments, the epigenetic editor comprises a configuration selected from

N']-[DNMT3A-DNMT3L]-[DBD]-[KRAB]-[C'
N']-[KRAB]-[DBD]-[DNMT3A-DNMT3L]-[C'
N']-[KRAB]-[DBD]-[DNMT3A]-[C'
N']-[DNMT3A]-[DBD]-[KRAB]-[C'
N']-[KRAB]-[DBD]-[DNMT3A]-[DNMT3L]-[C'
N']-[DNMT3A]- [DNMT3L]- [DBD]-[KRAB]-[C'
N']-[DNMT3A]-[DBD]-[C'
N']-[DBD]-[DNMT3A]-[C'
N']-[DNMT3L]-[DBD]-[C'
N']-[DBD]-[DNMT3L]-[C'

wherein [DNMT3A-DNMT3L] indicates that the DNMT3A and DNMT3L domains are directly fused via a peptide bond, and wherein the connecting structure]-[is any one of the linkers as described herein, a detectable tag, an affinity domain, a peptide bond, a nuclear localization signal, a promoter, and/or a regulatory sequence. The DBD, KRAB, DNMT3A, and DNMT3L domains may be any as described herein, in any combination. In particular embodiments, the DBD is a CRISPR-associated protein domain (e.g., dCas9) or a ZFP domain; the KRAB domain is derived from KOX1, ZIM3, ZFP28, or ZN627; the DNMT3A domain is a human DNMT3A domain; and the DNMT3L domain is a human or mouse DNMT3L domain; any combination of these components is also contemplated by the present disclosure.

[0141] In some embodiments, the epigenetic editor comprises a configuration selected from

N']-[DNMT3A]-[DBD]-[SETDB1]-[C'
N']-[DNMT3A]- [DNMT3L]- [DBD]-[SETDB1]-[C'
N']- [DNMT3A-DNMT3L]- [DBD]- [SETDB1]- [C'
N']-[SETDB1]-[DBD]-[DNMT3A]-[DNMT3L]-[C'
N']-[SETDB1]-[DBD]-[DNMT3A]-[C'

wherein [DNMT3A-DNMT3L] indicates that the DNMT3A and DNMT3L domains are directly fused via a peptide bond, and wherein the connecting structure]-[is any one of the linkers as described herein, a detectable tag, an affinity domain, a peptide bond, a nuclear localization signal, a promoter, and/or a regulatory sequence. The DBD, SETDB1, DNMT3A, and DNMT3L domains may be any as described herein, in any combination. In particular embodiments, the DBD is a CRISPR-associated protein domain (e.g., dCas9) or a ZFP domain; the SETDB1 domain is derived from human SETDB1, ZIM3, ZFP28, or ZN627; the DNMT3A domain is a human DNMT3A domain; and the DNMT3L domain is a human or mouse DNMT3L domain; any combination of these components is also contemplated by the present disclosure.

[0142] Particular constructs contemplated herein include:

DNMT3A-DNMT3L-XTEN80-NLS-**dCas9**-NLS-XTEN16-KOX1 KRAB (Configuration 1), and

DNMT3A-DNMT3L-XTEN80-NLS-**ZFP domain**-NLS-XTEN16-KOX1 KRAB (Configuration 2).

In particular embodiments, the DNMT3L and DNMT3A are both derived from human parental proteins. In particular embodiments, the DNMT3L and DNMT3A are derived from human and mouse parental proteins, respectively. In particular embodiments, the DNMT3L and DNMT3A are derived from mouse and human parental proteins, respectively. In particular embodiments, the DNMT3L and DNMT3A are both derived from mouse parental proteins. In some embodiments, the dCas9 is dSpCas9. In some embodiments, the KOX1 is human KOX1.

[0143] In particular embodiments, a fusion construct described herein may have Configuration 1 and comprise SEQ ID NO: 1080, or a sequence at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical thereto. In SEQ ID NO: 1080 below, the XTEN linkers are underlined, the NLS sequences are **bolded**, the DNMT3A sequence is *italicized*, the DNMT3L sequence is *underlined and italicized*, the dCas9 domain is ***bolded and italicized***, and the KOX1 KRAB domain is **underlined and bolded**:

MNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDS
ITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTG
RLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSA
AHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQ
HFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPL
KEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFR
NIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDR
CPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVR
GRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLR

*EYFKYFSQNSLPL*GGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS
APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE**PKKKRKV**Y***MDKKYSIGLAIGTNSVG***
WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
NRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ
LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFK
SNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE
FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY
PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAI
VDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL
INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI
EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD
FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL
TNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD**PKKKRK**
VSGSETPGTSESATPESTG**RTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNL**
VSLGYQLTKPDVILRLEKGEEP (SEQ ID NO: 1080)

[0144] In particular embodiments, a fusion construct described herein may have Configuration 2 and comprise SEQ ID NO: 1081, or a sequence at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical thereto. In SEQ ID NO: 1081 below, the XTEN linkers are underlined, the NLS sequences are **bolded and underlined**, the DNMT3A sequence is *italicized*, the DNMT3L sequence is *underlined and italicized*, the ZFP domain is **bolded**, and the KOX1 KRAB domain is **underlined and bolded**. Variable amino acids represented by Xs are the amino acids of the DNA-recognition helix of the zinc finger and ***XX*** in italics may be either TR, LR or LK.

MNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDS
ITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTG
RLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSA
AHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQ
HFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPL
KEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMEIYKTVSAWKRQPVRVLSLFR
NIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDR
CPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVR
GRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLR

*EYFKYFSQNSLPL*GGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE**PKKKRKVYSRPGERPFQCRICMRNFSXXXXXXXH*XX*THTGEKPFQCRICMRNFSXXXXXXXH*XX*TH[linker]PFQCRICMRNFSXXXXXXXH*XX*THTGEKPFQCRICMRNFSXXXXXXXH*XX*TH[linker]PFQCRICMRNFSXXXXXXXH*XX*THTGEKPFQCRICMRNFSXXXXXXXH*XX*THLRGSPKKKRKV**SGSETPGTSESATPESTG**RTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP** (SEQ ID NO: 1081)

In certain embodiments, the six "XXXXXXX" regions in SEQ ID NO: 1081 comprise, in order, the F1-F6 amino acid sequences shown in **Table 1**. [linker] represents a linker sequence. In some embodiments, one or both linker sequences may be TGSQKP (SEQ ID NO: 1085). In some embodiments, one or both linker sequences may be TGGGGSQKP (SEQ ID NO: 1086). In some embodiments, one linker sequence may have the amino acid sequence of SEQ ID NO: 1085 and the other linker sequence may have the amino acid sequence of SEQ ID NO: 1086.

[0145] Multiple epigenetic editors may be used to effect activation or repression of a target gene or multiple target genes. For example, an epigenetic editor fusion protein comprising a DNA-binding domain (e.g., a dCas9 domain) and an effector domain may be co-delivered with two or more guide polynucleotides (e.g., gRNAs), each targeting a different target DNA sequence. The target sites for two of the DNA-binding domains may be the same or in the vicinity of each other, or separated by, for example, about 100 base pairs, about 200 base pairs, about 300 base pairs, about 400 base pairs, about 500 base pairs, or about 600 or more base pairs. In addition, when targeting double-strand DNA, such as an endogenous gene locus, the guide polynucleotides may target the same or different strands (one or more to the positive strand and/or one or more to the negative strand).

V. Target Sequences

[0146] An epigenetic editor herein may be directed to an HBV target sequence to effect epigenetic modification of HBV or an HBV gene. As used herein, a "target sequence," a "target site," or a "target region" is a nucleic acid sequence present in a genome or gene of interest, e.g., in an HBV genome or an HBV gene; in some instances, the target sequence may be outside but in the vicinity of the gene of interest wherein methylation or binding by a repressor of the target sequence represses expression of the gene. In some embodiments, the target sequence may be a hypomethylated or hypermethylated nucleic acid sequence.

[0147] The target sequence may be in any part of a target gene. In some embodiments, the target sequence is part of or near a noncoding sequence of the gene. In some embodiments, the target sequence is part of an exon of the gene. In some embodiments, the target sequence is part of or near a transcriptional regulatory sequence of the gene, such as a promoter or an enhancer. In some embodiments, the target sequence is adjacent to, overlaps with, or encompasses a CpG island, e.g., a CpG island identified within the HBV genome. In some embodiments, the target sequence is outside of a CpG island. In certain embodiments, the target sequence is within about 3000, 2900, 2800, 2700, 2600, 2500, 2400, 2300, 2200, 2100, 2000, 1900, 1800, 1700, 1600, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700, 600, 500, 400, 300, 200, or 100 base pairs (bp) flanking an HBV TSS. In certain embodiments, the target sequence is within 500 bp flanking the *HBV* TSS. In certain embodiments, the target sequence is within 1000 bp flanking the *HBV* TSS.

[0148] In some embodiments, the target sequence may hybridize to a guide polynucleotide sequence (e.g., gRNA) complexed with a fusion protein comprising a polynucleotide guided DNA-binding domain (e.g., a CRISPR protein such as dCas9) and effector domain(s). The guide polynucleotide sequence may be designed to have complementarity to the target sequence, or identity to the opposing strand of the target sequence. In some embodiments, the guide polynucleotide comprises a spacer sequence that is about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to a protospacer sequence in the target sequence. In particular embodiments, the guide polynucleotide comprises a spacer sequence that is 100% identical to a protospacer sequence in the target sequence.

[0149] In some embodiments, where the DNA-binding domain of an epigenetic editor described herein is a zinc finger array, the target sequence may be recognized by said zinc finger array.

[0150] In some embodiments, where the DNA-binding domain of an epigenetic editor described herein is a TALE, the target sequence may be recognized by said TALE.

[0151] A target sequence described herein may be specific to one genotype of HBV, to one copy of am HBV target gene, or may be specific to one allele of an HBV target gene. In some embodiments, however, the target sequence may be conserved across two or more HBV genotypes, across two or more copies of an HBV gene, and across alleles of an HBV gene. Accordingly, the epigenetic modification and modulation of expression thereof may be specific to one copy or one allele of the target gene, or, in other embodiments, may be universal to different HBV genotypes, or HBV gene copies or alleles

[0152] In some embodiments, the target sequence is comprised in the following sequence:

```
>NC_003977.2 Hepatitis B virus (strain ayw) genome

AATTCCACAACCTTCCACCAAACTCTGCAAGATCCCAGAGTGAGAGGCCTGTATTTCCCTGCTGGTGGCTCCAGTTC
AGGAACAGTAAACCCTGTTCTGACTACTGCCTCTCCCTTATCGTCAATCTTCTCGAGGATTGGGGACCCTGCGCTGA
ACATGGAGAACATCACATCAGGATTCCTAGGACCCCTTCTCGTGTTACAGGCGGGGTTTTTCTTGTTGACAAGAATC
CTCACAATACCGCAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGGGGGAACTACCGTGTGTCTTGGCCA
```

AAATTCGCAGTCCCCAACCTCCAATCACTCACCAACCTCTTGTCCTCCAACTTGTCCTGGTTATCGCTGGATGTGTC
TGCGGCGTTTTATCATCTTCCTCTTCATCCTGCTGCTATGCCTCATCTTCTTGTTGGTTCTTCTGGACTATCAAGGT
ATGTTGCCCGTTTGTCCTCTAATTCCAGGATCCTCAACAACCAGCACGGGACCATGCCGGACCTGCATGACTACTGC
TCAAGGAACCTCTATGTATCCCTCCTGTTGCTGTACCAAACCTTCGGACGGAAATTGCACCTGTATTCCCATCCCAT
CATCCTGGGCTTTCGGAAAATTCCTATGGGAGTGGGCCTCAGCCCGTTTCTCCTGGCTCAGTTTACTAGTGCCATTT
GTTCAGTGGTTCGTAGGGCTTTCCCCCACTGTTTGGCTTTCAGTTATATGGATGATGTGGTATTGGGGGCCAAGTCT
GTACAGCATCTTGAGTCCCTTTTTACCGCTGTTACCAATTTTCTTTTGTCTTTGGGTATACATTTAAACCCTAACAA
AACAAAGAGATGGGGTTACTCTCTAAATTTTATGGGTTATGTCATTGGATGTTATGGGTCCTTGCCACAAGAACACA
TCATACAAAAAATCAAAGAATGTTTTAGAAAACTTCCTATTAACAGGCCTATTGATTGGAAAGTATGTCAACGAATT
GTGGGTCTTTTGGGTTTTGCTGCCCCTTTTACACAATGTGGTTATCCTGCGTTGATGCCTTTGTATGCATGTATTCA
ATCTAAGCAGGCTTTCACTTTCTCGCCAACTTACAAGGCCTTTCTGTGTAAACAATACCTGAACCTTTACCCCGTTG
CCCGGCAACGGCCAGGTCTGTGCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGTCATGGGCCATCAG
CGCATGCGTGGAACCTTTTCGGCTCCTCTGCCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTTGCTCGCAGCAG
GTCTGGAGCAAACATTATCGGGACTGATAACTCTGTTGTCCTATCCCGCAAATATACATCGTTTCCATGGCTGCTAG
GCTGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTTTACGTCCCGTCGGCGCTGAATCCTGCGGACGACCCT
TCTCGGGGTCGCTTGGGACTCTCTCGTCCCCTTCTCCGTCTGCCGTTCCGACCGACCACGGGGCGCACCTCTCTTTA
CGCGGACTCCCCGTCTGTGCCTTCTCATCTGCCGGACCGTGTGCACTTCGCTTCACCTCTGCACGTCGCATGGAGAC
CACCGTGAACGCCCACCAAATATTGCCCAAGGTCTTACATAAGAGGACTCTTGGACTCTCAGCAATGTCAACGACCG
ACCTTGAGGCATACTTCAAAGACTGTTTGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGGTTAAAGGTCTTT
GTACTAGGAGGCTGTAGGCATAAATTGGTCTGCGCACCAGCACCATGCAACTTTTTCACCTCTGCCTAATCATCTCT
TGTTCATGTCCTACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGGGCATGGACATCGACCCTTATAAAGA
ATTTGGAGCTACTGTGGAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTACGAGATCTTCTAGATACCG
CCTCAGCTCTGTATCGGGAAGCCTTAGAGTCTCCTGAGCATTGTTCACCTCACCATACTGCACTCAGGCAAGCAATT
CTTTGCTGGGGGGAACTAATGACTCTAGCTACCTGGGTGGGTGTTAATTTGGAAGATCCAGCGTCTAGAGACCTAGT
AGTCAGTTATGTCAACACTAATATGGGCCTAAAGTTCAGGCAACTCTTGTGGTTTCACATTTCTTGTCTCACTTTTG
GAAGAGAAACAGTTATAGAGTATTTGGTGTCTTTCGGAGTGTGGATTCGCACTCCTCCAGCTTATAGACCACCAAAT
GCCCCTATCCTATCAACACTTCCGGAGACTACTGTTGTTAGACGACGAGGCAGGTCCCCTAGAAGAAGAACTCCCTC
GCCTCGCAGACGAAGGTCTCAATCGCCGCGTCGCAGAAGATCTCAATCTCGGGAATCTCAATGTTAGTATTCCTTGG
ACTCATAAGGTGGGGAACTTTACTGGGCTTTATTCTTCTACTGTACCTGTCTTTAATCCTCATTGGAAAACACCATC
TTTTCCTAATATACATTTACACCAAGACATTATCAAAAAATGTGAACAGTTTGTAGGCCCACTCACAGTTAATGAGA
AAAGAAGATTGCAATTGATTATGCCTGCCAGGTTTTATCCAAAGGTTACCAAATATTTACCATTGGATAAGGGTATT
AAACCTTATTATCCAGAACATCTAGTTAATCATTACTTCCAAACTAGACACTATTTACACACTCTATGGAAGGCGGG
TATATTATATAAGAGAGAAACAACACATAGCGCCTCATTTTGTGGGTCACCATATTCTTGGGAACAAGATCTACAGC
ATGGGGCAGAATCTTTCCACCAGCAATCCTCTGGGATTCTTTCCCGACCACCAGTTGGATCCAGCCTTCAGAGCAAA
CACCGCAAATCCAGATTGGGACTTCAATCCCAACAAGGACACCTGGCCAGACGCCAACAAGGTAGGAGCTGGAGCAT
TCGGGCTGGGTTTCACCCCACCGCACGGAGGCCTTTTGGGGTGGAGCCCTCAGGCTCAGGGCATACTACAAACTTTG
CCAGCAAATCCGCCTCCTGCCTCCACCAATCGCCAGTCAGGAAGGCAGCCTACCCCGCTGTCTCCACCTTTGAGAAA
CACTCATCCTCAGGCCATGCAGTGG (SEQ ID No. 1082)

[0153] In some embodiments, the target sequence is comprised in the following sequence:

>U95551.1 Hepatitis B virus subtype ayw, complete genome

AATTCCACAACCTTTCACCAAACTCTGCAAGATCCCAGAGTGAGAGGCCTGTATTTCCCTGCTGGTGGCTCCAGTTC
AGGAGCAGTAAACCCTGTTCCGACTACTGCCTCTCCCTTATCGTCAATCTTCTCGAGGATTGGGGACCCTGCGCTGA
ACATGGAGAACATCACATCAGGATTCCTAGGACCCCTTCTCGTGTTACAGGCGGGGTTTTTCTTGTTGACAAGAATC
CTCACAATACCGCAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGGGGGAACTACCGTGTGTCTTGGCCA
AAATTCGCAGTCCCCAACCTCCAATCACTCACCAACCTCCTGTCCTCCAACTTGTCCTGGTTATCGCTGGATGTGTC
TGCGGCGTTTTATCATCTTCCTCTTCATCCTGCTGCTATGCCTCATCTTCTTGTTGGTTCTTCTGGACTATCAAGGT
ATGTTGCCCGTTTGTCCTCTAATTCCAGGATCCTCAACCACCAGCACGGGACCATGCCGAACCTGCATGACTACTGC
TCAAGGAACCTCTATGTATCCCTCCTGTTGCTGTACCAAACCTTCGGACGGAAATTGCACCTGTATTCCCATCCCAT
CATCCTGGGCTTTCGGAAAATTCCTATGGGAGTGGGCCTCAGCCCGTTTCTCCTGGCTCAGTTTACTAGTGCCATTT
GTTCAGTGGTTCGTAGGGCTTTCCCCCACTGTTTGGCTTTCAGTTATATGGATGATGTGGTATTGGGGGCCAAGTCT
GTACAGCATCTTGAGTCCCTTTTTACCGCTGTTACCAATTTTCTTTTGTCTTTGGGTATACATTTAAACCCTAACAA
AACAAAGAGATGGGGTTACTCTCTGAATTTTATGGGTTATGTCATTGGAAGTTATGGGTCCTTGCCACAAGAACACA
TCATACAAAAAATCAAAGAATGTTTTAGAAAACTTCCTATTAACAGGCCTATTGATTGGAAAGTATGTCAACGAATT
GTGGGTCTTTTGGGTTTTGCTGCCCCATTTACACAATGTGGTTATCCTGCGTTAATGCCCTTGTATGCATGTATTCA
ATCTAAGCAGGCTTTCACTTTCTCGCCAACTTACAAGGCCTTTCTGTGTAAACAATACCTGAACCTTTACCCCGTTG
CCCGGCAACGGCCAGGTCTGTGCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGTCATGGGCCATCAG
CGCGTGCGTGGAACCTTTTCGGCTCCTCTGCCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTTGCTCGCAGCAG
GTCTGGAGCAAACATTATCGGGACTGATAACTCTGTTGTCCTCTCCCGCAAATATACATCGTATCCATGGCTGCTAG

GCTGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTTTACGTCCCGTCGGCGCTGAATCCTGCGGACGACCCT
TCTCGGGGTCGCTTGGGACTCTCTCGTCCCCTTCTCCGTCTGCCGTTCCGACCGACCACGGGGCGCACCTCTCTTTA
CGCGGACTCCCCGTCTGTGCCTTCTCATCTGCCGGACCGTGTGCACTTCGCTTCACCTCTGCACGTCGCATGGAGAC
CACCGTGAACGCCCACCGAATGTTGCCCAAGGTCTTACATAAGAGGACTCTTGGACTCTCTGCAATGTCAACGACCG
ACCTTGAGGCATACTTCAAAGACTGTTTGTTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTAGATTAAAGGTCTTT
GTACTAGGAGGCTGTAGGCATAAATTGGTCTGCGCACCAGCACCATGCAACTTTTTCACCTCTGCCTAATCATCTCT
TGTTCATGTCCTACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGGGCATGGACATCGACCCTTATAAAGA
ATTTGGAGCTACTGTGGAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCCTTCAGTACGAGATCTTCTAGATACCG
CCTCAGCTCTGTATCGGGAAGCCTTAGAGTCTCCTGAGCATTGTTCACCTCACCATACTGCACTCAGGCAAGCAATT
CTTTGCTGGGGGGAACTAATGACTCTAGCTACCTGGGTGGGTGTTAATTTGGAAGATCCAGCATCTAGAGACCTAGT
AGTCAGTTATGTCAACACTAATATGGGCCTAAAGTTCAGGCAACTCTTGTGGTTTCACATTTCTTGTCTCACTTTTG
GAAGAGAAACCGTTATAGAGTATTTGGTGTCTTTCGGAGTGTGGATTCGCACTCCTCCAGCTTATAGACCACCAAAT
GCCCCTATCCTATCAACACTTCCGGAAACTACTGTTGTTAGACGACGAGGCAGGTCCCCTAGAAGAAGAACTCCCTC
GCCTCGCAGACGAAGGTCTCAATCGCCGCGTCGCAGAAGATCTCAATCTCGGGAACCTCAATGTTAGTATTCCTTGG
ACTCATAAGGTGGGGAACTTTACTGGTCTTTATTCTTCTACTGTACCTGTCTTTAATCCTCATTGGAAAACACCATC
TTTTCCTAATATACATTTACACCAAGACATTATCAAAAAATGTGAACAGTTTGTAGGCCCACTTACAGTTAATGAGA
AAAGAAGATTGCAATTGATTATGCCTGCTAGGTTTTATCCAAAGGTTACCAAATATTTACCATTGGATAAGGGTATT
AAACCTTATTATCCAGAACATCTAGTTAATCATTACTTCCAAACTAGACACTATTTACACACTCTATGGAAGGCGGG
TATATTATATAAGAGAGAAACAACACATAGCGCCTCATTTTGTGGGTCACCATATTCTTGGGAACAAGATCTACAGC
ATGGGGCAGAATCTTTCCACCAGCAATCCTCTGGGATTCTTTCCCGACCACCAGTTGGATCCAGCCTTCAGAGCAAA
CACAGCAAATCCAGATTGGGACTTCAATCCCAACAAGGACACCTGGCCAGACGCCAACAAGGTAGGAGCTGGAGCAT
TCGGGCTGGGTTTCACCCCACCGCACGGAGGCCTTTTGGGGTGGAGCCCTCAGGCTCAGGGCATACTACAAACTTTG
CCAGCAAATCCGCCTCCTGCCTCCACCAATCGCCAGACAGGAAGGCAGCCTACCCCGCTGTCTCCACCTTTGAGAAA
CACTCATCCTCAGGCCATGCAGTGG (SEQ ID No. 1083)

VI. Epigenetic Modifications

[0154] An epigenetic editor described herein may perform sequence-specific epigenetic modification(s) (e.g., alteration of chemical modification(s)) of a target gene that harbors the target sequence. Such epigenetic modulation may be safer and more easily reversible than modulation due to gene editing, e.g., with generation of DNA double-strand breaks. In some embodiments, the epigenetic modulation may reduce or silence the target gene. In some embodiments, the modification is at a specific site of the target sequence. In some embodiments, the modification is at a specific allele of the target gene. Accordingly, the epigenetic modification may result in modulated (e.g., reduced) expression of one copy of a target gene harboring a specific allele, and not the other copy of the target gene. In some embodiments, the specific allele is associated with a disease, condition, or disorder.

[0155] In some embodiments, the epigenetic modification reduces or abolishes transcription of the target gene harboring the target sequence. In some embodiments, the epigenetic modification reduces or abolishes transcription of a copy of the target gene harboring a specific allele recognized by the epigenetic editor. In some embodiments, the epigenetic editor reduces the level of or eliminates expression of a protein encoded by the target gene. In some embodiments, the epigenetic editor reduces the level of or eliminates expression of a protein encoded by a copy of the target gene harboring a specific allele recognized by the epigenetic editor. The target *HBV* gene may be epigenetically modified *in vitro*, *ex vivo*, or *in vivo*.

[0156] The effector domain of an epigenetic editor described herein may alter (e.g., deposit or remove) a chemical modification at a nucleotide of the target gene or at a histone associated with the target gene. The chemical modification may be altered at a single nucleotide or a single histone, or may be altered at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000 or more nucleotides.

[0157] In some embodiments, an effector domain of an epigenetic editor described herein may alter a CpG dinucleotide within the target gene. In some embodiments, all CpG dinucleotides within 2000, 1500, 1000, 500, or 200 bps flanking a target sequence (e.g., in an alteration site as described herein) are altered according to a modification type described herein, as compared to the original state of the gene or the gene in a comparable cell not contacted with the epigenetic editor. In some embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700 or more of the CpG dinucleotides are altered as compared to the original state of the gene or the gene in a comparable cell not contacted with the epigenetic editor. In some embodiments, at least 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% of the CpG dinucleotides are altered as compared to the original state of the gene or the gene in a comparable cell not contacted with the epigenetic editor. In some embodiments, one single CpG dinucleotide is altered, as compared to the original state of the gene or the gene in a comparable cell not contacted with the epigenetic editor.

[0158] An effector domain of an epigenetic editor described herein may alter a histone modification state of a histone associated with or bound to the target gene. For example, an effector domain may deposit a modification on one or more lysine residues of histone tails of histones associated with the target gene. In some embodiments, the effector domain may result in deacetylation of one or more histone tails of histones associated with the target gene, thereby reducing or silencing expression of the target gene. In some embodiments, the histone modification state is a methylation state. For example, the effector domain may result in a H3K9, H3K27 or H4K20 methylation (e.g. one or more of a H3K9me2, H3K9me3, H3K27me2, H3K27me3, and H4K20me3 methylation) at one or more histone tails associated with the target gene, thereby reducing or silencing expression of the target gene.

[0159] In some embodiments, all histone tails of histones bound to DNA nucleotides within 2000, 1500, 1000, 500, or 200 bps flanking the target sequence are altered according to a modification type as described herein, as compared to the original state of the chromosome or the chromosome in a comparable cell not contacted with the epigenetic editor. In some embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120 or more histone tails of the bound histones are altered as compared to the original state of the chromosome or the chromosome in a comparable cell not contacted with the epigenetic editor. In some embodiments, at least 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% of histone tails of the bound histones are altered as compared to the original state of the chromosome or the chromosome in a comparable cell not contacted with the epigenetic editor. For example, one single histone tail of the bound histones may be altered as compared to the original state of the chromosome or the chromosome in a comparable cell not contacted with the epigenetic editor. As another example, one single bound histone octamer may be altered as compared to the original state of the chromosome or the chromosome in a comparable cell not contacted with the epigenetic editor.

[0160] The chemical modification deposited at target gene DNA nucleotides or histone residues may be at or in close proximity to a target sequence in the target gene. In some embodiments, an effector domain of an epigenetic editor described herein alters a chemical modification state of a nucleotide or histone tail bound to a nucleotide 100-200, 200-300, 300-400, 400-55, 500-600, 600-700, or 700-800 nucleotides 5’ or 3’ to the target sequence in the target gene. In some embodiments, an effector domain alters a chemical modification state of a nucleotide or histone tail bound to a nucleotide within 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, or 2000 nucleotides flanking the target sequence. As used herein, “flanking” refers to nucleotide positions 5’ to the 5’ end of and 3’ to the 3’ end of a particular sequence, e.g. a target sequence.

[0161] In some embodiments, an effector domain mediates or induces a chemical modification change of a nucleotide or a histone tail bound to a nucleotide distant from a target sequence. Such modification may be initiated near the target sequence, and may subsequently spread to one or more nucleotides in the target gene distant from the target sequence. For example, an effector domain may initiate alteration of a chemical modification state of one or more nucleotides or one or more histone residues bound to one or more nucleotides within 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 nucleotides flanking the target sequence, and the chemical modification state alteration may spread to one or more nucleotides at least 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2500, 3000, or more nucleotides from the target sequence in the target gene, either upstream or downstream of the target sequence. In certain embodiments, the chemical modification may be initiated at less than 2, 3, 5, 10, 20, 30, 40, 50, or 100 nucleotides in the target gene and spread to at least 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, or more nucleotides in the target gene. In some embodiments, the chemical modification spreads to nucleotides in the entire target gene. Additional proteins or transcription factors, for example, transcription repressors, methyltransferases, or transcription regulation scaffold proteins, may be involved in the spreading of the chemical modification. Alternatively, the epigenetic editor alone may be involved.

[0162] In some embodiments, an epigenetic editor described herein reduces expression of a target gene by at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 99%, or more, as measured by transcription of the target gene in a cell, a tissue, or a subject as compared to a control cell, control tissue, or a control subject (e.g., in the absence of the epigenetic editor). In some embodiments, the epigenetic editors described herein reduces expression of a copy of target gene by at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 99%, or more, as measured by transcription of the copy of the target gene in a cell, a tissue, or a subject as compared to a control cell, control tissue, or a control subject. In certain embodiments, the copy of the target gene harbors a specific sequence or allele recognized by the epigenetic editor. In particular embodiments, the epigenetically modified copy encodes a functional protein, and accordingly an epigenetic editor disclosed herein may reduce or abolish expression and/or function of the protein. For example, an epigenetic editor described herein may reduce expression and/or function of a protein encoded by the target gene by at least 3-fold, at least 4-fold, at least 5-fold, at least 6-fold, at least 7-fold, at least 8-fold, at least 9-fold, at least 10-fold, at least 11-fold, at least 12-fold, at least 13-fold, at least 14-fold, at least 15-fold, at least 20-fold, at least 25-fold, at least 30-fold, at least 35-fold, at least 40-fold, at least 45-fold, at least 50-fold, at least 60-fold, at least 70-fold, at least 80-fold, at least 90-fold, or at least 100 fold in a cell, a tissue, or a subject as compared to a control cell, control tissue, or a control subject.

[0163] Modulation of target gene expression can be assayed by determining any parameter that is indirectly or directly affected by the expression of the target gene. Such parameters include, e.g., changes in RNA or protein levels; changes in protein activity; changes in product levels; changes in downstream gene expression; changes in transcription or activity of reporter genes such as, for example, luciferase, CAT, beta-galactosidase, or GFP; changes in signal transduction; changes in phosphorylation and dephosphorylation; changes in receptor-ligand interactions; changes in concentrations of second messengers such as, for example, cGMP, cAMP, IP3, and $Ca2^+$; changes in cell growth; changes in neovascularization; and/or changes in any functional effect of gene expression. Measurements can be made in vitro, in vivo, and/or ex vivo, and can be made by conventional methods, e.g., measurement of RNA or protein levels, measurement of RNA stability, and/or identification of downstream or reporter gene expression. Readout can be by way of, for example, chemiluminescence, fluorescence, colorimetric reactions, antibody binding, inducible markers, ligand binding assays, changes in intracellular second messengers such as cGMP and inositol triphosphate (IP3), changes in intracellular calcium levels; cytokine release, and the like.

[0164] Methods for determining the expression level of a gene, for example the target of an epigenetic editor, may include, e.g., determining the transcript level of a gene by reverse transcription PCR, quantitative RT-PCR, droplet digital PCR (ddPCR), Northern blot, RNA sequencing, DNA sequencing (e.g., sequencing of complementary deoxyribonucleic acid (cDNA) obtained from RNA); next generation (Next-Gen) sequencing, nanopore sequencing, pyrosequencing, or Nanostring sequencing. Levels of protein expressed from a gene may be determined, e.g., by Western blotting, enzyme linked immuno-absorbance assays, mass-spectrometry, immunohistochemistry, or flow cytometry analysis. Gene expression product levels may be normalized to an internal standard such as total messenger ribonucleic acid (mRNA) or the expression level of a particular gene, e.g., a housekeeping gene.

[0165] In some embodiments, the effect of an epigenetic editor in modulating target gene expression may be examined using a reporter system. For example, an epigenetic editor may be designed to target a reporter gene encoding a reporter protein, such as a fluorescent protein. Expression of the reporter gene in such a model system may be monitored by, e.g., flow cytometry, fluorescence-activated cell sorting (FACS), or fluorescence microscopy. In some embodiments, a population of cells may be transfected with a vector that harbors a reporter gene. The vector may be constructed such that the reporter gene is expressed when the vector transfects a cell. Suitable reporter genes include genes encoding fluorescent proteins, for example green, yellow, cherry, cyan or orange fluorescent proteins. The population of cells carrying the reporter system may be transfected with DNA, mRNA, or vectors encoding the epigenetic editor targeting the reporter gene.

VII. Pharmaceutical Compositions

[0166] Another aspect of the present disclosure is a pharmaceutical composition comprising as an active ingredient (or as the sole active ingredient) one or more epigenetic editors described herein or component(s) (e.g., fusion proteins and/or guide polynucleotides) thereof, or nucleic acid molecule(s) encoding said epigenetic editors or component(s) thereof. For example, a pharmaceutical composition may comprise nucleic acid molecule(s) encoding the fusion protein(s) (and guide polynucleotides, where applicable) of an epigenetic editor described herein. In some embodiments, separate pharmaceutical compositions comprise the fusion protein(s) and the guide polynucleotide(s). In some embodiments, multiple pharmaceutical compositions, each comprising one epigenetic editor, are administered simultaneously. A pharmaceutical composition may also comprise cells that have undergone epigenetic modification(s) mediated or induced by an epigenetic editor provided herein.

[0167] Generally, the epigenetic editors described herein or component(s) thereof, or nucleic acid molecule(s) encoding said epigenetic editors or component(s) thereof, of the present disclosure are suitable to be administered as a formulation in association with one or more pharmaceutically acceptable excipient(s), e.g., as described below.

[0168] The term "excipient" is used herein to describe any ingredient other than the compound(s) of the present disclosure. The choice of excipient(s) will to a large extent depend on factors such as the particular mode of administration, the effect of the excipient on solubility and stability, and the nature of the dosage form. As used herein, "pharmaceutically acceptable excipient" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like that are physiologically compatible. Some examples of pharmaceutically acceptable excipients are water, saline, phosphate buffered saline, dextrose, glycerol, ethanol and the like, as well as combinations thereof. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, or sodium chloride in the composition. Additional examples of pharmaceutically acceptable substances are wetting agents or minor amounts of auxiliary substances such as wetting or emulsifying agents, preservatives, or buffers, which enhance the shelf life or effectiveness of the antibody.

[0169] Formulations of a pharmaceutical composition suitable for parenteral administration typically comprise the active ingredient combined with a pharmaceutically acceptable carrier, such as sterile water or sterile isotonic saline. Such formulations may be prepared, packaged, or sold in a form suitable for bolus administration or for continuous administration. In some embodiments, the epigenetic editor or its component(s) are introduced to target cells in the form of nucleic acid molecule(s) encoding the epigenetic editor or its component(s); accordingly, the pharmaceutical compositions herein comprise the nucleic acid molecule(s). Such nucleic acid molecule(s) may be, for example, DNA, RNA or mRNA, and/or modified nucleic acid sequence(s) (e.g., with chemical modifications, a 5' cap, or one or more 3' modifications). In some embodiments, the nucleic acid molecule(s) may be delivered as naked DNA or RNA, for instance by means of transfection or electroporation, or can be conjugated to molecules (e.g., N-acetylgalactosamine) promoting uptake by target cells. In some embodiments, the nucleic acid molecule(s) may be in nucleic acid expression vector(s), which may include expression control sequences such as promoters, enhancers, transcription signal sequences, transcription termination sequences, introns, polyadenylation signals, Kozak consensus sequences, internal ribosome entry sites (IRES), etc. Such expression control sequences are well known in the art. A vector may also comprise a sequence encoding a signal peptide (e.g., for nuclear localization, nucleolar localization, or mitochondrial localization), associated with (e.g., inserted into or fused to) a sequence coding for a protein.

[0170] Examples of vectors include, but are not limited to, plasmid vectors; viral vectors based on vaccinia virus, poliovirus, adenovirus, adeno-associated virus, SV40, herpes simplex virus, human immunodeficiency virus, retrovirus (e.g., Murine Leukemia Virus, or spleen necrosis virus, vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus); and other recombinant vectors. In certain embodiments, the vector is a plasmid or a viral vector. Viral particles may also be used to deliver nucleic acid molecule(s) encoding epigenetic editors or component(s) thereof as described herein. For example, "empty" viral particles can be assembled to contain any suitable cargo. Viral vectors and viral particles may also be engineered to incorporate targeting ligands to alter target tissue specificity.

[0171] In certain embodiments, an epigenetic editor as described herein or component(s) thereof are encoded by nucleic acid sequence(s) present in one or more viral vectors, or a suitable capsid protein of any viral vector. Examples of viral vectors include adeno-associated viral vectors (e.g., derived from AAV3, AAV3b, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh8, AAV10, and/or variants thereof); retroviral vectors (e.g., Maloney murine leukemia virus, MML-V), adenoviral vectors (e.g., AD100), lentiviral vectors (e.g., HIV and FIV-based vectors), and herpesvirus vectors (e.g., HSV-2).

[0172] In some embodiments, delivery involves an adeno-associated virus (AAV) vector. AAV vector delivery may be particularly useful where the DNA-binding domain of an epigenetic editor fusion protein is a zinc finger array. Without wishing to be bound by any theory, the smaller size of zinc finger arrays compared to larger DNA-binding domains such as Cas protein domains may allow such a fusion protein to be conveniently packed in viral vectors such as an AAV vector.

[0173] Any AAV serotype, e.g., human AAV serotype, can be used for an AAV vector as described herein, including, but not limited to, AAV serotype 1 (AAV1), AAV serotype 2 (AAV2), AAV serotype 3 (AAV3), AAV serotype 4 (AAV4), AAV serotype 5 (AAV5), AAV serotype 6 (AAV6), AAV serotype 7 (AAV7), AAV serotype 8 (AAV8), AAV serotype 9 (AAV9), AAV serotype 10 (AAV10), and AAV serotype 11 (AAV11), as well as variants thereof. In some embodiments, an AAV variant has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to a wildtype AAV. In certain embodiments, the AAV variant may be engineered such that its capsid proteins have reduced immunogenicity or enhanced transduction ability in humans. In some instances, one or more regions of at least two different AAV serotype viruses are shuffled and reassembled to generate a chimeric variant. For example, a chimeric AAV may comprise inverted terminal repeats (ITRs) that are of a heterologous serotype compared to the serotype of the capsid. The resulting chimeric AAV can have a different antigenic reactivity or recognition compared to its parental serotypes. In some embodiments, a chimeric variant of an AAV includes amino acid sequences from 2, 3, 4, 5, or more different AAV serotypes.

[0174] Non-viral systems are also contemplated for delivery as described herein. Non-viral systems include, but are not limited to, nucleic acid transfection methods including electroporation, sonoporation, calcium phosphate transfection, microinjection, DNA biolistics, lipid-mediated transfection, transfection through heat shock, compacted DNA-mediated transfection, lipofection, cationic agent-mediated transfection, and transfection with liposomes, immunoliposomes, or cationic facial amphiphiles (CFAs). In certain embodiments, one or more mRNAs encoding epigenetic editor fusion proteins as described herein may be co-electroporated with one or more guide polynucleotides (e.g., gRNAs) as described herein. One important category of non-viral nucleic acid vectors is nanoparticles, which can be organic (e.g., lipid) or inorganic (e.g., gold). For instance, organic (e.g. lipid and/or polymer) nanoparticles can be suitable for use as delivery vehicles in certain embodiments of this disclosure.

[0175] In some embodiments, delivery is accomplished using a lipid nanoparticle (LNP). LNP compositions are typically sized on the order of micrometers or smaller and may include a lipid bilayer. In some embodiments, a LNP refers to any particle that has a diameter of less than 1000 nm, 500 nm, 250 nm, 200 nm, 150 nm, 100 nm, 75 nm, 50 nm, or 25 nm. Nanoparticle compositions encompass lipid nanoparticles (LNPs), liposomes (e.g., lipid vesicles), and lipoplexes.

[0176] An LNP as described herein may be made from cationic, anionic, or neutral lipids. In some embodiments, an LNP may comprise neutral lipids, such as the fusogenic phospholipid 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) or the membrane component cholesterol, as helper lipids to enhance transfection activity and nanoparticle stability. In some embodiments, an LNP may comprise hydrophobic lipids, hydrophilic lipids, or both hydrophobic and hydrophilic lipids. Any lipid or combination of lipids that are known in the art can be used to produce an LNP. The lipids may be combined in any molar ratios to produce the LNP. In some embodiments, the LNP is a liver-targeting (e.g., preferentially or specifically targeting the liver) LNP.

[0177] LNP formulations and methods of LNP delivery that can be used will be apparent to those skilled in the art based on the present disclosure and the state of the art. Non-limiting exemplary compositions and methods can be found in Shah, R., Eldridge, D., Palombo, E., and Harding, I., Lipid Nanoparticles: Production, Characterization and Stability, Springer, 2015, ISBN-13 978-3319107103; Ziegler, S., Lipid Nanoparticles: Advances in Research and Applications, Nova Science Pub., Inc, ISBN-13 978-1536186536; Mitchell, M.J., Billingsley, M.M., Haley, R.M. et al. *Engineering precision nanoparticles for drug delivery*, Nat Rev Drug Discov 20, 101–124 (2021); Hou, X., Zaks, T., Langer, R. et al. *Lipid nanoparticles for mRNA delivery*. Nat Rev Mater 6, 1078–1094 (2021); *Lipid-Nanoparticle-Based Delivery of CRISPR/Cas9 Genome-Editing Components*, Pardis Kazemian, Si-Yue Yu, Sarah B. Thomson, Alexandra Birkenshaw, Blair R. Leavitt, and Colin J. D. Ross. Molecular Pharmaceutics 2022 19 (6), 1669-1686; Cullis PR, Hope MJ. *Lipid Nanoparticle Systems for Enabling Gene Therapies*, Mol Ther. 2017 Jul 5;25(7):1467-1475; Hatit, M.Z.C., Lokugamage, M.P., Dobrowolski, C.N. et al. *Species-dependent in vivo mRNA delivery and cellular responses to nanoparticles*, Nat. Nanotechnol. 17, 310–318 (2022); Lam, K., Schreiner, P., Leung, A., Stainton, P., Reid, S., Yaworski, E., Lutwyche, P. and Heyes, J. (2023), *Optimizing Lipid Nanoparticles for Delivery in Primates*, Adv. Mater; Dilliard, S.A., Siegwart, D.J. *Passive, active and endogenous organ-targeted lipid and polymer nanoparticles for delivery of genetic drugs*, Nat Rev Mater (2023); Kasiewicz, L.N., et.al., *Lipid nanoparticles incorporating a GalNAc ligand enable in vivo liver ANGPTL3 editing in wild-type and somatic LDLR knockout non-human primates*, bioRxiv 2021.11.08.467731, doi: https://doi.org/10.1101/2021.11.08.467731; Tombácz, I., et.al., *Highly efficient CD4+ T cell targeting and genetic recombination using engineered CD4+ cell-*

homing mRNA-LNPs, Molecular Therapy, Volume 29, Issue 11, 2021, 3293-3304; Cheng, Q., Wei, T., Farbiak, L. et al. *Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR–Cas gene editing*, Nat. Nanotechnol. 15, 313–320 (2020); Zhang, Y., et.al., *Lipids and Lipid Derivatives for RNA Delivery*, Chemical Reviews 2021 121 (20); Lam, K., et.al, Unsaturated, *Trialkyl Ionizable Lipids are Versatile Lipid-Nanoparticle Components for Therapeutic and Vaccine Applications*, Adv. Mater. 2023, 35; Han, X., Zhang, H., Butowska, K. et al. *An ionizable lipid toolbox for RNA delivery*, Nat Commun 12, 7233 (2021); US Patent No. 9,364,435; US Patent No. 8,058,069; US Patent No. 8,822,668; US Patent No. 8,492,359; US Patent No. 11,141,378; US Patent No. 9,518,272; US Patent No. 9,404,127; US Patent No. 9,006,417; US Patent No. 7,901,708; US Patent No. 9,005,654; US Patent No. 9,878,042; US Patent No. 9,682,139; US Patent No. 8,642,076; US Patent No. 9,593,077; US Patent No. 9,415,109; US Patent No. 9,701,623; US Patent No. 10,369,226; US Patent No. 9,999,673; US Patent No. 9,301,923; US Patent No. 10,342,761; US Patent No. 10,137,201; International Publication No. WO2016081029A1; each of which are incorporated herein by reference in their entirety. The ordinarily skilled artisan will be able to identify an appropriate LNP and method of delivery based on the present disclosure and the state of the art. The present disclosure is not limited in this respect.

[0178] Other methods of delivery to target cells will be known to those skilled in the art and can be used with the compositions of the present disclosure.

[0179] Any type of cell may be targeted for delivery of an epigenetic editor or component(s) thereof as described herein. For example, the cells may be eukaryotic or prokaryotic. In some embodiments, the cells are mammalian (e.g., human) cells. Human cells may include, for example, hepatocytes, biliary epithelial cells (cholangiocytes), stellate cells, Kupffer cells, and liver sinusoidal endothelial cells.

[0180] In some embodiments, an epigenetic editor described herein, or component(s) thereof, are delivered to a host cell for transient expression, e.g., via a transient expression vector. Transient expression of the epigenetic editor or its component(s) may result in prolonged or permanent epigenetic modification of the target gene. For example, the epigenetic modification may be stable for at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. 11, or 12 weeks or more; or 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 months or more, after introduction of the epigenetic editor into the host cell. The epigenetic modification may be maintained after one or more mitotic and/or meiotic events of the host cell. In particular embodiments, the epigenetic modification is maintained across generations in offspring generated or derived from the host cell.

VIII. Therapeutic Uses of Epigenetic Editors

[0181] The present disclosure also provides methods for treating or preventing a condition in a subject, comprising administering to the subject an epigenetic editor or pharmaceutical composition as described herein. The epigenetic editor may effectuate an epigenetic modification of a target polynucleotide sequence in a target gene associated with a disease, condition, or disorder in the subject, thereby modulating expression of the target gene to treat or prevent the disease, condition, or disorder. In some embodiments, the epigenetic editor reduces the expression of the target gene to an extent sufficient to achieve a desired effect, e.g., a therapeutically relevant effect such as the prevention or treatment of the disease, condition, or disorder.

[0182] In some embodiments, a subject is administered a system for modulating (e.g., repressing) expression of HBV or of an HBV gene, wherein the system comprises (1) the fusion protein(s) and, where relevant, guide polynucleotide(s) of an epigenetic editor as described herein, or (2) nucleic acid molecules encoding said fusion protein(s) and, where relevant, guide polynucleotide(s).

[0183] "Treat," "treating" and "treatment" refer to a method of alleviating or abrogating a biological disorder and/or at least one of its attendant symptoms. As used herein, to "alleviate" a disease, disorder or condition means reducing the severity and/or occurrence frequency of the symptoms of the disease, disorder, or condition. Further, references herein to "treatment" include references to curative, palliative and prophylactic treatment. In some embodiments, as compared with an equivalent untreated control, alleviating a symptom may involve reduction of the symptom by at least 3%, 5%, 10%, 20%, 40%, 50%, 60%, 80%, 90%, 95%, 98%, 99%, 99.5%, 99.9%, or 100% as measured by any standard technique.

[0184] In some embodiments, the subject may be a mammal, e.g., a human. In some embodiments, the subject is selected from a non-human primate such as chimpanzee, cynomolgus monkey, or macaque, and other apes and monkey species.

[0185] In some embodiments, the human patient has a condition characterized by an HBV infection. In some embodiments, the patient has Hepatitis B.

[0186] In some embodiments, a patient to be treated with an epigenetic editor of the present disclosure has received prior treatment for the condition to be treated (e.g., HBV and/or HDV, or Hepatitis B). In other embodiments, the patient has not received such prior treatment. In some embodiments, the patient has failed on (or is refractory to) a prior treatment for the condition (*e.g.*, a prior HBV treatment).

[0187] An epigenetic editor of the present disclosure may be administered in a therapeutically effective amount to a patient with a condition described herein. "Therapeutically effective amount," as used herein, refers to an amount of the therapeutic agent being administered that will relieve to some extent one or more of the symptoms of the disorder being treated, and/or result in clinical endpoint(s) desired by healthcare professionals. An effective amount for therapy may be measured by its ability to stabilize disease progression and/or ameliorate symptoms in a patient, and preferably to reverse disease progression. The ability of an epigenetic editor of the present disclosure to reduce or silence *HBV* expression may be evaluated by *in vitro* assays, e.g., as described herein, as well as in suitable animal models that are predictive of the efficacy in humans. Suitable dosage regimens will be selected in order to provide an optimum therapeutic response in each particular situation, for example, administered as a single bolus or as a continuous infusion, and with possible adjustment of the dosage as indicated by the exigencies of each case.

[0188] An epigenetic editor of the present disclosure may be administered without additional therapeutic treatments, i.e., as a stand-alone therapy (monotherapy). Alternatively, treatment with an epigenetic editor of the present disclosure may include at least one additional therapeutic treatment (combination therapy). In some embodiments, the additional therapeutic agent is any known in the art to HBV and/or HDV. In some embodiments, therapeutic agents include, but are not limited to,antivirals such as entecavir, tenofovir, lamivudine, telvivudine, bictegravir, emtricitabine, or defovir, as well as immune modulators such as pegylated interferon and interferon alpha.

[0189] The epigenetic editors or components thereof (or nucleic acid molecules encoding the epigenetic editors or components thereof) of the present disclosure may be administered by any method accepted in the art (e.g., parenterally, intravenously, intradermally, or intramuscularly).

[0190] The epigenetic editors or components thereof (or nucleic acid molecules encoding the epigenetic editors or components thereof) of the present disclosure may be administered to a subject once, twice, three times, or 4, 5, 6, 7, 8, 9, 10, or more times. In some embodiments, the one, two, three, or 4, 5, 6, 7, 8, 9, 10, or more administrations of epigenetic editors or components thereof (or nucleic acid molecules encoding the epigenetic editors or components thereof) are in temporal proximity (e.g., within 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 4 weeks, 1 month or two months of each other). In some embodiments, a subject is re-dosed with the epigenetic editors or components thereof (or nucleic acid molecules encoding the epigenetic editors or components thereof) of the present disclosure for at least one more time after an initial dose. In some cases, a subject is administered with a subsequent dose of the epigenetic editors or components thereof (or nucleic acid molecules encoding the epigenetic editors or components thereof) of the present disclosure, which target a different DNA region of the HBV genome than the DNA region of the HBV genome that is targeted by the epigenetic editors or components thereof that the subject receives at the initial dose. In some cases, a subject is administered with multiple doses (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) of the same epigenetic editors or components thereof (or nucleic acid molecules encoding the epigenetic editors or components thereof) of the present disclosure. In some cases, a subject is administered with a single dose of different epigenetic editors or components thereof (or nucleic acid molecules encoding the epigenetic editors or components thereof) of the present disclosure, at least two of which target different DNA regions of the HBV genome. In some cases, a subject is administered with multiple doses (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) of different epigenetic editors or components thereof (or nucleic acid molecules encoding the epigenetic editors or components thereof) of the present disclosure, at least two of which target different DNA regions of the HBV genome. In some embodiments, redosing of the epigenetic editors or components thereof (or nucleic acid molecules encoding the epigenetic editors or components thereof) of the present disclosure has a better therapeutic efficacy than a single dose of the same, *e.g.*, more potent suppression of HBV replication, or more profound reduction in HBV DNA and/or HBV antigens (*e.g.*, HBsAg, HBeAg, and/or HBV core antigen (HBcAg)) present in the subject, *e.g.*, in the circulation system and/or liver of the subject.

XII. <u>Definitions</u>

[0191] The term "nucleic acid" as used herein refers to any oligonucleotide or polynucleotide containing nucleotides (e.g., deoxyribonucleotides or ribonucleotides) in either single- or double-strand form, and includes DNA and RNA. "Nucleotides" contain a sugar deoxyribose (DNA) or ribose (RNA), a base, and a phosphate group, and are linked together through the phosphate groups. "Bases" include purines and pyrimidines, which include natural compounds such as adenine, thymine, guanine, cytosine, uracil, inosine, and natural analogs; as well as synthetic derivatives of purines and pyrimidines, which include, but are not limited to, modified versions which place new reactive groups such as amines, alcohols, thiols, carboxylates, alkylhalides, etc. Nucleic acids may contain known nucleotide analogs and/or modified backbone residues or linkages, which may be synthetic, naturally occurring, and non-naturally occurring. Such nucleotide analogs, modified residues, and modified linkages are well known in the art, and may provide a nucleic acid molecule with enhanced cellular uptake, reduced immunogenicity, and/or increased stability in the presence of nucleases.

[0192] As used herein, an “isolated” or “purified” nucleic acid molecule is a nucleic acid molecule that exists apart from its native environment. For example, an “isolated” or “purified” nucleic acid molecule (1) has been separated away from the nucleic acids of the genomic DNA or cellular RNA of its source of origin; and/or (2) does not occur in nature. In some embodiments, an “isolated” or “purified” nucleic acid molecule is a recombinant nucleic acid molecule.

[0193] It will be understood that in addition to the specific proteins and nucleic acid molecules mentioned herein, the present disclosure also contemplates the use of variants, derivatives, homologs, and fragments thereof. A variant of any given sequence may have the specific sequence of residues (whether amino acid or nucleic acid residues) modified in such a manner that the polypeptide or polynucleotide in question substantially retains at least one of its endogenous functions. A variant sequence can be obtained by addition, deletion, substitution, modification, replacement and/or variation of at least one residue present in the naturally-occurring sequence (in some embodiments, no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, or 20 residues). For specific proteins described herein (e.g., KRAB, dCas9, DNMT3A, and DNMT3L proteins described herein), the present disclosure also contemplates any of the protein’s naturally occurring forms, or variants or homologs that retain at least one of its endogenous functions (e.g., at least 50%, 60%, 70%, 80%, 90%, 85%, 96%, 97%, 98%, or 99% of its function as compared to the specific protein described).

[0194] As used herein, a homologue of any polypeptide or nucleic acid sequence contemplated herein includes sequences having a certain homology with the wildtype amino acid and nucleic sequence. A homologous sequence may include a sequence, e.g. an amino acid sequence which may be at least 50%, 55%, 65%, 75%, 85%, 90%, 91%, 92%<93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the subject sequence. The term “percent identical” in the context of amino acid or nucleotide sequences refers to the percent of residues in two sequences that are the same when aligned for maximum correspondence. In some embodiments, the length of a reference sequence aligned for comparison purposes is at least 30%, (e.g., at least 40, 50, 60, 70, 80, or 90%, or 100%) of the reference sequence. Sequence identity may be measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e-3 and e-100 indicating a closely related sequence.

[0195] The percent identity of two nucleotide or polypeptide sequences is determined by, e.g., BLAST® using default parameters (available at the U.S. National Library of Medicine's National Center for Biotechnology Information website). In some embodiments, the length of a reference sequence aligned for comparison purposes is at least 30%, (e.g., at least 40, 50, 60, 70, 80, or 90%) of the reference sequence.

[0196] It will be understood that the numbering of the specific positions or residues in polypeptide sequences depends on the particular protein and numbering scheme used. Numbering might be different, e.g., in precursors of a mature protein and the mature protein itself, and differences in sequences from species to species may affect numbering. One of skill in the art will be able to identify the respective residue in any homologous protein and in the respective encoding nucleic acid by methods well known in the art, e.g., by sequence alignment and determination of homologous residues.

[0197] The term "modulate" or "alter" refers to a change in the quantity, degree, or extent of a function. For example, an epigenetic editor as described herein may modulate the activity of a promoter sequence by binding to a motif within the promoter, thereby inducing, enhancing, or suppressing transcription of a gene operatively linked to the promoter sequence. As other examples, an epigenetic editor as described herein may block RNA polymerase from transcribing a gene, or may inhibit translation of an mRNA transcript. The terms "inhibit," "repress," "suppress," "silence" and the like, when used in reference to an epigenetic editor or a component thereof as described herein, refers to decreasing or preventing the activity (e.g., transcription) of a nucleic acid sequence (e.g., a target gene) or protein relative to the activity of the nucleic acid sequence or protein in the absence of the epigenetic editor or component thereof. The term may include partially or totally blocking activity, or preventing or delaying activity. The inhibited activity may be, e.g., 10%, 20%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% less than that of a control, or may be, e.g., at least 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, or 10-fold less than that of a control.

[0198] 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, e.g., the limitations of the measurement system. For example, "about" can mean within one or more than one standard deviation, per the practice in the given value. Where particular values are described in the application and claims, unless otherwise stated, the term “about” should be assumed to mean an acceptable error range for the particular value.

[0199] Ranges provided herein are understood to be shorthand for all of the values within the range. For example, a range of 1 to 50 is understood to include any number, combination of numbers, or sub-range from the group consisting of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50, as well as all intervening decimal values between the aforementioned integers such as, for example, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, and 1.9. With respect to sub-ranges, “nested sub-ranges” that extend from either end point of the range are specifically contemplated. For example, a nested sub-range of an exemplary range of 1 to 50 may comprise 1 to 10, 1 to 20, 1 to 30, and 1 to 40 in one direction, or 50 to 40, 50 to 30, 50 to 20, and 50 to 10 in the other direction.

[0200] Unless otherwise defined herein, scientific and technical terms used in connection with the present disclosure shall have the meanings that are commonly understood by those of ordinary skill in the art. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present disclosure. In case of conflict, the present specification, including definitions, will control. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Throughout this specification and embodiments, the words “have” and “comprise,” or variations such as “has,” “having,” “comprises,” or “comprising,” will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers. The recitation of a listing of elements herein includes any of the elements singly or in any combination. The recitation of an embodiment herein includes that embodiment as a single embodiment, or in combination with any other embodiment(s) herein. All publications, patents, patent applications, and other references mentioned herein are incorporated by reference in their entirety. To the extent that references incorporated by reference contradict the disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material. Although a number of documents are cited herein, this citation does not constitute an admission that any of these documents forms part of the common general knowledge in the art.

[0201] In order that the present disclosure may be better understood, the following examples are set forth. These examples are for purposes of illustration only and are not to be construed as limiting the scope of the present disclosure in any manner.

EXAMPLES

Example 1: Selection of Target HBV Sequences for Epigenetic Silencing

[0202] Target sequences were manually and computationally designed using the representative HBV genome sequences (SEQ ID Nos. 1082, 1083) as a reference:

[0203] While target site design focused on CpG islands identified within the HBV genome, target sites outside of HBV CpG islands were also considered.

[0204] **Table 2** presents some representative target sites that were identified as suitable for targeting with an epigenetic repressor.

[0205] Target domains identified above that are adjacent to a PAM sequence, e.g., an *S. pyogenes* Cas9 PAM sequence, can be targeted by a CRISPR-based epigenetic repressor, e.g., an epigenetic repressor comprising a dCas9 DNA-binding domain. For example, target sites 1-143 are suitable for dCas9-based epigenetic repressor targeting. Figure 1 provides an overview over the position of the target sites identified in the HBV genome.

[0206] Target sites were analyzed for conservation across HBV genotypes A-E (Figures 2 and 3). Some target sites were identified that were well conserved across two or more, or in some cases all, HBV genotypes. Targeting such conserved sites allows for silencing different genotypes with the same epigenetic repressor.

Example 2: Guide RNA Assays in HepAD38 HBV cells

[0207] The HepAD38 cell line expresses the HBV genome under a doxycycline-inducible promoter (see, e.g., Ladner et al., Inducible expression of human hepatitis B virus (HBV) in stably transfected hepatoblastoma cells: a novel system for screening potential inhibitors of HBV replication. Antimicrob. Agents Chemother. 41:1715-1720(1997), incorporated herein by reference).

[0208] Results are shown in Figure 4A and B.

Example 3: Guide RNA Assays in HepG2-NTCP cells

[0209] HepG2 cells were engineered by lentiviral transduction to express the human NTCP receptor which is used by hepatitis B virus (HBV) to infect the cells.

[0210] HBV viral particles were produced using the HepAD38 cell line. HepAD38 is a subclone, derived from HepG2 cell line, that expresses HBV genome (genotype D subtype ayw) under the transcriptional control of a tetracycline-responsive promoter in a TET-OFF system.

[0211] A triple combination of Engineered Transcriptional Repressors (ETRs) consisting of three plasmids expressing dCas9-KRAB, dCas9-DNMT3A and dCas9-DNMT3L was used in combination with one or more of the designed sgRNAs.

[0212] LNPs were formulated using GENVOY ILM Lipid Mix (Precision Nanosystem) and the formulator Nanoassemblr Spark (Precision Nanosystem). LNPs were formulated according to the manufacturer's recommendations with Nitrogen:Phosphate (NP) ratio equal to 6 and flow rate ratio (FRR) 2:1. The RNA payload was diluted to a final concentration of 350 ng/uL in the PNI formulation buffer. The ETRs, dCas9-KRAB, dCas9-DNMT3A, dCas9-DNMT3L and each of the 121 sgRNA were mixed at 1:1:1:4 ratio. The RNA mix, the Genvoy lipid mix (25 mM) and PBS were loaded each in the dedicated chambers of the Spark cartridge and formulated. The quality of the formulated LNPs was evaluated quantifying the packaged mRNA using Quant-it™ RiboGreen RNA Assay Kit (Thermo Fisher) and sizing the LNP by Dynamic Light Scattering (Zetasizer, Malvern Panalytic).

[0213] HepG2-NTCP cells were plated at 20,000 cells/well in collagen coated 96 well plates. After 24h cells were infected with HBV at 5,000 multiplicity of genome equivalent (MGE) and 16h after viral inoculum was removed, cells were washed with PBS, and fresh media was added. Three days post-infection, using LNPs, each sgRNA and the mRNAs encoding each of the components of the triple constructs of ETRs (dCas9-KRAB, dCas9-DNMT3A, dCas9-DNMT3L) were delivered. Three days after, LNP was removed, medium was replaced, and cells were maintained in complete medium for three days.

[0214] Viral antigens HBeAg and HBsAg were quantified 6 days after LNP removal using ELISA assays. Data were normalized to a non-targeting guide designed against the mouse PCSK9 and control 3.2 gRNA was used as positive control. Cells viability assay were performed and normalized to non-targeting control.

[0215] The Table below provides amino acid sequences of exemplary epigenetic editors used in the gRNA screen (the ETR constructs):

TABLE 6: amino acid sequences of exemplary epigenetic editors

SEQ ID NO	Description	Amino acid sequence
476	dCas9:G:KRAB	MYPYDVPDYASPKKKRKVEASDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK

		HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQ LKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING IRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVN IVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLP KYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET RIDLSQLGGDSPKKKRKVGVD**GSGGG**ALSPQHSAVTQGSIIKNKEGMDAKSLT AWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKP DVILRLEKGEEPWLVEREIHQETHPDSETAFEIKSSV* YPYDVPDYA - HA-Tag **GSGGG** - Linker
477	dCas9:G:DNMT3A	MYPYDVPDYASPKKKRKVEASDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFK VLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLR KKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLAR GNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQ LKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING IRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVN IVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLP KYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET RIDLSQLGGDSPKKKRKVGVD**GSGGG**TYGLLRRREDWPSRLQMFFANNHDQEF DPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSI TVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGL YEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESN PVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSK VRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSR LARQRLLGRSWSVPVIRHLFAPLKEYFACV* YPYDVPDYA - HA-Tag **GSGGG** - Linker
478	dCas9:G:hDNMT3L	MYPYDVPDYASPKKKRKVEASDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFK VLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF

		SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLR KKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLAR GNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQ LKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING IRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVN IVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLP KYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET RIDLSQLGGDSPKKKRKVGVD**GSGGG**MAAIPALDPEAEPSMDVILVGSSELSS SVSPGTGRDLIAYEVKANQRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKF LDALFLYDDDGYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGK VHAMSNWVCYLCLPSSRSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWR RQPVRVLSLFEDIKKELTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPFD LVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPRPFFWMFVDNLVLNK EDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHWALVSEEELSL LAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSL* YPYDVPDYA - HA-Tag **GSGGG** - Linker
479	HA-Tag	YPYDVPDYA
480	linker	GSGGG

[0216] The Table below provides amino acid sequences and polynucleotide sequences of exemplary epigenetic editors.

TABLE 7: sequences of exemplary epigenetic editors

SEQ ID NO	Description	Sequence
481	PLA001 amino acid sequence	MPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATG LLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQE WGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDD RPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLP GMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPV FMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHL FAPLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMAAIPALDPEAEP SMDVILVGSSELSSSVSPGTGRDLIAYEVKANQRNIEDICICCGSLQVHTQ HPLFEGGICAPCKDKFLDALFLYDDDGYQSYCSICCSGETLLICGNPDCTR CYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPSSRSGLLQRRRKWRSQLKA FYDRESENPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLESGSDPGQ LKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQ YARPKPGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQN AVRVWSNIPAIRSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLP LREYFKYFSTELTSSLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESG

		PGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSELEDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSPKKKRKVGVDGSSGSETPGTSESATPESTGDSVAFEDVAVNFTLEEWALLDPSQKNLYRDVMRETFRNLASVGKQWEDQNIEDPFKIPRRNISHIPERLCESKEGGQGEESADYKDDDDKAPKKKRKVPKKKRKV
482	PLA001 polynucleotide sequence	ATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCA

		GTGTGGCGCCGGCAGCCCGTGAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACC

		CAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCCATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCACCGGTGACTCCGTTGCTTTCGAGGACGTGGCCGTGAACTTCACACTTGAGGAATGGGCCTTGCTCGACCCAAGTCAGAAGAATCTGTACAGAGACGTGATGCGGGAGACATTCAGGAATCTCGCCAGTGTCGGAAAGCAGTGGGAAGACCAGAACATCGAAGATCCTTTCAAGATACCACGGCGCAATATCTCCCACATTCCTGAGAGGCTGTGTGAATCTAAGGAAGGCGGACAAGGTGAGGAAAGCGCTGATTACAAAGATGATGACGATAAAGCCCCCAAGAAGAAAAGGAAGGTCCCAAAGAAAAAAAGAAAGGTGTGA
483	PLA002 Amino acid sequence	MPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMAAIPALDPEAEPSMDVILVGSSELSSSVSPGTGRDLIAYEVKANQRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKFLDALFLYDDDGYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPSSRSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSELEDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV

		EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSPKKKRKVGVDGSSGSETPGTSESATPESTGMNNSQGRVTFEDVTVNFTQGEWQRLNPEQRNLYRDVMLENYSNLVSVGQGETTKPDVILRLEQGKEPWLEEEEVLGSGRAEKNGDIGGQIWKPKDVKESLSADYKDDDDKAPKKKRKVPKKKRKV
484	PLA002 polynucleotide sequence	ATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAG

		TGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGAT

		ATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCCATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCACCGGTATGAACAATTCACAGGGGAGAGTGACATTCGAAGACGTGACCGTGAACTTCACCCAGGGAGAATGGCAGCGCTTGAACCCAGAACAAAGGAACCTCTATCGGGACGTGATGCTGGAAAACTACTCAAATTTGGTGAGCGTTGGGCAGGGTGAGACCACTAAGCCTGACGTGATCCTGAGATTGGAACAGGGCAAGGAGCCTTGGCTCGAGGAAGAGGAAGTCCTGGGCTCAGGGAGGGCCGAGAAAAACGGTGATATAGGAGGCCAGATATGGAAGCCTAAGGACGTCAAGGAGAGCCTGAGCGCTGATTACAAAGATGATGACGATAAAGCCCCCAAGAAGAAAAGGAAGGTCCCAAAGAAAAAAAGAAAGGTGTGA
492	PLA003 amino acid sequence	MPKKKRKVPKKKRKVYNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMAAIPALDPEAEPSMDVILVGSSELSSSVSPGTGRDLIAYEVKANQRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKFLDALFLYDDDGYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPSSRSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSLGGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSELEDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAK

		AILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSPKKKRKVGVDGSSGSETPGTSESATPESTGMNNSQGRVTFEDVTVNFTQGEWQRLNPEQRNLYRDVMLENYSNLVSVGQGETTKPDVILRLEQGKEPWLEEEEVLGSGRAEKNGDIGGQIWKPKDVKESLSAPKKKRKVPKKKRKV
493	PLA003 full plasmid sequence	GGGCGCTCGAGCAGGTTCAGAAGGAGATCAAAAACCCCCAAGGATCAAACATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACAT

		GCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGG

		ACGCCATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGG TGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCG AAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCA AGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCG GCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAA CCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACA CTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCC TGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAG TGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCG TCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCG TGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCA TGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGC GGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGG GCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATA TCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC TGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACC CTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGG TGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAG AGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCA TCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCA TCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAA TGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCT CCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGG GCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGC ACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGA TCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACC GGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCC TGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCG ACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCC ACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGG GAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGGAGTCGACGGATCCAGCG GCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCACCGGTA TGAACAATTCACAGGGGAGAGTGACATTCGAAGACGTGACCGTGAACTTCA CCCAGGGAGAATGGCAGCGCTTGAACCCAGAACAAAGGAACCTCTATCGGG ACGTGATGCTGGAAAACTACTCAAATTTGGTGAGCGTTGGGCAGGGTGAGA CCACTAAGCCTGACGTGATCCTGAGATTGGAACAGGGCAAGGAGCCTTGGC TCGAGGAAGAGGAAGTCCTGGGCTCAGGGAGGGCCGAGAAAAACGGTGATA TAGGAGGCCAGATATGGAAGCCTAAGGACGTCAAGGAGAGCCTGAGCGCTC CCAAGAAGAAAAGGAAGGTCCCAAAGAAAAAAAGAAAGGTGTGAGGATCCT GAGTCTAGAAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGT ATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATG CCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTG TATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGG CAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGG GGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTC CCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACA GGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCA TCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGG ACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCC CGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCT CAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCCTGTTAATTAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCTTGA AGAGCCTAGTGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGT ATTTCACACCGCATAATCCAGCACAGTGGCGGCCCGTTTAAACCCGCTGAT CAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCC CCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAAT AAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGG GGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCA GGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCA GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGG

		GCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCT GCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA ATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCC GACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCG CTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCC TTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTC GGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCA GCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGT AAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAG AGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTA CGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGT TACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGC TGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAA AGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG GAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGAT CTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAG TATATATGAGTAAACTTGGTCTGACAGTTAGAAAAACTCATCGAGCATCAA ATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAA AAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGAT GGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACA ACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACC ATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTT CCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGC ATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAAACGAAATAC GCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCG CAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTC TTCTAATACCTGGAATGCTGTTTTCCCAGGGATCGCAGTGGTGAGTAACCA TGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAA TTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAAC GCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATA CAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTT ATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCA AGACGTTTCCCGTTGAATATGGCTCATACTCTTCCTTTTTCAATATTATTG AAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT TTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCC ACCTGACGTCGATCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCAC TCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCC TGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTAC AACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAG GCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTG ATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAG CCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGG CTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTT ACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTAC GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCA GTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGT CATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGG ATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAA CAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGT CTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCT TATCGAAATTAATACGACTCACTATAAG
494	PLA003 plasmid coding sequence	ATGCCAAAAAAGAAGAGAAAGGTACCGAAGAAAAAAAGAAAGGTATACAAT CACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAG AAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGC CTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCC GAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAG ATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAG TGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCC

		ATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGAGGACCCTCCTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTAC

		ATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCCATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCACCGGTATGAACAATTCACAGGGGAGAGTGACATTCGAAGACGTGACCGTGAACTTC

		ACCCAGGGAGAATGGCAGCGCTTGAACCCAGAACAAAGGAACCTCTATCGGGACGTGATGCTGGAAAACTACTCAAATTTGGTGAGCGTTGGGCAGGGTGAGACCACTAAGCCTGACGTGATCCTGAGATTGGAACAGGGCAAGGAGCCTTGGCTCGAGGAAGAGGAAGTCCTGGGCTCAGGGAGGGCCGAGAAAAACGGTGATATAGGAGGCCAGATATGGAAGCCTAAGGACGTCAAGGAGAGCCTGAGCGCTCCCAAGAAGAAAAGGAAGGTCCCAAAGAAAAAAAGAAAGGTGTGA

[0217] Table 8 below lists components of the fusion polypeptide PLA001 and their corresponding amino acid position in the fusion polypeptide sequence (SEQ ID No. 481) set forth in Table 7.

TABLE 8: annotation of PLA001 amino acid sequence

	Type	Start	End	Length
SV40 NLS	CDS	2	8	7
SV40 NLS	CDS	9	15	7
DNMT3A	CDS	17	317	301
Linker	CDS	318	344	27
DNMT3L full-length	CDS	345	730	386
XTEN80	CDS	731	810	80
dCas9	CDS	811	2180	1370
NLS	CDS	2181	2187	7
XTEN16	CDS	2188	2208	21
ZN627	CDS	2211	2290	80
FLAG	CDS	2293	2300	8
SV40 NLS	CDS	2302	2308	7
SV40 NLS	CDS	2309	2315	7

[0218] Table 9 below lists components of the polynucleotide encoding the fusion polypeptide PLA001 and their corresponding nucleotide position in the polynucleotide sequence (SEQ ID No. 482) set forth in Table 7.

TABLE 9: annotation of PLA001 polynucleotide sequence

Name	Type	Minimum	Maximum	Length
SV40 NLS	CDS	4	24	21
SV40 NLS	CDS	25	44	20
DNMT3A	CDS	49	951	903

Linker	CDS	952	1032	81
DNMT3L full-length	CDS	1033	2190	1158
XTEN80	CDS	2191	2430	240
dCas9	CDS	2431	6540	4110
NLS	CDS	6541	6561	21
XTEN16	CDS	6562	6624	63
ZN627	CDS	6631	6870	240
FLAG	CDS	6877	6900	24
SV40 NLS	CDS	6904	6924	21
SV40 NLS	CDS	6925	6945	21

[0219] Table 10 below lists components of the fusion polypeptide PLA002 and their corresponding amino acid position in the fusion polypeptide sequence (SEQ ID No. 483) set forth in Table 7.

TABLE 10: annotation of PLA002 amino acid sequence

Name	Type	Minimum	Maximum	Length
SV40 NLS	CDS	2	8	7
SV40 NLS	CDS	9	15	7
DNMT3A	CDS	17	317	301
Linker	CDS	318	344	27
DNMT3L full-length	CDS	345	730	386
XTEN80	CDS	731	810	80
dCas9	CDS	811	2180	1370
NLS	CDS	2181	2187	7
XTEN16	CDS	2188	2208	21
ZIM3	CDS	2211	2310	100
FLAG	CDS	2313	2320	8
SV40 NLS	CDS	2322	2328	7
SV40 NLS	CDS	2329	2335	7

[0220] Table 11 below lists components of the polynucleotide encoding the fusion polypeptide PLA002 and their corresponding nucleotide position in the polynucleotide sequence (SEQ ID No. 484) set forth in Table 7.

TABLE 11: annotation of PLA002 polynucleotide sequence

Name	Type	Minimum	Maximum	Length
SV40 NLS	CDS	4	24	21
SV40 NLS	CDS	25	45	21
DNMT3A	CDS	49	951	903
Linker	CDS	952	1032	81
DNMT3L full-length	CDS	1033	2190	1158
XTEN80	CDS	2191	2430	240
dCas9	CDS	2431	6540	4110
NLS	CDS	6541	6561	21
XTEN16	CDS	6562	6624	63
ZIM3	CDS	6631	6930	300
FLAG	CDS	6937	6960	24
SV40 NLS	CDS	6964	6984	21
SV40 NLS	CDS	6985	7005	21
stop	terminator	7006	7008	3

TABLE 12. Annotation of PLA003 amino acid sequence

Name	Type	Minimum	Maximum	Length
SV40 NLS	CDS	2	8	7
SV40 NLS	CDS	9	15	7
DNMT3A	CDS	17	317	301
Linker	CDS	318	344	27
DNMT3L full-length	CDS	345	730	386
XTEN80	CDS	731	810	80
dCas9	CDS	811	2180	1370
NLS	CDS	2181	2187	7
XTEN16	CDS	2188	2208	21
ZIM3	CDS	2211	2310	100
SV40 NLS	CDS	2313	2319	7
SV40 NLS	CDS	2320	2326	7

TABLE 13. Annotation of PLA003 polynucleotide sequence

Name	Type	Minimum	Maximum	Length
SV40 NLS	CDS	4	24	21
SV40 NLS	CDS	25	45	21
DNMT3A	CDS	49	951	903
Linker	CDS	952	1032	81
DNMT3L full-length	CDS	1033	2190	1158
XTEN80	CDS	2191	2430	240
dCas9	CDS	2431	6540	4110
NLS	CDS	6541	6561	21

XTEN16	CDS	6562	6624	63
ZIM3	CDS	6631	6930	300
SV40 NLS	CDS	6937	6957	21
SV40 NLS	CDS	6958	6978	21
stop	terminator	6979	6981	3

[0221] Table 14 below provides gRNA sequence tested.

TABLE 14: Exemplary gRNA sequences

SEQ IDs	Target domain sequence		SEQ IDs	gRNA sequence
333	CCTGCTGGTGGCTCCAGTTC		1093	CCUGCUGGUGGCUCCAGUUCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
334	CTGAACTGGAGCCACCAGCA		1094	CUGAACUGGAGCCACCAGCAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
335	CCTGAACTGGAGCCACCAGC		1095	CCUGAACUGGAGCCACCAGCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
336	CCTCGAGAAGATTGACGATA		1096	CCUCGAGAAGAUUGACGAUAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
337	TCGTCAATCTTCTCGAGGAT		1097	UCGUCAAUCUUCUCGAGGAUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
338	CGTCAATCTTCTCGAGGATT		1098	CGUCAAUCUUCUCGAGGAUUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
339	GTCAATCTTCTCGAGGATTG		1099	GUCAAUCUUCUCGAGGAUUGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
340	AACATGGAGAACATCACATC		1100	AACAUGGAGAACAUCACAUCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
341	AACATCACATCAGGATTCCT		1101	AACAUCACAUCAGGAUUCCUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
342	CTAGACTCTGCGGTATTGTG		1102	CUAGACUCUGCGGUAUUGUGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
343	TACCGCAGAGTCTAGACTCG		1103	UACCGCAGAGUCUAGACUCGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
344	CGCAGAGTCTAGACTCGTGG		1104	CGCAGAGUCUAGACUCGUGGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
345	CACCACGAGTCTAGACTCTG		1105	CACCACGAGUCUAGACUCUGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
346	TGGACTTCTCTCAATTTTCT		1106	UGGACUUCUCUCAAUUUUCUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
347	GGACTTCTCTCAATTTTCTA		1107	GGACUUCUCUCAAUUUUCUAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
348	GACTTCTCTCAATTTTCTAG		1108	GACUUCUCUCAAUUUUCUAGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
349	ACTTCTCTCAATTTTCTAGG		1109	ACUUCUCUCAAUUUUCUAGGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
350	CGAATTTTGGCCAAGACACA		1110	CGAAUUUUGGCCAAGACACAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
351	AGGTTGGGGACTGCGAATTT		1111	AGGUUGGGGACUGCGAAUUUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
352	GGCATAGCAGCAGGATGAAG		1112	GGCAUAGCAGCAGGAUGAAGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
353	AGAAGATGAGGCATAGCAGC		1113	AGAAGAUGAGGCAUAGCAGCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
354	GCTATGCCTCATCTTCTTGT		1114	GCUAUGCCUCAUCUUCUUGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
355	GAAGAACCAACAAGAAGATG		1115	GAAGAACCAACAAGAAGAUGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
356	CATCTTCTTGTTGGTTCTTC		1116	CAUCUUCUUGUUGGUUCUUCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
357	CCCGTTTGTCCTCTAATTCC		1117	CCCGUUUGUCCUCUAAUUCCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
358	CCTGGAATTAGAGGACAAAC		1118	CCUGGAAUUAGAGGACAAACGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
359	TCCTGGAATTAGAGGACAAA		1119	UCCUGGAAUUAGAGGACAAAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
360	TACTAGTGCCATTTGTTCAG		1120	UACUAGUGCCAUUUGUUCAGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
361	CCATTTGTTCAGTGGTTCGT		1121	CCAUUUGUUCAGUGGUUCGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU

362	CATTTGTTCAGTGGTTCGTA		1122	CAUUUGUUCAGUGGUUCGUAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
363	CCTACGAACCACTGAACAAA		1123	CCUACGAACCACUGAACAAAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
364	TTTCAGTTATATGGATGATG		1124	UUUCAGUUAUAUGGAUGAUGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
365	CAAAAGAAAATTGGTAACAG		1125	CAAAAGAAAAUUGGUAACAGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
366	TACCAATTTTCTTTTGTCTT		1126	UACCAAUUUUCUUUUGUCUUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
367	ACCAATTTTCTTTTGTCTTT		1127	ACCAAUUUUCUUUUGUCUUUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
368	ACCCAAAGACAAAAGAAAAT		1128	ACCCAAAGACAAAAGAAAAUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
369	TGACATACTTTCCAATCAAT		1129	UGACAUACUUUCCAAUCAAUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
370	CACTTTCTCGCCAACTTACA		1130	CACUUUCUCGCCAACUUACAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
371	CACAGAAAGGCCTTGTAAGT		1131	CACAGAAAGGCCUUGUAAGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
372	TGAACCTTTACCCCGTTGCC		1132	UGAACCUUUACCCCGUUGCCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
373	GGGCAACGGGGTAAAGGTTC		1133	GGGCAACGGGGUAAAGGUUCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
374	TTTACCCCGTTGCCCGGCAA		1134	UUUACCCCGUUGCCCGGCAAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
375	GTTGCCGGGCAACGGGGTAA		1135	GUUGCCGGGCAACGGGGUAAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
376	CCCGTTGCCCGGCAACGGCC		1136	CCCGUUGCCCGGCAACGGCCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
377	CTGGCCGTTGCCGGGCAACG		1137	CUGGCCGUUGCCGGGCAACGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
378	CCTGGCCGTTGCCGGGCAAC		1138	CCUGGCCGUUGCCGGGCAACGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
379	ACCTGGCCGTTGCCGGGCAA		1139	ACCUGGCCGUUGCCGGGCAAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
380	GCACAGACCTGGCCGTTGCC		1140	GCACAGACCUGGCCGUUGCCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
381	GGCACAGACCTGGCCGTTGC		1141	GGCACAGACCUGGCCGUUGCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
382	GCAAACACTTGGCACAGACC		1142	GCAAACACUUGGCACAGACCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
383	GGGTTGCGTCAGCAAACACT		1143	GGGUUGCGUCAGCAAACACUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
384	TTTGCTGACGCAACCCCCAC		1144	UUUGCUGACGCAACCCCCACGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
385	CTGACGCAACCCCCACTGGC		1145	CUGACGCAACCCCCACUGGCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
386	TGACGCAACCCCCACTGGCT		1146	UGACGCAACCCCCACUGGCUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
387	GACGCAACCCCCACTGGCTG		1147	GACGCAACCCCCACUGGCUGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
388	AACCCCCACTGGCTGGGGCT		1148	AACCCCCACUGGCUGGGGCUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
389	TCCTCTGCCGATCCATACTG		1149	UCCUCUGCCGAUCCAUACUGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
390	TCCGCAGTATGGATCGGCAG		1150	UCCGCAGUAUGGAUCGGCAGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
391	AGGAGTTCCGCAGTATGGAT		1151	AGGAGUUCCGCAGUAUGGAUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
392	CGGCTAGGAGTTCCGCAGTA		1152	CGGCUAGGAGUUCCGCAGUAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
393	TGCGAGCAAAACAAGCGGCT		1153	UGCGAGCAAAACAAGCGGCUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
394	CCGCTTGTTTTGCTCGCAGC		1154	CCGCUUGUUUUGCUCGCAGCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU

395	CCTGCTGCGAGCAAAACAAG		1155	CCUGCUGCGAGCAAAACAAGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
396	TGTTTTGCTCGCAGCAGGTC		1156	UGUUUUGCUCGCAGCAGGUCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
397	GCAGCACAGCCTAGCAGCCA		1157	GCAGCACAGCCUAGCAGCCAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
398	TGCTAGGCTGTGCTGCCAAC		1158	UGCUAGGCUGUGCUGCCAACGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
399	GCTGCCAACTGGATCCTGCG		1159	GCUGCCAACUGGAUCCUGCGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
400	CTGCCAACTGGATCCTGCGC		1160	CUGCCAACUGGAUCCUGCGCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
401	CGTCCCGCGCAGGATCCAGT		1161	CGUCCCGCGCAGGAUCCAGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
402	AAACAAAGGACGTCCCGCGC		1162	AAACAAAGGACGUCCCGCGCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
403	GTCCTTTGTTTACGTCCCGT		1163	GUCCUUUGUUUACGUCCCGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
404	CGCCGACGGGACGTAAACAA		1164	CGCCGACGGGACGUAAACAAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
405	TGCCGTTCCGACCGACCACG		1165	UGCCGUUCCGACCGACCACGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
406	AGGTGCGCCCCGTGGTCGGT		1166	AGGUGCGCCCCGUGGUCGGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
407	AGAGAGGTGCGCCCCGTGGT		1167	AGAGAGGUGCGCCCCGUGGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
408	GTAAAGAGAGGTGCGCCCCG		1168	GUAAAGAGAGGUGCGCCCCGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
409	GGGGCGCACCTCTCTTTACG		1169	GGGGCGCACCUCUCUUUACGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
410	CGGGGAGTCCGCGTAAAGAG		1170	CGGGGAGUCCGCGUAAAGAGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
411	CAGATGAGAAGGCACAGACG		1171	CAGAUGAGAAGGCACAGACGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
412	GTCTGTGCCTTCTCATCTGC		1172	GUCUGUGCCUUCUCAUCUGCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
413	GGCAGATGAGAAGGCACAGA		1173	GGCAGAUGAGAAGGCACAGAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
414	GCAGATGAGAAGGCACAGAC		1174	GCAGAUGAGAAGGCACAGACGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
415	ACACGGTCCGGCAGATGAGA		1175	ACACGGUCCGGCAGAUGAGAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
416	GAAGCGAAGTGCACACGGTC		1176	GAAGCGAAGUGCACACGGUCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
417	GAGGTGAAGCGAAGTGCACA		1177	GAGGUGAAGCGAAGUGCACAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
418	CTTCACCTCTGCACGTCGCA		1178	CUUCACCUCUGCACGUCGCAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
419	GGTCTCCATGCGACGTGCAG		1179	GGUCUCCAUGCGACGUGCAGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
420	TGCCCAAGGTCTTACATAAG		1180	UGCCCAAGGUCUUACAUAAGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
421	GTCCTCTTATGTAAGACCTT		1181	GUCCUCUUAUGUAAGACCUUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
422	AGTCCTCTTATGTAAGACCT		1182	AGUCCUCUUAUGUAAGACCUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
423	GTCTTACATAAGAGGACTCT		1183	GUCUUACAUAAGAGGACUCUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
424	AATGTCAACGACCGACCTTG		1184	AAUGUCAACGACCGACCUUGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
425	TTTGAAGTATGCCTCAAGGT		1185	UUUGAAGUAUGCCUCAAGGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
426	AGTCTTTGAAGTATGCCTCA		1186	AGUCUUUGAAGUAUGCCUCAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
427	AAGACTGTTTGTTTAAAGAC		1187	AAGACUGUUUGUUUAAAGACGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU

428	AGACTGTTTGTTTAAAGACT		1188	AGACUGUUUGUUUAAAGACUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
429	CTGTTTGTTTAAAGACTGGG		1189	CUGUUUGUUUAAAGACUGGGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
430	GTTTAAAGACTGGGAGGAGT		1190	GUUUAAAGACUGGGAGGAGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
431	TCTTTGTACTAGGAGGCTGT		1191	UCUUUGUACUAGGAGGCUGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
432	AGGAGGCTGTAGGCATAAAT		1192	AGGAGGCUGUAGGCAUAAAUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
433	GTGAAAAAGTTGCATGGTGC		1193	GUGAAAAAGUUGCAUGGUGCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
434	GCAGAGGTGAAAAAGTTGCA		1194	GCAGAGGUGAAAAAGUUGCAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
435	AACAAGAGATGATTAGGCAG		1195	AACAAGAGAUGAUUAGGCAGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
436	GACATGAACAAGAGATGATT		1196	GACAUGAACAAGAGAUGAUUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
437	AGCTTGGAGGCTTGAACAGT		1197	AGCUUGGAGGCUUGAACAGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
438	CAAGCCTCCAAGCTGTGCCT		1198	CAAGCCUCCAAGCUGUGCCUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
439	AAGCCTCCAAGCTGTGCCTT		1199	AAGCCUCCAAGCUGUGCCUUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
440	CCTCCAAGCTGTGCCTTGGG		1200	CCUCCAAGCUGUGCCUUGGGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
441	CCACCCAAGGCACAGCTTGG		1201	CCACCCAAGGCACAGCUUGGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
442	AGCTGTGCCTTGGGTGGCTT		1202	AGCUGUGCCUUGGGUGGCUUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
443	AAGCCACCCAAGGCACAGCT		1203	AAGCCACCCAAGGCACAGCUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
444	GCTGTGCCTTGGGTGGCTTT		1204	GCUGUGCCUUGGGUGGCUUUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
445	CTGTGCCTTGGGTGGCTTTG		1205	CUGUGCCUUGGGUGGCUUUGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
446	TAGCTCCAAATTCTTTATAA		1206	UAGCUCCAAAUUCUUUAUAAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
447	GTAGCTCCAAATTCTTTATA		1207	GUAGCUCCAAAUUCUUUAUAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
448	TAAAGAATTTGGAGCTACTG		1208	UAAAGAAUUUGGAGCUACUGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
449	ATGACTCTAGCTACCTGGGT		1209	AUGACUCUAGCUACCUGGGUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
450	CACATTTCTTGTCTCACTTT		1210	CACAUUUCUUGUCUCACUUUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
451	TAGTTTCCGGAAGTGTTGAT		1211	UAGUUUCCGGAAGUGUUGAUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
452	CGTCTAACAACAGTAGTTTC		1212	CGUCUAACAACAGUAGUUUCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
453	ACTACTGTTGTTAGACGACG		1213	ACUACUGUUGUUAGACGACGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
454	CTGTTGTTAGACGACGAGGC		1214	CUGUUGUUAGACGACGAGGCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
455	CGAGGGAGTTCTTCTTCTAG		1215	CGAGGGAGUUCUUCUUCUAGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
456	GCGAGGGAGTTCTTCTTCTA		1216	GCGAGGGAGUUCUUCUUCUAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
457	GGCGAGGGAGTTCTTCTTCT		1217	GGCGAGGGAGUUCUUCUUCUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
458	CTCCCTCGCCTCGCAGACGA		1218	CUCCCUCGCCUCGCAGACGAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
459	GACCTTCGTCTGCGAGGCGA		1219	GACCUUCGUCUGCGAGGCGAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
460	AGACCTTCGTCTGCGAGGCG		1220	AGACCUUCGUCUGCGAGGCGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU

461	GATTGAGACCTTCGTCTGCG		1221	GAUUGAGACCUUCGUCUGCGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
462	GATTGAGATCTTCTGCGACG		1222	GAUUGAGAUCUUCUGCGACGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
463	GTCGCAGAAGATCTCAATCT		1223	GUCGCAGAAGAUCUCAAUCUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
464	TCGCAGAAGATCTCAATCTC		1224	UCGCAGAAGAUCUCAAUCUCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
465	ATATGGTGACCCACAAAATG		1225	AUAUGGUGACCCACAAAAUGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
466	TTTGTGGGTCACCATATTCT		1226	UUUGUGGGUCACCAUAUUCUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
467	TTGTGGGTCACCATATTCTT		1227	UUGUGGGUCACCAUAUUCUUGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
468	GCTGGATCCAACTGGTGGTC		1228	GCUGGAUCCAACUGGUGGUCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
469	CACCCCAAAAGGCCTCCGTG		1229	CACCCCAAAAGGCCUCCGUGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
470	CCTTTTGGGGTGGAGCCCTC		1230	CCUUUUGGGGUGGAGCCCUCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
471	CCTGAGGGCTCCACCCCAAA		1231	CCUGAGGGCUCCACCCCAAAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
472	GGGGTGGAGCCCTCAGGCTC		1232	GGGGUGGAGCCCUCAGGCUCGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
473	GGGTGGAGCCCTCAGGCTCA		1233	GGGUGGAGCCCUCAGGCUCAGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
474	CGATTGGTGGAGGCAGGAGG		1234	CGAUUGGUGGAGGCAGGAGGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
475	CTCATCCTCAGGCCATGCAG		1235	CUCAUCCUCAGGCCAUGCAGGUUUAAGAGCUAAGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU

TABLE 15: Exemplary target domain sequences and effect on HbeAg and HbsAg expression

SEQ IDs	Associated guide RNA name (if applicable)	Target domain sequence	HbeAg(%expression of non targeting control)	HbsAg (%expression of non targeting control)
334	gRNA#001	CTGAACTGGAGCCACCAGCA	27.77203753	23.4507853
335	gRNA#002	CCTGAACTGGAGCCACCAGC	41.3794605	42.3814023
333		CCTGCTGGTGGCTCCAGTTC	65.36067834	43.2303179
336		CCTCGAGAAGATTGACGATA	82.8943107	72.648219
337		TCGTCAATCTTCTCGAGGAT	45.82985382	59.7223204
338		CGTCAATCTTCTCGAGGATT	70.38176383	73.1313979
339		GTCAATCTTCTCGAGGATTG	51.92713248	54.330978
340		AACATGGAGAACATCACATC	79.31612772	80.8981286
341		AACATCACATCAGGATTCCT	41.40633262	37.5509299
342		CTAGACTCTGCGGTATTGTG	48.56267424	41.5330827
345	gRNA#003	CACCACGAGTCTAGACTCTG	44.43853541	40.8553881
343		TACCGCAGAGTCTAGACTCG	49.18078863	56.151898
344		CGCAGAGTCTAGACTCGTGG	52.41583101	57.2264647
346		TGGACTTCTCTCAATTTTCT	49.58564481	51.1350719
347		GGACTTCTCTCAATTTTCTA	76.16671739	79.1684976
348		GACTTCTCTCAATTTTCTAG	49.79317156	54.1540479
349		ACTTCTCTCAATTTTCTAGG	69.66968253	77.4650531
350		CGAATTTTGGCCAAGACACA	53.53282063	54.0024954
371	gRNA#004	CACAGAAAGGCCTTGTAAGT	42.35590319	41.6928086
370		CACTTTCTCGCCAACTTACA	53.25960148	55.120666
373	gRNA#005	GGGCAACGGGGTAAAGGTTC	36.54111842	42.8120918
375	gRNA#006	GTTGCCGGGCAACGGGGTAA	41.20322042	38.1885911
377		CTGGCCGTTGCCGGGCAACG	57.27834882	60.830473
372		TGAACCTTTACCCCGTTGCC	48.16509881	60.952804
378		CCTGGCCGTTGCCGGGCAAC	56.34234102	65.50842
379		ACCTGGCCGTTGCCGGGCAA	54.10829257	53.324749
374		TTTACCCCGTTGCCCGGCAA	56.72089131	62.6906255
380		GCACAGACCTGGCCGTTGCC	42.46818432	47.3720079
381		GGCACAGACCTGGCCGTTGC	72.65381719	77.2400091
376		CCCGTTGCCCGGCAACGGCC	50.93018919	61.086777
382		GCAAACACTTGGCACAGACC	57.0196485	69.491449
383		GGGTTGCGTCAGCAAACACT	49.73518831	54.7510029
384		TTTGCTGACGCAACCCCCAC	41.79724731	50.0362297
385		CTGACGCAACCCCCACTGGC	36.90727137	36.8247762
386		TGACGCAACCCCCACTGGCT	46.49501492	59.6959921
387		GACGCAACCCCCACTGGCTG	40.09200943	51.4756937
388		AACCCCCACTGGCTGGGGCT	61.82883278	79.8761795

390	gRNA#007	TCCGCAGTATGGATCGGCAG	26.33655968	33.7255842
391	gRNA#008	AGGAGTTCCGCAGTATGGAT	28.49512897	40.080391
389	gRNA#009	TCCTCTGCCGATCCATACTG	28.45399116	42.735093
392		CGGCTAGGAGTTCCGCAGTA	56.5241517	66.9060644
393	gRNA#010	TGCGAGCAAAACAAGCGGCT	41.5479747	40.5350018
395		CCTGCTGCGAGCAAAACAAG	36.4525077	50.516964
394		CCGCTTGTTTTGCTCGCAGC	108.4014077	90.5082399
396		TGTTTTGCTCGCAGCAGGTC	68.78508191	75.7537996
397		GCAGCACAGCCTAGCAGCCA	78.73231487	68.3785588
398		TGCTAGGCTGTGCTGCCAAC	59.52249922	69.0333267
401		CGTCCCGCGCAGGATCCAGT	52.51634701	49.5876502
399		GCTGCCAACTGGATCCTGCG	75.81794218	89.0162904
400		CTGCCAACTGGATCCTGCGC	77.79441236	73.9461516
402		AAACAAAGGACGTCCCGCGC	67.52500576	72.6685954
404		CGCCGACGGGACGTAAACAA	77.77475148	70.288774
403		GTCCTTTGTTTACGTCCCGT	94.99070926	103.867949
406		AGGTGCGCCCCGTGGTCGGT	68.80565242	65.4335257
407		AGAGAGGTGCGCCCCGTGGT	42.18514493	55.1199635
408		GTAAAGAGAGGTGCGCCCCG	53.39922155	55.7151401
410		CGGGGAGTCCGCGTAAAGAG	52.63946411	66.9249801
409		GGGGCGCACCTCTCTTTACG	72.81702761	66.4993545
411	gRNA#011	CAGATGAGAAGGCACAGACG	32.31425506	44.762352
413		GGCAGATGAGAAGGCACAGA	59.89738685	59.5785052
415		ACACGGTCCGGCAGATGAGA	41.29188182	52.515655
412		GTCTGTGCCTTCTCATCTGC	70.71073836	72.0049046
416		GAAGCGAAGTGCACACGGTC	31.51588976	59.2847924
417		GAGGTGAAGCGAAGTGCACA	53.23795933	54.7085711
419		GGTCTCCATGCGACGTGCAG	98.80315853	94.871871
418		CTTCACCTCTGCACGTCGCA	76.66072308	76.4195077
421		GTCCTCTTATGTAAGACCTT	50.06169791	63.8903663
422		AGTCCTCTTATGTAAGACCT	54.84793515	62.0058784
420		TGCCCAAGGTCTTACATAAG	65.64906417	79.7359246
423		GTCTTACATAAGAGGACTCT	65.0201597	62.5458243
424		AATGTCAACGACCGACCTTG	53.64938718	65.5805852
425		TTTGAAGTATGCCTCAAGGT	68.9199506	80.763234
426	gRNA#012	AGTCTTTGAAGTATGCCTCA	30.45840615	47.6679105
427		AAGACTGTTTGTTTAAAGAC	75.19137394	74.1370789
428		AGACTGTTTGTTTAAAGACT	66.21290133	75.2309845
429		CTGTTTGTTTAAAGACTGGG	63.52924235	72.0972239
430		GTTTAAAGACTGGGAGGAGT	52.01423199	66.8961386
431		TCTTTGTACTAGGAGGCTGT	51.48581844	68.9533809
432		AGGAGGCTGTAGGCATAAAT	37.69681736	56.2655965

433		GTGAAAAAGTTGCATGGTGC	82.88524703	98.0043703
434		GCAGAGGTGAAAAAGTTGCA	31.73533955	53.6210823
435	gRNA#013	AACAAGAGATGATTAGGCAG	30.51551968	43.8402184
436	gRNA#014	GACATGAACAAGAGATGATT	15.37394867	25.9017005
437		AGCTTGGAGGCTTGAACAGT	84.06388656	100.433196
441	gRNA#015	CCACCCAAGGCACAGCTTGG	22.57628478	29.4502561
443		AAGCCACCCAAGGCACAGCT	38.69686132	57.447646
438		CAAGCCTCCAAGCTGTGCCT	57.03790348	55.3144232
439		AAGCCTCCAAGCTGTGCCTT	101.2197916	108.433992
442		AGCTGTGCCTTGGGTGGCTT	62.50798441	75.5245296
444		GCTGTGCCTTGGGTGGCTTT	63.60985011	68.2127614
445		CTGTGCCTTGGGTGGCTTTG	58.80930094	60.2093595
446		TAGCTCCAAATTCTTTATAA	81.50792369	102.062484
447		GTAGCTCCAAATTCTTTATA	57.5300482	84.4089935
448		TAAAGAATTTGGAGCTACTG	55.34840957	67.1682598
449		ATGACTCTAGCTACCTGGGT	70.72899714	69.314819
450		CACATTTCTTGTCTCACTTT	135.7647935	119.430868
451		TAGTTTCCGGAAGTGTTGAT	52.38647155	59.8621336
452		CGTCTAACAACAGTAGTTTC	84.81350809	79.1119745
453		ACTACTGTTGTTAGACGACG	50.34753433	57.5139945
454		CTGTTGTTAGACGACGAGGC	47.03375963	53.0434947
455		CGAGGGAGTTCTTCTTCTAG	36.81318989	50.1844755
456		GCGAGGGAGTTCTTCTTCTA	68.04429109	71.2738682
457	gRNA#016	GGCGAGGGAGTTCTTCTTCT	35.40374342	49.4263836
459		GACCTTCGTCTGCGAGGCGA	28.35732375	53.108582
460		AGACCTTCGTCTGCGAGGCG	41.45363172	58.2048965
461		GATTGAGACCTTCGTCTGCG	63.13599738	73.3793991
458		CTCCCTCGCCTCGCAGACGA	41.73812486	56.4066766
462		GATTGAGATCTTCTGCGACG	134.1434937	133.039909
463		GTCGCAGAAGATCTCAATCT	44.87633493	58.0732445
464		TCGCAGAAGATCTCAATCTC	70.59684886	75.0458487
465	gRNA#017	ATATGGTGACCCACAAAATG	41.36374656	46.043276
466		TTTGTGGGTCACCATATTCT	66.33644682	65.6466534
467	gRNA#018	TTGTGGGTCACCATATTCTT	48.06595023	41.7714626
468		GCTGGATCCAACTGGTGGTC	65.83430344	69.3357339
469		CACCCCAAAAGGCCTCCGTG	21.63462413	23.5507547
471	gRNA#019	CCTGAGGGCTCCACCCCAAA	45.40727826	44.6869573
470		CCTTTTGGGGTGGAGCCCTC	50.06807456	31.73417
472		GGGGTGGAGCCCTCAGGCTC	64.29444481	64.1755302
473		GGGTGGAGCCCTCAGGCTCA	44.19826805	53.1051257
474		CGATTGGTGGAGGCAGGAGG	65.52555289	60.9306557
475	gRNA#020	CTCATCCTCAGGCCATGCAG	35.40063237	17.5286587

[0222] *In vitro* silencing was observed in an HepG2-NTCP infection model with gRNAs targeting CpG islands with ETRs (Figure 5A–Figure 5B). A primary screen was conducted using LNPs of quality within expected parameters and a pilot experiment with a single guide (Figure 6–Figure 8). Results demonstrated that 48 gRNAs showed less than 50% expression of HBeAg at day 6 compared to non-targeting control (Figure 9) and 28 gRNAs showed less than 50% expression of HBsAg at day 6 compared to non-targeting control (Figure 10). HBsAg and HBeAg expression was positively correlated as shown in Figure 11.

Example 4: Zinc finger repressors for silencing HBV

[0223] Zinc finger repressors targeting epigenetic target sites identified in the HBV genome were designed. Table 1 above provides amino acid sequences of zinc finger and its corresponding motif sequences and target sequences of the zinc finger.

[0224] Zinc finger repressors described in Table 1 are tested in an HBV infection model, e.g., in HepG2 cells as described herein, and efficient repression of HBV is confirmed for the zinc finger repressors provided in Table 1.

Example 5: Further *In vitro* Evaluation of gRNAs

[0225] A CRISPR-Off single construct encoding PLA002, consisting of KRAB, DNMT3A, DNMT3L, and dCas9, was used in combination with one or more of the designed sgRNAs for the *in vitro* assays described in this example.

[0226] HepG2-NTCP cells were infected with HBV for 4 days, following procedures similar as those in Example 3, and were then transfected with CRISPR-off construct and individual exemplary gRNAs (as indicated in Table 13) formulated in a research-grade LNP. At Day 6 post-transfection HBsAg and HBeAg protein expression in the supernatant was evaluated by ELISA, as depicted in Figure 12A. Results from this experiment are shown in Figure 12B. All of the tested gRNAs led to reduction of HBsAg and HBeAg levels in the supernatant. Positive control used in this experiment is a gRNA against HBV genome that was previously shown to reduce antigens ~50%.

[0227] In another experiment, the integrated HBV cell line, PLC/PRF/5, was used to evaluate activity of gRNAs. The PLC/PRF/5 cells were transfected with CRISPR-off (PLA002) and individual gRNAs using a commercial lipid-based transfection reagent. As depicted in Figure 13A, four days after transfection HBsAg protein expression in the supernatant was evaluated by ELISA. Results from this experiment are shown in Figure 13B. Target conservation was evaluated in silico and target conservation was defined as 100% gRNA-DNA match.

[0228] In a further experiment, primary human hepatocytes (PHH) derived from humanized mice were infected with HBV for 4 days and then transfected with CRISPR-off (PLA002) and individual gRNAs formulated in a research-grade LNP, GenVoy LNPs. As depicted in Figure 14A, at Day 6 post-infection HBsAg and HBeAg protein expression in the supernatant was evaluated by ELISA. Results from this experiment are shown in Figure 14B. Positive control used in this experiment is a HBV gRNA that was previously shown to reduce antigens ~50%. The data suggested strong *in vitro* silencing by certain gRNAs at Day 6 after transfection. In a second PHH experiment, depicted in Figure 14C, post-infection HBsAg and HBeAg protein expression in the supernatant was evaluated by ELISA at Day 12 after delivery of 100 ng of payload (1:1 effector to guide RNA ratio) in research-grade LNPs . Epigenetic editors repress HBsAg and HBeAg secretion in HBV infected PHH cells at this time point, as well. Results are shown in Figure 14D.

[0229] Sequences of the exemplary gRNAs that were tested in this example are listed in Table 13.

Example 6: Evaluation of ZFP in HepG2-NTCP Cells

[0230] In this example, ZF-off single constructs encoding a fusion protein consisting of KRAB, DNMT3A, DNMT3L, and an exemplary zinc finger motif of choice, were tested. Sequences of the exemplary zinc fingers that were tested in this example are listed in Table 20, as are sequences for plasmids yielding a subset of the ZF-off single construct fusion proteins.

[0231] Certain exemplary ZF-off constructs were formulated in a research-grade LNP. HepG2-NTCP cells were infected with HBV for 4 days and then transfected with the ZF-off loaded LNPs. As depicted in Figure 15A, at Day 6 post-infection HBsAg and HBeAg protein expression in the supernatant was evaluated by ELISA. Figure 15B shows the results as measured by percentage reduction in HBV antigens as compared to non-targeting control. Positive control used in this experiment is a HBV gRNA previously shown to reduce antigens ~50%. Figure 16A shows the results of the top ten ZF-off constructs that lead to the most reduction in HBV antigens. Figure 16B shows the results for all constructs in the screen.

[0232] Table 16 and 17 below show the raw data from these experiments, listed with the mRNA number yielding the zinc finger motif.

Table 16. % HBsAg expression relative to non-targeting control

Trial#	1	2	3	4	5	6	7	8

Non-targ control	100	100	100	100				
Pos control	54	59	68	61	75	79	65	86
mRNA0001	10	19	25	23				
mRNA0002	12	2	8	12				
mRNA0003	10	11	14	15				
mRNA0004	10	28	13	39				
mRNA0005	3	5	1	8				
mRNA0006	4	12	8	19				
mRNA0007	97	86	60	66				
mRNA0008	68	69	65	64				
mRNA0009	65	67	74	98				
mRNA0010	84	69	66	73				
mRNA0011	67	50	60	59				
mRNA0012	59	61	70	92				
mRNA0013	97	70	66	71				
mRNA0014	60	81	66	74				
mRNA0015	81	73	77	129				
mRNA0016	120	78	71	77				
mRNA0017	75	77	82	82				
mRNA0018	78	84	93	131				
mRNA0019	107	107	77	100				
mRNA0020	77	99	60	116				
mRNA0021	32	49	68	66				
mRNA0022	71	66	51	56				
mRNA0023	65	71	76	41				
mRNA0024	109	89	86	92				
mRNA0025	86	92	90	82				
mRNA0026	77	88	81	104				
mRNA0027	128	77	80	81				
mRNA0028	71	67	59	66				
mRNA0029	48	47	40	57				
mRNA0030	109	82	76	75				
mRNA0031	46	32	41	27				
mRNA0032	50	59	52	73				
mRNA0033	61	62	46	50				
mRNA0034	51	24	41	25				
mRNA0035	30	25	24	34				
mRNA0036	16	22	19	19				
mRNA0037	54	43	42	46				
mRNA0038	19	23	13	29				
mRNA0039	28	46	37	36				
mRNA0040	88	78	83	80				
mRNA0041	103	92	100					
mRNA0042	99	91	99					

mRNA0043	93	89	97					
mRNA0044	98	100	95					
mRNA0045	100	96	95					
mRNA0046	94	83	92					
mRNA0047	97	77	99					
mRNA0048	96	94	90					
mRNA0049	88	87	89					
mRNA0050	87	87	85					
mRNA0051	106	104	114					
mRNA0052	104	101	107					
mRNA0053	88	86	92					
mRNA0054	98	102	91					
mRNA0055	101	96	100					
mRNA0056	99	107	108					
mRNA0057	101	102	104					
mRNA0058	110	104	102					
mRNA0059	100	91	98					
mRNA0060	94	103	100					
mRNA0061	104	96	103					
mRNA0062	106	98	104					
mRNA0063	96	86	99					

Table 17. % HBeAg expression relative to non-targeting control

Trial#	100	100	100	100				
Non-targ control	100	100	100	100				
Pos control	26	36	41	53	43	43	34	54
mRNA0001	12	19	22	23				
mRNA0002	15	8	17	20				
mRNA0003	11	9	13	12				
mRNA0004	10	17	9	27				
mRNA0005	1	1	-1	3				
mRNA0006	5	8	7	13				
mRNA0007	95	78	59	65				
mRNA0008	64	67	60	65				
mRNA0009	65	64	81	98				
mRNA0010	84	68	69	70				
mRNA0011	65	51	51	67				
mRNA0012	64	61	74	96				
mRNA0013	92	74	73	79				
mRNA0014	58	85	58	76				
mRNA0015	82	83	78	124				
mRNA0016	108	81	72	80				
mRNA0017	72	77	72	80				

mRNA0018	55	55	71	93				
mRNA0019	71	79	51	87				
mRNA0020	34	36	32	52				
mRNA0021	32	40	55	55				
mRNA0022	77	64	53	65				
mRNA0023	60	69	72	43				
mRNA0024	98	76	87	84				
mRNA0025	91	86	82	92				
mRNA0026	78	97	87	102				
mRNA0027	117	62	68	74				
mRNA0028	75	59	58	71				
mRNA0029	31	32	22	45				
mRNA0030	124	86	79	77				
mRNA0031	42	23	27	20				
mRNA0032	46	57	57	82				
mRNA0033	56	51	44	76				
mRNA0034	42	21	41	18				
mRNA0035	22	22	24	39				
mRNA0036	13	17	16	13				
mRNA0037	50	35	34	35				
mRNA0038	12	16	13	25				
mRNA0039	29	45	39	36				
mRNA0040	93	73	80	82				
mRNA0041	80	63	111					
mRNA0042	114	94	98					
mRNA0043	98	91	99					
mRNA0044	91	115	108					
mRNA0045	71	55	62					
mRNA0046	76	66	63					
mRNA0047	55	55	45					
mRNA0048	66	63	78					
mRNA0049	83	59	52					
mRNA0050	51	55	49					
mRNA0051	55	49	49					
mRNA0052	56	57	66					
mRNA0053	92	60	57					
mRNA0054	50	55	56					
mRNA0055	83	88	74					
mRNA0056	61	69	112					
mRNA0057	106	73	65					
mRNA0058	66	65	65					
mRNA0059	69	66	71					
mRNA0060	59	94	101					
mRNA0061	111	81	68					

mRNA0062	28	33	41					
mRNA0063	65	55	31					

Example 7. Dose Response Testing of Viral Antigens in HepG2-NTCP cells

[0233] In this example, top ZF fusion proteins were tested in 5-point dose response assay for HBsAg and HBeAg. The 5 dosage points were 200ng, 150ng, 100ng, 50ng, and 25ng. Experimental schematic and results are shown in Figure 17.

Example 8. Testing for Durable Repression of HBsAg in HepG2.2.15 cells

[0234] In this example, top ZF fusion proteins were tested for durable repression of HBsAg. Active ZFPs showed durable silencing through Day 27 with 50ng total treatment. Experimental schematic and results are shown in Figure 18.

Example 9. Testing of Silencing of HBsAg in a Second Model for int-HBV

[0235] In this example, top ZF fusion proteins were tested for repression of HBsAg in PLC/PRF/5 cells. A subset of the ZFPs silenced HBsAg in this second model. Experimental schematic and results are shown in Figure 19.

Example 10. Testing ZF Fusion Proteins and CRISPR-off with guide RNAs for Specificity

[0236] In this example, ZF fusion proteins targeting HBV exhibiting significant silencing were profiled for specificity in HepG2-NTCP at day 19. All comparisons were performed against a non-targeting ZFP control. An exemplary result for the ZF fusion protein with mRNA0001 zinc finger motif is shown in Figure 20A. CRISPR-off with guide RNAs were similarly profiled. HepG2-NTCP cells were transfected with 100ng of total payload using GenVoyTM LNP at a 1:1 gRNA:effector ratio. Cells were split every 3-4 days and collected at day 15 post-treatment for specificity assessments, including RNA-seq and methylation array. DESeq2 was used to identify differential gene expression. As shown in Figure 20B, little to no changes were observed above chosen thresholds (absolute[log2[fold change]]>1 and -log10[adjusted p-value]>5) as expected for effectors targeting HBV DNA. For methylation array, the Infinium MethylationEPIC v2.0 array was used, and DMRs were identified *in silico*. EE3, EE4, and EE5 had a result of DMR=0. Results are shown in Figures 20C-20D.

Example 11. Stable HBV Silencing via Epigenetic Editing in Non-Transgenic Mouse Model of Persistent HBV infection

[0237] A non-transgenic model of persistent HBV infection (AAV-HBV) in immunocompetent mice was used, which was established by administering an adeno-associated viral vector (AAV) that contains HBV Genotype D DNA into the mice. The administration of the AAV-HBV vector resulted in expression of hepatitis B surface antigen (HBsAg), hepatitis B e antigen (HBeAg), and high levels of serum HBV DNA in the mice.

[0238] The CRISPR-off and ZF-off constructs are tested. Constructs are delivered via IV administration of mRNA/gRNA (CRISPR-Off) or mRNA (ZF-Off) formulated into a lipid nanoparticle (LNP) at 2.5 mg/kg and 0.5 mg/kg for CRISPR-Off and ZF-Off, respectively. Some constructs are formulated in LNP compositions as described in US20220402862A1 and/or US20230203480A1. A subset of the mice are re-dosed at two weeks after the first dose; a second subset are re-dosed at one month after the first dose. The readouts are circulating viral DNA, HBsAg, and HBeAg, tested using mouse plasma at one or more time points (such as 7, 14, 28, and 35 days). A durable and significant reduction in the levels of one or more of HBV DNA, HBsAg, and HBeAg is observed for some constructs.

[0239] Longer-term durability is tested over three to six months using the HBV DNA, HBsAg, and HBeAg markers. Progressive and durable reduction in one or more of these markers is seen with delivery of some constructs. The mice are sacrificed and livers are collected for further analysis, and durable silencing is confirmed by at least 2 log reduction of HBsAg and HBV DNA.

Example 12: Stable HBV Silencing via Epigenetic Editing in Transgenic Mice Expressing viral HBV DNA

[0240] A transgenic mouse model of persistent HBV infection (Tg-HBV) was used, whose genome was engineered to integrate HBV Genotype A DNA, resulting in expression of HBsAg and HBeAg, and circulating viral DNA in the mice.

[0241] The CRISPR-off and ZF-off constructs are tested. Constructs are delivered via IV administration of mRNA/gRNA (CRISPR-Off) or mRNA (ZF-Off) formulated into LNP at 2.5 mg/kg and 0.5 mg/kg for CRISPR-Off and ZF-Off, respectively. Some constructs are formulated in LNP compositions as described in US20220402862A1 and/or US20230203480A1. A subset of the mice are re-dosed at two weeks after the first dose; a second subset are re-dosed at one month after the first dose. The readouts are circulating viral DNA, HBsAg, and HBeAg, tested using mouse plasma at one or more time points (such as 7, 14, 28, and 35 days). A durable and significant reduction in the levels of one or more of HBV DNA, HBsAg, and HBeAg is observed for some constructs.

[0242] Longer-term durability is tested over three to six months using the HBV DNA, HBsAg, and HBeAg markers. Progressive and durable reduction in one or more of these markers is seen with delivery of some constructs. The mice are sacrificed and livers are collected for further analysis, and durable silencing is confirmed by at least 2 log reduction of HBsAg and HBV DNA.

Example 13. CRISPR-Off Guide RNA Multiplexing Study in AAV-HBV and Tg-HBV Mouse Models

[0243] AAV-HBV and Tg-HBV mice are injected with a single administration of one, two, or three guide RNAs with a CRISPR-Off fusion protein in LNPs at 1.5 mg/kg in accordance with Table 18. Samples are included with CRISPR-Off from each of PLA002 and PLA003. HBV DNA, HBsAg, and HBeAg are assayed in plasma at one or more time points, and the mouse liver is collected for further analysis. Durable silencing is confirmed by at least 2 log reduction of HBsAg and HBV DNA.

[0244] **Table 18. CRISPR-Off Multiplexing sample groups**

Group	Guide RNA 1	Guide RNA 2	Guide RNA 3
1	gRNA#008	gRNA#011	-
2	gRNA#008	gRNA#003	-
3	gRNA#008	gRNA#015	-
4	gRNA#008	gRNA#011	gRNA#015
5	gRNA#008	gRNA#011	gRNA#003
6	gRNA#008	-	-
7	Vehicle	-	-

Example 14. Zinc Finger Protein Multiplexing Study in AAV-HBV and Tg-HBV Mouse Models

[0245] AAV-HBV and Tg-HBV mice are injected with a single administration at 0.5 mg/kg of one, two, or three ZF fusion proteins in LNPs (schematic, Figure 21) in accordance with Table 19. HBV DNA, HBsAg, and HBeAg are assayed in plasma at one or more time points, and the mouse liver is collected for further analysis. Durable silencing is confirmed by at least 2 log reduction of HBsAg and HBV DNA.

Table 19. ZFP Multiplexing sample groups.

Group	ZF_Off-1	ZF_Off-2	ZF_Off-3
1	mRNA0004	mRNA0021	-
2	mRNA0004	mRNA0003	-
3	mRNA0004	mRNA0038	-
4	mRNA0004	mRNA0021	mRNA0003
5	mRNA0004	mRNA0038	mRNA0003
6	mRNA0004	mRNA0021	mRNA0038
7	mRNA0004	mRNA0001	-
8	mRNA0004	mRNA0039	-
9	mRNA0004	-	-
10	Vehicle	-	-

SEQUENCES

[0246] The SEQ ID NOs (SEQ) of nucleotide (nt) and amino acid (aa) sequences described in the present disclosure are listed in Table 20 below.

TABLE 20. Sequence listing.

SEQ	Description	Sequence
1	*S. pyogenes* WT Cas9 Sequence (nt)	ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAAT

SEQ	Description	Sequence
		CTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAA GATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGCG CAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATT TTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCA ATGATTAAACGCTACGATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGA CAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCA GGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTA GAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGC AAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCAT GCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATT GAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGT CGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAA GTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAA AATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTT TATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTT TCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACC GTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATT TCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATT ATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTT TTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCT CACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGA CGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTA GATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGAT AGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTA CATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACT GTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTT ATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGT ATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCT GTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGA GACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCAC ATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCT GATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAA AACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTA ACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAA TTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAAT ACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCT AAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAAT TACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAA TATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAA ATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCT AATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGC CCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTT GCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTA CAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATT GCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCT TATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTT AAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGAC TTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAA TATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTA CAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGT CATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAG CAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTT ATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAA CCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCT CCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAA GAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATT GATTTGAGTCAGCTAGGAGGTGACTGA
2	*S. pyogenes* WT Cas9 Sequence (aa)	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG

SEQ	Description	Sequence
		RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
3	SaCas9	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
4	*F. novicida* WT Cpf1	MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN
5	CasX	MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISNNAANNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEKRNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFVALTFERREVVDPSNIKPVNLIGVDRGENIPAVIALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFYHAVTHDAVLVFENLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKTLAQYTSKTCSNCGFTITTADYDGMLVRLKKTSDGWATTLNNKELKAEGQITYYNRYKRQTVEKELSAELDRLSEESGNNDISKWTK

SEQ	Description	Sequence
		GRRDEALFLLKKRFSHRPVQEQFVCLDCGHEVHADEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA
6	CasY	MRKKLFKGYILHNKRLVYTGKAAIRSIKYPLVAPNKTALNNLSEKIIYDYEHLFGPLNVASYARNSNRYSLVDFWIDSLRAGVIWQSKSTSLIDLISKLEGSKSPSEKIFEQIDFELKNKLDKEQFKDIILLNTGIRSSSNVRSLRGRFLKCFKEEFRDTEEVIACVDKWSKDLIVEGKSILVSKQFLYWEEEFGIKIFPHFKDNHDLPKLTFFVEPSLEFSPHLPLANCLERLKKFDISRESLLGLDNNFSAFSNYFNELFNLLSRGEIKKIVTAVLAVSKSWENEPELEKRLHFLSEKAKLLGYPKLTSSWADYRMIIGGKIKSWHSNYTEQLIKVREDLKKHQIALDKLQEDLKKVVDSSLREQIEAQREALLPLLDTMLKEKDFSDDLELYRFILSDFKSLLNGSYQRYIQTEEERKEDRDVTKKYKDLYSNLRNIPRFFGESKKEQFNKFINKSLPTIDVGLKILEDIRNALETVSVRKPPSITEEYVTKQLEKLSRKYKINAFNSNRFKQITEQVLRKYNNGELPKISEVFYRYPRESHVAIRILPVKISNPRKDISYLLDKYQISPDWKNSNPGEVVDLIEIYKLTLGWLLSCNKDFSMDFSSYDLKLFPEAASLIKNFGSCLSGYYLSKMIFNCITSEIKGMITLYTRDKFVVRYVTQMIGSNQKFPLLCLVGEKQTKNFSRNWGVLIEEKGDLGEEKNQEKCLIFKDKTDFAKAKEVEIFKNNIWRIRTSKYQIQFLNRLFKKTKEWDLMNLVLSEPSLVLEEEWGVSWDKDKLLPLLKKEKSCEERLYYSLPLNLVPATDYKEQSAEIEQRNTYLGLDVGEFGVAYAVVRIVRDRIELLSWGFLKDPALRKIRERVQDMKKKQVMAVFSSSSTAVARVREMAIHSLRNQIHSIALAYKAKIIYEISISNFETGGNRMAKIYRSIKVSDVYRESGADTLVSEMIWGKKNKQMGNHISSYATSYTCCNCARTPFELVIDNDKEYEKGGDEFIFNVGDEKKVRGFLQKSLLGKTIKGKEVLKSIKEYARPPIREVLLEGEDVEQLLKRRGNSYIYRCPFCGYKTDADIQAALNIACRGYISDNAKDAVKEGERKLDYILEVRKLWEKNGAVLRSAKFL
7	CasPhi	MADTPTLFTQFLRHHLPGQRFRKDILKQAGRILANKGEDATIAFLRGKSEESPPDFQPPVKCPIIACSRPLTEWPIYQASVAIQGYVYGQSLAEFEASDPGCSKDGLLGWFDKTGVCTDYFSVQGLNLIFQNARKRYIGVQTKVTNRNEKRHKKLKRINAKRIAEGLPELTSDEPESALDETGHLIDPPGLNTNIYCYQQVSPKPLALSEVNQLPTAYAGYSTSGDDPIQPMVTKDRLSISKGQPGYIPEHQRALLSQKKHRRMRGYGLKARALLVIVRIQDDWAVIDLRSLLRNAYWRRIVQTKEPSTITKLLKLVTGDPVLDATRMVATFTYKPGIVQVRSAKCLKNKQGSKLFSERYLNETVSVTSIDLGSNNLVAVATYRLVNGNTPELLQRFTLPSHLVKDFERYKQAHDTLEDSIQKTAVASLPQGQQTEIRMWSMYGFREAQERVCQELGLADGSIPWNVMTATSTILTDLFLARGGDPKKCMFTSEPKKKKNSKQVLYKIRDRAWAKMYRTLLSKETREAWNKALWGLKRGSPDYARLSKRKEELARRCVNYTISTAEKRAQCGRTIVALEDLNIGFFHGRGKQEPGWVGLFTRKKENRWLMQALHKAFLELAHHRGYHVIEVNPAYTSQTCPVCRHCDPDNRDQHNREAFHCIGCGFRGNADLDVATHNIAMVAITGESLKRARGSVASKTPQPLAAE
8	Cas12f1 (Cas14a)	MIKVYRYEIVKPLDLDWKEFGTILRQLQQETRFALNKATQLAWEWMGFSSDYKDNHGEYPKSKDILGYTNVHGYAYHTIKTKAYRLNSGNLSQTIKRATDRFKAYQKEILRGDMSIPSYKRDIPLDLIKENISVNRMNHGDYIASLSLLSNPAKQEMNVKRKISVIIIVRGAGKTIMDRILSGEYQVSASQIIHDDRKNKWYLNISYDFEPQTRVLDLNKIMGIDLGVAVAVYMAFQHTPARYKLEGGEIENFRRQVESRRISMLRQGKYAGGARGGHGRDKRIKPIEQLRDKIANFRDTTNHRYSRYIVDMAIKEGCGTIQMEDLTNIRDIGSRFLQNWTYYDLQQKIIYKAEEAGIKVIKIDPQYTSQRCSECGNIDSGNRIGQAIFKCRACGYEANADYNAARNIAIPNIDKIIAESIKSGGS
9	Cas12f2 (Cas14b)	NAMIAQKTIKIKLNPTKEQIIKLNSIIEEYIKVSNFTAKKIAEIQESFTDSGLTQGTCSECGKEKTYRKYHLLKKDNKLFCITCYKRKYSQFTLQKVEFQNKTGLRNVAKLPKTYYTNAIRFASDTFSGFDEIIKKKQNRLNSIQNRLNFWKELLYNPSNRNEIKIKVVKYAPKTDTREHPHYYSEAEIKGRIKRLEKQLKKFKMPKYPEFTSETISLQRELYSWKNPDELKISSITDKNESMNYYGKEYLKRYIDLINSQTPQILLEKENNSFYLCFPITKNIEMPKIDDTFEPVGIDWGITRNIAVVSILDSKTKKPKFVKFYSAGYILGKRKHYKSLRKHFGQKKRQDKINKLGTKEDRFIDSNIHKLAFLIVKEIRNHSNKPIILMENITDNREEAEKSMRQNILLHSVKSRLQNYIAYKALWNNIPTNLVKPEHTSQICNRCGHQDRENRPKGSKLFKCVKCNYMSNADFNASINIARKFYIGEYEPFYKDNEKMKSGVNSISM
10	Cas12f3 (Cas14c)	MEVQKTVMKTLSLRILRPLYSQEIEKEIKEEEKERRKQAGGTGELDGGFYKKLEKKHSEMFSFDRLNLLLNQLQREIAKVYNHAISELYIATIAQGNKSNKHYISSIVYNRAYGYFYNAYIALGICSKVEANFRSNELLTQQSALPTAKSDNFPIVLHKQKGAEGEDGGFRISTEGSDLIFEIPIPFYEYNGENRKEPYKWVKKGGQKPVLKLILSTFRRQRNKGWAKDEGTDAEIRKVTEGKYQVSQIEINRGKKLGEHQKWFANFSIEQPIYERKPNRSIVGGLDVGIRSPLVCAINNSFSRYSVDSNDVFKFSKQVFAFRRRLLSKNSLKRKHGHAAHKLEPITEMTEKNDKFRKKIIERWAKEVTNFFVKNQVGIVQIEDLSTMKDREDHFFNQYLRGFWPYYQMQTLIENKLKEYGIEVKRVQAKYTSQLCSNPNCRYWNNYFNFEYRKVNKFPKFKCEKCNLEISADYNAARNLSTPDIEKFVAKATKGINLPEK
11	C2c8	MKVLEFKIHPTEEQVSKIDQSLAACKLLWNLSIALKEESKQRYYRKKHKFDEFSPEIWGLSYSGHYDEKEFKTLKDKEKKLLIGNPCCKIAYFKKTSNGKEYTPLNSIPIRRFMNAENIDKDAVNYLNRKKLAFYFRENTAKFIGEIETEFKKGFFKSVIKPAYDAAKKGIRGIPRFKGRRDKVETLVNGQPETIKIKSNGVIVSSKIGLLKIRGLDRLQGKAPRMAKITRKATGYYLQLTIETDDTIYKESDKCVGLDMGAVAIFTDDLGRQSEAKRYAKIQKKRLNRLQRQASRQKDNSNNQRKTYAKLARVHEKIARQRKGRNAQLAHKITSEYQSVILEDLNLKNMTAAAKPKEREDGDGYKQNGKKRKSGLNKALLDNAIGQLRTFIENKANERGRKIIRVNPKHTSQTCPNCGNIDKANRVSQSKFKCVSCGYEAHADQNAAANILIRGLRDEFLRAIGSLYKFPVSMIGKYPG

SEQ	Description	Sequence
		LAGEFTPDLDANQESIGDAPIENAEHSISKQMKQEGNRTPTQPENGSQSLIFLSAPPQPCGDSHGTNNPKALPNKASKRSSKKPRGAIPENPDQLTIWDLLD
12	dSpCas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
13	dSaCas9	MKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
14	inactive FnCpf1	MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN
15	dNmeCas9	MAAFKPNSINYILGLAIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRTRRLLKREGVLQAANFDENGLIKSLPNTPWQLRAAALDR

SEQ	Description	Sequence
		KLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQRSDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDAALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSAKRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDKYYLVPIYSWQVAKGILPDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR
16	dCjCas9	MARILAFAIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGKVHKINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQDEKMLEIDAIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK
17	dSt1Cas9	MGSDLVLGLAIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDAILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF
18	dSt3Cas9	MTKPYSIGLAIGTNSVGWAVITDNYKVPSKKMKVLGNTSKKYIKKNLLGVLLFDSGITAEGRRLKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSFLVPDDKRDSKYPIFGNLVEEKVYHDEFPTIYHLRKYLADSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSKNNDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKKDRILKLFPGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLETLLGYIGDDYSDVFLKAKKLYDAILLSGFLTVTDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNISLKTYNEVFKDDTKNGYAGYIDGKTNQEDFYVYLKNLLAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQAKFYPFLAKNKERIEKILTFRIPYYVGPLARGNSDFAWSIRKRNEKITPWNFEDVIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVRFIAESMRDYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIELKGIEKQFNSSLSTYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLSRRHYTGWGKLSAKLINGIRDEKSGNTILDYLIDDGISNRNFMQLIHDDALSFKKKIQKAQIIGDEDKGNIKEVVKSLPGSPAIKKGILQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQGKSNSQQRLKRLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLYYLQNGKDMYTGDDLDIDRL

SEQ	Description	Sequence
		SNYDIDHIIPQAFLKDNSIDNKVLVSSASARGKSDDFPSLEVVKKRKTFWYQLLKSKLISQRKFDNLTKAERGGLLPEDKAGFIQRQLVETRQITKHVARLLDEKFNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYKVREINDFHHAHDAYLNAVIASALLKKYPKLEPEFVYGDYPKYNSFRERKSATEKVYFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLATVRRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANLSSKPKPNSNENLVGAKEYLDPKKYGGYAGISNSFAVLVKGTIEKGAKKKITNVLEFQGISILDRINYRKDKLNFLLEKGYKDIELIIELPKYSLFELSDGSRRMLASILSTNNKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKEFEELFYYILEFNENYVGAKKNGKLLNSAFQSWQNHSIDELCSSFIGPTGSERKGLFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKDATLIHQSVTGLYETRIDLAKLGEG
19	dLbCpf1	MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVVHPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLEYAQTSVKH
20	inactive AsCpf1	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIARGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN
21	inactive enAsCpf1	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYRNRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLARGWDVNREKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSD

SEQ	Description	Sequence
		EARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIARGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN
22	inactive HFAsCpf1	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYRNRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLALAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLARGWDVNREKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIARGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN
23	inactive RVRAsCpf1	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLARGWDVNVEKNRGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIARGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN
24	inactive RRAsCpf1	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLARGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPRCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHP

SEQ	Description	Sequence
		ETPIIGIARGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN
25	dCasX	MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISNNAANNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPEKDSDEAVTYSLGKFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEKRNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFVALTFERREVVDPSNIKPVNLIGVARGENIPAVIALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFYHAVTHDAVLVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKTLAQYTSKTCSNCGFTITTADYDGMLVRLKKTSDGWATTLNNKELKAEGQITYYNRYKRQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFLLKKRFSHRPVQEQFVCLDCGHEVHAAEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA
26	dCasPhi	MPKPAVESEFSKVLKKHFPGERFRSSYMKRGGKILAAQGEEAVVAYLQGKSEEEPPNFQPPAKCHVVTKSRDFAEWPIMKASEAIQRYIYALSTTERAACKPGKSSESHAAWFAATGVSNHGYSHVQGLNLIFDHTLGRYDGVLKKVQLRNEKARARLESINASRADEGLPEIKAEEEEVATNETGHLLQPPGINPSFYVYQTISPQAYRPRDEIVLPPEYAGYVRDPNAPIPLGVVRNRCDIQKGCPGYIPEWQREAGTAISPKTGKAVTVPGLSPKKNKRMRRYWRSEKEKAQDALLVTVRIGTDWVVIDVRGLLRNARWRTIAPKDISLNALLDLFTGDPVIDVRRNIVTFTYTLDACGTYARKWTLKGKQTKATLDKLTATQTVALVAIALGQTNPISAGISRVTQENGALQCEPLDRFTLPDDLLKDISAYRIAWDRNEEELRARSVEALPEAQQAEVRALDGVSKETARTQLCADFGLDPKRLPWDKMSSNTTFISEALLSNSVSRDQVFFTPAPKKGAKKKAPVEVMRKDRTWARAYKPRLSVEAQKLKNEALWALKRTSPEYLKLSRRKEELCRRSINYVIEKTRRRTQCQIVIPVIEDLNVRFFHGSGKRLPGWDNFFTAKKENRWFIQGLHKAFSDLRTHRSFYVFEVRPERTSITCPKCGHCEVGNRDGEAFQCLSCGKTCNADLDVATHNLTQVALTGKTMPKREEPRDAQGTAPARKTKKASKSKAPPAEREDQTPAQEPSQTS
27	inactive VRER SpCas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD
28	inactive EQR SpCas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH

SEQ	Description	Sequence
		AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDA IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFESPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD
29	inactive VQR SpCas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDA IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD
30	inactive SPG SpCas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDA IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFLWPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD
31	inactive SpRY Cas9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE RTRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

SEQ	Description	Sequence
		LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDKLIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTRLGAPRAFKYFDTTIDPKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD
32	inactive KKH dSaCas9	MKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
33	mRNA0001	SRPGERPFQCRICMRNFSKKFNLLQHTRTHTGEKPFQCRICMRNFSRQDNLNSHLRTHTGSQKPFQCRICMRNFSRSHNLKLHTRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHTGSQKPFQCRICMRNFSRNTNLTRHTRTHTGEKPFQCRICMRNFSIKHNLARHLRTHLRGS
34	mRNA0002	SRPGERPFQCRICMRNFSKKFNLLQHTRTHTGEKPFQCRICMRNFSRKDYLISHLRTHTGSQKPFQCRICMRNFSRSHNLKLHTRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHTGSQKPFQCRICMRNFSRQDNLGRHLRTHTGEKPFQCRICMRNFSVVNNLNRHLKTHLRGS
35	mRNA0003	SRPGERPFQCRICMRNFSKKFNLLQHTRTHTGEKPFQCRICMRNFSRKDYLISHLRTHTGSQKPFQCRICMRNFSRSHNLRLHTRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHTGSQKPFQCRICMRNFSRQDNLGRHLRTHTGEKPFQCRICMRNFSVVNNLNRHLKTHLRGS
36	mRNA0004	SRPGERPFQCRICMRNFSRRHILDRHTRTHTGEKPFQCRICMRNFSRQDNLGRHLRTHTGSQKPFQCRICMRNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSRRDGLAGHLKTHTGSQKPFQCRICMRNFSVHHNLVRHLRTHTGEKPFQCRICMRNFSISHNLARHLKTHLRGS
37	mRNA0005	SRPGERPFQCRICMRNFSRREVLENHLRTHTGEKPFQCRICMRNFSRRDNLNRHLKTHTGSQKPFQCRICMRNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSRRDGLAGHLKTHTGSQKPFQCRICMRNFSVHHNLVRHLRTHTGEKPFQCRICMRNFSISHNLARHLKTHLRGS
38	mRNA0006	SRPGERPFQCRICMRNFSRRAVLDRHTRTHTGEKPFQCRICMRNFSRQDNLGRHLRTHTGSQKPFQCRICMRNFSQSTTLKRHLRTHTGEKPFQCRICMRNFSRRDGLAGHLKTHTGSQKPFQCRICMRNFSVHHNLVRHLRTHTGEKPFQCRICMRNFSISHNLARHLKTHLRGS
39	mRNA0064	SRPGERPFQCRICMRNFSRQEHLVRHLRTHTGEKPFQCRICMRNFSEGGNLMRHLKTHTGSQKPFQCRICMRNFSSDRRDLDHTRTHTGEKPFQCRICMRNFSSFQSYLEHLRTHTGSQKPFQCRICMRNFSRPNHLAIHTRTHTGEKPFQCRICMRNFSQSPHLKRHLRTHLRGS
40	mRNA0007	SRPGERPFQCRICMRNFSRREHLVRHLRTHTGEKPFQCRICMRNFSDPSNLQRHLKTHTGSQKPFQCRICMRNFSSDRRDLDHTRTHTGEKPFQCRICMRNFSSFQSYLEHLRTHTGSQKPFQCRICMRNFSRPNHLAIHTRTHTGEKPFQCRICMRNFSQSPHLKRHLRTHLRGS
41	mRNA0008	SRPGERPFQCRICMRNFSRREHLVRHLRTHTGEKPFQCRICMRNFSDMGNLGRHLKTHTGSQKPFQCRICMRNFSSDRRDLDHTRTHTGEKPFQCRICMRNFSSFQSYLEHLRTHTGSQKPFQCRICMRNFSRPNHLAIHTRTHTGEKPFQCRICMRNFSQSPHLKRHLRTHLRGS
42	mRNA0009	SRPGERPFQCRICMRNFSKKDHLHRHTRTHTGEKPFQCRICMRNFSQKEILTRHLRTHTGSQKPFQCRICMRNFSQSAHLKRHLRTHTGEKPFQCRICMRNFSETGSLRRHLKTHTGGGGSQKPFQCRICMRNFSQSHSLKSHLRTHTGEKPFQCRICMRNFSESGHLKRHLKTHLRGS

SEQ	**Description**	**Sequence**
43	mRNA0010	SRPGERPFQCRICMRNFSKKDHLHRHTRTHTGEKPFQCRICMRNFSQKEILTRHLRTHTGSQKPFQCRICMRNFSQSAHLKRHLRTHTGEKPFQCRICMRNFSDRTPLNRHLKTHTGGGGSQKPFQCRICMRNFSQSHSLKSHLRTHTGEKPFQCRICMRNFSESGHLKRHLKTHLRGS
44	mRNA0011	SRPGERPFQCRICMRNFSKTDHLARHTRTHTGEKPFQCRICMRNFSQKEILTRHLRTHTGSQKPFQCRICMRNFSQSAHLKRHLRTHTGEKPFQCRICMRNFSETGSLRRHLKTHTGGGGSQKPFQCRICMRNFSQKHHLVTHLRTHTGEKPFQCRICMRNFSENSKLRRHLKTHLRGS
45	mRNA0012	SRPGERPFQCRICMRNFSQAGNLVRHLRTHTGEKPFQCRICMRNFSQNSHLRRHLKTHTGGGGSQKPFQCRICMRNFSDLSTLRRHTRTHTGEKPFQCRICMRNFSQNEHLKVHLRTHTGSQKPFQCRICMRNFSGGTALRMHTRTHTGEKPFQCRICMRNFSQRSSLVRHLRTHLRGS
46	mRNA0013	SRPGERPFQCRICMRNFSQRGNLQRHLRTHTGEKPFQCRICMRNFSQTTHLSRHLKTHTGGGGSQKPFQCRICMRNFSDGSTLRRHTRTHTGEKPFQCRICMRNFSQKTHLAVHLRTHTGSQKPFQCRICMRNFSGGTALRMHTRTHTGEKPFQCRICMRNFSQRSSLVRHLRTHLRGS
47	mRNA0014	SRPGERPFQCRICMRNFSQRGNLQRHLRTHTGEKPFQCRICMRNFSQTTHLSRHLKTHTGGGGSQKPFQCRICMRNFSDLSTLRRHTRTHTGEKPFQCRICMRNFSQNEHLKVHLRTHTGSQKPFQCRICMRNFSGGSALSMHTRTHTGEKPFQCRICMRNFSQRSSLVRHLRTHLRGS
48	mRNA0015	SRPGERPFQCRICMRNFSDRGNLTRHLRTHTGEKPFQCRICMRNFSQARSLRAHLKTHTGGGGSQKPFQCRICMRNFSEKASLIKHTRTHTGEKPFQCRICMRNFSDHSSLKRHLRTHTGSQKPFQCRICMRNFSRRFILSRHTRTHTGEKPFQCRICMRNFSRNDSLKCHLRTHLRGS
49	mRNA0016	SRPGERPFQCRICMRNFSDRGNLTRHLRTHTGEKPFQCRICMRNFSQARSLRAHLKTHTGGGGSQKPFQCRICMRNFSDKSSLRKHTRTHTGEKPFQCRICMRNFSDHSSLKRHLRTHTGSQKPFQCRICMRNFSRNFILQRHTRTHTGEKPFQCRICMRNFSRNDTLIIHLRTHLRGS
50	mRNA0017	SRPGERPFQCRICMRNFSDRGNLTRHLRTHTGEKPFQCRICMRNFSQARSLRAHLKTHTGGGGSQKPFQCRICMRNFSCNGSLKKHTRTHTGEKPFQCRICMRNFSDHSSLKRHLRTHTGSQKPFQCRICMRNFSRNFILQRHTRTHTGEKPFQCRICMRNFSRNDTLIIHLRTHLRGS
51	mRNA0018	SRPGERPFQCRICMRNFSRTDTLARHLRTHTGEKPFQCRICMRNFSRTDSLPRHLKTHTGGGGSQKPFQCRICMRNFSDHSSLKRHLRTHTGEKPFQCRICMRNFSQPHGLAHHLKTHTGSQKPFQCRICMRNFSQSAHLKRHLRTHTGEKPFQCRICMRNFSVGNSLSRHLKTHLRGS
52	mRNA0019	SRPGERPFQCRICMRNFSRTDTLARHLRTHTGEKPFQCRICMRNFSRTDSLPRHLKTHTGGGGSQKPFQCRICMRNFSDHSSLKRHLRTHTGEKPFQCRICMRNFSQPHGLRHHLKTHTGSQKPFQCRICMRNFSQSAHLKRHLRTHTGEKPFQCRICMRNFSVGNSLSRHLKTHLRGS
53	mRNA0020	SRPGERPFQCRICMRNFSRTDTLARHLRTHTGEKPFQCRICMRNFSRLDMLARHLKTHTGGGGSQKPFQCRICMRNFSDHSSLKRHLRTHTGEKPFQCRICMRNFSQPHGLSTHLKTHTGSQKPFQCRICMRNFSQQAHLVRHTRTHTGEKPFQCRICMRNFSVHESLKRHLRTHLRGS
54	mRNA0021	SRPGERPFQCRICMRNFSRADNLGRHLRTHTGEKPFQCRICMRNFSRNTHLSYHLKTHTGSQKPFQCRICMRNFSRGDGLRRHLRTHTGEKPFQCRICMRNFSRRDNLNRHLKTHTGSQKPFQCRICMRNFSRARNLTLHTRTHTGEKPFQCRICMRNFSDPSSLKRHLRTHLRGS
55	mRNA0022	SRPGERPFQCRICMRNFSRADNLGRHLRTHTGEKPFQCRICMRNFSRNTHLSYHLKTHTGSQKPFQCRICMRNFSRKLGLLRHTRTHTGEKPFQCRICMRNFSRQDNLGRHLRTHTGSQKPFQCRICMRNFSRARNLTLHTRTHTGEKPFQCRICMRNFSDPSSLKRHLRTHLRGS
56	mRNA0023	SRPGERPFQCRICMRNFSRADNLGRHLRTHTGEKPFQCRICMRNFSRNTHLSYHLKTHTGSQKPFQCRICMRNFSRKLGLLRHTRTHTGEKPFQCRICMRNFSRQDNLGRHLRTHTGSQKPFQCRICMRNFSRRRNLQLHTRTHTGEKPFQCRICMRNFSDHSSLKRHLRTHLRGS
57	mRNA0024	SRPGERPFQCRICMRNFSQQSSLLRHTRTHTGEKPFQCRICMRNFSRREHLVRHLRTHTGSQKPFQCRICMRNFSGLTALRTHTRTHTGEKPFQCRICMRNFSERAKLIRHLRTHTGGGGSQKPFQCRICMRNFSAKRDLDRHTRTHTGEKPFQCRICMRNFSVNSSLTRHLRTHLRGS
58	mRNA0025	SRPGERPFQCRICMRNFSQQSSLLRHTRTHTGEKPFQCRICMRNFSRREHLVRHLRTHTGSQKPFQCRICMRNFSGLTALRTHTRTHTGEKPFQCRICMRNFSERAKLIRHLRTHTGGGGSQKPFQCRICMRNFSLRKDLVRHTRTHTGEKPFQCRICMRNFSVRHSLTRHLRTHLRGS
59	mRNA0026	SRPGERPFQCRICMRNFSQASALSRHTRTHTGEKPFQCRICMRNFSRREHLVRHLRTHTGSQKPFQCRICMRNFSGLTALRTHTRTHTGEKPFQCRICMRNFSERAKLIRHLRTHTGGGGSQKPFQCRICMRNFSAKRDLDRHTRTHTGEKPFQCRICMRNFSVNSSLTRHLRTHLRGS
60	mRNA0061	SRPGERPFQCRICMRNFSRGRNLEMHTRTHTGEKPFQCRICMRNFSDSSVLRRHLRTHTGGGGSQKPFQCRICMRNFSQNANLKRHTRTHTGEKPFQCRICMRNFSQKHHLAVHLRTHTGSQKPFQCRICMRNFSQRSNLARHLRTHTGEKPFQCRICMRNFSQKVHLEAHLKTHLRGS
61	mRNA0027	SRPGERPFQCRICMRNFSRRRNLDVHTRTHTGEKPFQCRICMRNFSDSSVLRRHLRTHTGGGGSQKPFQCRICMRNFSQNANLKRHTRTHTGEKPFQCRICMRNFSQKHHLAVHLRTHTGSQKPFQCRICMRNFSQRSNLARHLRTHTGEKPFQCRICMRNFSQKVHLEAHLKTHLRGS
62	mRNA0065	SRPGERPFQCRICMRNFSRGRNLAIHTRTHTGEKPFQCRICMRNFSDSSVLRRHLRTHTGGGGSQKPFQCRICMRNFSLKSNLHRHTRTHTGEKPFQCRICMRNFSLKQHLVVHLRTHTGSQKPFQCRICMRNFSLKTNLARHTRTHTGEKPFQCRICMRNFSQKCHLKAHLRTHLRGS
63	mRNA0028	SRPGERPFQCRICMRNFSDGSNLRRHLRTHTGEKPFQCRICMRNFSRIDNLDGHLKTHTGSQKPFQCRICMRNFSQRRYLVEHTRTHTGEKPFQCRICMRNFSQQTNLARHLRTHTGGGGSQKPFQCRICMRNFSQRSDLTRHLRTHTGEKPFQCRICMRNFSRGDNLNRHLKTHLRGS
64	mRNA0029	SRPGERPFQCRICMRNFSDPSNLQRHLRTHTGEKPFQCRICMRNFSRRDNLPKHLKTHTGSQKPFQCRICMRNFSTTFNLRVHTRTHTGEKPFQCRICMRNFSQTQNLTRHLRTHTGGGGSQKPFQCRICMRNFSHKETLNRHLRTHTGEKPFQCRICMRNFSREDNLGRHLKTHLRGS

SEQ	Description	Sequence
65	mRNA0030	SRPGERPFQCRICMRNFSDPSNLQRHLRTHTGEKPFQCRICMRNFSRRDNLPKHLKTHTGSQKPFQCRICMRNFSQRRYLVEHTRTHTGEKPFQCRICMRNFSQQTNLARHLRTHTGGGGSQKPFQCRICMRNFSQRSDLTRHLRTHTGEKPFQCRICMRNFSRGDNLNRHLKTHLRGS
66	mRNA0031	SRPGERPFQCRICMRNFSQQTNLTRHLRTHTGEKPFQCRICMRNFSANRTLVHHLKTHTGSQKPFQCRICMRNFSEEANLRRHTRTHTGEKPFQCRICMRNFSRGEHLTRHLRTHTGSQKPFQCRICMRNFSTNSSLTRHLRTHTGEKPFQCRICMRNFSRIDNLIRHLKTHLRGS
67	mRNA0032	SRPGERPFQCRICMRNFSQQTNLTRHLRTHTGEKPFQCRICMRNFSANRTLVHHLKTHTGSQKPFQCRICMRNFSEEANLRRHTRTHTGEKPFQCRICMRNFSRREHLVRHLRTHTGSQKPFQCRICMRNFSMTSSLRRHTRTHTGEKPFQCRICMRNFSRQDNLGRHLRTHLRGS
68	mRNA0033	SRPGERPFQCRICMRNFSQQTNLTRHLRTHTGEKPFQCRICMRNFSANRTLVHHLKTHTGSQKPFQCRICMRNFSEEANLRRHTRTHTGEKPFQCRICMRNFSRGEHLTRHLRTHTGSQKPFQCRICMRNFSMTSSLRRHTRTHTGEKPFQCRICMRNFSRQDNLGRHLRTHLRGS
69	mRNA0034	SRPGERPFQCRICMRNFSRATHLTRHTRTHTGEKPFQCRICMRNFSRADVLKGHLRTHTGSQKPFQCRICMRNFSQRSSLVRHLRTHTGEKPFQCRICMRNFSRKDALHVHLKTHTGSQKPFQCRICMRNFSVHHNLVRHLRTHTGEKPFQCRICMRNFSISHNLARHLKTHLRGS
70	mRNA0035	SRPGERPFQCRICMRNFSRATHLTRHTRTHTGEKPFQCRICMRNFSRADVLKGHLRTHTGSQKPFQCRICMRNFSQSSSLVRHLRTHTGEKPFQCRICMRNFSRKERLATHLKTHTGSQKPFQCRICMRNFSVRHNLTRHLRTHTGEKPFQCRICMRNFSISHNLARHLKTHLRGS
71	mRNA0036	SRPGERPFQCRICMRNFSKKDHLHRHTRTHTGEKPFQCRICMRNFSRKESLTVHLRTHTGSQKPFQCRICMRNFSQSSSLVRHLRTHTGEKPFQCRICMRNFSRKERLATHLKTHTGSQKPFQCRICMRNFSVHHNLVRHLRTHTGEKPFQCRICMRNFSISHNLARHLKTHLRGS
72	mRNA0037	SRPGERPFQCRICMRNFSRVDHLHRHLRTHTGEKPFQCRICMRNFSRREHLSGHLKTHTGGGGSQKPFQCRICMRNFSQSSSLVRHLRTHTGEKPFQCRICMRNFSRKERLATHLKTHTGSQKPFQCRICMRNFSVAHNLTRHLRTHTGEKPFQCRICMRNFSISHNLARHLKTHLRGS
73	mRNA0038	SRPGERPFQCRICMRNFSRKHHLGRHTRTHTGEKPFQCRICMRNFSRREHLTIHLRTHTGGGGSQKPFQCRICMRNFSQSSSLVRHLRTHTGEKPFQCRICMRNFSRKERLATHLKTHTGSQKPFQCRICMRNFSVAHNLTRHLRTHTGEKPFQCRICMRNFSISHNLARHLKTHLRGS
74	mRNA0039	SRPGERPFQCRICMRNFSRVDHLHRHLRTHTGEKPFQCRICMRNFSRSDHLSLHLKTHTGGGGSQKPFQCRICMRNFSQSSSLVRHLRTHTGEKPFQCRICMRNFSRKERLATHLKTHTGSQKPFQCRICMRNFSVAHNLTRHLRTHTGEKPFQCRICMRNFSISHNLARHLKTHLRGS
75	mRNA0040	SRPGERPFQCRICMRNFSKTDHLARHTRTHTGEKPFQCRICMRNFSQKEILTRHLRTHTGSQKPFQCRICMRNFSQSAHLKRHLRTHTGEKPFQCRICMRNFSETGSLRRHLKTHTGSQKPFQCRICMRNFSQSSSLVRHLRTHTGEKPFQCRICMRNFSQTNTLGRHLKTHLRGS
76	mRNA0041	SRPGERPFQCRICMRNFSKKDHLHRHTRTHTGEKPFQCRICMRNFSQKEILTRHLRTHTGSQKPFQCRICMRNFSQSAHLKRHLRTHTGEKPFQCRICMRNFSETGSLRRHLKTHTGSQKPFQCRICMRNFSQSSSLVRHLRTHTGEKPFQCRICMRNFSQGGTLRRHLKTHLRGS
77	mRNA0042	SRPGERPFQCRICMRNFSKKDHLHRHTRTHTGEKPFQCRICMRNFSQKEILTRHLRTHTGSQKPFQCRICMRNFSQSAHLKRHLRTHTGEKPFQCRICMRNFSDPTSLNRHLKTHTGSQKPFQCRICMRNFSQSSSLVRHLRTHTGEKPFQCRICMRNFSQTNTLGRHLKTHLRGS
78	mRNA0043	SRPGERPFQCRICMRNFSQQTNLTRHLRTHTGEKPFQCRICMRNFSVGGNLARHLKTHTGSQKPFQCRICMRNFSKRYNLYQHTRTHTGEKPFQCRICMRNFSRQDNLNTHLRTHTGSQKPFQCRICMRNFSRSHNLKLHTRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHLRGS
79	mRNA0044	SRPGERPFQCRICMRNFSQQTNLTRHLRTHTGEKPFQCRICMRNFSVGGNLSRHLKTHTGSQKPFQCRICMRNFSKRYNLYQHTRTHTGEKPFQCRICMRNFSRQDNLNTHLRTHTGSQKPFQCRICMRNFSRSHNLRLHTRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHLRGS
80	mRNA0045	SRPGERPFQCRICMRNFSQQTNLTRHLRTHTGEKPFQCRICMRNFSVGGNLSRHLKTHTGSQKPFQCRICMRNFSKKFNLLQHTRTHTGEKPFQCRICMRNFSRRDNLKSHLRTHTGSQKPFQCRICMRNFSRSHNLKLHTRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHLRGS
81	mRNA0046	SRPGERPFQCRICMRNFSDKSSLRKHTRTHTGEKPFQCRICMRNFSDHSSLKRHLRTHTGSQKPFQCRICMRNFSRNFILQRHTRTHTGEKPFQCRICMRNFSRNDTLIIHLRTHTGGGGSQKPFQCRICMRNFSTSTLLKRHTRTHTGEKPFQCRICMRNFSLKEHLTRHLRTHLRGS
82	mRNA0047	SRPGERPFQCRICMRNFSCNGSLKKHTRTHTGEKPFQCRICMRNFSDHSSLKRHLRTHTGSQKPFQCRICMRNFSRNFILARHTRTHTGEKPFQCRICMRNFSRQDILVVHLRTHTGGGGSQKPFQCRICMRNFSHKSSLTRHLRTHTGEKPFQCRICMRNFSESGHLKRHLKTHLRGS
83	mRNA0048	SRPGERPFQCRICMRNFSCNGSLKKHTRTHTGEKPFQCRICMRNFSDHSSLKRHLRTHTGSQKPFQCRICMRNFSRNFILARHTRTHTGEKPFQCRICMRNFSRQDILVVHLRTHTGGGGSQKPFQCRICMRNFSTSTLLKRHTRTHTGEKPFQCRICMRNFSLKEHLTRHLRTHLRGS
84	mRNA0049	SRPGERPFQCRICMRNFSTNNNLARHTRTHTGEKPFQCRICMRNFSRTDSLTLHLRTHTGSQKPFQCRICMRNFSQREHLTTHLRTHTGEKPFQCRICMRNFSRRDNLNRHLKTHTGSQKPFQCRICMRNFSRRQKLTIHTRTHTGEKPFQCRICMRNFSHKSSLTRHLRTHLRGS
85	mRNA0050	SRPGERPFQCRICMRNFSTNNNLARHTRTHTGEKPFQCRICMRNFSRTDSLTLHLRTHTGSQKPFQCRICMRNFSQREHLTTHLRTHTGEKPFQCRICMRNFSRGDNLKRHLKTHTGSQKPFQCRICMRNFSRRQKLTIHTRTHTGEKPFQCRICMRNFSHKSSLTRHLRTHLRGS
86	mRNA0066	SRPGERPFQCRICMRNFSTNNNLARHTRTHTGEKPFQCRICMRNFSRTDSLTLHLRTHTGSQKPFQCRICMRNFSQREHLNGHLRTHTGEKPFQCRICMRNFSRGDNLARHLKTHTGSQKPFQCRICMRNFSRRQKLTIHTRTHTGEKPFQCRICMRNFSHKSSLTRHLRTHLRGS

SEQ	Description	Sequence
87	mRNA0051	SRPGERPFQCRICMRNFSQQTNLTRHLRTHTGEKPFQCRICMRNFSANRTLVHHLKTHTGSQKPFQCRICMRNFSDPANLRRHTRTHTGEKPFQCRICMRNFSRQEHLVRHLRTHTGGGGSQKPFQCRICMRNFSMKHHLGRHLRTHTGEKPFQCRICMRNFSQNSHLRRHLKTHLRGS
88	mRNA0052	SRPGERPFQCRICMRNFSQQTNLTRHLRTHTGEKPFQCRICMRNFSANRTLVHHLKTHTGSQKPFQCRICMRNFSEEANLRRHTRTHTGEKPFQCRICMRNFSRREHLVRHLRTHTGGGGSQKPFQCRICMRNFSMKHHLGRHLRTHTGEKPFQCRICMRNFSQNSHLRRHLKTHLRGS
89	mRNA0067	SRPGERPFQCRICMRNFSQQTNLTRHLRTHTGEKPFQCRICMRNFSANRTLVHHLKTHTGSQKPFQCRICMRNFSDPANLRRHTRTHTGEKPFQCRICMRNFSRQEHLVRHLRTHTGGGGSQKPFQCRICMRNFSLKQHLVRHLRTHTGEKPFQCRICMRNFSQGGHLARHLKTHLRGS
90	mRNA0068	SRPGERPFQCRICMRNFSRNTHLARHTRTHTGEKPFQCRICMRNFSRADVLKGHLRTHTGSQKPFQCRICMRNFSQRSSLVRHLRTHTGEKPFQCRICMRNFSRKDALHVHLKTHTGGGGSQKPFQCRICMRNFSQNEHLKVHLRTHTGEKPFQCRICMRNFSQNSHLRRHLKTHLRGS
91	mRNA0053	SRPGERPFQCRICMRNFSRNTHLARHTRTHTGEKPFQCRICMRNFSRADVLKGHLRTHTGSQKPFQCRICMRNFSQSSSLVRHLRTHTGEKPFQCRICMRNFSRKERLATHLKTHTGGGGSQKPFQCRICMRNFSQKTHLAVHLRTHTGEKPFQCRICMRNFSQGGHLKRHLKTHLRGS
92	mRNA0054	SRPGERPFQCRICMRNFSRNTHLARHTRTHTGEKPFQCRICMRNFSRADVLKGHLRTHTGSQKPFQCRICMRNFSQSSSLVRHLRTHTGEKPFQCRICMRNFSRKERLATHLKTHTGGGGSQKPFQCRICMRNFSQKTHLAVHLRTHTGEKPFQCRICMRNFSQNSHLRRHLKTHLRGS
93	mRNA0055	SRPGERPFQCRICMRNFSHKSSLTRHLRTHTGEKPFQCRICMRNFSESGHLKRHLKTHTGSQKPFQCRICMRNFSRRRNLTLHTRTHTGEKPFQCRICMRNFSDRSSLKRHLRTHTGSQKPFQCRICMRNFSQPHSLAVHLRTHTGEKPFQCRICMRNFSQKPHLSRHLKTHLRGS
94	mRNA0056	SRPGERPFQCRICMRNFSHKSSLTRHLRTHTGEKPFQCRICMRNFSEGGHLKRHLKTHTGSQKPFQCRICMRNFSRRRNLQLHTRTHTGEKPFQCRICMRNFSDHSSLKRHLRTHTGSQKPFQCRICMRNFSRRQHLQYHTRTHTGEKPFQCRICMRNFSQSAHLKRHLRTHLRGS
95	mRNA0057	SRPGERPFQCRICMRNFSHKSSLTRHLRTHTGEKPFQCRICMRNFSEGGHLKRHLKTHTGSQKPFQCRICMRNFSRRRNLTLHTRTHTGEKPFQCRICMRNFSDRSSLKRHLRTHTGSQKPFQCRICMRNFSRRQHLQYHTRTHTGEKPFQCRICMRNFSQSAHLKRHLRTHLRGS
96	mRNA0058	SRPGERPFQCRICMRNFSGHTALRNHTRTHTGEKPFQCRICMRNFSQSGTLHRHLRTHTGGGGSQKPFQCRICMRNFSDHSSLKRHLRTHTGEKPFQCRICMRNFSAMRSLMGHLKTHTGSQKPFQCRICMRNFSRRSRLVRHTRTHTGEKPFQCRICMRNFSRGEHLTRHLRTHLRGS
97	mRNA0059	SRPGERPFQCRICMRNFSGHTALRNHTRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHTGGGGSQKPFQCRICMRNFSDHSSLKRHLRTHTGEKPFQCRICMRNFSQQRSLVGHLKTHTGSQKPFQCRICMRNFSEAHHLSRHLRTHTGEKPFQCRICMRNFSRTEHLARHLKTHLRGS
98	mRNA0060	SRPGERPFQCRICMRNFSGHTALRNHTRTHTGEKPFQCRICMRNFSQSTTLKRHLRTHTGGGGSQKPFQCRICMRNFSDHSSLKRHLRTHTGEKPFQCRICMRNFSAMRSLMGHLKTHTGSQKPFQCRICMRNFSRQSRLQRHTRTHTGEKPFQCRICMRNFSRREHLVRHLRTHLRGS
99	mRNA0062	SRPGERPFQCRICMRNFSQGETLKRHLRTHTGEKPFQCRICMRNFSRADNLRRHLKTHTGSQKPFQCRICMRNFSDKANLTRHLRTHTGEKPFQCRICMRNFSDQGNLIRHLKTHTGGGGSQKPFQCRICMRNFSHRHVLINHTRTHTGEKPFQCRICMRNFSTNSSLTRHLRTHLRGS
100	mRNA0063	SRPGERPFQCRICMRNFSQGETLKRHLRTHTGEKPFQCRICMRNFSRADNLRRHLKTHTGSQKPFQCRICMRNFSDSSNLRRHLRTHTGEKPFQCRICMRNFSDQGNLIRHLKTHTGGGGSQKPFQCRICMRNFSHKSSLTRHLRTHTGEKPFQCRICMRNFSIRTSLKRHLKTHLRGS
101	mRNA0069	SRPGERPFQCRICMRNFSQGETLKRHLRTHTGEKPFQCRICMRNFSRADNLRRHLKTHTGSQKPFQCRICMRNFSEQGNLLRHLRTHTGEKPFQCRICMRNFSDGGNLGRHLKTHTGGGGSQKPFQCRICMRNFSHRHVLINHTRTHTGEKPFQCRICMRNFSTNSSLTRHLRTHLRGS
102	HBV target sequence	GATGAGGCATAGCAGCAG
103	HBV target sequence	GATGATTAGGCAGAGGTG
104	HBV target sequence	GGATTCAGCGCCGACGGG
105	HBV target sequence	GGCAGTAGTCGGAACAGGG
106	HBV target sequence	GTAAACTGAGCCAGGAGAA
107	HBV target sequence	ACGGTGGTCTCCATGCGAC
108	HBV target sequence	GCTGGATGTGTCTGCGGCG
109	HBV target sequence	GTCTGCGAGGCGAGGGAG
110	HBV target sequence	GTTGCCGGGCAACGGGGTA
111	HBV target sequence	CGAGAAAGTGAAAGCCTGC

SEQ	Description	Sequence
112	HBV target sequence	GAGGCTTGAACAGTAGGAC
113	HBV target sequence	GAGGTTGGGGACTGCGAA
114	HBV target sequence	GATGATGTGGTATTGGGG
115	HBV target sequence	GATGATGTGGTATTGGGGG
116	HBV target sequence	GCAGTAGTCGGAACAGGG
117	HBV target sequence	GCATAGCAGCAGGATGAA
118	HBV target sequence	GGCGTTCACGGTGGTCTCC
119	HBV target sequence	GTTGGTGAGTGATTGGAG
120	HBV target sequence	GGAGGTTGGGGACTGCGAA
121	HBV target sequence	GGATGATGTGGTATTGGGG
122	HBV target sequence	GGATGTGTCTGCGGCGTT
123	HBV target sequence	GGGGGTTGCGTCAGCAAAC
124	HBV target sequence	GTTGTTAGACGACGAGGCA
125	F1	KKFNLLQ
126	F1	RRHILDR
127	F1	RREVLEN
128	F1	RRAVLDR
129	F1	RQEHLVR
130	F1	RREHLVR
131	F1	KKDHLHR
132	F1	KTDHLAR
133	F1	QAGNLVR
134	F1	QRGNLQR
135	F1	DRGNLTR
136	F1	RTDTLAR
137	F1	RADNLGR
138	F1	QQSSLLR
139	F1	QASALSR
140	F1	RGRNLEM
141	F1	RRRNLDV
142	F1	RGRNLAI
143	F1	DGSNLRR
144	F1	DPSNLQR
145	F1	QQTNLTR
146	F1	RATHLTR
147	F1	RVDHLHR
148	F1	RKHHLGR
149	F1	DKSSLRK
150	F1	CNGSLKK
151	F1	TNNNLAR
152	F1	RNTHLAR
153	F1	HKSSLTR
154	F1	GHTALRN
155	F1	QGETLKR
156	F2	RQDNLNS
157	F2	RKDYLIS
158	F2	RQDNLGR
159	F2	RRDNLNR
160	F2	EGGNLMR
161	F2	DPSNLQR
162	F2	DMGNLGR
163	F2	QKEILTR

SEQ	Description	Sequence
164	F2	QNSHLRR
165	F2	QTTHLSR
166	F2	QARSLRA
167	F2	RTDSLPR
168	F2	RLDMLAR
169	F2	RNTHLSY
170	F2	RREHLVR
171	F2	DSSVLRR
172	F2	RIDNLDG
173	F2	RRDNLPK
174	F2	ANRTLVH
175	F2	RADVLKG
176	F2	RKESLTV
177	F2	RREHLSG
178	F2	RREHLTI
179	F2	RSDHLSL
180	F2	VGGNLAR
181	F2	VGGNLSR
182	F2	DHSSLKR
183	F2	RTDSLTL
184	F2	ESGHLKR
185	F2	EGGHLKR
186	F2	QSGTLHR
187	F2	QSTTLKR
188	F2	RADNLRR
189	F3	RSHNLKL
190	F3	RSHNLRL
191	F3	QSTTLKR
192	F3	SDRRDLD
193	F3	QSAHLKR
194	F3	DLSTLRR
195	F3	DGSTLRR
196	F3	EKASLIK
197	F3	DKSSLRK
198	F3	CNGSLKK
199	F3	DHSSLKR
200	F3	RGDGLRR
201	F3	RKLGLLR
202	F3	GLTALRT
203	F3	QNANLKR
204	F3	LKSNLHR
205	F3	QRRYLVE
206	F3	TTFNLRV
207	F3	EEANLRR
208	F3	QRSSLVR
209	F3	QSSSLVR
210	F3	KRYNLYQ
211	F3	KKFNLLQ
212	F3	RNFILQR
213	F3	RNFILAR
214	F3	QREHLTT
215	F3	QREHLNG
216	F3	DPANLRR
217	F3	RRRNLTL
218	F3	RRRNLQL
219	F3	DKANLTR
220	F3	DSSNLRR
221	F3	EQGNLLR
222	F4	QSTTLKR
223	F4	RRDGLAG
224	F4	SFQSYLE
225	F4	ETGSLRR
226	F4	DRTPLNR
227	F4	QNEHLKV

SEQ	Description	Sequence
228	F4	QKTHLAV
229	F4	DHSSLKR
230	F4	QPHGLAH
231	F4	QPHGLRH
232	F4	QPHGLST
233	F4	RRDNLNR
234	F4	RQDNLGR
235	F4	ERAKLIR
236	F4	QKHHLAV
237	F4	LKQHLVV
238	F4	QQTNLAR
239	F4	QTQNLTR
240	F4	RGEHLTR
241	F4	RREHLVR
242	F4	RKDALHV
243	F4	RKERLAT
244	F4	DPTSLNR
245	F4	RQDNLNT
246	F4	RRDNLKS
247	F4	RNDTLII
248	F4	RQDILVV
249	F4	RGDNLKR
250	F4	RGDNLAR
251	F4	RQEHLVR
252	F4	DRSSLKR
253	F4	AMRSLMG
254	F4	QQRSLVG
255	F4	DQGNLIR
256	F4	DGGNLGR
257	F5	RNTNLTR
258	F5	RQDNLGR
259	F5	VHHNLVR
260	F5	RPNHLAI
261	F5	QSHSLKS
262	F5	QKHHLVT
263	F5	GGTALRM
264	F5	GGSALSM
265	F5	RRFILSR
266	F5	RNFILQR
267	F5	QSAHLKR
268	F5	QQAHLVR
269	F5	RARNLTL
270	F5	RRRNLQL
271	F5	AKRDLDR
272	F5	LRKDLVR
273	F5	QRSNLAR
274	F5	LKTNLAR
275	F5	QRSDLTR
276	F5	HKETLNR
277	F5	TNSSLTR
278	F5	MTSSLRR
279	F5	VRHNLTR
280	F5	VAHNLTR
281	F5	QSSSLVR
282	F5	RSHNLKL
283	F5	RSHNLRL
284	F5	TSTLLKR
285	F5	HKSSLTR
286	F5	RRQKLTI
287	F5	MKHHLGR
288	F5	LKQHLVR
289	F5	QNEHLKV
290	F5	QKTHLAV
291	F5	QPHSLAV

SEQ	Description	Sequence
292	F5	RRQHLQY
293	F5	RRSRLVR
294	F5	EAHHLSR
295	F5	RQSRLQR
296	F5	HRHVLIN
297	F6	IKHNLAR
298	F6	VVNNLNR
299	F6	ISHNLAR
300	F6	QSPHLKR
301	F6	ESGHLKR
302	F6	ENSKLRR
303	F6	QRSSLVR
304	F6	RNDSLKC
305	F6	RNDTLII
306	F6	VGNSLSR
307	F6	VHESLKR
308	F6	DPSSLKR
309	F6	DHSSLKR
310	F6	VNSSLTR
311	F6	VRHSLTR
312	F6	QKVHLEA
313	F6	QKCHLKA
314	F6	RGDNLNR
315	F6	REDNLGR
316	F6	RIDNLIR
317	F6	RQDNLGR
318	F6	QTNTLGR
319	F6	QGGTLRR
320	F6	QSTTLKR
321	F6	LKEHLTR
322	F6	HKSSLTR
323	F6	QNSHLRR
324	F6	QGGHLAR
325	F6	QGGHLKR
326	F6	QKPHLSR
327	F6	QSAHLKR
328	F6	RGEHLTR
329	F6	RTEHLAR
330	F6	RREHLVR
331	F6	TNSSLTR
332	F6	IRTSLKR
495	ZIM3	MNNSQGRVTFEDVTVNFTQGEWQRLNPEQRNLYRDVMLENYSNLVSVGQGETTKPDVILRLEQGKEPWLEEEEVLGSGRAEKNGDIGGQIWKPKDVKESL
496	ZNF436	MAATLLMAGSQAPVTFEDMAMYLTREEWRPLDAAQRDLYRDVMQENYGNVVSLDFEIRSENEVNPKQEISEDVQFGTTSERPAENAEENPESEEGFESGDRSERQW
497	ZNF257	MLENYRNLVFLGIAVSKPDLITCLEQGKEPCNMKRHEMVAKPPVMCSHIAEDLCPERDIKYFFQKVILRRYDKCEHENLQLRKGCKSVDECKVCK
498	ZNF675	MGLLTFRDVAIEFSLEEWQCLDTAQRNLYKNVILENYRNLVFLGIAVSKQDLITCLEQEKEPLTVKRHEMVNEPPVMCSHFAQEFWPEQNIKDSF
499	ZNF490	MLQMQNSEHHGQSIKTQTDSISLEDVAVNFTLEEWALLDPGQRNIYRDVMRATFKNLACIGEKWKDQDIEDEHKNQGRNLRSPMVEALCENKEDCPCGKSTSQIPDLNTNLETPTG
500	ZNF320	MALSQGLLTFRDVAIEFSQEEWKCLDPAQRTLYRDVMLENYRNLVSLDISSKCMMNTLSSTGQGNTEVIHTGTLQRQASYHIGAFCSQEIEKDIHDFVFQ
501	ZNF331	MAQGLVTFADVAIDFSQEEWACLNSAQRDLYWDVMLENYSNLVSLDLESAYENKSLPTKKNIHEIRASKRNSDRRSKSLGRNWICEGTLERPQRSRGR
502	ZNF816	MLREEATKKSKEKEPGMALPQGRLTFRDVAIEFSLEEWKCLNPAQRALYRAVMLENYRNLEFVDSSLKSMMEFSSTRHSITGEVIHTGTLQRHKSHHIGDFCFPEMKKDIHHFEFQWQ
503	ZNF680	MPGPPGSLEMGPLTFRDVAIEFSLEEWQCLDTAQRNLYRKVMFENYRNLVFLGIAVSKPHLITCLEQGKEPWNRKRQEMVAKPPVIYSHFTEDLWPEHSIKDSF
504	ZNF41	MSPPWSPALAAEGRGSSCEASVSFEDVTVDFSKEEWQHLDPAQRRLYWDVTLENYSHLLSVGYQIPKSEAAFKLEQGEGPWMLEGEAPHQSCSGEAIGKMQQQGIPGGIFFHC
505	ZNF189	MASPSPPPESKEEWDYLDPAQRSLYKDVMMENYGNLVSLDVLNRDKDEEPTVKQEIEEIEEEVEPQGVIVTRIKSEIDQDPMGRETFELVGRLDKQRGIFLWEIPRESL
506	ZNF528	MALTQGPLKFMDVAIEFSQEEWKCLDPAQRTLYRDVMLENYRNLVSLGICLPDLSVTSMLEQKRDPWTLQSEEKIANDPDGRECIKGVNTERSSKLGSN

SEQ	Description	Sequence
507	ZNF543	MAASAQVSVTFEDVAVTFTQEEWGQLDAAQRTLYQEVMLETCGLLMSLGCPLFKPELIYQLDHRQELWMATKDLSQSSYPGDNTKPKTTEPTFSHLALPE
508	ZNF554	MFSQEERMAAGYLPRWSQELVTFEDVSMDFSQEEWELLEPAQKNLYREVMLENYRNVVSLEALKNQCTDVGIKEGPLSPAQTSQVTSLSSWTGYLLFQPVASSHLEQREALWIEEKGTPQASCSDWMTVLRNQDSTYKKVALQE
509	ZNF140	MSQGSVTFRDVAIDFSQEEWKWLQPAQRDLYRCVMLENYGHLVSLGLSISKPDVVSLLEQGKEPWLGKREVKRDLFSVSESSGEIKDFSPKNVIYDD
510	ZNF610	MEEAQKRKAKESGMALPQGRLTFMDVAIEFSQEEWKSLDPGQRALYRDVMLENYRNLVFLGRSCVLGSNAENKPIKNQLGLTLESHLSELQLFQAGRKIYRSNQVEKFTNHR
511	ZNF264	MAAAVLTDRAQVSVTFDDVAVTFTKEEWGQLDLAQRTLYQEVMLENCGLLVSLGCPVPKAELICHLEHGQEPWTRKEDLSQDTCPGDKGKPKTTEPTTCEPALSE
512	ZNF350	MIQAQESITLEDVAVDFTWEEWQLLGAAQKDLYRDVMLENYSNLVAVGYQASKPDALFKLEQGEQLWTIEDGIHSGACSDIWKVDHVLERLQSESLVNR
513	ZNF8	MEGVAGVMSVGPPAARLQEPVTFRDVAVDFTQEEWGQLDPTQRILYRDVMLETFGHLLSIGPELPKPEVISQLEQGTELWVAERGTTQGCHPAWEPRSESQASRKEEGLPEE
514	ZNF582	MSLGSELFRDVAIVFSQEEWQWLAPAQRDLYRDVMLETYSNLVSLGLAVSKPDVISFLEQGKEPWMVERVVSGGLCPVLESRYDTKELFPKQHVYEV
515	ZNF30	MAHKYVGLQYHGSVTFEDVAIAFSQQEWESLDSSQRGLYRDVMLENYRNLVSMAGHSRSKPHVIALLEQWKEPEVTVRKDGRRWCTDLQLEDDTIGCKEMPTSEN
516	ZNF324	MAFEDVAVYFSQEEWGLLDTAQRALYRRVMLDNFALVASLGLSTSRPRVVIQLERGEEPWVPSGTDTTLSRTTYRRRNPGSWSLTEDRDVSG
517	ZNF98	MLENYRNLVFVGIAASKPDLITCLEQGKEPWNVKRHEMVTEPPVVYSYFAQDLWPKQGKKNYFQKVILRTYKKCGRENLQLRKYCKSMDECKVHKECYNGLNQC
518	ZNF669	MHFRRPDPCREPLASPIQDSVAFEDVAVNFTQEEWALLDSSQKNLYREVMQETCRNLASVGSQWKDQNIEDHFEKPGKDIRNHIVQRLCESKEDGQYGEVVSQIPNLDLNENISTGLKPCECSICGK
519	ZNF677	MALSQGLFTFKDVAIEFSQEEWECLDPAQRALYRDVMLENYRNLLSLDEDNIPPEDDISVGFTSKGLSPKENNKEELYHLVILERKESHGINNFDLKEVWENMPKFDSLW
520	ZNF596	MTFEDIIVDFTQEEWALLDTSQRKLFQDVMLENISHLVSIGKQLCKSVVLSQLEQVEKLSTQRISLLQGREVGIKHQEIPFIHHIYQKGTSTISTMRS
521	ZNF214	MAVTFEDVTIIFTWEEWKFLDSSQKRLYREVMWENYTNVMSVENWNESYKSQEEKFRYLEYENFSYWQGWWNAGAQMYENQNYGETVQGTDSKDLTQQDRSQC
522	ZNF37A	MITSQGSVSFRDVTVGFTQEEWQHLDPAQRTLYRDVMLENYSHLVSVGYCIPKPEVILKLEKGEEPWILEEKFPSQSHLELINTSRNYSIMKFNEFNKG
523	ZNF34	MFEDVAVYLSREEWGRLGPAQRGLYRDVMLETYGNLVSLGVGPAGPKPGVISQLERGDEPWVLDVQGTSGKEHLRVNSPALGTRTEYKELTSQETFGEEDPQGSEPVEACDHIS
524	ZNF250	METYGNVVSLGLPGSKPDIISQLERGEDPWVLDRKGAKKSQGLWSDYSDNLKYDHTTACTQQDSLSCPWECETKGESQNTDLSPKPLISEQTVILGKTPLGRIDQENNETKQ
525	ZNF547	MAEMNPAQGHVVFEDVAIYFSQEEWGHLDEAQRLLYRDVMLENLALLSSLGCCHGAEDEEAPLEPGVSVGVSQVMAPKPCLSTQNTQPCETCSSLLKDILRL
526	ZNF273	MLDNYRNLVFLGIAVSKPDLITCLEQGKEPCNMKRHAMVAKPPVVCSHFAQDLWPKQGLKDS
527	ZNF354A	MAAGQREARPQVSLTFEDVAVLFTRDEWRKLAPSQRNLYRDVMLENYRNLVSLGLPFTKPKVISLLQQGEDPWEVEKDGSGVSSLGSKSSHKTTKSTQTQDSSFQ
528	ZFP82	MALRSVMFSDVSIDFSPEEWEYLDLEQKDLYRDVMLENYSNLVSLGCFISKPDVISSLEQGKEPWKVVRKGRRQYPDLETKYETKKLSLENDIYEIN
529	ZNF224	MTTFKEAMTFKDVAVVFTEEELGLLDLAQRKLYRDVMLENFRNLLSVGHQAFHRDTFHFLREEKIWMMKTAIQREGNSGDKIQTEMETVSEAGTHQEW
530	ZNF33A	MFQVEQKSQESVSFKDVTVGFTQEEWQHLDPSQRALYRDVMLENYSNLVSVGYCVHKPEVIFRLQQGEEPWKQEEEFPSQSFPEVWTADHLKERSQENQSKHL
531	ZNF45	MTKSKEAVTFKDVAVVFSEEELQLLDLAQRKLYRDVMLENFRNVVSVGHQSTPDGLPQLEREEKLWMMKMATQRDNSSGAKNLKEMETLQEVGLRYLP
532	ZNF175	MSQKPQVLGPEKQDGSCEASVSFEDVTVDFSREEWQQLDPAQRCLYRDVMLELYSHLFAVGYHIPNPEVIFRMLKEKEPRVEEAEVSHQRCQEREFGLEIPQKEISKKASFQ
533	ZNF595	MELVTFRDVAIEFSPEEWKCLDPAQQNLYRDVMLENYRNLVSLGFVISNPDLVTCLEQIKEPCNLKIHETAAKPPAICSPFSQDLSPVQGIEDSF
534	ZNF184	MSTLLQGGHNLLSSASFQESVTFKDVIVDFTQEEWKQLDPGQRDLFRDVTLENYTHLVSIGLQVSKPDVISQLEQGTEPWIMEPSIPVGTCADWETRLENSVSAPEPDISEE
535	ZNF419	MDPAQVPVAADLLTDHEEGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLENFTLLASLGLASSKTHEITQLESWEEPFMPAWEVVTSAIPRGCWHGAEAEEAPEQIASVG
536	ZFP28-1	MKKLEAVGTGIEPKAMSQGLVTFGDVAVDFSQEEWEWLNPIQRNLYRKVMLENYRNLASLGLCVSKPDVISSLEQGKEPWTVKRKMTRAWCPDLKAVWKIKELPLKKDFCEG
537	ZFP28-2	MSLLGEHWDYDALFETQPGLVTIKNLAVDFRQQLHPAQKNFCKNGIWENNSDLGSAGHCVAKPDLVSLLEQEKEPWMVKRELTGSLFSGQRSVHETQELFPKQDSYAE
538	ZNF18	MLALAASQPARLEERLIRDRDLGASLLPAAPQEQWRQLDSTQKEQYWDLILETYGKMVSGAGISHPKSDLTNSIEFGEELAGIYLHVNEKIPRPTCIGDRQENDKENLNLENH

SEQ	Description	Sequence
539	ZNF213	MEGRPGETTDTCFVSGVHGPVALGDIPFYFSREEWGTLDPAQRDLFWDIKRENSRNTTLG FGLKGQSEKSLLQEMVPVVPGQTGSDVTVSWSPEEAEAWESENRPRAALGPVVGARRGRP PTRRRQFRDLA
540	ZNF394	MVAVVRALQRALDGTSSQGMVTFEDTAVSLTWEEWERLDPARRDFCRESAQKDSGSTVPP SLESRVENKELIPMQQILEEAEPQGQLQEAFQGKRPLFSKCGSTHEDRVEKQSGDP
541	ZFP1	MNKSQGSVSFTDVTVDFTQEEWEQLDPSQRILYMDVMLENYSNLLSVEVWKADDQMERDH RNPDEQARQFLILKNQTPIEERGDLFGKALNLNTDFVSLRQVPYKYDLYEKTL
542	ZFP14	MAHGSVTFRDVAIDFSQEEWEFLDPAQRDLYRDVMWENYSNFISLGPSISKPDVITLLDE ERKEPGMVVREGTRRYCPDLESRYRTNTLSPEKDIYEIYSFQWDIMER
543	ZNF416	MAAAVLRDSTSVPVTAEAKLMGFTQGCVTFEDVAIYFSQEEWGLLDEAQRLLYRDVMLEN FALITALVCWHGMEDEETPEQSVSVEGVPQVRTPEASPSTQKIQSCDMCVPFLTDILHLT DLPGQELYLTGACAVFHQDQK
544	ZNF557	MLPPTAASQREGHTEGGELVNELLKSWLKGLVTFEDVAVEFTQEEWALLDPAQRTLYRDV MLENCRNLASLGNQVDKPRLISQLEQEDKVMTEERGILSGTCPDVENPFKAKGLTPKLHV FRKEQSRNMKMER
545	ZNF566	MAQESVMFSDVSVDFSQEEWECLNDDQRDLYRDVMLENYSNLVSMGHSISKPNVISYLEQ GKEPWLADRELTRGQWPVLESRCETKKLFLKKEIYEIESTQWEIMEK
546	ZNF729	MPGAPGSLEMGPLTFRDVTIEFSLEEWQCLDTVQQNLYRDVMLENYRNLVFLGMAVFKPD LITCLKQGKEPWNMKRHEMVTKPPVMRSHFTQDLWPDQSTKDSFQEVILRTYAR
547	ZIM2	MAGSQFPDFKHLGTFLVFEELVTFEDVLVDFSPEELSSLSAAQRNLYREVMLENYRNLVS LGHQFSKPDIISRLEEEESYAMETDSRHTVICQGE
548	ZNF254	MPGPPRSLEMGLLTFRDVAIEFSLEEWQHLDIAQQNLYRNVMLENYRNLAFLGIAVSKPD LITCLEQGKEPWNMKRHE
549	ZNF764	MAPPLAPLPPRDPNGAGPEWREPGAVSFADVAVYFCREEWGCLRPAQRALYRDVMRETYG HLSALGIGGNKPALISWVEEEAELWGPAAQDPE
550	ZNF785	MGPPLAPRPAHVPGEAGPRRTRESRPGAVSFADVAVYFSPEEWECLRPAQRALYRDVMRE TFGHLGALGFSVPKPAFISWVEGEVEAWSPEAQDPDGESS
551	ZNF10 (KOX1)	MDAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKP DVILRLEKGEEPWLVEREIHQETHPDSETAFEIKSSVSSRSIFKDKQSCDIKMEGMARND LWYLSLEEVWKCRDQLDKYQENPERHLRQVAFTQKKVLTQERVSESGKYGGNCLLPAQLV LREYFHKRDSHTKSLKHDLVLNGHQDSCASNSNECGQTFCQNIHLIQFARTHTGDKSYKC PDNDNSLTHGSSLGISKGIHREKPYECKECGKFFSWRSNLTRHQLIHTGEKPYECKECGK SFSRSSHLIGHQKTHTGEEPYECKECGKSFSWFSHLVTHQRTHTGDKLYTCNQCGKSFVH SSRLIRHQRTHTGEKPYECPECGKSFRQSTHLILHQRTHVRVRPYECNECGKSYSQRSHL VVHHRIHTGLKPFECKDCGKCFSRSSHLYSHQRTHTGEKPYECHDCGKSFSQSSALIVHQ RIHTGEKPYECCQCGKAFIRKNDLIKHQRIHVGEETYKCNQCGIIFSQNSPFIVHQIAHT GEQFLTCNQCGTALVNTSNLIGYQTNHIRENAY
552	CBX5 (chromoshadow domain)	MGKKTKRTADSSSSEDEEEYVVEKVLDRRVVKGQVEYLLKWKGFSEEHNTWEPEKNLDCP ELISEFMKKYKKMKEGENNKPREKSESNKRKSNFSNSADDIKSKKKREQSNDIARGFERG LEPEKIIGATDSCGDLMFLMKWKDTDEADLVLAKEANVKCPQIVIAFYEERLTWHAYPED AENKEKETAKS
553	RYBP (YAF2_RYBP component of PRC1)	MTMGDKKSPTRPKRQAKPAADEGFWDCSVCTFRNSAEAFKCSICDVRKGTSTRKPRINSQ LVAQQVAQQYATPPPPKKEKKEKVEKQDKEKPEKDKEISPSVTKKNTNKKTKPKSDILKD PPSEANSIQSANATTKTSETNHTSRPRLKNVDRSTAQQLAVTVGNVTVIITDFKEKTRSS STSSSTVTSSAGSEQQNQSSSGSESTDKGSSRSSTPKGDMSAVNDESF
554	YAF2 (YAF2_RYBP component of PRC1)	MGDKKSPTRPKRQPKPSSDEGYWDCSVCTFRNSAEAFKCMMCDVRKGTSTRKPRPVSQLV AQQVTQQFVPPTQSKKEKKDKVEKEKSEKETTSKKNSHKKTRPRLKNVDRSSAQHLEVTV GDLTVIITDFKEKTKSPPASSAASADQHSQSGSSSDNTERGMSRSSSPRGEASSLNGESH
555	MGA (component of PRC1.6)	MEEKQQIILANQDGGTVAGAAPTFFVILKQPGNGKTDQGILVTNQDACALASSVSSPVKS KGKICLPADCTVGGITVTLDNNSMWNEFYHRSTEMILTKQGRRMFPYCRYWITGLDSNLK YILVMDISPVDNHRYKWNGRWWEPSGKAEPHVLGRVFIHPESPSTGHYWMHQPVSFYKLK LTNNTLDQEGHIILHSMHRYLPRLHLVPAEKAVEVIQLNGPGVHTFTFPQTEFFAVTAYQ NIQITQLKIDYNPFAKGFRDDGLNNKPQRDGKQKNSSDQEGNNISSSSGHRVRLTEGQGS EIQPGDLDPLSRGHETSGKGLEKTSLNIKRDFLGFMDTDSALSEVPQLKQEISECLIASS FEDDSRVASPLDQNGSFNVVIKEEPLDDYDYELGECPEGVTVKQEETDEETDVYSNSDDD PILEKQLKRHNKVDNPEADHLSSKWLPSSPSGVAKAKMFKLDTGKMPVVYLEPCAVTRST VKISELPDNMLSTSRKDKSSMLAELEYLPTYIENSNETAFCLGKESENGLRKHSPDLRVV QKYPLLKEPQWKYPDISDSISTERILDDSKDSVGDSLSGKEDLGRKRTTMLKIATAAKVV NANQNASPNVPGKRGRPRKLKLCKAGRPPKNTGKSLISTKNTPVSPGSTFPDVKPDLEDV DGVLFVSFESKEALDIHAVDGTTEESSSLQASTTNDSGYRARISQLEKELIEDLKTLRHK QVIHPGLQEVGLKLNSVDPTMSIDLKYLGVQLPLAPATSFPFWNLTGTNPASPDAGFPFV SRTGKTNDFTKIKGWRGKFHSASASRNEGGNSESSLKNRSAFCSDKLDEYLENEGKLMET SMGFSSNAPTSPVVYQLPTKSTSYVRTLDSVLKKQSTISPSTSYSLKPHSVPPVSRKAKS QNRQATFSGRTKSSYKSILPYPVSPKQKYSHVILGDKVTKNSSGIISENQANNFVVPTLD ENIFPKQISLRQAQQQQQQQQGSRPPGLSKSQVKLMDLEDCALWEGKPRTYITEERADVS LTTLLTAQASLKTKPIHTIIRKRAPPCNNDFCRLGCVCSSLALEKRQPAHCRRPDCMFGC TCLKRKVVLVKGGSKTKHFQRKAAHRDPVFYDTLGEEAREEEEGIREEEEQLKEKKKRKK

SEQ	Description	Sequence
		LEYTICETEPEQPVRHYPLWVKVEGEVDPEPVYIPTPSVIEPMKPLLLPQPEVLSPTVKGKLLTGIKSPRSYTPKPNPVIREEDKDPVYLYFESMMTCARVRVYERKKEDQRQPSSSSSPSPSFQQQTSCHSSPENHNNAKEPDSEQQPLKQLTCDLEDDSDKLQEKSWKSSCNEGESSSTSYMHQRSPGGPTKLIEIISDCNWEEDRNKILSILSQHINSNMPQSLKVGSFIIELASQRKSRGEKNPPVYSSRVKISMPSCQDQDDMAEKSGSETPDGPLSPGKMEDISPVQTDALDSVRERLHGGKGLPFYAGLSPAGKLVAYKRKPSSSTSGLIQVASNAKVAASRKPRTLLPSTSNSKMASSSGTATNRPGKNLKAFVPAKRPIAARPSPGGVFTQFVMSKVGALQQKIPGVSTPQTLAGTQKFSIRPSPVMVVTPVVSSEPVQVCSPVTAAVTTTTPQVFLENTTAVTPMTAISDVETKETTYSSGATTTGVVEVSETNTSTSVTSTQSTATVNLTKTTGITTPVASVAFPKSLVASPSTITLPVASTASTSLVVVTAAASSSMVTTPTSSLGSVPIILSGINGSPPVSQRPENAAQIPVATPQVSPNTVKRAGPRLLLIPVQQGSPTLRPVSNTQLQGHRMVLQPVRSPSGMNLFRHPNGQIVQLLPLHQLRGSNTQPNLQPVMFRNPGSVMGIRLPAPSKPSETPPSSTSSSAFSVMNPVIQAVGSSSAVNVITQAPSLLSSGASFVSQAGTLTLRISPPEPQSFASKTGSETKITYSSGGQPVGTASLIPLQSGSFALLQLPGQKPVPSSILQHVASLQMKRESQNPDQKDETNSIKREQETKKVLQSEGEAVDPEANVIKQNSGAATSEETLNDSLEDRGDHLDEECLPEEGCATVKPSEHSCITGSHTDQDYKDVNEEYGARNRKSSKEKVAVLEVRTISEKASNKTVQNLSKVQHQKLGDVKVEQQKGFDNPEENSSEFPVTFKEESKFELSGSKVMEQQSNLQPEAKEKECGDSLEKDRERWRKHLKGPLTRKCVGASQECKKEADEQLIKETKTCQENSDVFQQEQGISDLLGKSGITEDARVLKTECDSWSRISNPSAFSIVPRRAAKSSRGNGHFQGHLLLPGEQIQPKQEKKGGRSSADFTVLDLEEDDEDDNEKTDDSIDEIVDVVSDYQSEEVDDVEKNNCVEYIEDDEEHVDIETVEELSEEINVAHLKTTAAHTQSFKQPSCTHISADEKAAERSRKAPPIPLKLKPDYWSDKLQKEAEAFAYYRRTHTANERRRRGEMRDLFEKLKITLGLLHSSKVSKSLILTRAFSEIQGLTDQADKLIGQKNLLTRKRNILIRKVSSLSGKTEEVVLKKLEYIYAKQQALEAQKRKKKMGSDEFDISPRISKQQEGSSASSVDLGQMFINNRRGKPLILSRKKDQATENTSPLNTPHTSANLVMTPQGQLLTLKGPLFSGPVVAVSPDLLESDLKPQVAGSAVALPENDDLFMMPRIVNVTSLATEGGLVDMGGSKYPHEVPDSKPSDHLKDTVRNEDNSLEDKGRISSRGNRDGRVTLGPTQVFLANKDSGYPQIVDVSNMQKAQEFLPKKISGDMRGIQYKWKESESRGERVKSKDSSFHKLKMKDLKDSSIEMELRKVTSAIEEAALDSSELLTNMEDEDDTDETLTSLLNEIAFLNQQLNDDSVGLAELPSSMDTEFPGDARRAFISKVPPGSRATFQVEHLGTGLKELPDVQGESDSISPLLLHLEDDDFSENEKQLAEPASEPDVLKIVIDSEIKDSLLSNKKAIDGGKNTSGLPAEPESVSSPPTLHMKTGLENSNSTDTLWRPMPKLAPLGLKVANPSSDADGQSLKVMPCLAPIAAKVGSVGHKMNLTGNDQEGRESKVMPTLAPVVAKLGNSGASPSSAGK
556	CBX1 (chromoshadow)	MGKKQNKKKVEEVLEEEEEEYVVEKVLDRRVVKGKVEYLLKWKGFSDEDNTWEPEENLDCPDLIAEFLQSQKTAHETDKSEGGKRKADSDSEDKGEESKPKKKKEESEKPRGFARGLEPERIIGATDSSGELMFLMKWKNSDEADLVPAKEANVKCPQVVISFYEERLTWHSYPSEDDDKKDDKN
557	SCMH1 (SAM_1/SPM)	MLVCYSVLACEILWDLPCSIMGSPLGHFTWDKYLKETCSVPAPVHCFKQSYTPPSNEFKISMKLEAQDPRNTTSTCIATVVGLTGARLRLRLDGSDNKNDFWRLVDSAEIQPIGNCEKNGGMLQPPLGFRLNASSWPMFLLKTLNGAEMAPIRIFHKEPPSPSHNFFKMGMKLEAVDRKNPHFICPATIGEVRGSEVLVTFDGWRGAFDYWCRFDSRDIFPVGWCSLTGDNLQPPGTKVVIPKNPYPASDVNTEKPSIHSSTKTVLEHQPGQRGRKPGKKRGRTPKTLISHPISAPSKTAEPLKFPKKRGPKPGSKRKPRTLLNPPPASPTTSTPEPDTSTVPQDAATIPSSAMQAPTVCIYLNKNGSTGPHLDKKKVQQLPDHFGPARASVVLQQAVQACIDCAYHQKTVFSFLKQGHGGEVISAVFDREQHTLNLPAVNSITYVLRFLEKLCHNLRSDNLFGNQPFTQTHLSLTAIEYSHSHDRYLPGETFVLGNSLARSLEPHSDSMDSASNPTNLVSTSQRHRPLLSSCGLPPSTASAVRRLCSRGVLKGSNERRDMESFWKLNRSPGSDRYLESRDASRLSGRDPSSWTVEDVMQFVREADPQLGPHADLFRKHEIDGKALLLLRSDMMMKYMGLKLGPALKLSYHIDRLKQGKF
558	MPP8 (Chromodomain)	MEQVAEGARVTAVPVSAADSTEELAEVEEGVGVVGEDNDAAARGAEAFGDSEEDGEDVFEVEKILDMKTEGGKVLYKVRWKGYTSDDDTWEPEIHLEDCKEVLLEFRKKIAENKAKAVRKDIQRLSLNNDIFEANSDSDQQSETKEDTSPKKKKKKLRQREEKSPDDLKKKKAKAGKLKDKSKPDLESSLESLVFDLRTKKRISEAKEELKESKKPKKDEVKETKELKKVKKGEIRDLKTKTREDPKENRKTKKEKFVESQVESESSVLNDSPFPEDDSEGLHSDSREEKQNTKSARERAGQDMGLEHGFEKPLDSAMSAEEDTDVRGRRKKKTPRKAEDTRENRKLENKNAFLEKKTVPKKQRNQDRSKSAAELEKLMPVSAQTPKGRRLSGEERGLWSTDSAEEDKETKRNESKEKYQKRHDSDKEEKGRKEPKGLKTLKEIRNAFDLFKLTPEEKNDVSENNRKREEIPLDFKTIDDHKTKENKQSLKERRNTRDETDTWAYIAAEGDQEVLDSVCQADENSDGRQQILSLGMDLQLEWMKLEDFQKHLDGKDENFAATDAIPSNVLRDAVKNGDYITVKVALNSNEEYNLDQEDSSGMTLVMLAAAGGQDDLLRLLITKGAKVNGRQKNGTTALIHAAEKNFLTTVAILLEAGAFVNVQQSNGETALMKACKRGNSDIVRLVIECGADCNILSKHQNSALHFAKQSNNVLVYDLLKNHLETLSRVAEETIKDYFEARLALLEPVFPIACHRLCEGPDFSTDFNYKPPQNIPEGSGILLFIFHANFLGKEVIARLCGPCSVQAVVLNDKFQLPVFLDSHFVYSFSPVAGPNKLFIRLTEAPSAKVKLLIGAYRVQLQ
559	SUMO3 (Rad60-SLD)	MSEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGGVPESSLAGHSF
560	HERC2 (Cyt-b5)	MPSESFCLAAQARLDSKWLKTDIQLAFTRDGLCGLWNEMVKDGEIVYTGTESTQNGELPPRKDDSVEPSGTKKEDLNDKEKKDEEETPAPIYRAKSILDSWVWGKQPDVNELKECLSVLV

SEQ	Description	Sequence
		KEQQALAVQSATTTLSALRLKQRLVILERYFIALNRTVFQENVKVKWKSSGISLPPVDKK SSRPAGKGVEGLARVGSRAALSFAFAFLRRAWRSGEDADLCSELLQESLDALRALPEASL FDESTVSSVWLEVVERATRFLRSVVTGDVHGTPATKGPGSIPLQDQHLALAILLELAVQR GTLSQMLSAILLLLQLWDSGAQETDNERSAQGTSAPLLPLLQRFQSIICRKDAPHSEGDM HLLSGPLSPNESFLRYLTLPQDNELAIDLRQTAVVVMAHLDRLATPCMPPLCSSPTSHKG SLQEVIGWGLIGWKYYANVIGPIQCEGLANLGVTQIACAEKRFLILSRNGRVYTQAYNSD TLAPQLVQGLASRNIVKIAAHSDGHHYLALAATGEVYSWGCGDGGRLGHGDTVPLEEPKV ISAFSGKQAGKHVVHIACGSTYSAAITAEGELYTWGRGNYGRLGHGSSEDEAIPMLVAGL KGLKVIDVACGSGDAQTLAVTENGQVWSWGDGDYGKLGRGGSDGCKTPKLIEKLQDLDVV KVRCGSQFSIALTKDGQVYSWGKGDNQRLGHGTEEHVRYPKLLEGLQGKKVIDVAAGSTH CLALTEDSEVHSWGSNDQCQHFDTLRVTKPEPAALPGLDTKHIVGIACGPAQSFAWSSCS EWSIGLRVPFVVDICSMTFEQLDLLLRQVSEGMDGSADWPPPQEKECVAVATLNLLRLQL HAAISHQVDPEFLGLGLGSILLNSLKQTVVTLASSAGVLSTVQSAAQAVLQSGWSVLLPT AEERARALSALLPCAVSGNEVNISPGRRFMIDLLVGSLMADGGLESALHAAITAEIQDIE AKKEAQKEKEIDEQEANASTFHRSRTPLDKDLINTGICESSGKQCLPLVQLIQQLLRNIA SQTVARLKDVARRISSCLDFEQHSRERSASLDLLLRFQRLLISKLYPGESIGQTSDISSP ELMGVGSLLKKYTALLCTHIGDILPVAASIASTSWRHFAEVAYIVEGDFTGVLLPELVVS IVLLLSKNAGLMQEAGAVPLLGGLLEHLDRFNHLAPGKERDDHEELAWPGIMESFFTGQN CRNNEEVTLIRKADLENHNKDGGFWTVIDGKVYDIKDFQTQSLTGNSILAQFAGEDPVVA LEAALQFEDTRESMHAFCVGQYLEPDQEIVTIPDLGSLSSPLIDTERNLGLLLGLHASYL AMSTPLSPVEIECAKWLQSSIFSGGLQTSQIHYSYNEEKDEDHCSSPGGTPASKSRLCSH RRALGDHSQAFLQAIADNNIQDHNVKDFLCQIERYCRQCHLTTPIMFPPEHPVEEVGRLL LCCLLKHEDLGHVALSLVHAGALGIEQVKHRTLPKSVVDVCRVVYQAKCSLIKTHQEQGR SYKEVCAPVIERLRFLFNELRPAVCNDLSIMSKFKLLSSLPRWRRIAQKIIRERRKKRVP KKPESTDDEEKIGNEESDLEEACILPHSPINVDKRPIAIKSPKDKWQPLLSTVTGVHKYK WLKQNVQGLYPQSPLLSTIAEFALKEEPVDVEKMRKCLLKQLERAEVRLEGIDTILKLAS KNFLLPSVQYAMFCGWQRLIPEGIDIGEPLTDCLKDVDLIPPFNRMLLEVTFGKLYAWAV QNIRNVLMDASAKFKELGIQPVPLQTITNENPSGPSLGTIPQARFLLVMLSMLTLQHGAN NLDLLLNSGMLALTQTALRLIGPSCDNVEEDMNASAQGASATVLEETRKETAPVQLPVSG PELAAMMKIGTRVMRGVDWKWGDQDGPPPGLGRVIGELGEDGWIRVQWDTGSTNSYRMGK EGKYDLKLAELPAAAQPSAEDSDTEDDSEAEQTERNIHPTAMMFTSTINLLQTLCLSAGV HAEIMQSEATKTLCGLLRMLVESGTTDKTSSPNRLVYREQHRSWCTLGFVRSIALTPQVC GALSSPQWITLLMKVVEGHAPFTATSLQRQILAVHLLQAVLPSWDKTERARDMKCLVEKL FDFLGSLLTTCSSDVPLLRESTLRRRRVRPQASLTATHSSTLAEEVVALLRTLHSLTQWN GLINKYINSQLRSITHSFVGRPSEGAQLEDYFPDSENPEVGGLMAVLAVIGGIDGRLRLG GQVMHDEFGEGTVTRITPKGKITVQFSDMRTCRVCPLNQLKPLPAVAFNVNNLPFTEPML SVWAQLVNLAGSKLEKHKIKKSTKQAFAGQVDLDLLRCQQLKLYILKAGRALLSHQDKLR QILSQPAVQETGTVHTDDGAVVSPDLGDMSPEGPQPPMILLQQLLASATQPSPVKAIFDK QELEAAALAVCQCLAVESTHPSSPGFEDCSSSEATTPVAVQHIRPARVKRRKQSPVPALP IVVQLMEMGFSRRNIEFALKSLTGASGNASSLPGVEALVGWLLDHSDIQVTELSDADTVS DEYSDEEVVEDVDDAAYSMSTGAVVTESQTYKKRADFLSNDDYAVYVRENIQVGMMVRCC RAYEEVCEGDVGKVIKLDRDGLHDLNVQCDWQQKGGTYWVRYIHVELIGYPPPSSSSHIK IGDKVRVKASVTTPKYKWGSVTHQSVGVVKAFSANGKDIIVDFPQQSHWTGLLSEMELVP SIHPGVTCDGCQMFPINGSRFKCRNCDDFDFCETCFKTKKHNTRHTFGRINEPGQSAVFC GRSGKQLKRCHSSQPGMLLDSWSRMVKSLNVSSSVNQASRLIDGSEPCWQSSGSQGKHWI RLEIFPDVLVHRLKMIVDPADSSYMPSLVVVSGGNSLNNLIELKTININPSDTTVPLLND CTEYHRYIEIAIKQCRSSGIDCKIHGLILLGRIRAEEEDLAAVPFLASDNEEEEDEKGNS GSLIRKKAAGLESAATIRTKVFVWGLNDKDQLGGLKGSKIKVPSFSETLSALNVVQVAGG SKSLFAVTVEGKVYACGEATNGRLGLGISSGTVPIPRQITALSSYVVKKVAVHSGGRHAT ALTVDGKVFSWGEGDDGKLGHFSRMNCDKPRLIEALKTKRIRDIACGSSHSAALTSSGEL YTWGLGEYGRLGHGDNTTQLKPKMVKVLLGHRVIQVACGSRDAQTLALTDEGLVFSWGDG DFGKLGRGGSEGCNIPQNIERLNGQGVCQIECGAQFSLALTKSGVVWTWGKGDYFRLGHG SDVHVRKPQVVEGLRGKKIVHVAVGALHCLAVTDSGQVYAWGDNDHGQQGNGTTTVNRKP TLVQGLEGQKITRVACGSSHSVAWTTVDVATPSVHEPVLFQTARDPLGASYLGVPSDADS SAASNKISGASNSKPNRPSLAKILLSLDGNLAKQQALSHILTALQIMYARDAVVGALMPA AMIAPVECPSFSSAAPSDASAMASPMNGEECMLAVDIEDRLSPNPWQEKREIVSSEDAVT PSAVTPSAPSASARPFIPVTDDLGAASIIAETMTKTKEDVESQNKAAGPEPQALDEFTSL LIADDTRVVVDLLKLSVCSRAGDRGRDVLSAVLSGMGTAYPQVADMLLELCVTELEDVAT DSQSGRLSSQPVVVESSHPYTDDTSTSGTVKIPGAEGLRVEFDRQCSTERRHDPLTVMDG VNRIVSVRSGREWSDWSSELRIPGDELKWKFISDGSVNGWGWRFTVYPIMPAAGPKELLS DRCVLSCPSMDLVTCLLDFRLNLASNRSIVPRLAASLAACAQLSALAASHRMWALQRLRK LLTTEFGQSININRLLGENDGETRALSFTGSALAALVKGLPEALQRQFEYEDPIVRGGKQ LLHSPFFKVLVALACDLELDTLPCCAETHKWAWFRRYCMASRVAVALDKRTPLPRLFLDE VAKKIRELMADSENMDVLHESHDIFKREQDEQLVQWMNRRPDDWTLSAGGSGTIYGWGHN HRGQLGGIEGAKVKVPTPCEALATLRPVQLIGGEQTLFAVTADGKLYATGYGAGGRLGIG GTESVSTPTLLESIQHVFIKKVAVNSGGKHCLALSSEGEVYSWGEAEDGKLGHGNRSPCD RPRVIESLRGIEVVDVAAGGAHSACVTAAGDLYTWGKGRYGRLGHSDSEDQLKPKLVEAL QGHRVVDIACGSGDAQTLCLTDDDTVWSWGDGDYGKLGRGGSDGCKVPMKIDSLTGLGVV

SEQ	Description	Sequence
		KVECGSQFSVALTKSGAVYTWGKGDYHRLGHGSDDHVRRPRQVQGLQGKKVIAIATGSLHCVCCTEDGEVYTWGDNDEGQLGDGTTNAIQRPRLVAALQGKKVNRVACGSAHTLAWSTSKPASAGKLPAQVPMEYNHLQEIPIIALRNRLLLLHHLSELFCPCIPMFDLEGSLDETGLGPSVGFDTLRGILISQGKEAAFRKVVQATMVRDRQHGPVVELNRIQVKRSRSKGGLAGPDGTKSVFGQMCAKMSSFGPDSLLLPHRVWKVKFVGESVDDCGGGYSESIAEICEELQNGLTPLLIVTPNGRDESGANRDCYLLSPAARAPVHSSMFRFLGVLLGIAIRTGSPLSLNLAEPVWKQLAGMSLTIADLSEVDKDFIPGLMYIRDNEATSEEFEAMSLPFTVPSASGQDIQLSSKHTHITLDNRAEYVRLAINYRLHEFDEQVAAVREGMARVVPVPLLSLFTGYELETMVCGSPDIPLHLLKSVATYKGIEPSASLIQWFWEVMESFSNTERSLFLRFVWGRTRLPRTIADFRGRDFVIQVLDKYNPPDHFLPESYTCFFLLKLPRYSCKQVLEEKLKYAIHFCKSIDTDDYARIALTGEPAADDSSDDSDNEDVDSFASDSTQDYLTGH
561	BIN1 (SH3_9)	MAEMGSKGVTAGKIASNVQKKLTRAQEKVLQKLGKADETKDEQFEQCVQNFNKQLTEGTRLQKDLRTYLASVKAMHEASKKLNECLQEVYEPDWPGRDEANKIAENNDLLWMDYHQKLVDQALLTMDTYLGQFPDIKSRIAKRGRKLVDYDSARHHYESLQTAKKKDEAKIAKPVSLLEKAAPQWCQGKLQAHLVAQTNLLRNQAEEELIKAQKVFEEMNVDLQEELPSLWNSRVGFYVNTFQSIAGLEENFHKEMSKLNQNLNDVLVGLEKQHGSNTFTVKAQPSDNAPAKGNKSPSPPDGSPAATPEIRVNHEPEPAGGATPGATLPKSPSQLRKGPPVPPPPKHTPSKEVKQEQILSLFEDTFVPEISVTTPSQFEAPGPFSEQASLLDLDFDPLPPVTSPVKAPTPSGQSIPWDLWEPTESPAGSLPSGEPSAAEGTFAVSWPSQTAEPGPAQPAEASEVAGGTQPAAGAQEPGETAASEAASSSLPAVVVETFPATVNGTVEGGSGAGRLDLPPGFMFKVQAQHDYTATDTDELQLKAGDVVLVIPFQNPEEQDEGWLMGVKESDWNQHKELEKCRGVFPENFTERVP
562	PCGF2 (RING finger protein domain)	MHRTTRIKITELNPHLMCALCGGYFIDATTIVECLHSFCKTCIVRYLETNKYCPMCDVQVHKTRPLLSIRSDKTLQDIVYKLVPGLFKDEMKRRRDFYAAYPLTEVPNGSNEDRGEVLEQEKGALSDDEIVSLSIEFYEGARDRDEKKGPLENGDGDKEKTGVRFLRCPAAMTVMHLAKFLRNKMDVPSKYKVEVLYEDEPLKEYYTLMDIAYIYPWRRNGPLPLKYRVQPACKRLTLATVPTPSEGTNTSGASECESVSDKAPSPATLPATSSSLPSPATPSHGSPSSHGPPATHPTSPTPPSTASGATTAANGGSLNCLQTPSSTSRGRKMTVNGAPVPPLT
563	TOX (HMG box)	MDVRFYPPPAQPAAAPDAPCLGPSPCLDPYYCNKFDGENMYMSMTEPSQDYVPASQSYPGPSLESEDFNIPPITPPSLPDHSLVHLNEVESGYHSLCHPMNHNGLLPFHPQNMDLPEITVSNMLGQDGTLLSNSISVMPDIRNPEGTQYSSHPQMAAMRPRGQPADIRQQPGMMPHGQLTTINQSQLSAQLGLNMGGSNVPHNSPSPPGSKSATPSPSSSVHEDEGDDTSKINGGEKRPASDMGKKPKTPKKKKKKDPNEPQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWDGLGEEQKQVYKKKTEAAKKEYLKQLAAYRASLVSKSYSEPVDVKTSQPPQLINSKPSVFHGPSQAHSALYLSSHYHQQPGMNPHLTAMHPSLPRNIAPKPNNQMPVTVSIANMAVSPPPPLQISPPLHQHLNMQQHQPLTMQQPLGNQLPMQVQSALHSPTMQQGFTLQPDYQTIINPTSTAAQVVTQAMEYVRSGCRNPPPQPVDWNNDYCSSGGMQRDKALYLT
564	FOXA1 (HNF3A C-terminal domain)	MLGTVKMEGHETSDWNSYYADTQEAYSSVPVSNMNSGLGSMNSMNTYMTMNTMTTSGNMTPASFNMSYANPGLGAGLSPGAVAGMPGGSAGAMNSMTAAGVTAMGTALSPSGMGAMGAQQAASMNGLGPYAAAMNPCMSPMAYAPSNLGRSRAGGGGDAKTFKRSYPHAKPPYSYISLITMAIQQAPSKMLTLSEIYQWIMDLFPYYRQNQQRWQNSIRHSLSFNDCFVKVARSPDKPGKGSYWTLHPDSGNMFENGCYLRRQKRFKCEKQPGAGGGGGSGSGGSGAKGGPESRKDPSGASNPSADSPLHRGVHGKTGQLEGAPAPGPAASPQTLDHSGATATGGASELKTPASSTAPPISSGPGALASVPASHPAHGLAPHESQLHLKGDPHYSFNHPFSINNLMSSSEQQHKLDFKAYEQALQYSPYGSTLPASLPLGSASVTTRSPIEPSALEPAYYQGVYSRPVLNTS
565	FOXA2 (HNF3B C-terminal domain)	MLGAVKMEGHEPSDWSSYYAEPEGYSSVSNMNAGLGMNGMNTYMSMSAAAMGSGSGNMSAGSMNMSSYVGAGMSPSLAGMSPGAGAMAGMGGSAGAAGVAGMGPHLSPSLSPLGGQAAGAMGGLAPYANMNSMSPMYGQAGLSRARDPKTYRRSYTHAKPPYSYISLITMAIQQSPNKMLTLSEIYQWIMDLFPFYRQNQQRWQNSIRHSLSFNDCFLKVPRSPDKPGKGSFWTLHPDSGNMFENGCYLRRQKRFKCEKQLALKEAAGAAGSGKKAAAGAQASQAQLGEAAGPASETPAGTESPHSSASPCQEHKRGGLGELKGTPAAALSPPEPAPSPGQQQQAAAHLLGPPHHPGLPPEAHLKPEHHYAFNHPFSINNLMSSEQQHHHSHHHHQPHKMDLKAYEQVMHYPGYGSPMPGSLAMGPVTNKTGLDASPLAADTSYYQGVYSRPIMNSS
566	IRF2BP1 (IRF-2BP1_2 N-terminal domain)	MASVQASRRQWCYLCDLPKMPWAMVWDFSEAVCRGCVNFEGADRIELLIDAARQLKRSHVLPEGRSPGPPALKHPATKDLAAAAAQGPQLPPPQAQPQPSGTGGGVSGQDRYDRATSSGRLPLPSPALEYTLGSRLANGLGREEAVAEGARRALLGSMPGLMPPGLLAAAVSGLGSRGLTLAPGLSPARPLFGSDFEKEKQQRNADCLAELNEAMRGRAEEWHGRPKAVREQLLALSACAPFNVRFKKDHGLVGRVFAFDATARPPGYEFELKLFTEYPCGSGNVYAGVLAVARQMFHDALREPGKALASSGFKYLEYERRHGSGEWRQLGELLTDGVRSFREPAPAEALPQQYPEPAPAALCGPPPRAPSRNLAPTPRRRKASPEPEGEAAGKMTTEEQQRHWVAPGGPYSAETPGVPSPIAALKNVAEALGHSPKDPGGGGGPVRAGGASPAASSTAQPPTQHRLVARNGEAEVSPTAGAEAVSGGGSGTGATPGAPLCCTLCRERLEDTHFVQCPSVPGHKFCFPCSREFIKAQGPAGEVYCPSGDKCPLVGSSVPWAFMQGEIATILAGDIKVKKERDP
567	IRF2BP2 (IRF-2BP1_2 N-terminal domain)	MAAAVAVAAASRRQSCYLCDLPRMPWAMIWDFTEPVCRGCVNYEGADRVEFVIETARQLKRAHGCFPEGRSPPGAAASAAAKPPPLSAKDILLQQQQQLGHGGPEAAPRAPQALERYPLAAAAERPPRLGSDFGSSRPAASLAQPPTPQPPPVNGILVPNGFSKLEEPPELNRQSPNPRRGHAVPPTLVPLMNGSATPLPTALGLGGRAAASLAAVSGTAAASLGSAQPTDLGAHKRPASVSSSAAVEHEQREAAAKEKQPPPPAHRGPADSLSTAAGAAELSAEGAGKSRGSGEQDWVN

SEQ	Description	Sequence
		RPKTVRDTLLALHQHGHSGPFESKFKKEPALTAGRLLGFEANGANGSKAVARTARKRKPS PEPEGEVGPPKINGEAQPWLSTSTEGLKIPMTPTSSFVSPPPPTASPHSNRTTPPEAAQN GQSPMAALILVADNAGGSHASKDANQVHSTTRRNSNSPPSPSSMNQRRLGPREVGGQGAG NTGGLEPVHPASLPDSSLATSAPLCCTLCHERLEDTHFVQCPSVPSHKFCFPCSRQSIKQ QGASGEVYCPSGEKCPLVGSNVPWAFMQGEIATILAGDVKVKKERDS
568	IRF2BPL IRF-2BP1_2 N-terminal domain	MSAAQVSSSRRQSCYLCDLPRMPWAMIWDFSEPVCRGCVNYEGADRIEFVIETARQLKRA HGCFQDGRSPGPPPPVGVKTVALSAKEAAAAAAAAAAAAAAAQQQQQQQQQQQQQQQQQQ QQQQQQQLNHVDGSSKPAVLAAPSGLERYGLSAAAAAAAAAAAAVEQRSRFEYPPPPVSL GSSSHTARLPNGLGGPNGFPKPTPEEGPPELNRQSPNSSSAAASVASRRGTHGGLVTGLP NPGGGGGPQLTVPPNLLPQTLLNGPASAAVLPPPPPHALGSRGPPTPAPPGAPGGPACLG GTPGVSATSSSASSSTSSSVAEVGVGAGGKRPGSVSSTDQERELKEKQRNAEALAELSES LRNRAEEWASKPKMVRDTLLTLAGCTPYEVRFKKDHSLLGRVFAFDAVSKPGMDYELKLF IEYPTGSGNVYSSASGVAKQMYQDCMKDFGRGLSSGFKYLEYEKKHGSGDWRLLGDLLPE AVRFFKEGVPGADMLPQPYLDASCPMLPTALVSLSRAPSAPPGTGALPPAAPSGRGAAAS LRKRKASPEPPDSAEGALKLGEEQQRQQWMANQSEALKLTMSAGGFAAPGHAAGGPPPPP PPLGPHSNRTTPPESAPQNGPSPMAALMSVADTLGTAHSPKDGSSVHSTTASARRNSSSP VSPASVPGQRRLASRNGDLNLQVAPPPPSAHPGMDQVHPQNIPDSPMANSGPLCCTICHE RLEDTHFVQCPSVPSHKFCFPCSRESIKAQGATGEVYCPSGEKCPLVGSNVPWAFMQGEI ATILAGDVKVKKERDP
569	HOXA13 (homeodomain)	MTASVLLHPRWIEPTVMFLYDNGGGLVADELNKNMEGAAAAAAAAAAAAAAGAGGGGFPH PAAAAAGGNFSVAAAAAAAAAAAANQCRNLMAHPAPLAPGAASAYSSAPGEAPPSAAAAA AAAAAAAAAAAAASSSGGPGPAGPAGAEAAKQCSPCSAAAQSSSGPAALPYGYFGSGYYP CARMGPHPNAIKSCAQPASAAAAAAFADKYMDTAGPAAEEFSSRAKEFAFYHQGYAAGPY HHHQPMPGYLDMPVVPGLGGPGESRHEPLGLPMESYQPWALPNGWNGQMYCPKEQAQPPH LWKSTLPDVVSHPSDASSYRRGRKKRVPYTKVQLKELEREYATNKFITKDKRRRISATTN LSERQVTIWFQNRRVKEKKVINKLKTTS
570	HOXB13 (homeodomain)	MEPGNYATLDGAKDIEGLLGAGGGRNLVAHSPLTSHPAAPTLMPAVNYAPLDLPGSAEPP KQCHPCPGVPQGTSPAPVPYGYFGGGYYSCRVSRSSLKPCAQAATLAAYPAETPTAGEEY PSRPTEFAFYPGYPGTYQPMASYLDVSVVQTLGAPGEPRHDSLLPVDSYQSWALAGGWNS QMCCQGEQNPPGPFWKAAFADSSGQHPPDACAFRRGRKKRIPYSKGQLRELEREYAANKF ITKDKRRKISAATSLSERQITIWFQNRRVKEKKVLAKVKNSATP
571	HOXC13 (homeodomain)	MTTSLLLHPRWPESLMYVYEDSAAESGIGGGGGGGGGGTGGAGGGCSGASPGKAPSMDGL GSSCPASHCRDLLPHPVLGRPPAPLGAPQGAVYTDIPAPEAARQCAPPPAPPTSSSATLG YGYPFGGSYYGCRLSHNVNLQQKPCAYHPGDKYPEPSGALPGDDLSSRAKEFAFYPSFAS SYQAMPGYLDVSVVPGISGHPEPRHDALIPVEGYQHWALSNGWDSQVYCSKEQSQSAHLW KSPFPDVVPLQPEVSSYRRGRKKRVPYTKVQLKELEKEYAASKFITKEKRRRISATTNLS ERQVTIWFQNRRVKEKKVVSKSKAPHLHST
572	HOXA11 (homeodomain)	MDFDERGPCSSNMYLPSCTYYVSGPDFSSLPSFLPQTPSSRPMTYSYSSNLPQVQPVREV TFREYAIEPATKWHPRGNLAHCYSAEELVHRDCLQAPSAAGVPGDVLAKSSANVYHHPTP AVSSNFYSTVGRNGVLPQAFDQFFETAYGTPENLASSDYPGDKSAEKGPPAATATSAAAA AAATGAPATSSSDSGGGGGCRETAAAAEEKERRRRPESSSSPESSSGHTEDKAGGSSGQR TRKKRCPYTKYQIRELEREFFFSVYINKEKRLQLSRMLNLTDRQVKIWFQNRRMKEKKIN RDRLQYYSANPLL
573	HOXC11 (homeodomain)	MFNSVNLGNFCSPSRKERGADFGERGSCASNLYLPSCTYYMPEFSTVSSFLPQAPSRQIS YPYSAQVPPVREVSYGLEPSGKWHHRNSYSSCYAAADELMHRECLPPSTVTEILMKNEGS YGGHHHPSAPHATPAGFYSSVNKNSVLPQAFDRFFDNAYCGGGDPPAEPPCSGKGEAKGE PEAPPASGLASRAEAGAEAEAEEENTNPSSSGSAHSVAKEPAKGAAPNAPRTRKKRCPYS KFQIRELEREFFFNVYINKEKRLQLSRMLNLTDRQVKIWFQNRRMKEKKLSRDRLQYFSG NPLL
574	HOXC10 (homeodomain)	MTCPRNVTPNSYAEPLAAPGGGERYSRSAGMYMQSGSDFNCGVMRGCGLAPSLSKRDEGS SPSLALNTYPSYLSQLDSWGDPKAAYRLEQPVGRPLSSCSYPPSVKEENVCCMYSAEKRA KSGPEAALYSHPLPESCLGEHEVPVPSYYRASPSYSALDKTPHCSGANDFEAPFEQRASL NPRAEHLESPQLGGKVSFPETPKSDSQTPSPNEIKTEQSLAGPKGSPSESEKERAKAADS SPDTSDNEAKEEIKAENTTGNWLTAKSGRKKRCPYTKHQTLELEKEFLFNMYLTRERRLE ISKTINLTDRQVKIWFQNRRMKLKKMNRENRIRELTSNFNFT
575	HOXA10 (homeodomain)	MSARKGYLLPSPNYPTTMSCSESPAANSFLVDSLISSGRGEAGGGGGGAGGGGGGGYYAH GGVYLPPAADLPYGLQSCGLFPTLGGKRNEAASPGSGGGGGGLGPGAHGYGPSPIDLWLD APRSCRMEPPDGPPPPPQQQPPPPPQPPQPAPQATSCSFAQNIKEESSYCLYDSADKCPK VSATAAELAPFPRGPPPDGCALGTSSGVPVPGYFRLSQAYGTAKGYGSGGGGAQQLGAGP FPAQPPGRGFDLPPALASGSADAARKERALDSPPPPTLACGSGGGSQGDEEAHASSSAAE ELSPAPSESSKASPEKDSLGNSKGENAANWLTAKSGRKKRCPYTKHQTLELEKEFLFNMY LTRERRLEISRSVHLTDRQVKIWFQNRRMKLKKMNRENRIRELTANFNFS
576	HOXB9 (homeodomain)	MSISGTLSSYYVDSIISHESEDAPPAKFPSGQYASSRQPGHAEHLEFPSCSFQPKAPVFG ASWAPLSPHASGSLPSVYHPYIQPQGVPPAESRYLRTWLEPAPRGEAAPGQGQAAVKAEP LLGAPGELLKQGTPEYSLETSAGREAVLSNQRPGYGDNKICEGSEDKERPDQTNPSANWL HARSSRKKRCPYTKYQTLELEKEFLFNMYLTRDRRHEVARLLNLSERQVKIWFQNRRMKM KKMNKEQGKE

SEQ	Description	Sequence
577	HOXA9 (homeodomain)	MATTGALGNYYVDSFLLGADAADELSVGRYAPGTLGQPPRQAATLAEHPDFSPCSFQSKATVFGASWNPVHAAGANAVPAAVYHHHHHHPYVHPQAPVAAAAPDGRYMRSWLEPTPGALSFAGLPSSRPYGIKPEPLSARRGDCPTLDTHTLSLTDYACGSPPVDREKQPSEGAFSENNAENESGGDKPPIDPNNPAANWLHARSTRKKRCPYTKHQTLELEKEFLFNMYLTRDRRYEVARLLNLTERQVKIWFQNRRMKMKKINKDRAKDE
578	ZFP28_HUMAN	NKKLEAVGTGIEPKAMSQGLVTFGDVAVDFSQEEWEWLNPIQRNLYRKVMLENYRNLASLGLCVSKPDVISSLEQGKEPW
579	ZN334_HUMAN	KMKKFQIPVSFQDLTVNFTQEEWQQLDPAQRLLYRDVMLENYSNLVSVGYHVSKPDVIFKLEQGEEPWIVEEFSNQNYPD
580	ZN568_HUMAN	CSQESALSEEEEDTTRPLETVTFKDVAVDLTQEEWEQMKPAQRNLYRDVMLENYSNLVTVGCQVTKPDVIFKLEQEEEPW
581	ZN37A_HUMAN	ITSQGSVSFRDVTVGFTQEEWQHLDPAQRTLYRDVMLENYSHLVSVGYCIPKPEVILKLEKGEEPWILEEKFPSQSHLEL
582	ZN181_HUMAN	PQVTFNDVAIDFTHEEWGWLSSAQRDLYKDVMVQNYENLVSVAGLSVTKPYVITLLEDGKEPWMMEKKLSKGMIPDWESR
583	ZN510_HUMAN	PLRFSTLFQEQQKMNISQASVSFKDVTIEFTQEEWQQMAPVQKNLYRDVMLENYSNLVSVGYCCFKPEVIFKLEQGEEPW
584	ZN862_HUMAN	QDPSAEGLSEEVPVVFEELPVVFEDVAVYFTREEWGMLDKRQKELYRDVMRMNYELLASLGPAAAKPDLISKLERRAAPW
585	ZN140_HUMAN	SQGSVTFRDVAIDFSQEEWKWLQPAQRDLYRCVMLENYGHLVSLGLSISKPDVVSLLEQGKEPWLGKREVKRDLFSVSES
586	ZN208_HUMAN	GSLTFRDVAIEFSLEEWQCLDTAQQNLYRNVMLENYRNLVFLGIAAFKPDLIIFLEEGKESWNMKRHEMVEESPVICSHF
587	ZN248_HUMAN	NKSQEQVSFKDVCVDFTQEEWYLLDPAQKILYRDVILENYSNLVSVGYCITKPEVIFKIEQGEEPWILEKGFPSQCHPER
588	ZN571_HUMAN	PHLLVTFRDVAIDFSQEEWECLDPAQRDLYRDVMLENYSNLISLDLESSCVTKKLSPEKEIYEMESLQWENMGKRINHHL
589	ZN699_HUMAN	EEERKTAELQKNRIQDSVVFEDVAVDFTQEEWALLDLAQRNLYRDVMLENFQNLASLGYPLHTPHLISQWEQEEDLQTVK
590	ZN726_HUMAN	GLLTFRDVAIEFSLEEWQCLDTAQKNLYRNVMLENYRNLAFLGIAVSKPDLIICLEKEKEPWNMKRDEMVDEPPGICPHF
591	ZIK1_HUMAN	RAPTQVTVSPETHMDLTKGCVTFEDIAIYFSQDEWGLLDEAQRLLYLEVMLENFALVASLGCGHGTEDEETPSDQNVSVG
592	ZNF2_HUMAN	AAVSPTTRCQESVTFEDVAVVFTDEEWSRLVPIQRDLYKEVMLENYNSIVSLGLPVPQPDVIFQLKRGDKPWMVDLHGSE
593	Z705F_HUMAN	HSLEKVTFEDVAIDFTQEEWDMMDTSKRKLYRDVMLENISHLVSLGYQISKSYIILQLEQGKELWREGRVFLQDQNPDRE
594	ZNF14_HUMAN	DSVSFEDVAVNFTLEEWALLDSSQKKLYEDVMQETFKNLVCLGKKWEDQDIEDDHRNQGKNRRCHMVERLCESRRGSKCG
595	ZN471_HUMAN	NVEVVKVMPQDLVTFKDVAIDFSQEEWQWMNPAQKRLYRSMMLENYQSLVSLGLCISKPYVISLLEQGREPWEMTSEMTR
596	ZN624_HUMAN	TQPDEDLHLQAEETQLVKESVTFKDVAIDFTLEEWRLMDPTQRNLHKDVMLENYRNLVSLGLAVSKPDMISHLENGKGPW
597	ZNF84_HUMAN	TMLQESFSFDDLSVDFTQKEWQLLDPSQKNLYKDVMLENYSSLVSLGYEVMKPDVIFKLEQGEEPWVGDGEIPSSDSPEV
598	ZNF7_HUMAN	EVVTFGDVAVHFSREEWQCLDPGQRALYREVMLENHSSVAGLAGFLVFKPELISRLEQGEEPWVLDLQGAEGTEAPRTSK
599	ZN891_HUMAN	RNAEEERMIAVFLTTWLQEPMTFKDVAVEFTQEEWMMLDSAQRSLYRDVMLENYRNLTSVEYQLYRLTVISPLDQEEIRN
600	ZN337_HUMAN	GPQGARRQAFLAFGDVTVDFTQKEWRLLSPAQRALYREVTLENYSHLVSLGILHSKPELIRRLEQGEVPWGEERRRRPGP
601	Z705G_HUMAN	HSLKKLTFEDVAIDFTQEEWAMMDTSKRKLYRDVMLENISHLVSLGYQISKSYIILQLEQGKELWREGRVFLQDQNPNRE
602	ZN529_HUMAN	MPEVEFPDQFFTVLTMDHELVTLRDVVINFSQEEWEYLDSAQRNLYWDVMMENYSNLLSLDLESRNETKHLSVGKDIIQN
603	ZN729_HUMAN	PGAPGSLEMGPLTFRDVTIEFSLEEWQCLDTVQQNLYRDVMLENYRNLVFLGMAVFKPDLITCLKQGKEPWNMKRHEMVT
604	ZN419_HUMAN	RDPAQVPVAADLLTDHEEGYVTFEDVAVYFSQEEWRLLDDAQRLLYRNVMLENFTLLASLGLASSKTHEITQLESWEEPF
605	Z705A_HUMAN	HSLKKVTFEDVAIDFTQEEWAMMDTSKRKLYRDVMLENISHLVSLGYQISKSYIILQLEQGKELWREGREFLQDQNPDRE
606	ZNF45_HUMAN	TKSKEAVTFKDVAVVFSEEELQLLDLAQRKLYRDVMLENFRNVVSVGHQSTPDGLPQLEREEKLWMMKMATQRDNSSGAK
607	ZN302_HUMAN	SQVTFSDVAIDFSHEEWACLDSAQRDLYKDVMVQNYENLVSVGLSVTKPYVIMLLEDGKEPWMMEKKLSKAYPFPLSHSV

SEQ	Description	Sequence
608	ZN486_HUMAN	PGPLRSLEMESLQFRDVAVEFSLEEWHCLDTAQQNLYRDVMLENYRHLVFLGIIVSKPDLITCLEQGIKPLTMKRHEMIA
609	ZN621_HUMAN	LQTTWPQESVTFEDVAVYFTQNQWASLDPAQRALYGEVMLENYANVASLVAFPFPKPALISHLERGEAPWGPDPWDTEIL
610	ZN688_HUMAN	APLLAPRPGETRPGCRKPGTVSFADVAVYFSPEEWGCLRPAQRALYRDVMQETYGHLGALGFPGPKPALISWMEQESEAW
611	ZN33A_HUMAN	NKVEQKSQESVSFKDVTVGFTQEEWQHLDPSQRALYRDVMLENYSNLVSVGYCVHKPEVIFRLQQGEEPWKQEEEFPSQS
612	ZN554_HUMAN	CFSQEERMAAGYLPRWSQELVTFEDVSMDFSQEEWELLEPAQKNLYREVMLENYRNVVSLEALKNQCTDVGIKEGPLSPA
613	ZN878_HUMAN	DSVAFEDVAVNFTQEEWALLDPSQKNLYREVMQETLRNLTSIGKKWNNQYIEDEHQNPRRNLRRLIGERLSESKESHQHG
614	ZN772_HUMAN	MGPAQVPMNSEVIVDPIQGQVNFEDVFVYFSQEEWVLLDEAQRLLYRDVMLENFALMASLGHTSFMSHIVASLVMGSEPW
615	ZN224_HUMAN	TTFKEAMTFKDVAVVFTEEELGLLDLAQRKLYRDVMLENFRNLLSVGHQAFHRDTFHFLREEKIWMMKTAIQREGNSGDK
616	ZN184_HUMAN	DSTLLQGGHNLLSSASFQEAVTFKDVIVDFTQEEWKQLDPGQRDLFRDVTLENYTHLVSIGLQVSKPDVISQLEQGTEPW
617	ZN544_HUMAN	EARSMLVPPQASVCFEDVAMAFTQEEWEQLDLAQRTLYREVTLETWEHIVSLGLFLSKSDVISQLEQEEDLCRAEQEAPR
618	ZNF57_HUMAN	DSVVFEDVAVDFTLEEWALLDSAQRDLYRDVMLETFRNLASVDDGTQFKANGSVSLQDMYGQEKSKEQTIPNFTGNNSCA
619	ZN283_HUMAN	EESHGALISSCNSRTMTDGLVTFRDVAIDFSQEEWECLDPAQRDLYVDVMLENYSNLVSLDLESKTYETKKIFSENDIFE
620	ZN549_HUMAN	VITPQIPMVTEEFVKPSQGHVTFEDIAVYFSQEEWGLLDEAQRCLYHDVMLENFSLMASVGCLHGIEAEEAPSEQTLSAQ
621	ZN211_HUMAN	VQLRPQTRMATALRDPASGSVTFEDVAVYFSWEEWDLLDEAQKHLYFDVMLENFALTSSLGCWCGVEHEETPSEQRISGE
622	ZN615_HUMAN	MQAQESLTLEDVAVDFTWEEWQFLSPAQKDLYRDVMLENYSNLVAVGYQASKPDALSKLERGEETCTTEDEIYSRICSEI
623	ZN253_HUMAN	GPLQFRDVAIEFSLEEWHCLDTAQRNLYRDVMLENYRNLVFLGIVVSKPDLVTCLEQGKKPLTMERHEMIAKPPVMSSHF
624	ZN226_HUMAN	NMFKEAVTFKDVAVAFTEEELGLLGPAQRKLYRDVMVENFRNLLSVGHPPFKQDVSPIERNEQLWIMTTATRRQGNLGEK
625	ZN730_HUMAN	GALTFRDVAIEFSLEEWQCLDTEQQNLYRNVMLDNYRNLVFLGIAVSKPDLITCLEQEKEPWNLKTHDMVAKPPVICSHI
626	Z585A_HUMAN	SPQKSSALAPEDHGSSYEGSVSFRDVAIDFSREEWRHLDPSQRNLYRDVMLETYSHLLSVGYQVPEAEVVMLEQGKEPWA
627	ZN732_HUMAN	ELLTFRDVAIEFSPEEWKCLDPAQQNLYRDVMLENYRNLISLGVAISNPDLVIYLEQRKEPYKVKIHETVAKHPAVCSHF
628	ZN681_HUMAN	EPLKFRDVAIEFSLEEWQCLDTIQQNLYRNVMLENYRNLVFLGIVVSKPDLITCLEQEKEPWTRKRHRMVAEPPVICSHF
629	ZN667_HUMAN	PSARGKSKSKAPITFGDLAIYFSQEEWEWLSPIQKDLYEDVMLENYRNLVSLGLSFRRPNVITLLEKGKAPWMVEPVRRR
630	ZN649_HUMAN	TKAQESLTLEDVAVDFTWEEWQFLSPAQKDLYRDVMLENYSNLVSVGYQAGKPDALTKLEQGEPLWTLEDEIHSPAHPEI
631	ZN470_HUMAN	SQEEVEVAGIKLCKAMSLGSVTFTDVAIDFSQDEWEWLNLAQRSLYKKVMLENYRNLVSVGLCISKPDVISLLEQEKDPW
632	ZN484_HUMAN	TKSLESVSFKDVTVDFSRDEWQQLDLAQKSLYREVMLENYFNLISVGCQVPKPEVIFSLEQEEPCMLDGEIPSQSRPDGD
633	ZN431_HUMAN	SGCPGAERNLLVYSYFEKETLTFRDVAIEFSLEEWECLNPAQQNLYMNVMLENYKNLVFLGVAVSKQDPVTCLEQEKEPW
634	ZN382_HUMAN	PLQGSVSFKDVTVDFTQEEWQQLDPAQKALYRDVMLENYCHFVSVGFHMAKPDMIRKLEQGEELWTQRIFPSYSYLEEDG
635	ZN254_HUMAN	PGPPRSLEMGLLTFRDVAIEFSLEEWQHLDIAQQNLYRNVMLENYRNLAFLGIAVSKPDLITCLEQGKEPWNMKRHEMVD
636	ZN124_HUMAN	SGHPGSWEMNSVAFEDVAVNFTQEEWALLDPSQKNLYRDVMQETFRNLASIGNKGEDQSIEDQYKNSSRNLRHIISHSGN
637	ZN607_HUMAN	SYGSITFGDVAIDFSHQEWEYLSLVQKTLYQEVMMENYDNLVSLAGHSVSKPDLITLLEQGKEPWMIVREETRGECTDLD
638	ZN317_HUMAN	DLFVCSGLEPHTPSVGSQESVTFQDVAVDFTEKEWPLLDSSQRKLYKDVMLENYSNLTSLGYQVGKPSLISHLEQEEEPR
639	ZN620_HUMAN	FQTAWRQEPVTFEDVAVYFTQNEWASLDSVQRALYREVMLENYANVASLAFPFTTPVLVSQLEQGELPWGLDPWEPMGRE
640	ZN141_HUMAN	ELLTFRDVAIEFSPEEWKCLDPDQQNLYRDVMLENYRNLVSLGVAISNPDLVTCLEQRKEPYNVKIHKIVARPPAMCSHF

SEQ	Description	Sequence
641	ZN584_HUMAN	AGEAEAQLDPSLQGLVMFEDVTVYFSREEWGLLNVTQKGLYRDVMLENFALVSSLGLAPSRSPVFTQLEDDEQSWVPSWV
642	ZN540_HUMAN	AHALVTFRDVAIDFSQKEWECLDTTQRKLYRDVMLENYNNLVSLGYSGSKPDVITLLEQGKEPCVVARDVTGRQCPGLLS
643	ZN75D_HUMAN	KRIKHWKMASKLILPESLSLLTFEDVAVYFSEEEWQLLNPLEKTLYNDVMQDIYETVISLGLKLKNDTGNDHPISVSTSE
644	ZN555_HUMAN	DSVVFEDVAVDFTLEEWALLDSAQRDLYRDVMLETFQNLASVDDETQFKASGSVSQQDIYGEKIPKESKIATFTRNVSWA
645	ZN658_HUMAN	NMSQASVSFQDVTVEFTREEWQHLGPVERTLYRDVMLENYSHLISVGYCITKPKVISKLEKGEEPWSLEDEFLNQRYPGY
646	ZN684_HUMAN	ISFQESVTFQDVAVDFTAEEWQLLDCAERTLYWDVMLENYRNLISVGCPITKTKVILKVEQGQEPWMVEGANPHESSPES
647	RBAK_HUMAN	NTLQGPVSFKDVAVDFTQEEWQQLDPDEKITYRDVMLENYSHLVSVGYDTTKPNVIIKLEQGEEPWIMGGEFPCQHSPEA
648	ZN829_HUMAN	HPEEEERMHDELLQAVSKGPVMFRDVSIDFSQEEWECLDADQMNLYKEVMLENFSNLVSVGLSNSKPAVISLLEQGKEPW
649	ZN582_HUMAN	SLGSELFRDVAIVFSQEEWQWLAPAQRDLYRDVMLETYSNLVSLGLAVSKPDVISFLEQGKEPWMVERVVSGGLCPVLES
650	ZN112_HUMAN	TKFQEMVTFKDVAVVFTEEELGLLDSVQRKLYRDVMLENFRNLLLVAHQPFKPDLISQLEREEKLLMVETETPRDGCSGR
651	ZN716_HUMAN	AKRPGPPGSREMGLLTFRDIAIEFSLAEWQCLDHAQQNLYRDVMLENYRNLVSLGIAVSKPDLITCLEQNKEPQNIKRNE
652	HKR1_HUMAN	TCMVHRQTMSCSGAGGITAFVAFRDVAVYFTQEEWRLLSPAQRTLHREVMLETYNHLVSLEIPSSKPKLIAQLERGEAPW
653	ZN350_HUMAN	IQAQESITLEDVAVDFTWEEWQLLGAAQKDLYRDVMLENYSNLVAVGYQASKPDALFKLEQGEQLWTIEDGIHSGACSDI
654	ZN480_HUMAN	AQKRRKRKAKESGMALPQGHLTFRDVAIEFSQAEWKCLDPAQRALYKDVMLENYRNLVSLGISLPDLNINSMLEQRREPW
655	ZN416_HUMAN	DSTSVPVTAEAKLMGFTQGCVTFEDVAIYFSQEEWGLLDEAQRLLYRDVMLENFALITALVCWHGMEDEETPEQSVSVEG
656	ZNF92_HUMAN	GPLTFRDVKIEFSLEEWQCLDTAQRNLYRDVMLENYRNLVFLGIAVSKPDLITWLEQGKEPWNLKRHEMVDKTPVMCSHF
657	ZN100_HUMAN	SGCPGAERSLLVQSYFEKGPLTFRDVAIEFSLEEWQCLDSAQQGLYRKVMLENYRNLVFLAGIALTKPDLITCLEQGKEP
658	ZN736_HUMAN	GVLTFRDVAVEFSPEEWECLDSAQQRLYRDVMLENYGNLVSLGLAIFKPDLMTCLEQRKEPWKVKRQEAVAKHPAGSFHF
659	ZNF74_HUMAN	KENLEDISGWGLPEARSKESVSFKDVAVDFTQEEWGQLDSPQRALYRDVMLENYQNLLALGPPLHKPDVISHLERGEEPW
660	CBX1_HUMAN	EESEKPRGFARGLEPERIIGATDSSGELMFLMKWKNSDEADLVPAKEANVKCPQVVISFYEERLTWHSYPSEDDDKKDDK
661	ZN443_HUMAN	ASVALEDVAVNFTREEWALLGPCQKNLYKDVMQETIRNLDCVVMKWKDQNIEDQYRYPRKNLRCRMLERFVESKDGTQCG
662	ZN195_HUMAN	TLLTFRDVAIEFSLEEWKCLDLAQQNLYRDVMLENYRNLFSVGLTVCKPGLITCLEQRKEPWNVKRQEAADGHPEMGFHH
663	ZN530_HUMAN	AAALRAPTQQVFVAFEDVAIYFSQEEWELLDEMQRLLYRDVMLENFAVMASLGCWCGAVDEGTPSAESVSVEELSQGRTP
664	ZN782_HUMAN	NTFQASVSFQDVTVEFSQEEWQHMGPVERTLYRDVMLENYSHLVSVGYCFTKPELIFTLEQGEDPWLLEKEKGFLSRNSP
665	ZN791_HUMAN	DSVAFEDVSVSFSQEEWALLAPSQKKLYRDVMQETFKNLASIGEKWEDPNVEDQHKNQGRNLRSHTGERLCEGKEGSQCA
666	ZN331_HUMAN	AQGLVTFADVAIDFSQEEWACLNSAQRDLYWDVMLENYSNLVSLDLESAYENKSLPTEKNIHEIRASKRNSDRRSKSLGR
667	Z354C_HUMAN	AVDLLSAQEPVTFRDVAVFFSQDEWLHLDSAQRALYREVMLENYSSLVSLGIPFSMPKLIHQLQQGEDPCMVEREVPSDT
668	ZN157_HUMAN	SPQRFPALIPGEPGRSFEGSVSFEDVAVDFTRQEWHRLDPAQRTMHKDVMLETYSNLASVGLCVAKPEMIFKLERGEELW
669	ZN727_HUMAN	RVLTFRDVAVEFSPEEWECLDSAQQRLYRDVMLENYGNLFSLGLAIFKPDLITYLEQRKEPWNARRQKTVAKHPAGSLHF
670	ZN550_HUMAN	AETKDAAQMLVTFKDVAVTFTREEWRQLDLAQRTLYREVMLETCGLLVSLGHRVPKPELVHLLEHGQELWIVKRGLSHAT
671	ZN793_HUMAN	IEYQIPVSFKDVVVGFTQEEWHRLSPAQRALYRDVMLETYSNLVSVGYEGTKPDVILRLEQEEAPWIGEAACPGCHCWED
672	ZN235_HUMAN	TKFQEAVTFKDVAVAFTEEELGLLDSAQRKLYRDVMLENFRNLVSVGHQSFKPDMISQLEREEKLWMKELQTQRGKHSGD
673	ZNF8_HUMAN	DEGVAGVMSVGPPAARLQEPVTFRDVAVDFTQEEWGQLDPTQRILYRDVMLETFGHLLSIGPELPKPEVISQLEQGTELW

SEQ	Description	Sequence
674	ZN724_HUMAN	GPLTFMDVAIEFSVEEWQCLDTAQQNLYRNVMLENYRNLVFLGIAVSKPDLITCLEQGKEPWNMERHEMVAKPPGMCCYF
675	ZN573_HUMAN	HQVGLIRSYNSKTMTCFQELVTFRDVAIDFSRQEWEYLDPNQRDLYRDVMLENYRNLVSLGGHSISKPVVVDLLERGKEP
676	ZN577_HUMAN	NATIVMSVRREQGSSSGEGSLSFEDVAVGFTREEWQFLDQSQKVLYKEVMLENYINLVSIGYRGTKPDSLFKLEQGEPPG
677	ZN789_HUMAN	FPPARGKELLSFEDVAMYFTREEWGHLNWGQKDLYRDVMLENYRNMVLLGFQFPKPEMICQLENWDEQWILDLPRTGNRK
678	ZN718_HUMAN	ELLTFKDVAIEFSPEEWKCLDTSQQNLYRDVMLENYRNLVSLGVSISNPDLVTSLEQRKEPYNLKIHETAARPPAVCSHF
679	ZN300_HUMAN	MKSQGLVSFKDVAVDFTQEEWQQLDPSQRTLYRDVMLENYSHLVSMGYPVSKPDVISKLEQGEEPWIIKGDISNWIYPDE
680	ZN383_HUMAN	AEGSVMFSDVSIDFSQEEWDCLDPVQRDLYRDVMLENYGNLVSMGLYTPKPQVISLLEQGKEPWMVGRELTRGLCSDLES
681	ZN429_HUMAN	GPLTFTDVAIEFSLEEWQCLDTAQQNLYRNVMLENYRNLVFLGIAVSKPDLITCLEKEKEPCKMKRHEMVDEPPVVCSHF
682	ZN677_HUMAN	ALSQGLFTFKDVAIEFSQEEWECLDPAQRALYRDVMLENYRNLLSLDEDNIPPEDDISVGFTSKGLSPKENNKEELYHLV
683	ZN850_HUMAN	NMEGLVMFQDLSIDFSQEEWECLDAAQKDLYRDVMMENYSSLVSLGLSIPKPDVISLLEQGKEPWMVSRDVLGGWCRDSE
684	ZN454_HUMAN	AVSHLPTMVQESVTFKDVAILFTQEEWGQLSPAQRALYRDVMLENYSNLVSLGLLGPKPDTFSQLEKREVWMPEDTPGGF
685	ZN257_HUMAN	GPLTIRDVTVEFSLEEWHCLDTAQQNLYRDVMLENYRNLVFLGIAVSKPDLITCLEQGKEPCNMKRHEMVAKPPVMCSHI
686	ZN264_HUMAN	AAAVLTDRAQVSVTFDDVAVTFTKEEWGQLDLAQRTLYQEVMLENCGLLVSLGCPVPKAELICHLEHGQEPWTRKEDLSQ
687	ZFP82_HUMAN	ALRSVMFSDVSIDFSPEEWEYLDLEQKDLYRDVMLENYSNLVSLGCFISKPDVISSLEQGKEPWKVVRKGRRQYPDLETK
688	ZFP14_HUMAN	AHGSVTFRDVAIDFSQEEWEFLDPAQRDLYRDVMWENYSNFISLGPSISKPDVITLLDEERKEPGMVVREGTRRYCPDLE
689	ZN485_HUMAN	APRAQIQGPLTFGDVAVAFTRIEWRHLDAAQRALYRDVMLENYGNLVSVGLLSSKPKLITQLEQGAEPWTEVREAPSGTH
690	ZN737_HUMAN	GPLQFRDVAIEFSLEEWHCLDTAQRNLYRNVMLENYRNLVFLGIVVSKPDLITCLEQGKKPLTMKKHEMVANPSVTCSHF
691	ZNF44_HUMAN	TLPRGQPEVLEWGLPKDQDSVAFEDVAVNFTHEEWALLGPSQKNLYRDVMRETIRNLNCIGMKWENQNIDDQHQNLRRNP
692	ZN596_HUMAN	PSPDSMTFEDIIVDFTQEEWALLDTSQRKLFQDVMLENISHLVSIGKQLCKSVVLSQLEQVEKLSTQRISLLQGREVGIK
693	ZN565_HUMAN	EESREIRAGQIVLKAMAQGLVTFRDVAIEFSLEEWKCLEPAQRDLYREVTLENFGHLASLGLSISKPDVVSLLEQGKEPW
694	ZN543_HUMAN	AASAQVSVTFEDVAVTFTQEEWGQLDAAQRTLYQEVMLETCGLLMSLGCPLFKPELIYQLDHRQELWMATKDLSQSSYPG
695	ZFP69_HUMAN	RESLEDEVTPGLPTAESQELLTFKDISIDFTQEEWGQLAPAHQNLYREVMLENYSNLVSVGYQLSKPSVISQLEKGEEPW
696	SUMO1_HUMAN	EGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQRQGVPMNSLRFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGG
697	ZNF12_HUMAN	NKSLGPVSFKDVAVDFTQEEWQQLDPEQKITYRDVMLENYSNLVSVGYHIIKPDVISKLEQGEEPWIVEGEFLLQSYPDE
698	ZN169_HUMAN	SPGLLTTRKEALMAFRDVAVAFTQKEWKLLSSAQRTLYREVMLENYSHLVSLGIAFSKPKLIEQLEQGDEPWREENEHLL
699	ZN433_HUMAN	MFQDSVAFEDVAVTFTQEEWALLDPSQKNLCRDVMQETFRNLASIGKKWKPQNIYVEYENLRRNLRIVGERLFESKEGHQ
700	SUMO3_HUMAN	ENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG
701	ZNF98_HUMAN	PGPLGSLEMGVLTFRDVALEFSLEEWQCLDTAQQNLYRNVMLENYRNLVFVGIAASKPDLITCLEQGKEPWNVKRHEMVT
702	ZN175_HUMAN	LSQKPQVLGPEKQDGSCEASVSFEDVTVDFSREEWQQLDPAQRCLYRDVMLELYSHLFAVGYHIPNPEVIFRMLKEKEPR
703	ZN347_HUMAN	ALTQGQVTFRDVAIEFSQEEWTCLDPAQRTLYRDVMLENYRNLASLGISCFDLSIISMLEQGKEPFTLESQVQIAGNPDG
704	ZNF25_HUMAN	NKFQGPVTLKDVIVEFTKEEWKLLTPAQRTLYKDVMLENYSHLVSVGYHVNKPNAVFKLKQGKEPWILEVEFPHRGFPED
705	ZN519_HUMAN	ELLTFRDVAIEFSPEEWKCLDPAQQNLYRDVMLENYRNLVSLAVYSYYNQGILPEQGIQDSFKKATLGRYGSCGLENICL
706	Z585B_HUMAN	SPQKSSALAPEDHGSSYEGSVSFRDVAIDFSREEWRHLDLSQRNLYRDVMLETYSHLLSVGYQVPKPEVVMLEQGKEPWA

SEQ	Description	Sequence
707	ZIM3_HUMAN	NNSQGRVTFEDVTVNFTQGEWQRLNPEQRNLYRDVMLENYSNLVSVGQGETTKPDVILRLEQGKEPWLEEEEVLGSGRAE
708	ZN517_HUMAN	AMALPMPGPQEAVVFEDVAVYFTRIEWSCLAPDQQALYRDVMLENYGNLASLGFLVAKPALISLLEQGEEPGALILQVAE
709	ZN846_HUMAN	DSSQHLVTFEDVAVDFTQEEWTLLDQAQRDLYRDVMLENYKNLIIILAGSELFKRSLMSGLEQMEELRTGVTGVLQELDLQ
710	ZN230_HUMAN	TTFKEAVTFKDVAVFFTEEELGLLDPAQRKLYQDVMLENFTNLLSVGHQPFHPFHFLREEKFWMMETATQREGNSGGKTI
711	ZNF66_HUMAN	GPLQFRDVAIEFSLEEWHCLDMAQRNLYRDVMLENYRNLVFLGIVVSKPDLITHLEQGKKPSTMQRHEMVANPSVLCSHF
712	ZFP1_HUMAN	NKSQGSVSFTDVTVDFTQEEWEQLDPSQRILYMDVMLENYSNLLSVEVWKADDQMERDHRNPDEQARQFLILKNQTPIEE
713	ZN713_HUMAN	EEEEMNDGSQMVRSQESLTFQDVAVDFTREEWDQLYPAQKNLYRDVMLENYRNLVALGYQLCKPEVIAQLELEEEWVIER
714	ZN816_HUMAN	EEATKKSKEKEPGMALPQGRLTFRDVAIEFSLEEWKCLNPAQRALYRAVMLENYRNLEFVDSSLKSMMEFSSTRHSITGE
715	ZN426_HUMAN	EKTPAGRIVADCLTDCYQDSVTFDDVAVDFTQEEWTLLDSTQRSLYSDVMLENYKNLATVGGQIIKPSLISWLEQEESRT
716	ZN674_HUMAN	AMSQESLTFKDVFVDFTLEEWQQLDSAQKNLYRDVMLENYSHLVSVGHLVGKPDVIFRLGPGDESWMADGGTPVRTCAGE
717	ZN627_HUMAN	DSVAFEDVAVNFTLEEWALLDPSQKNLYRDVMRETFRNLASVGKQWEDQNIEDPFKIPRRNISHIPERLCESKEGGQGEE
718	ZNF20_HUMAN	MFQDSVAFEDVAVSFTQEEWALLDPSQKNLYRDVMQETFKNLTSVGKTWKVQNIEDEYKNPRRNLSLMREKLCESKESHH
719	Z587B_HUMAN	AVVATLRLSAQGTVTFEDVAVKFTQEEWNLLSEAQRCLYRDVTLENLALMSSLGCWCGVEDEAAPSKQSIYIQRETQVRT
720	ZN316_HUMAN	EEEEEDEDEDDLLTAGCQELVTFEDVAVYFSLEEWERLEADQRGLYQEVMQENYGILVSLGYPIPKPDLIFRLEQGEEPW
721	ZN233_HUMAN	TKFQEMVTFKDVAVVFTREELGLLDLAQRKLYQDVMLENFRNLLSVGYQPFKLDVILQLGKEDKLRMMETEIQGDGCSGH
722	ZN611_HUMAN	EEAAQKRKGKEPGMALPQGRLTFRDVAIEFSLAEWKCLNPSQRALYREVMLENYRNLEAVDISSKCMMKEVLSTGQGNTE
723	ZN556_HUMAN	DTVVFEDVVVDFTLEEWALLNPAQRKLYRDVMLETFKHLASVDNEAQLKASGSISQQDTSGEKLSLKQKIEKFTRKNIWA
724	ZN234_HUMAN	TTFKEGLTFKDVAVVFTEEELGLLDPVQRNLYQDVMLENFRNLLSVGHHPFKHDVFLLEKEKKLDIMKTATQRKGKSADK
725	ZN560_HUMAN	SALQQEFWKIQTSNGIQMDLVTFDSVAVEFTQEEWTLLDPAQRNLYSDVMLENYKNLSSVGYQLFKPSLISWLEEEEELS
726	ZNF77_HUMAN	DCVIFEEVAVNFTPEEWALLDHAQRSLYRDVMLETCRNLASLDCYIYVRTSGSSSQRDVFGNGISNDEEIVKFTGSDSWS
727	ZN682_HUMAN	ELLTFRDVTIEFSLEEWEFLNPAQQSLYRKVMLENYRNLVSLGLTVSKPELISRLEQRQEPWNVKRHETIAKPPAMSSHY
728	ZN614_HUMAN	IKTQESLTLEDVAVEFSWEEWQLLDTAQKNLYRDVMVENYNHLVSLGYQTSKPDVLSKLAHGQEPWTTDAKIQNKNCPGI
729	ZN785_HUMAN	PAHVPGEAGPRRTRESRPGAVSFADVAVYFSPEEWECLRPAQRALYRDVMRETFGHLGALGFSVPKPAFISWVEGEVEAW
730	ZN445_HUMAN	GCPGDQVTPTRSLTAQLQETMTFKDVEVTFSQDEWGWLDSAQRNLYRDVMLENYRNMASLVGPFTKPALISWLEAREPWG
731	ZFP30_HUMAN	ARDLVMFRDVAVDFSQEEWECLNSYQRNLYRDVILENYSNLVSLAGCSISKPDVITLLEQGKEPWMVVRDEKRRWTLDLE
732	ZN225_HUMAN	TTLKEAVTFKDVAVVFTEEELRLLDLAQRKLYREVMLENFRNLLSVGHQSLHRDTFHFLKEEKFWMMETATQREGNLGGK
733	ZN551_HUMAN	SPPSPRSSMAAVALRDSAQGMTFEDVAIYFSQEEWELLDESQRFLYCDVMLENFAHVTSLGYCHGMENEAIASEQSVSIQ
734	ZN610_HUMAN	DEEAQKRKAKESGMALPQGRLTFMDVAIEFSQEEWKSLDPGQRALYRDVMLENYRNLVFLGICLPDLSIISMLKQRREPL
735	ZN528_HUMAN	ALTQGPLKFMDVAIEFSQEEWKCLDPAQRTLYRDVMLENYRNLVSLGICLPDLSVTSMLEQKRDPWTLQSEEKIANDPDG
736	ZN284_HUMAN	TMFKEAVTFKDVAVVFTEEELGLLDVSQRKLYRDVMLENFRNLLSVGHQLSHRDTFHFQREEKFWIMETATQREGNSGGK
737	ZN418_HUMAN	QGTVAFEDVAVNFSQEEWSLLSEVQRCLYHDVMLENWVLISSLGCWCGSEDEEAPSKKSISIQRVSQVSTPGAGVSPKKA
738	MPP8_HUMAN	AEAFGDSEEDGEDVFEVEKILDMKTEGGKVLYKVRWKGYTSDDDTWEPEIHLEDCKEVLLEFRKKIAENKAKAVRKDIQR
739	ZN490_HUMAN	VLQMQNSEHHGQSIKTQTDSISLEDVAVNFTLEEWALLDPGQRNIYRDVMRATFKNLACIGEKWKDQDIEDEHKNQGRNL

SEQ	Description	Sequence
740	ZN805_HUMAN	AMALTDPAQVSVTFDDVAVTFTQEEWGQLDLAQRTLYQEVMLENCGLLVSLGCPVPRPELIYHLEHGQEPWTRKEDLSQG
741	Z780B_HUMAN	VHGSVTFRDVAIDFSQEEWECLQPDQRTLYRDVMLENYSHLISLGSSISKPDVITLLEQEKEPWIVVSKETSRWYPDLES
742	ZN763_HUMAN	DPVACEDVAVNFTQEEWALLDISQRKLYREVMLETFRNLTSIGKKWKDQNIEYEYQNPRRNFRSLIEGNVNEIKEDSHCG
743	ZN285_HUMAN	IKFQERVTFKDVAVVFTKEELALLDKAQINLYQDVMLENFRNLMLVRDGIKNNILNLQAKGLSYLSQEVLHCWQIWKQRI
744	ZNF85_HUMAN	GPLTFRDVAIEFSLKEWQCLDTAQRNLYRNVMLENYRNLVFLGITVSKPDLITCLEQGKEAWSMKRHEIMVAKPTVMCSH
745	ZN223_HUMAN	TMSKEAVTFKDVAVVFTEEELGLLDLAQRKLYRDVMLENFRNLLSVGHQPFHRDTFHFLREEKFWMMDIATQREGNSGGK
746	ZNF90_HUMAN	GPLEFRDVAIEFSLEEWHCLDTAQQNLYRDVMLENYRHLVFLGIVVTKPDLITCLEQGKKPFTVKRHEMIAKSPVMCFHF
747	ZN557_HUMAN	GHTEGGELVNELLKSWLKGLVTFEDVAVEFTQEEWALLDPAQRTLYRDVMLENCRNLASLGNQVDKPRLISQLEQEDKVM
748	ZN425_HUMAN	AEPASVTVTFDDVALYFSEQEWEILEKWQKQMYKQEMKTNYETLDSLGYAFSKPDLITWMEQGRMLLISEQGCLDKTRRT
749	ZN229_HUMAN	HSQASAISQDREEKIMSQEPLSFKDVAVVFTEEELELLDSTQRQLYQDVMQENFRNLLSVGERNPLGDKNGKDTEYIQDE
750	ZN606_HUMAN	GSLEEGRRATGLPAAQVQEPVTFKDVAVDFTQEEWGQLDLVQRTLYRDVMLETYGHLLSVGNQIAKPEVISLLEQGEEPW
751	ZN155_HUMAN	TTFKEAVTFKDVAVVFTEEELGLLDPAQRKLYRDVMLENFRNLLSVGHQPFHQDTCHFLREEKFWMMGTATQREGNSGGK
752	ZN222_HUMAN	AKLYEAVTFKDVAVIFTEEELGLLDPAQRKLYRDVMLENFRNLLSVGGKIQTEMETVPEAGTHEEFSCKQIWEQIASDLT
753	ZN442_HUMAN	RSDLFLPDSQTNEERKQYDSVAFEDVAVNFTQEEWALLGPSQKSLYRDVMWETIRNLDCIGMKWEDTNIEDQHRNPRRSL
754	ZNF91_HUMAN	PGTPGSLEMGLLTFRDVAIEFSPEEWQCLDTAQQNLYRNVMLENYRNLAFLGIALSKPDLITYLEQGKEPWNMKQHEMVD
755	ZN135_HUMAN	TPGVRVSTDPEQVTFEDVVVGFSQEEWGQLKPAQRTLYRDVMLDTFRLLVSVGHWLPKPNVISLLEQEAELWAVESRLPQ
756	ZN778_HUMAN	EQTQAAGMVAGWLINCYQDAVTFDDVAVDFTQEEWTLLDPSQRDLYRDVMLENYENLASVEWRLKTKGPALRQDRSWFRA
757	RYBP_HUMAN	PSEANSIQSANATTKTSETNHTSRPRLKNVDRSTAQQLAVTVGNVTVIITDFKEKTRSSSTSSSTVTSSAGSEQQNQSSS
758	ZN534_HUMAN	ALTQGQLSFSDVAIEFSQEEWKCLDPGQKALYRDVMLENYRNLVSLGEDNVRPEACICSGICLPDLSVTSMLEQKRDPWT
759	ZN586_HUMAN	AAAAALRAPAQSSVTFEDVAVNFSLEEWSLLNEAQRCLYRDVMLETLTLISSLGCWHGGEDEAAPSKQSTCIHIYKDQGG
760	ZN567_HUMAN	AQGSVSFNDVTVDFTQEEWQHLDHAQKTLYMDVMLENYCHLISVGCHMTKPDVILKLERGEEPWTSFAGHTCLEENWKAE
761	ZN440_HUMAN	DPVAFKDVAVNFTQEEWALLDISQRKLYREVMLETFRNLTSLGKRWKDQNIEYEHQNPRRNFRSLIEEKVNEIKDDSHCG
762	ZN583_HUMAN	SKDLVTFGDVAVNFSQEEWEWLNPAQRNLYRKVMLENYRSLVSLGVSVSKPDVISLLEQGKEPWMVKKEGTRGPCPDWEY
763	ZN441_HUMAN	DSVAFEDVAINFTCEEWALLGPSQKSLYRDVMQETIRNLDCIGMIWQNHDIEEDQYKDLRRNLRCHMVERACEIKDNSQC
764	ZNF43_HUMAN	GPLTFMDVAIEFCLEEWQCLDIAQQNLYRNVMLENYRNLVFLGIAVSKPDLITCLEQEKEPWEPMRRHEMVAKPPVMCSH
765	CBX5_HUMAN	QSNDIARGFERGLEPEKIIGATDSCGDLMFLMKWKDTDEADLVLAKEANVKCPQIVIAFYEERLTWHAYPEDAENKEKET
766	ZN589_HUMAN	ALPAKDSAWPWEEKPRYLGPVTFEDVAVLFTEAEWKRLSLEQRNLYKEVMLENLRNLVSLAESKPEVHTCPSCPLAFGSQ
767	ZNF10_HUMAN	DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPWLVEREIHQ
768	ZN563_HUMAN	DAVAFEDVAVNFTQEEWALLGPSQKNLYRYVMQETIRNLDCIRMIWEEQNTEDQYKNPRRNLRCHMVERFSESKDSSQCG
769	ZN561_HUMAN	EKTKVERMVEDYLASGYQDSVTFDDVAVDFTPEEWALLDTTEKYLYRDVMLENYMNLASVEWEIQPRTKRSSLQQGFLKN
770	ZN136_HUMAN	DSVAFEDVDVNFTQEEWALLDPSQKNLYRDVMWETMRNLASIGKKWKDQNIKDHYKHRGRNLRSHMLERLYQTKDGSQRG
771	ZN630_HUMAN	IESQEPVTFEDVAVDFTQEEWQQLNPAQKTLHRDVMLETYNHLVSVGCSGIKPDVIFKLEHGKDPWIIESELSRWIYPDR
772	ZN527_HUMAN	AVGLCKAMSQGLVTFRDVALDFSQEEWEWLKPSQKDLYRDVMLENYRNLVWLGLSISKPNMISLLEQGKEPWMVERKMSQ

SEQ	**Description**	**Sequence**
773	ZN333_HUMAN	DKVEEEAMAPGLPTACSQEPVTFADVAVVFTPEEWVFLDSTQRSLYRDVMLENYRNLASVADQLCKPNALSYLEERGEQW
774	Z324B_HUMAN	TFEDVAVYFSQEEWGLLDTAQRALYRHVMLENFTLVTSLGLSTSRPRVVIQLERGEEPWVPSGKDMTLARNTYGRLNSGS
775	ZN786_HUMAN	AEPPRLPLTFEDVAIYFSEQEWQDLEAWQKELYKHVMRSNYETLVSLDDGLPKPELISWIEHGGEPFRKWRESQKSGNII
776	ZN709_HUMAN	DSVVFEDVAVNFTQEEWALLGPSQKKLYRDVMQETFVNLASIGENWEEKNIEDHKNQGRKLRSHMVERLCERKEGSQFGE
777	ZN792_HUMAN	AAAALRDPAQGCVTFEDVTIYFSQEEWVLLDEAQRLLYCDVMLENFALIASLGLISFRSHIVSQLEMGKEPWVPDSVDMT
778	ZN599_HUMAN	AAPALALVSFEDVVVTFTGEEWGHLDLAQRTLYQEVMLETCRLLVSLGHPVPKPELIYLLEHGQELWTVKRGLSQSTCAG
779	ZN613_HUMAN	IKSQESLTLEDVAVEFTWEEWQLLGPAQKDLYRDVMLENYSNLVSVGYQASKPDALFKLEQGEPWTVENEIHSQICPEIK
780	ZF69B_HUMAN	GESLESRVTLGSLTAESQELLTFKDVSVDFTQEEWGQLAPAHRNLYREVMLENYGNLVSVGCQLSKPGVISQLEKGEEPW
781	ZN799_HUMAN	ASVALEDVAVNFTREEWALLGPCQKNLYKDVMQETIRNLDCVGMKWKDQNIEDQYRYPRKNLRCRMLERFVESKDGTQCG
782	ZN569_HUMAN	TESQGTVTFKDVAIDFTQEEWKRLDPAQRKLYRNVMLENYNNLITVGYPFTKPDVIFKLEQEEEPWVMEEEVLRRHWQGE
783	ZN564_HUMAN	DSVASEDVAVNFTLEEWALLDPSQKKLYRDVMRETFRNLACVGKKWEDQSIEDWYKNQGRILRNHMEEGLSESKEYDQCG
784	ZN546_HUMAN	EETQGELTSSCGSKTMANVSLAFRDVSIDLSQEEWECLDAVQRDLYKDVMLENYSNLVSLGYTIPKPDVITLLEQEKEPW
785	ZFP92_HUMAN	AAILLTTRPKVPVSFEDVSVYFTKTEWKLLDLRQKVLYKRVMLENYSHLVSLGFSFSKPHLISQLERGEGPWVADIPRTW
786	YAF2_HUMAN	KDKVEKEKSEKETTSKKNSHKKTRPRLKNVDRSSAQHLEVTVGDLTVIITDFKEKTKSPPASSAASADQHSQSGSSSDNT
787	ZN723_HUMAN	GPLTFTDVAIKFSLEEWQFLDTAQQNLYRDVMLENYRNLVFLGVGVSKPDLITCLEQGKEPWNMKRHKMVAKPPVVCSHF
788	ZNF34_HUMAN	RKPNPQAMAALFLSAPPQAEVTFEDVAVYLSREEWGRLGPAQRGLYRDVMLETYGNLVSLGVGPAGPKPGVISQLERGDE
789	ZN439_HUMAN	LSLSPILLYTCEMFQDPVAFKDVAVNFTQEEWALLDISQKNLYREVMLETFWNLTSIGKKWKDQNIEYEYQNPRRNFRSV
790	ZFP57_HUMAN	AAGEPRSLLFFQKPVTFEDVAVNFTQEEWDCLDASQRVLYQDVMSETFKNLTSVARIFLHKPELITKLEQEEEQWRETRV
791	ZNF19_HUMAN	AAMPLKAQYQEMVTFEDVAVHFTKTEWTGLSPAQRALYRSVMLENFGNLTALGYPVPKPALISLLERGDMAWGLEAQDDP
792	ZN404_HUMAN	ARVPLTFSDVAIDFSQEEWEYLNSDQRDLYRDVMLENYTNLVSLDFNFTTESNKLSSEKRNYEVNAYHQETWKRNKTFNL
793	ZN274_HUMAN	ASRLPTAWSCEPVTFEDVTLGFTPEEWGLLDLKQKSLYREVMLENYRNLVSVEHQLSKPDVVSQLEEAEDFWPVERGIPQ
794	CBX3_HUMAN	SKKKRDAADKPRGFARGLDPERIIGATDSSGELMFLMKWKDSDEADLVLAKEANMKCPQIVIAFYEERLTWHSCPEDEAQ
795	ZNF30_HUMAN	AHKYVGLQYHGSVTFEDVAIAFSQQEWESLDSSQRGLYRDVMLENYRNLVSMGHSRSKPHVIALLEQWKEPEVTVRKDGR
796	ZN250_HUMAN	AAARLLPVPAGPQPLSFQAKLTFEDVAVLLSQDEWDRLCPAQRGLYRNVMMETYGNVVSLGLPGSKPDIISQLERGEDPW
797	ZN570_HUMAN	AVGLLKAMYQELVTFRDVAVDFSQEEWDCLDSSQRHLYSNVMLENYRILVSLGLCFSKPSVILLLEQGKAPWMVKRELTK
798	ZN675_HUMAN	GLLTFRDVAIEFSLEEWQCLDTAQRNLYKNVILENYRNLVFLGIAVSKQDLITCLEQEKEPLTVKRHEMVNEPPVMCSHF
799	ZN695_HUMAN	GLLAFRDVALEFSPEEWECLDPAQRSLYRDVMLENYRNLISLGEDSFNMQFLFHSLAMSKPELIICLEARKEPWNVNTEK
800	ZN548_HUMAN	NLTEGRVVFEDVAIYFSQEEWGHLDEAQRLLYRDVMLENLALLSSLGSWHGAEDEEAPSQQGFSVGVSEVTASKPCLSSQ
801	ZN132_HUMAN	GPAQHTSWPCGSAVPTLKSMVTFEDVAVYFSQEEWELLDAAQRHLYHSVMLENLELVTSLGSWHGVEGEGAHPKQNVSVE
802	ZN738_HUMAN	SGYPGAERNLLEYSYFEKGPLTFRDVVIEFSQEEWQCLDTAQQDLYRKVMLENFRNLVFLGIDVSKPDLITCLEQGKDPW
803	ZN420_HUMAN	ARKLVMFRDVAIDFSQEEWECLDSAQRDLYRDVMLENYSNLVSLDLPSRCASKDLSPEKNTYETELSQWEMSDRLENCDL
804	ZN626_HUMAN	GPLQFRDVAIEFSLEEWHCLDTAQRNLYRNVMLENYSNLVFLGITVSKPDLITCLEQGRKPLTMKRNEMIAKPSVMCSHF
805	ZN559_HUMAN	VAGWLTNYSQDSVTFEDVAVDFTQEEWTLLDQTQRNLYRDVMLENYKNLVAVDWESHINTKWSAPQQNFLQGKTSSVVEM

SEQ	Description	Sequence
806	ZN460_HUMAN	AAAWMAPAQESVTFEDVAVTFTQEEWGQLDVTQRALYVEVMLETCGLLVALGDSTKPETVEPIPSHLALPEEVSLQEQLA
807	ZN268_HUMAN	VLEWLFISQEQPKITKSWGPLSFMDVFVDFTWEEWQLLDPAQKCLYRSVMLENYSNLVSLGYQHTKPDIIFKLEQGEELC
808	ZN304_HUMAN	AAAVLMDRVQSCVTFEDVFVYFSREEWELLEEAQRFLYRDVMLENFALVATLGFWCEAEHEAPSEQSVSVEGVSQVRTAE
809	ZIM2_HUMAN	AGSQFPDFKHLGTFLVFEELVTFEDVLVDFSPEELSSLSAAQRNLYREVMLENYRNLVSLGHQFSKPDIISRLEEEESYA
810	ZN605_HUMAN	IQSQISFEDVAVDFTLEEWQLLNPTQKNLYRDVMLENYSNLVFLEVWLDNPKMWLRDNQDNLKSMERGHKYDVFGKIFNS
811	ZN844_HUMAN	DLVAFEDVAVNFTQEEWSLLDPSQKNLYREVMQETLRNLASIGEKWKDQNIEDQYKNPRNNLRSLLGERVDENTEENHCG
812	SUMO5_HUMAN	KDEDIKLRVIGQDSSEIHFKVKMTTPLKKLKKSYCQRQGVPVNSLRFLFEGQRIADNHTPEELGMEEEDVIEVYQEQIGG
813	ZN101_HUMAN	DSVAFEDVAVNFTQEEWALLSPSQKNLYRDVTLETFRNLASVGIQWKDQDIENLYQNLGIKLRSLVERLCGRKEGNEHRE
814	ZN783_HUMAN	RNFWILRLPPGSKGEAPKVPVTFDDVAVYFSELEWGKLEDWQKELYKHVMRGNYETLVSLDYAISKPDILTRIERGEEPC
815	ZN417_HUMAN	AAAAPRRPTQQGTVTFEDVAVNFSQEEWCLLSEAQRCLYRDVMLENLALISSLGCWCGSKDEEAPCKQRISVQRESQSRT
816	ZN182_HUMAN	SGEDSGSFYSWQKAKREQGLVTFEDVAVDFTQEEWQYLNPPQRTLYRDVMLETYSNLVFVGQQVTKPNLILKLEVEECPA
817	ZN823_HUMAN	DSVAFEDVAVNFTQEEWALLGPSQKSLYRNVMQETIRNLDCIEMKWEDQNIGDQCQNAKRNLRSHTCEIKDDSQCGETFG
818	ZN177_HUMAN	AAGWLTTWSQNSVTFQEVAVDFSQEEWALLDPAQKNLYKDVMLENFRNLASVGYQLCRHSLISKVDQEQLKTDERGILQG
819	ZN197_HUMAN	ENPRNQLMALMLLTAQPQELVMFEEVSVCFTSEEWACLGPIQRALYWDVMLENYGNVTSLEWETMTENEEVTSKPSSSQR
820	ZN717_HUMAN	LETYNSLVSLQELVSFEEVAVHFTWEEWQDLDDAQRTLYRDVMLETYSSLVSLGHCITKPEMIFKLEQGAEPWIVEETPN
821	ZN669_HUMAN	RHFRRPEPCREPLASPIQDSVAFEDVAVNFTQEEWALLDSSQKNLYREVMQETCRNLASVGSQWKDQNIEDHFEKPGKDI
822	ZN256_HUMAN	AAAELTAPAQGIVTFEDVAVYFSWKEWGLLDEAQKCLYHDVMLENLTLTTSLGGSGAGDEEAPYQQSTSPQRVSQVRIPK
823	ZN251_HUMAN	AATFQLPGHQEMPLTFQDVAVYFSQAEGRQLGPQQRALYRDVMLENYGNVASLGFPVPKPELISQLEQGKELWVLNLLGA
824	CBX4_HUMAN	RSEAGEPPSSLQVKPETPASAAVAVAAAAAPTTTAEKPPAEAQDEPAESLSEFKPFFGNIIITDVTANCLTVTFKEYVTV
825	PCGF2_HUMAN	HRTTRIKITELNPHLMCALCGGYFIDATTIVECLHSFCKTCIVRYLETNKYCPMCDVQVHKTRPLLSIRSDKTLQDIVYK
826	CDY2_HUMAN	ASQEFEVEAIVDKRQDKNGNTQYLVRWKGYDKQDDTWEPEQHLMNCEKCVHDFNRRQTEKQKKLTWTTTSRIFSNNARRR
827	CDYL2_HUMAN	ASGDLYEVERIVDKRKNKKGKWEYLIRWKGYGSTEDTWEPEHHLLHCEEFIDEFNGLHMSKDKRIKSGKQSSTSKLLRDS
828	HERC2_HUMAN	TLIRKADLENHNKDGGFWTVIDGKVYDIKDFQTQSLTGNSILAQFAGEDPVVALEAALQFEDTRESMHAFCVGQYLEPDQ
829	ZN562_HUMAN	EKTKIGTMVEDHRSNSYQDSVTFDDVAVEFTPEEWALLDTTQKYLYRDVMLENYMNLASVDFFFCLTSEWEIQPRTKRSS
830	ZN461_HUMAN	AHELVMFRDVAIDVSQEEWECLNPAQRNLYKEVMLENYSNLVSLGLSVSKPAVISSLEQGKEPWMVVREETGRWCPGTWK
831	Z324A_HUMAN	AFEDVAVYFSQEEWGLLDTAQRALYRRVMLDNFALVASLGLSTSRPRVVIQLERGEEPWVPSGTDTTLSRTTYRRRNPGS
832	ZN766_HUMAN	AQLRRGHLTFRDVAIEFSQEEWKCLDPVQKALYRDVMLENYRNLVSLGICLPDLSIISMMKQRTEPWTVENEMKVAKNPD
833	ID2_HUMAN	SDHSLGISRSKTPVDDPMSLLYNMNDCYSKLKELVPSIPQNKKVSKMEILQHVIDYILDLQIALDSHPTIVSLHHQRPGQ
834	TOX_HUMAN	KDPNEPQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWDGLGEEQKQVYKKKTEAAKKEYLKQLAAYRASLVSK
835	ZN274_HUMAN	QEEKQEDAAICPVTVLPEEPVTFQDVAVDFSREEWGLLGPTQRTEYRDVMLETFGHLVSVGWETTLENKELAPNSDIPEE
836	SCMH1_HUMAN	DASRLSGRDPSSWTVEDVMQFVREADPQLGPHADLFRKHEIDGKALLLLRSDMMMKYMGLKLGPALKLSYHIDRLKQGKF
837	ZN214_HUMAN	AVTFEDVTIIFTWEEWKFLDSSQKRLYREVMWENYTNVMSVENWNESYKSQEEKFRYLEYENFSYWQGWWNAGAQMYENQ
838	CBX7_HUMAN	ELSAIGEQVFAVESIRKKRVRKGKVEYLVKWKGWPPKYSTWEPEEHILDPRLVMAYEEKEERDRASGYRKRGPKPKRLLL

SEQ	Description	Sequence
839	ID1_HUMAN	GGAGARLPALLDEQQVNVLLYDMNGCYSRLKELVPTLPQNRKVSKVEILQHVIDYIRDLQLELNSESEVGTPGGRGLPVR
840	CREM_HUMAN	VVMAASPGSLHSPQQLAEEATRKRELRLMKNREAAKECRRRKKEYVKCLESRVAVLEVQNKKLIEELETLKDICSPKTDY
841	SCX_HUMAN	GGGPGGRPGREPRQRHTANARERDRTNSVNTAFTALRTLIPTEPADRKLSKIETLRLASSYISHLGNVLLAGEACGDGQP
842	ASCL1_HUMAN	SGFGYSLPQQQPAAVARRNERERNRVKLVNLGFATLREHVPNGAANKKMSKVETLRSAVEYIRALQQLLDEHDAVSAAFQ
843	ZN764_HUMAN	APLPPRDPNGAGPEWREPGAVSFADVAVYFCREEWGCLRPAQRALYRDVMRETYGHLSALGIGGNKPALISWVEEEAELW
844	SCML2_HUMAN	KQGFSKDPSTWSVDEVIQFMKHTDPQISGPLADLFRQHEIDGKALFLLKSDVMMKYMGLKLGPALKLCYYIEKLKEGKYS
845	TWST1_HUMAN	SGGGSPQSYEELQTQRVMANVRERQRTQSLNEAFAALRKIIPTLPSDKLSKIQTLKLAARYIDFLYQVLQSDELDSKMAS
846	CREB1_HUMAN	IAPGVVMASSPALPTQPAEEAARKREVRLMKNREAARECRRKKKEYVKCLENRVAVLENQNKTLIEELKALKDLYCHKSD
847	TERF1_HUMAN	SRIPVSKSQPVTPEKHRARKRQAWLWEEDKNLRSGVRKYGEGNWSKILLHYKFNNRTSVMLKDRWRTMKKLKLISSDSED
848	ID3_HUMAN	SLAIARGRGKGPAAEEPLSLLDDMNHCYSRLRELVPGVPRGTQLSQVEILQRVIDYILDLQVVLAEPAPGPPDGPHLPIQ
849	CBX8_HUMAN	GSGPPSSGGGLYRDMGAQGGRPSLIARIPVARILGDPEEESWSPSLTNLEKVVVTDVTSNFLTVTIKESNTDQGFFKEKR
850	CBX4_HUMAN	ELPAVGEHVFAVESIEKKRIRKGRVEYLVKWRGWSPKYNTWEPEENILDPRLLIAFQNRERQEQLMGYRKRGPKPKPLVV
851	GSX1_HUMAN	VDSSSNQLPSSKRMRTAFTSTQLLELEREFASNMYLSRLRRIEIATYLNLSEKQVKIWFQNRRVKHKKEGKGSNHRGGGG
852	NKX22_HUMAN	TPGGGGDAGKKRKRRVLFSKAQTYELERRFRQQRYLSAPEREHLASLIRLTPTQVKIWFQNHRYKMKRARAEKGMEVTPL
853	ATF1_HUMAN	QTVVMTSPVTLTSQTTKTDDPQLKREIRLMKNREAARECRRKKKEYVKCLENRVAVLENQNKTLIEELKTLKDLYSNKSV
854	TWST2_HUMAN	KGSPSAQSFEELQSQRILANVRERQRTQSLNEAFAALRKIIPTLPSDKLSKIQTLKLAARYIDFLYQVLQSDEMDNKMTS
855	ZNF17_HUMAN	NLTEDYMVFEDVAIHFSQEEWGILNDVQRHLHSDVMLENFALLSSVGCWHGAKDEEAPSKQCVSVGVSQVTTLKPALSTQ
856	TOX3_HUMAN	KDPNEPQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWDSLGEEQKQVYKRKTEAAKKEYLKALAAYRASLVSK
857	TOX4_HUMAN	KDPNEPQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWDSLGEEQKQVYKRKTEAAKKEYLKALAAYKDNQECQ
858	ZMYM3_HUMAN	LDGSTWDFCSEDCKSKYLLWYCKAARCHACKRQGKLLETIHWRGQIRHFCNQQCLLRFYSQQNQPNLDTQSGPESLLNSQ
859	I2BP1_HUMAN	ASVQASRRQWCYLCDLPKMPWAMVWDFSEAVCRGCVNFEGADRIELLIDAARQLKRSHVLPEGRSPGPPALKHPATKDLA
860	RHXF1_HUMAN	MEGPQPENMQPRTRRTKFTLLQVEELESVFRHTQYPDVPTRRELAENLGVTEDKVRVWFKNKRARCRRHQRELMLANELR
861	SSX2_HUMAN	PKIMPKKPAEEGNDSEEVPEASGPQNDGKELCPPGKPTTSEKIHERSGPKRGEHAWTHRLRERKQLVIYEEISDPEEDDE
862	I2BPL_HUMAN	SAAQVSSSRRQSCYLCDLPRMPWAMIWDFSEPVCRGCVNYEGADRIEFVIETARQLKRAHGCFQDGRSPGPPPPVGVKTV
863	ZN680_HUMAN	PGPPGSLEMGPLTFRDVAIEFSLEEWQCLDTAQRNLYRKVMFENYRNLVFLGIAVSKPHLITCLEQGKEPWNRKRQEMVA
864	CBX1_HUMAN	NKKKVEEVLEEEEEEYVVEKVLDRRVVKGKVEYLLKWKGFSDEDNTWEPEENLDCPDLIAEFLQSQKTAHETDKSEGGKR
865	TRI68_HUMAN	LANVVEKVRLLRLHPGMGLKGDLCERHGEKLKMFCKEDVLIMCEACSQSPEHEAHSVVPMEDVAWEYKWELHEALEHLKK
866	HXA13_HUMAN	VVSHPSDASSYRRGRKKRVPYTKVQLKELEREYATNKFITKDKRRRISATTNLSERQVTIWFQNRRVKEKKVINKLKTTS
867	PHC3_HUMAN	ENSDLLPVAQTEPSIWTVDDVWAFIHSLPGCQDIADEFRAQEIDGQALLLLKEDHLMSAMNIKLGPALKICARINSLKES
868	TCF24_HUMAN	AGPGGGSRSGSGRPAAANAARERSRVQTLRHAFLELQRTLPSVPPDTKLSKLDVLLLATTYIAHLTRSLQDDAEAPADAG
869	CBX3_HUMAN	QNGKSKKVEEAEPEEFVVEKVLDRRVVNGKVEYFLKWKGFTDADNTWEPEENLDCPELIEAFLNSQKAGKEKDGTKRKSL
870	HXB13_HUMAN	QHPPDACAFRRGRKKRIPYSKGQLRELEREYAANKFITKDKRRKISAATSLSERQITIWFQNRRVKEKKVLAKVKNSATP
871	HEY1_HUMAN	SMSPTTSSQILARKRRRGIIEKRRRDRINNSLSELRRLVPSAFEKQGSAKLEKAEILQMTVDHLKMLHTAGGKGYFDAHA

SEQ	Description	Sequence
872	PHC2_HUMAN	LVGMGHHFLPSEPTKWNVEDVYEFIRSLPGCQEIAEEFRAQEIDGQALLLLKEDHLMSAMNIKLGPALKIYARISMLKDS
873	ZNF81_HUMAN	PANEDAPQPGEHGSACEVSVSFEDVTVDFSREEWQQLDSTQRRLYQDVMLENYSHLLSVGFEVPKPEVIFKLEQGEGPWT
874	FIGLA_HUMAN	GYSSTENLQLVLERRRVANAKERERIKNLNRGFARLKALVPFLPQSRKPSKVDILKGATEYIQVLSDLLEGAKDSKKQDP
875	SAM11_HUMAN	EEAPAPEDVTKWTVDDVCSFVGGLSGCGEYTRVFREQGIDGETLPLLTEEHLLTNMGLKLGPALKIRAQVARRLGRVFYV
876	KMT2B_HUMAN	GGTLAHTPRRSLPSHHGKKMRMARCGHCRGCLRVQDCGSCVNCLDKPKFGGPNTKKQCCVYRKCDKIEARKMERLAKKGR
877	HEY2_HUMAN	LNSPTTTSQIMARKKRRGIIEKRRRDRINNSLSELRRLVPTAFEKQGSAKLEKAEILQMTVDHLKMLQATGGKGYFDAHA
878	JDP2_HUMAN	QPVKSELDEEEERRKRRREKNKVAAARCRNKKKERTEFLQRESERLELMNAELKTQIEELKQERQQLILMLNRHRPTCIV
879	HXC13_HUMAN	LQPEVSSYRRGRKKRVPYTKVQLKELEKEYAASKFITKEKRRRISATTNLSERQVTIWFQNRRVKEKKVVSKSKAPHLHS
880	ASCL4_HUMAN	LPVPLDSAFEPAFLRKRNERERQRVRCVNEGYARLRDHLPRELADKRLSKVETLRAAIDYIKHLQELLERQAWGLEGAAG
881	HHEX_HUMAN	SPFLQRPLHKRKGGQVRFSNDQTIELEKKFETQKYLSPPERKRLAKMLQLSERQVKTWFQNRRAKWRRLKQENPQSNKKE
882	HERC2_HUMAN	IAIATGSLHCVCCTEDGEVYTWGDNDEGQLGDGTTNAIQRPRLVAALQGKKVNRVACGSAHTLAWSTSKPASAGKLPAQV
883	GSX2_HUMAN	GGSDASQVPNGKRMRTAFTSTQLLELEREFSSNMYLSRLRRIEIATYLNLSEKQVKIWFQNRRVKHKKEGKGTQRNSHAG
884	BIN1_HUMAN	RLDLPPGFMFKVQAQHDYTATDTDELQLKAGDVVLVIPFQNPEEQDEGWLMGVKESDWNQHKELEKCRGVFPENFTERVP
885	ETV7_HUMAN	GICKLPGRLRIQPALWSREDVLHWLRWAEQEYSLPCTAEHGFEMNGRALCILTKDDFRHRAPSSGDVLYELLQYIKTQRR
886	ASCL3_HUMAN	PNYRGCEYSYGPAFTRKRNERERQRVKCVNEGYAQLRHHLPEEYLEKRLSKVETLRAAIKYINYLQSLLYPDKAETKNNP
887	PHC1_HUMAN	LHGINPVFLSSNPSRWSVEEVYEFIASLQGCQEIAEEFRSQEIDGQALLLLKEEHLMSAMNIKLGPALKICAKINVLKET
888	OTP_HUMAN	QAGQQQGQQKQKRHRTRFTPAQLNELERSFAKTHYPDIFMREELALRIGLTESRVQVWFQNRRAKWKKRKKTTNVFRAPG
889	I2BP2_HUMAN	AAAVAVAAASRRQSCYLCDLPRMPWAMIWDFTEPVCRGCVNYEGADRVEFVIETARQLKRAHGCFPEGRSPPGAAASAAA
890	VGLL2_HUMAN	FSSQTPASIKEEEGSPEKERPPEAEYINSRCVLFTYFQGDISSVVDEHFSRALSQPSSYSPSCTSSKAPRSSGPWRDCSF
891	HXA11_HUMAN	DKAGGSSGQRTRKKRCPYTKYQIRELEREFFFSVYINKEKRLQLSRMLNLTDRQVKIWFQNRRMKEKKINRDRLQYYSAN
892	PDLI4_HUMAN	GAPLSGLQGLPECTRCGHGIVGTIVKARDKLYHPECFMCSDCGLNLKQRGYFFLDERLYCESHAKARVKPPEGYDVVAVY
893	ASCL2_HUMAN	RRPATAETGGGAAAVARRNERERNRVKLVNLGFQALRQHVPHGGASKKLSKVETLRSAVEYIRALQRLLAEHDAVRNALA
894	CDX4_HUMAN	TVQVTGKTRTKEKYRVVYTDHQRLELEKEFHCNRYITIQRKSELAVNLGLSERQVKIWFQNRRAKERKMIKKKISQFENS
895	ZN860_HUMAN	EEAAQKRKEKEPGMALPQGHLTFRDVAIEFSLEEWKCLDPTQRALYRAMMLENYRNLHSVDISSKCMMKKFSSTAQGNTE
896	LMBL4_HUMAN	DIRASQVARWTVDEVAEFVQSLLGCEEHAKCFKKEQIDGKAFLLLTQTDIVKVMKIKLGPALKIYNSILMFRHSQELPEE
897	PDIP3_HUMAN	LSPLEGTKMTVNNLHPRVTEEDIVELFCVCGALKRARLVHPGVAEVVFVKKDDAITAYKKYNNRCLDGQPMKCNLHMNGN
898	NKX25_HUMAN	DNAERPRARRRRKPRVLFSQAQVYELERRFKQQRYLSAPERDQLASVLKLTSTQVKIWFQNRRYKCKRQRQDQTLELVGL
899	CEBPB_HUMAN	SQVKSKAKKTVDKHSDEYKIRRERNNIAVRKSRDKAKMRNLETQHKVLELTAENERLQKKVEQLSRELSTLRNLFKQLPE
900	ISL1_HUMAN	KRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEFALREDGLFCRADHDVVERASLGAGDPL
901	CDX2_HUMAN	SLGSQVKTRTKDKYRVVYTDHQRLELEKEFHYSRYITIRRKAELAATLGLSERQVKIWFQNRRAKERKINKKKLQQQQQ
902	PROP1_HUMAN	QGGQRGRPHSRRRHRTTFSPVQLEQLESAFGRNQYPDIWARESLARDTGLSEARIQVWFQNRRAKQRKQERSLLQPLAHL
903	SIN3B_HUMAN	DALTYLDQVKIRFGSDPATYNGFLEIMKEFKSQSIDTPGVIRRVSQLFHEHPDLIVGFNAFLPLGYRIDIPKNGKLNIQS
904	SMBT1_HUMAN	RLHLDSNPLKWSVADVVRFIRSTDCAPLARIFLDQEIDGQALLLLTLPTVQECMDLKLGPAIKLCHHIERIKFAFYEQFA

SEQ	Description	Sequence
905	HXC11_HUMAN	AKGAAPNAPRTRKKRCPYSKFQIRELEREFFFNVYINKEKRLQLSRMLNLTDRQVKIWFQNRRMKEKKLSRDRLQYFSGN
906	HXC10_HUMAN	TTGNWLTAKSGRKKRCPYTKHQTLELEKEFLFNMYLTRERRLEISKTINLTDRQVKIWFQNRRMKLKKMNRENRIRELTS
907	PRS6A_HUMAN	YLVSNVIELLDVDPNDQEEDGANIDLDSQRKGKCAVIKTSTRQTYFLPVIGLVDAEKLKPGDLVGVNKDSYLILETLPTE
908	VSX1_HUMAN	KASPTLGKRKKRRHRTVFTAHQLEELEKAFSEAHYPDVYAREMLAVKTELPEDRIQVWFQNRRAKWRKREKRWGGSSVMA
909	NKX23_HUMAN	EESERPKPRSRRKPRVLFSQAQVFELERRFKQQRYLSAPEREHLASSLKLTSTQVKIWFQNRRYKCKRQRQDKSLELGAH
910	MTG16_HUMAN	VVPGSRQEEVIDHKLTEREWAEEWKHLNNLLNCIMDMVEKTRRSLTVLRRCQEADREELNHWARRYSDAEDTKKGPAPAA
911	HMX3_HUMAN	ESPEKKPACRKKKTRTVFSRSQVFQLESTFDMKRYLSSSERAGLAASLHLTETQVKIWFQNRRNKWKRQLAAELEAANLS
912	HMX1_HUMAN	RGGVGVGGGRKKKTRTVFSRSQVFQLESTFDLKRYLSSAERAGLAASLQLTETQVKIWFQNRRNKWKRQLAAELEAASLS
913	KIF22_HUMAN	ELLAHGRQKILDLLNEGSARDLRSLQRIGPKKAQLIVGWRELHGPFSQVEDLERVEGITGKQMESFLKANILGLAAGQRC
914	CSTF2_HUMAN	ESPYGETISPEDAPESISKAVASLPPEQMFELMKQMKLCVQNSPQEARNMLLQNPQLAYALLQAQVVMRIVDPEIALKIL
915	CEBPE_HUMAN	AGPLHKGKKAVNKDSLEYRLRRERNNIAVRKSRDKAKRRILETQQKVLEYMAENERLRSRVEQLTQELDTLRNLFRQIPE
916	DLX2_HUMAN	IRIVNGKPKKVRKPRTIYSSFQLAALQRRFQKTQYLALPERAELAASLGLTQTQVKIWFQNRRSKFKKMWKSGEIPSEQH
917	ZMYM3_HUMAN	TVYQFCSPSCWTKFQRTSPEGGIHLSCHYCHSLFSGKPEVLDWQDQVFQFCCRDCCEDFKRLRGVVSQCEHCRQEKLLHE
918	PPARG_HUMAN	TMVDTEMPFWPTNFGISSVDLSVMEDHSHSFDIKPFTTVDFSSISTPHYEDIPFTRTDPVVADYKYDLKLQEYQSAIKVE
919	PRIC1_HUMAN	GRHHAELLKPRCSACDEIIFADECTEAEGRHWHMKHFCCLECETVLGGQRYIMKDGRPFCCGCFESLYAEYCETCGEHIG
920	UNC4_HUMAN	DPDKESPGCKRRRTRTNFTGWQLEELEKAFNESHYPDVFMREALALRLDLVESRVQVWFQNRRAKWRKKENTKKGPGRPA
921	BARX2_HUMAN	TEQPTPRQKKPRRSRTIFTELQLMGLEKKFQKQKYLSTPDRLDLAQSLGLTQLQVKTWYQNRRMKWKKMVLKGGQEAPTK
922	ALX3_HUMAN	SMELAKNKSKKRRNRTTFSTFQLEELEKVFQKTHYPDVYAREQLALRTDLTEARVQVWFQNRRAKWRKRERYGKIQEGRN
923	TCF15_HUMAN	GGGGGAGPVVVVRQRQAANARERDRTQSVNTAFTALRTLIPTEPVDRKLSKIETVRLASSYIAHLANVLLLGDSADDGQP
924	TERA_HUMAN	IDDTVEGITGNLFEVYLKPYFLEAYRPIRKGDIFLVRGGMRAVEFKVVETDPSPYCIVAPDTVIHCEGEPIKREDEEESL
925	VSX2_HUMAN	SALNQTKKRKKRRHRTIFTSYQLEELEKAFNEAHYPDVYAREMLAMKTELPEDRIQVWFQNRRAKWRKREKCWGRSSVMA
926	HXD12_HUMAN	DGLPWGAAPGRARKKRKPYTKQQIAELENEFLVNEFINRQKRKELSNRLNLSDQQVKIWFQNRRMKKKRVVLREQALALY
927	CDX1_HUMAN	GGGGSGKTRTKDKYRVVYTDHQRLELEKEFHYSRYITIRRKSELAANLGLTERQVKIWFQNRRAKERKVNKKKQQQQQPP
928	TCF23_HUMAN	TRAGGLALGRSEASPENAARERSRVRTLRQAFLALQAALPAVPPDTKLSKLDVLVLAASYIAHLTRTLGHELPGPAWPPF
929	ALX1_HUMAN	KCDSNVSSSKKRRHRTTFTSLQLEELEKVFQKTHYPDVYVREQLALRTELTEARVQVWFQNRRAKWRKRERYGQIQQAKS
930	HXA10_HUMAN	NAANWLTAKSGRKKRCPYTKHQTLELEKEFLFNMYLTRERRLEISRSVHLTDRQVKIWFQNRRMKLKKMNRENRIRELTA
931	RX_HUMAN	LSEEEQPKKKHRRNRTTFTTYQLHELERAFEKSHYPDVYSREELAGKVNLPEVRVQVWFQNRRAKWRRQEKLEVSSMKLQ
932	CXXC5_HUMAN	HMAGLAEYPMQGELASAISSGKKKRKRCGMCAPCRRRINCEQCSSCRNRKTGHQICKFRKCEELKKKPSAALEKVMLPTG
933	SCML1_HUMAN	SITKHPSTWSVEAVVLFLKQTDPLALCPLVDLFRSHEIDGKALLLLTSDVLLKHLGVKLGTAVKLCYYIDRLKQGKCFEN
934	NFIL3_HUMAN	ACRRKREFIPDEKKDAMYWEKRRKNNEAAKRSREKRRLNDLVLENKLIALGEENATLKAELLSLKLKFGLISSTAYAQEI
935	DLX6_HUMAN	EIRFNGKGKKIRKPRTIYSSLQLQALNHRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLLKQGSNPHESD
936	MTG8_HUMAN	GLHGTRQEEMIDHRLTDREWAEEWKHLDHLLNCIMDMVEKTRRSLTVLRRCQEADREELNYWIRRYSDAEDLKKGGGSSS
937	CBX8_HUMAN	ELSAVGERVFAAEALLKRRIRKGRMEYLVKWKGWSQKYSTWEPEENILDARLLAAFEEREREMELYGPKKRGPKPKTFLL

SEQ	Description	Sequence
938	CEBPD_HUMAN	AREKSAGKRGPDRGSPEYRQRRERNNIAVRKSRDKAKRRNQEMQQKLVELSAENEKLHQRVEQLTRDLAGLRQFFKQLPS
939	SEC13_HUMAN	SGGCDNLIKLWKEEEDGQWKEEQKLEAHSDWVRDVAWAPSIGLPTSTIASCSQDGRVFIWTCDDASSNTWSPKLLHKFND
940	FIP1_HUMAN	VKGVDLDAPGSINGVPLLEVDLDSFEDKPWRKPGADLSDYFNYGFNEDTWKAYCEKQKRIRMGLEVIPVTSTTNKITAED
941	ALX4_HUMAN	KADSESNKGKKRRNRTTFTSYQLEELEKVFQKTHYPDVYAREQLAMRTDLTEARVQVWFQNRRAKWRKRERFGQMQQVRT
942	LHX3_HUMAN	TAKQREAEATAKRPRTTITAKQLETLKSAYNTSPKPARHVREQLSSETGLDMRVVQVWFQNRRAKEKRLKKDAGRQRWGQ
943	PRIC2_HUMAN	GRHHAECLKPRCAACDEIIFADECTEAEGRHWHMKHFCCFECETVLGGQRYIMKEGRPYCCHCFESLYAEYCDTCAQHIG
944	MAGI3_HUMAN	IIGGDRPDEFLQVKNVLKDGPAAQDGKIAPGDVIVDINGNCVLGHTHADVVQMFQLVPVNQYVNLTLCRGYPLPDDSEDP
945	NELL1_HUMAN	CCPECDTRVTSQCLDQNGHKLYRSGDNWTHSCQQCRCLEGEVDCWPLTCPNLSCEYTAILEGECCPRCVSDPCLADNITY
946	PRRX1_HUMAN	LNSEEKKKRKQRRNRTTFNSSQLQALERVFERTHYPDAFVREDLARRVNLTEARVQVWFQNRRAKFRRNERAMLANKNAS
947	MTG8R_HUMAN	GLNGGYQDELVDHRLTEREWADEWKHLDHALNCIMEMVEKTRRSMAVLRRCQESDREELNYWKRRYNENTELRKTGTELV
948	RAX2_HUMAN	GPGEEAPKKKHRRNRTTFTTYQLHQLERAFEASHYPDVYSREELAAKVHLPEVRVQVWFQNRRAKWRRQERLESGSGAVA
949	DLX3_HUMAN	VRMVNGKPKKVRKPRTIYSSYQLAALQRRFQKAQYLALPERAELAAQLGLTQTQVKIWFQNRRSKFKKLYKNGEVPLEHS
950	DLX1_HUMAN	EVRFNGKGKKIRKPRTIYSSLQLQALNRRFQQTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKLMKQGGAALEGS
951	NKX26_HUMAN	GRSEQPKARQRRKPRVLFSQAQVLALERRFKQQRYLSAPEREHLASALQLTSTQVKIWFQNRRYKCKRQRQDKSLELAGH
952	NAB1_HUMAN	LPRTLGELQLYRILQKANLLSYFDAFIQQGGDDVQQLCEAGEEEFLEIMALVGMASKPLHVRRLQKALRDWVTNPGLFNQ
953	SAMD7_HUMAN	NLSLDEDIQKWTVDDVHSFIRSLPGCSDYAQVFKDHAIDGETLPLLTEEHLRGTMGLKLGPALKIQSQVSQHVGSMFYKK
954	PITX3_HUMAN	SPEDGSLKKKQRRQRTHFTSQQLQELEATFQRNRYPDMSTREEIAVWTNLTEARVRVWFKNRRAKWRKRERSQQAELCKG
955	WDR5_HUMAN	SNLLVSASDDKTLKIWDVSSGKCLKTLKGHSNYVFCCNFNPQSNLIVSGSFDESVRIWDVKTGKCLKTLPAHSDPVSAVH
956	MEOX2_HUMAN	GNYKSEVNSKPRKERTAFTKEQIRELEAEFAHHNYLTRLRRYEIAVNLDLTERQVKVWFQNRRMKWKRVKGGQQGAAARE
957	NAB2_HUMAN	LPRTLGELQLYRVLQRANLLSYYETFIQQGGDDVQQLCEAGEEEFLEIMALVGMATKPLHVRRLQKALREWATNPGLFSQ
958	DHX8_HUMAN	PEEPTIGDIYNGKVTSIMQFGCFVQLEGLRKRWEGLVHISELRREGRVANVADVVSKGQRVKVKVLSFTGTKTSLSMKDV
959	FOXA2_HUMAN	YAFNHPFSINNLMSSEQQHHHSHHHHQPHKMDLKAYEQVMHYPGYGSPMPGSLAMGPVTNKTGLDASPLAADTSYYQGVY
960	CBX6_HUMAN	TAAAGPAPPTAPEPAGASSEPEAGDWRPEMSPCSNVVVTDVTSNLLTVTIKEFCNPEDFEKVAAGVAGAAGGGGSIGASK
961	EMX2_HUMAN	FLLHNALARKPKRIRTAFSPSQLLRLEHAFEKNHYVVGAERKQLAHSLSLTETQVKVWFQNRRTKFKRQKLEEEGSDSQQ
962	CPSF6_HUMAN	KRIALYIGNLTWWTTDEDLTEAVHSLGVNDILEIKFFENRANGQSKGFALVGVGSEASSKKLMDLLPKRELHGQNPVVTP
963	HXC12_HUMAN	SGAPWYPINSRSRKKRKPYSKLQLAELEGEFLVNEFITRQRRRELSDRLNLSDQQVKIWFQNRRMKKKRLLLREQALSFF
964	KDM4B_HUMAN	SDNLYPESITSRDCVQLGPPSEGELVELRWTDGNLYKAKFISSVTSHIYQVEFEDGSQLTVKRGDIFTLEEELPKRVRSR
965	LMBL3_HUMAN	GIPASKVSKWSTDEVSEFIQSLPGCEEHGKVFKDEQIDGEAFLLMTQTDIVKIMSIKLGPALKIFNSILMFKAAEKNSHN
966	PHX2A_HUMAN	EPSGLHEKRKQRRIRTTFTSAQLKELERVFAETHYPDIYTREELALKIDLTEARVQVWFQNRRAKFRKQERAASAKGAAG
967	EMX1_HUMAN	LLLHGPFARKPKRIRTAFSPSQLLRLERAFEKNHYVVGAERKQLAGSLSLSETQVKVWFQNRRTKYKRQKLEEEGPESEQ
968	NC2B_HUMAN	SSGNDDDLTIPRAAINKMIKETLPNVRVANDARELVVNCCTEFIHLISSEANEICNKSEKKTISPEHVIQALESLGFGSY
969	DLX4_HUMAN	ERRPQAPAKKLRKPRTIYSSLQLQHLNQRFQHTQYLALPERAQLAAQLGLTQTQVKIWFQNKRSKYKKLLKQNSGGQEGD
970	SRY_HUMAN	NVQDRVKRPMNAFIVWSRDQRRKMALENPRMRNSEISKQLGYQWKMLTEAEKWPFFQEAQKLQAMHREKYPNYKYRPRRK

SEQ	Description	Sequence
971	ZN777_HUMAN	EITRLAVWAAVQAVERKLEAQAMRLLTLEGRTGTNEKKIADCEKTAVEFANHLESKWVVLGTLLQEYGLLQRRLENMENL
972	NELL1_HUMAN	CEKDIDECSEGIIECHNHSRCVNLPGWYHCECRSGFHDDGTYSLSGESCIDIDECALRTHTCWNDSACINLAGGFDCLCP
973	ZN398_HUMAN	AAISLWTVVAAVQAIERKVEIHSRRLLHLEGRTGTAEKKLASCEKTVTELGNQLEGKWAVLGTLLQEYGLLQRRLENLEN
974	GATA3_HUMAN	GQNRPLIKPKRRLSAARRAGTSCANCQTTTTTLWRRNANGDPVCNACGLYYKLHNINRPLTMKKEGIQTRNRKMSSKSKK
975	BSH_HUMAN	HAELPGKHCRRRKARTVFSDSQLSGLEKRFEIQRYLSTPERVELATALSLSETQVKTWFQNRRMKHKKQLRKSQDEPKAP
976	SF3B4_HUMAN	QDATVYVGGLDEKVSEPLLWELFLQAGPVVNTHMPKDRVTGQHQGYGFVEFLSEEDADYAIKIMNMIKLYGKPIRVNKAS
977	TEAD1_HUMAN	PIDNDAEGVWSPDIEQSFQEALAIYPPCGRRKIILSDEGKMYGRNELIARYIKLRTGKTRTRKQVSSHIQVLARRKSRDF
978	TEAD3_HUMAN	GLDNDAEGVWSPDIEQSFQEALAIYPPCGRRKIILSDEGKMYGRNELIARYIKLRTGKTRTRKQVSSHIQVLARKKVREY
979	RGAP1_HUMAN	DSVGTPQSNGGMRLHDFVSKTVIKPESCVPCGKRIKFGKLSLKCRDCRVVSHPECRDRCPLPCIPTLIGTPVKIGEGMLA
980	PHF1_HUMAN	SAPHSMTASSSSVSSPSPGLPRRSAPPSPLCRSLSPGTGGGVRGGVGYLSRGDPVRVLARRVRPDGSVQYLVEWGGGGIF
981	FOXA1_HUMAN	GDPHYSFNHPFSINNLMSSSEQQHKLDFKAYEQALQYSPYGSTLPASLPLGSASVTTRSPIEPSALEPAYYQGVYSRPVL
982	GATA2_HUMAN	GQNRPLIKPKRRLSAARRAGTCCANCQTTTTTLWRRNANGDPVCNACGLYYKLHNVNRPLTMKKEGIQTRNRKMSNKSKK
983	FOXO3_HUMAN	DSLSGSSLYSTSANLPVMGHEKFPSDLDLDMFNGSLECDMESIIRSELMDADGLDFNFDSLISTQNVVGLNVGNFTGAKQ
984	ZN212_HUMAN	TEISLWTVVAAIQAVEKKMESQAARLQSLEGRTGTAEKKLADCEKMAVEFGNQLEGKWAVLGTLLQEYGLLQRRLENVEN
985	IRX4_HUMAN	MDSGTRRKNATRETTSTLKAWLQEHRKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKENKMTWPPRNKCADEKR
986	ZBED6_HUMAN	NIEKQIYLPSTRAKTSIVWHFFHVDPQYTWRAICNLCEKSVSRGKPGSHLGTSTLQRHLQARHSPHWTRANKFGVASGEE
987	LHX4_HUMAN	AKQNDDSEAGAKRPRTTITAKQLETLKNAYKNSPKPARHVREQLSSETGLDMRVVQVWFQNRRAKEKRLKKDAGRHRWGQ
988	SIN3A_HUMAN	DALSYLDQVKLQFGSQPQVYNDFLDIMKEFKSQSIDTPGVISRVSQLFKGHPDLIMGFNTFLPPGYKIEVQTNDMVNVTT
989	RBBP7_HUMAN	DDHTVCLWDINAGPKEGKIVDAKAIFTGHSAVVEDVAWHLLHESLFGSVADDQKLMIWDTRSNTTSKPSHLVDAHTAEVN
990	NKX61_HUMAN	GSILLDKDGKRKHTRPTFSGQQIFALEKTFEQTKYLAGPERARLAYSLGMTESQVKVWFQNRRTKWRKKHAAEMATAKKK
991	TRI68_HUMAN	DPTALVEAIVEEVACPICMTFLREPMSIDCGHSFCHSCLSGLWEIPGESQNWGYTCPLCRAPVQPRNLRPNWQLANVVEK
992	R51A1_HUMAN	QSLPKKVSLSSDTTRKPLEIRSPSAESKKPKWVPPAASGGSRSSSSPLVVVSVKSPNQSLRLGLSRLARVKPLHPNATST
993	MB3L1_HUMAN	AKSSQRKQRDCVNQCKSKPGLSTSIPLRMSSYTFKRPVTRITPHPGNEVRYHQWEESLEKPQQVCWQRRLQGLQAYSSAG
994	DLX5_HUMAN	VRMVNGKPKKVRKPRTIYSSFQLAALQRRFQKTQYLALPERAELAASLGLTQTQVKIWFQNKRSKIKKIMKNGEMPPEHS
995	NOTC1_HUMAN	LQCNNHACGWDGGDCSLNFNDPWKNCTQSLQCWKYFSDGHCDSQCNSAGCLFDGFDCQRAEGQCNPLYDQYCKDHFSDGH
996	TERF2_HUMAN	ETWVEEDELFQVQAAPDEDSTTNITKKQKWTVEESEWVKAGVQKYGEGNWAAISKNYPFVNRTAVMIKDRWRTMKRLGMN
997	ZN282_HUMAN	AEISLWTVVAAIQAVERKVDAQASQLLNLEGRTGTAEKKLADCEKTAVEFGNHMESKWAVLGTLLQEYGLLQRRLENLEN
998	RGS12_HUMAN	LEKRTLFRLDLVPINRSVGLKAKPTKPVTEVLRPVVARYGLDLSGLLVRLSGEKEPLDLGAPISSLDGQRVVLEEKDPSR
999	ZN840_HUMAN	PNCLSSSMQLPHGGGRHQELVRFRDVAVVFSPEEWDHLTPEQRNLYKDVMLDNCKYLASLGNWTYKAHVMSSLKQGKEPW
1000	SPI2B_HUMAN	DDYKEGDLRIMPESSESPPTEREPGGVVDGLIGKHVEYTKEDGSKRIGMVIHQVEAKPSVYFIKFDDDFHIYVYDLVKKS
1001	PAX7_HUMAN	SEPDLPLKRKQRRSRTTFTAEQLEELEKAFERTHYPDIYTREELAQRTKLTEARVQVWFSNRRARWRKQAGANQLAAFNH
1002	NKX62_HUMAN	AGGVLDKDGKKKHSRPTFSGQQIFALEKTFEQTKYLAGPERARLAYSLGMTESQVKVWFQNRRTKWRKRHAVEMASAKKK
1003	ASXL2_HUMAN	DVMSFSVTVTTIPASQAMNPSSHGQTIPVQAFSEENSIEGTPSKCYCRLKAMIMCKGCGAFCHDDCIGPSKLCVSCLVVR

SEQ	Description	Sequence
1004	FOXO1_HUMAN	GGYSSVSSCNGYGRMGLLHQEKLPSDLDGMFIERLDCDMESIIRNDLMDGDTLDFNFDNVLPNQSFPHSVKTTTHSWVSG
1005	GATA3_HUMAN	GGSPTGFGCKSRPKARSSTGRECVNCGATSTPLWRRDGTGHYLCNACGLYHKMNGQNRPLIKPKRRLSAARRAGTSCANC
1006	GATA1_HUMAN	GQNRPLIRPKKRLIVSKRAGTQCTNCQTTTTTLWRRNASGDPVCNACGLYYKLHQVNRPLTMRKDGIQTRNRKASGKGKK
1007	ZMYM5_HUMAN	PVALLRKQNFQPTAQQQLTKPAKITCANCKKPLQKGQTAYQRKGSAHLFCSTTCLSSFSHKRTQNTRSIICKKDASTKKA
1008	ZN783_HUMAN	TEITLWTVVAAIQALEKKVDSCLTRLLTLEGRTGTAEKKLADCEKTAVEFGNQLEGKWAVLGTLLQEYGLLQRRLENVEN
1009	SPI2B_HUMAN	KKQRGRPSSQPRRNIVGCRISHGWKEGDEPITQWKGTVLDQVPINPSLYLVKYDGIDCVYGLELHRDERVLSLKILSDRV
1010	LRP1_HUMAN	WTCDLDDDCGDRSDESASCAYPTCFPLTQFTCNNGRCININWRCDNDNDCGDNSDEAGCSHSCSSTQFKCNSGRCIPEHW
1011	MIXL1_HUMAN	PKGAAAPSASQRRKRTSFSAEQLQLLELVFRRTRYPDIHLRERLAALTLLPESRIQVWFQNRRAKSRRQSGKSFQPLARP
1012	SGT1_HUMAN	KIKYDWYQTESQVVITLMIKNVQKNDVNVEFSEKELSALVKLPSGEDYNLKLELLHPIIPEQSTFKVLSTKIEIKLKKPE
1013	LMCD1_HUMAN	DPSKEVEYVCELCKGAAPPDSPVVYSDRAGYNKQWHPTCFVCAKCSEPLVDLIYFWKDGAPWCGRHYCESLRPRCSGCDE
1014	CEBPA_HUMAN	GSGAGKAKKSVDKNSNEYRVRRERNNIAVRKSRDKAKQRNVETQQKVLELTSDNDRLRKRVEQLSRELDTLRGIFRQLPE
1015	GATA2_HUMAN	GPASSFTPKQRSKARSCSEGRECVNCGATATPLWRRDGTGHYLCNACGLYHKMNGQNRPLIKPKRRLSAARRAGTCCANC
1016	SOX14_HUMAN	KPSDHIKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSEAEKRPYIDEAKRLRAQHMKEHPDYKYRPRRK
1017	WTIP_HUMAN	LYSGFQQTADKCSVCGHLIMEMILQALGKSYHPGCFRCSVCNECLDGVPFTVDVENNIYCVRDYHTVFAPKCASCARPIL
1018	PRP19_HUMAN	HPSQDLVFSASPDATIRIWSVPNASCVQVVRAHESAVTGLSLHATGDYLLSSSDDQYWAFSDIQTGRVLTKVTDETSGCS
1019	CBX6_HUMAN	ELSAVGERVFAAESIIKRRIRKGRIEYLVKWKGWAIKYSTWEPEENILDSRLIAAFEQKERERELYGPKKRGPKPKTFLL
1020	NKX11_HUMAN	RTGSDSKSGKPRRARTAFTYEQLVALENKFKATRYLSVCERLNLALSLSLTETQVKIWFQNRRTKWKKQNPGADTSAPTG
1021	RBBP4_HUMAN	VWDLSKIGEEQSPEDAEDGPPELLFIHGGHTAKISDFSWNPNEPWVICSVSEDNIMQVWQMAENIYNDEDPEGSVDPEGQ
1022	DMRT2_HUMAN	ERCTPAGGGAEPRKLSRTPKCARCRNHGVVSCLKGHKRFCRWRDCQCANCLLVVERQRVMAAQVALRRQQATEDKKGLSG
1023	SMCA2_HUMAN	SQPGALIPGDPQAMSQPNRGPSPFSPVQLHQLRAQILAYKMLARGQPLPETLQLAVQGKRTLPGLQQQQQQQQQQQQQQQ
1024	ZNF10	MDAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEPWLVEREIHQETHPDSETAFEIKSSVSSRSIFKDKQSCDIKMEGMARNDLWYLSLEEVWKCRDQLDKYQENPERHLRQVAFTQKKVLTQERVSESGKYGGNCLLPAQLVLREYFHKRDSHTKSLKHDLVLNGHQDSCASNSNECGQTFCQNIHLIQFARTHTGDKSYKCPDNDNSLTHGSSLGISKGIHREKPYECKECGKFFSWRSNLTRHQLIHTGEKPYECKECGKSFSRSSHLIGHQKTHTGEEPYECKECGKSFSWFSHLVTHQRTHTGDKLYTCNQCGKSFVHSSRLIRHQRTHTGEKPYECPECGKSFRQSTHLILHQRTHVRVRPYECNECGKSYSQRSHLVVHHRIHTGLKPFECKDCGKCFSRSSHLYSHQRTHTGEKPYECHDCGKSFSQSSALIVHQRIHTGEKPYECCQCGKAFIRKNDLIKHQRIHVGEETYKCNQCGIIFSQNSPFIVHQIAHTGEQFLTCNQCGTALVNTSNLIGYQTNHIRENAY
1025	EED_HUMAN	MSEREVSTAPAGTDMPAAKKQKLSSDENSNPDLSGDENDDAVSIESGTNTERPDTPTNTPNAPGRKSWGKGKWKSKKCKYSFKCVNSLKEDHNQPLFGVQFNWHSKEGDPLVFATVGSNRVTLYECHSQGEIRLLQSYVDADADENFYTCAWTYDSNTSHPLLAVAGSRGIIRIINPITMQCIKHYVGHGNAINELKFHPRDPNLLLSVSKDHALRLWNIQTDTLVAIFGGVEGHRDEVLSADYDLLGEKIMSCGMDHSLKLWRINSKRMMNAIKESYDYNPNKTNRPFISQKIHFPDFSTRDIHRNYVDCVRWLGDLILSKSCENAIVCWKPGKMEDDIDKIKPSESNVTILGRFDYSQCDIWYMRFSMDFWQKMLALGNQVGKLYVWDLEVEDPHKAKCTTLTHHKCGAAIRQTSFSRDSSILIAVCDDASIWRWDRLR
1026	RCOR1_HUMAN	MPAMVEKGPEVSGKRRGRNNAAASASAAAASAAASAACASPAATAASGAAASSASAAAASAAAAPNNGQNKSLAAAAPNGNSSSNSWEEGSSGSSSDEEHGGGGMRVGPQYQAVVPDFDPAKLARRSQERDNLGMLVWSPNQNLSEAKLDEYIAIAKEKHGYNMEQALGMLFWHKHNIEKSLADLPNFTPFPDEWTVEDKVLFEQAFSFHGKTFHRIQQMLPDKSIASLVKFYYSWKKTRTKTSVMDRHARKQKREREESEDELEEANGNNPIDIEVDQNKESKKEVPPTETVPQVKKEKHSTQAKNRAKRKPPKGMFLSQEDVEAVSANATAATTVLRQLDMELVSVKRQIQNIKQTNSALKEKLDGGIEPYRLPEVIQKCNARWTTEEQLLAVQAIRKYGRDFQAISDVIGNKSVVQVKNFFVNYRRRFNIDEVLQEWEAEHGKEETNGPSNQKPVKSPDNSIKMPEEEDEAPVLDVRYASAS

SEQ	Description	Sequence
1027	human DNMT1	MPARTAPARVPTLAVPAISLPDDVRRRLKDLERDSLTEKECVKEKLNLLHEFLQTEIKNQLCDLETKLRKEELSEEGYLAKVKSLLNKDLSLENGAHAYNREVNGRLENGNQARSEARRVGMADANSPPKPLSKPRTPRRSKSDGEAKPEPSPSPRITRKSTRQTTITSHFAKGPAKRKPQEESERAKSDESIKEEDKDQDEKRRRVTSRERVARPLPAEEPERAKSGTRTEKEEERDEKEEKRLRSQTKEPTPKQKLKEEPDREARAGVQADEDEDGDEKDEKKHRSQPKDLAAKRRPEEKEPEKVNPQISDEKDEDEKEEKRRKTTPKEPTEKKMARAKTVMNSKTHPPKCIQCGQYLDDPLKYGQHPPDAVDEPQMLTNEKLSIFDANESGFESYEALPQHKLTCFSVYCKHGHLCPIDTGLIEKNIELFFSGSAKPIYDDDPSLEGGVNGKNLGPINEWWITGFDGGEKALIGFSTSFAEYILMDPSPEYAPIFGLMQEKIYISKIVVEFLQSNSDSTYEDLINKIETTVPPSGLNLNRFTEDSLLRHAQFVVEQVESYDEAGDSDEQPIFLTPCMRDLIKLAGVTLGQRRAQARRQTIRHSTREKDRGPTKATTTKLVYQIFDTFFAEQIEKDDREDKENAFKRRRCGVCEVCQQPECGKCKACKDMVKFGGSGRSKQACQERRCPNMAMKEADDDEEVDDNIPEMPSPKKMHQGKKKKQNKNRISWVGEAVKTDGKKSYYKKVCIDAETLEVGDCVSVIPDDSSKPLYLARVTALWEDSSNGQMFHAHWFCAGTDTVLGATSDPLELFLVDECEDMQLSYIHSKVKVIYKAPSENWAMEGGMDPESLLEGDDGKTYFYQLWYDQDYARFESPPKTQPTEDNKFKFCVSCARLAEMRQKEIPRVLEQLEDLDSRVLYYSATKNGILYRVGDGVYLPPEAFTFNIKLSSPVKRPRKEPVDEDLYPEHYRKYSDYIKGSNLDAPEPYRIGRIKEIFCPKKSNGRPNETDIKIRVNKFYRPENTHKSTPASYHADINLLYWSDEEAVVDFKAVQGRCTVEYGEDLPECVQVYSMGGPNRFYFLEAYNAKSKSFEDPPNHARSPGNKGKGKGKGKGKPKSQACEPSEPEIEIKLPKLRTLDVFSGCGGLSEGFHQAGISDTLWAIEMWDPAAQAFRLNNPGSTVFTEDCNILLKLVMAGETTNSRGQRLPQKGDVEMLCGGPPCQGFSGMNRFNSRTYSKFKNSLVVSFLSYCDYYRPRFFLLENVRNFVSFKRSMVLKLTLRCLVRMGYQCTFGVLQAGQYGVAQTRRRAIILAAAPGEKLPLFPEPLHVFAPRACQLSVVVDDKKFVSNITRLSSGPFRTITVRDTMSDLPEVRNGASALEISYNGEPQSWFQRQLRGAQYQPILRDHICKDMSALVAARMRHIPLAPGSDWRDLPNIEVRLSDGTMARKLRYTHHDRKNGRSSSGALRGVCSCVEAGKACDPAARQFNTLIPWCLPHTGNRHNHWAGLYGRLEWDGFFSTTVTNPEPMGKQGRVLHPEQHRVVSVRECARSQGFPDTYRLFGNILDKHRQVGNAVPPPLAKAIGLEIKLCMLAKARESASAKIKEEEAAKD
1028	human DNMT3A	MPAMPSSGPGDTSSSAAEREEDRKDGEEQEEPRGKEERQEPSTTARKVGRPGRKRKHPPVESGDTPKDPAVISKSPSMAQDSGASELLPNGDLEKRSEPQPEEGSPAGGQKGGAPAEGEGAAETLPEASRAVENGCCTPKEGRGAPAEAGKEQKETNIESMKMEGSRGRLRGGLGWESSLRQRPMPRLTFQAGDPYYISKRKRDEWLARWKREAEKKAKVIAGMNAVEENQGPGESQKVEEASPPAVQQPTDPASPTVATTPEPVGSDAGDKNATKAGDDEPEYEDGRGFGIGELVWGKLRGFSWWPGRIVSWWMTGRSRAAEGTRWVMWFGDGKFSVVCVEKLMPLSSFCSAFHQATYNKQPMYRKAIYEVLQVASSRAGKLFPVCHDSDESDTAKAVEVQNKPMIEWALGGFQPSGPKGLEPPEEEKNPYKEVYTDMWVEPEAAAYAPPPPAKKPRKSTAEKPKVKEIIDERTRERLVYEVRQKCRNIEDICISCGSLNVTLEHPLFVGGMCQNCKNCFLECAYQYDDDGYQSYCTICCGGREVLMCGNNNCCRCFCVECVDLLVGPGAAQAAIKEDPWNCYMCGHKGTYGLLRREDWPSRLQMFFANNHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV
1029	human DNMT3A catalytic domain	NHDQEFDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV
1030	human DNMT3B	MKGDTRHLNGEEDAGGREDSILVNGACSDQSSDSPPILEAIRTPEIRGRRSSSRLSKREVSSLLSYTQDLTGDGDGEDGDGSDTPVMPKLFRETRTRSESPAVRTRNNNSVSSRERHRPSPRSTRGRQGRNHVDESPVEFPATRSLRRRATASAGTPWPSPPSSYLTIDLTDDTEDTHGTPQSSSTPYARLAQDSQQGGMESPQVEADSGDGDSSEYQDGKEFGIGDLVWGKIKGFSWWPAMVVSWKATSKRQAMSGMRWVQWFGDGKFSEVSADKLVALGLFSQHFNLATFNKLVSYRKAMYHALEKARVRAGKTFPSSPGDSLEDQLKPMLEWAHGGFKPTGIEGLKPNNTQPVVNKSKVRRAGSRKLESRKYENKTRRRTADDSATSDYCPAPKRLKTNCYNNGKDRGDEDQSREQMASDVANNKSSLEDGCLSCGRKNPVSFHPLFEGGLCQTCRDRFLELFYMYDDDGYQSYCTVCCEGRELLLCSNTSCCRCFCVECLEVLVGTGTAAEAKLQEPWSCYMCLPQRCHGVLRRRKDWNVRLQAFFTSDTGLEYEAPKLYPAIPAARRRPIRVLSLFDGIATGYLVLKELGIKVGKYVASEVCEESIAVGTVKHEGNIKYVNDVRNITKKNIEEWGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEFYHLLNYSRPKEGDDRPFFWMFENVVAMKVGDKRDISRFLECNPVMIDAIKVSAAHRARYFWGNLPGMNRPVIASKNDKLELQDCLEYNRIAKLKKVQTITTKSNSIKQGKNQLFPVVMNGKEDVLWCTELERIFGFPVHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE
1031	mouse DNMT3C	MRGGSRHLSNEEDVSGCEDCIIISGTCSDQSSDPKTVPLTQVLEAVCTVENRGCRTSSQPSKRKASSLISYVQDLTGDGDEDRDGEVGGSSGSGTPVMPQLFCETRIPSKTPAPLSWQANTSASTPWLSPASPYPIIDLTDEDVIPQSISTPSVDWSQDSHQEGMDTTQVDAESRDGGNI

SEQ	Description	Sequence
		EYQVSADKLLLSQSCILAAFYKLVPYRESIYRTLEKARVRAGKACPSSPGESLEDQLKPMLEWAHGGFKPTGIEGLKPNKKQPENKSRRRTTNDPAASESSPPKRLKTNSYGGKDRGEDEESREQMASDVTNNKGNLEDHCLSCGRKDPVSFHPLFEGGLCQSCRDRFLELFYMYDEDGYQSYCTVCCEGRELLLCSNTSCCRCFCVECLEVLVGAGTAEDVKLQEPWSCYMCLPQRCHGVLRRRKDWNMRLQDFFTTDPDLEEFEPPKLYPAIPAAKRRPIRVLSLFDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKHEGQIKYVDDIRNITKEHIDEWGPFDLVIGGSPCNDLSCVNPVRKGLFEGTGRLFFEFYRLLNYSCPEEEDDRPFFWMFENVVAMEVGDKRDISRFLECNPVMIDAIKVSAAHRARYFWGNLPGMNRPVMASKNDKLELQDCLEFSRTAKLKKVQTITTKSNSIRQGKNQLFPVVMNGKDDVLWCTELERIFGFPEHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDHFACE
1032	human DNMT3L	MAAIPALDPEAEPSMDVILVGSSELSSSVSPGTGRDLIAYEVKANQRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKFLDALFLYDDDGYQSYCSICCSGETLLICGNPDCTRCYCFECVDSLVGPGTSGKVHAMSNWVCYLCLPSSRSGLLQRRRKWRSQLKAFYDRESENPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSL
1033	human DNMT3L catalytic domain	NPLEMFETVPVWRRQPVRVLSLFEDIKKELTSLGFLESGSDPGQLKHVVDVTDTVRKDVEEWGPFDLVYGATPPLGHTCDRPPSWYLFQFHRLLQYARPKPGSPRPFFWMFVDNLVLNKEDLDVASRFLEMEPVTIPDVHGGSLQNAVRVWSNIPAIRSRHWALVSEEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLREYFKYFSTELTSSL
1034	mouse DNMT3L	MGSRETPSSCSKTLETLDLETSDSSSPDADSPLEEQWLKSSPALKEDSVDVVLEDCKEPLSPSSPPTGREMIRYEVKVNRRSIEDICLCCGTLQVYTRHPLFEGGLCAPCKDKFLESLFLYDDDGHQSYCTICCSGGTLFICESPDCTRCYCFECVDILVGPGTSERINAMACWVCFLCLPFSRSGLLQRRKRWRHQLKAFHDQEGAGPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPL
1035	mouse DNMT3L catalytic domain	GPMEIYKTVSAWKRQPVRVLSLFRNIDKVLKSLGFLESGSGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKHAPLTPKEEEYLQAQVRSRSKLDAPKVDLLVKNCLLPLREYFKYFSQNSLPL
1036	human TRDMT1 (DNMT2)	MEPLRVLELYSGVGGMHHALRESCIPAQVVAAIDVNTVANEVYKYNFPHTQLLAKTIEGITLEEFDRLSFDMILMSPPCQPFTRIGRQGDMTDSRTNSFLHILDILPRLQKLPKYILLENVKGFEVSSTRDLLIQTIENCGFQYQEFLLSPTSLGIPNSRLRYFLIAKLQSEPLPFQAPGQVLMEFPKIESVHPQKYAMDVENKIQEKNVEPNISFDGSIQCSGKDAILFKLETAEEIHRKNQQDSDLSVKMLKDFLEDDTDVNQYLLPPKSLLRYALLLDIVQPTCRRSVCFTKGYGSYIEGTGSVLQTAEDVQVENIYKSLTNLSQEEQITKLLILKLRYFTPKEIANLLGFPPEFGFPEKITVKQRYRLLGNSLNVHVVAKLIKILYE
1037	*M. penetrans* M MpeI	MNSNKDKIKVIKVFEAFAGIGSQFKALKNIARSKNWEIQHSGMVEWFVDAIVSYVAIHSKNFNPKIEQLDKDILSISNDSKMPISEYGIKKINNTIKASYLNYAKKHFNNLFDIKKVNKDNFPKNIDIFTYSFPCQDLSVQGLQKGIDKELNTRSGLLWEIERILEEIKNSFSKEEMPKYLLMENVKNLLSHKNKKNYNTWLKQLEKFGYKSKTYLLNSKNFDNCQNRERVFCLSIRDDYLEKTGFKFKELEKVKNPPKKIKDILVDSSNYKYLNLNKYETTTFRETKSNIISRSLKNYTTFNSENYVYNINGIGPTLTASGANSRIKIETQQGVRYLTPLECFKYMQFDVNDFKKVQSTNLISENKMIYIAGNSIPVKILEAIFNTLEFVNNEE
1038	*S. monobiae* M SssI	MSKVENKTKKLRVFEAFAGIGAQRKALEKVRKDEYEIVGLAEWYVPAIVMYQAIHNNFHTKLEYKSVSREEMIDYLENKTLSWNSKNPVSNGYWKRKKDDELKIIYNAIKLSEKEGNIFDIRDLYKRTLKNIDLLTYSFPCQDLSQQGIQKGMKRGSGTRSGLLWEIERALDSTEKNDLPKYLLMENVGALLHKKNEEELNQWKQKLESLGYQNSIEVLNAADFGSSQARRRVFMISTLNEFVELPKGDKKPKSIKKVLNKIVSEKDILNNLLKYNLTEFKKTKSNINKASLIGYSKFNSEGYVYDPEFTGPTLTASGANSRIKIKDGSNIRKMNSDETFLYIGFDSQDGKRVNEIEFLTENQKIFVCGNSISVEVLEAIIDKIGG
1039	*H. parainfluenzae* M HpaII	MKDVLDDNLLEEPAAQYSLFEPESNPNLREKFTFIDLFAGIGGFRIAMQNLGGKCIFSSEWDEQAQKTYEANFGDLPYGDITLEETKAFIPEKFDILCAGFPCQAFSIAGKRGGFEDTRGTLFFDVAEIIRRHQPKAFFLENVKGLKNHDKGRTLKTILNVLREDLGYFVPEPAIVNAKNFGVPQNRERIYIVGFHKSTGVNSFSYPEPLDKIVTFADIREEKTVPTKYYLSTQYIDTLRKHKERHESKGNGFGYEIIPDDGIANAIVVGGMGRERNLVIDHRITDFTPTTNIKGEVNREGIRKMTPREWARLQGFPDSYVIPVSDASAYKQFGNSVAVPAIQATGKKILEKLGNLYD
1040	*A. luteus* M AluI	MSKANAKYSFVDLFAGIGGFHAALAATGGVCEYAVEIDREAAAVYERNWNKPALGDITDDANDEGVTLRGYDGPIDVLTGGFPCQPFSKSGAQHGMAETRGTLFWNIARIIEEREPTVLILENVRNLVGPRHRHEWLTIIETLRFFGYEVSGAPAIFSPHLLPAWMGGTPQVRERVFITATLVPERMRDERIPRTETGEIDAEAIGPKPVATMNDRFPIKKGGTELFHPGDRKSGWNLLTSGIIREGDPEPSNVDLRLTETETLWIDAWDDLESTIRRATGRPLEGFPYWADSWTDFRELSRLVVIRGFQAPEREVVGDRKRYVARTDMPEGFVPASVTRPAIDETLPAWKQSHLRRNYDFFERHFAEVVAWAYRWGVYTDLFPASRRKLEWQAQDAPRLWDTVMHFRPSGIRAKRPTYL

SEQ	Description	Sequence
		PALVAITQTSIVGPLERRLSPRETARLQGLPEWFDFGEQRAAATYKQMGNGVNVGVVRHILREHVRRDRALLKLTPAGQRIINAVLADEPDATVGALGAAE
1041	*H. aegyptius* M HaeIII	MNLISLFSGAGGLDLGFQKAGFRIICANEYDKSIWKTYESNHSAKLIKGDISKISSDEFPKCDGIIGGPPCQSWSEGGSLRGIDDPRGKLFYEYIRILKQKKPIFFLAENVKGMMAQRHNKAVQEFIQEFDNAGYDVHIILLNANDYGVAQDRKRVFYIGFRKELNINYLPPIPHLIKPTFKDVIWDLKDNPIPALDKNKTNGNKCIYPNHEYFIGSYSTIFMSRNRVRQWNEPAFTVQASGRQCQLHPQAPVMLKVSKNLNKFVEGKEHLYRRLTVRECARVQGFPDDFIFHYESLNDGYKMIGNAVPVNLAYEIAKTIKSALEICKGN
1042	*H. haemolyticus* M HhaI	MIEIKDKQLTGLRFIDLFAGLGGFRLALESCGAECVYSNEWDKYAQEVYEMNFGEKPEGDITQVNEKTIPDHDILCAGFPCQAFSISGKQKGFEDSRGTLFFDIARIVREKKPKVVFMENVKNFASHDNGNTLEVVKNTMNELDYSFHAKVLNALDYGIPQKRERIYMICFRNDLNIQNFQFPKPFELNTFVKDLLLPDSEVEHLVIDRKDLVMTNQEIEQTTPKTVRLGIVGKGGQGERIYSTRGIAITLSAYGGGIFAKTGGYLVNGKTRKLHPRECARVMGYPDSYKVHPSTSQAYKQFGNSVVINVLQYIAYNIGSSLNFKPY
1043	*Moraxella M* MspI	MKPEILKLIRSKLDLTQKQASEIIEVSDKTWQQWESGKTEMHPAYYSFLQEKLKDKINFEELSAQKTLQKKIFDKYNQNQITKNAEELAEITHIEERKDAYSSDFKFIDLFSGIGGIRQSFEVNGGKCVFSSEIDPFAKFTYYTNFGVVPFGDITKVEATTIPQHDILCAGFPCQPFSHIGKREGFEHPTQGTMFHEIVRIIETKKTPVLFLENVPGLINHDDGNTLKVIIETLEDMGYKVHHTVLDASHFGIPQKRKRFYLVAFLNQNIHFEFPKPPMISKDIGEVLESDVTGYSISEHLQKSYLFKKDDGKPSLIDKNTTGAVKTLVSTYHKIQRLTGTFVKDGETGIRLLTTNECKAIMGFPKDFVIPVSRTQMYRQMGNSVVVPVVTKIAEQISLALKTVNQQSPQENFELELV
1044	*Ascobolus* Masc1	MSERRYEAGMTVALHEGSFLKIQRVYIRQYHADNRREHMLVGPLFRRTKYLKALSKKVNEVAIVHESIHVPVQDVIGVRELIITNRPFPECRKGDEHTGRLVCRWVYNLDERAKGREYKKQRYIRRITEAEADPEYRVEDRVLRRRWFQEGYIGDEISYKEHGNGDIVDIRSESPLQVLDGWGGDLVDLENGEETSIPGPCRSASSYGRLMKPPLAQAADSNTSRKYTFGDTFCGGGGVSLGARQAGLEVKWAFDMNPNAGANYRRNFPNTDFFLAEAEQFIQLSVGISQHVDILHLSPPCQTFSRAHTIAGKNDENNEASFFAVVNLIKAVRPRLFTVEETDGIMDRQSRQFIDTALMGITELGYSFRICVLNAIEYGVCQNRKRLIIIGAAPGEELPPFPLPTHQDFFSKDPRRDLLPAVTLDDALSTITPESTDHHLNHVWQPAEWKTPYDAHRPFKNAIRAGGGEYDIYPDGRRKFTVRELACIQGFPDEYEFVGTLTDKRRIIGNAVPPPLSAAIMSTLRQWMTEKDFERME
1045	*Arabidopsis* MET1	MVENGAKAAKRKKRPLPEIQEVEDVPRTRRPRRAAACTSFKEKSIRVCEKSATIEVKKQQIVEEEFLALRLTALETDVEDRPTRRLNDFVLFDSDGVPQPLEMLEIHDIFVSGAILPSDVCTDKEKEKGVRCTSFGRVEHWSISGYEDGSPVIWISTELADYDCRKPAASYRKVYDYFYEKARASVAVYKKLSKSSGGDPDIGLEELLAAVVRSMSSGSKYFSSGAAIIDFVISQGDFIYNQLAGLDETAKKHESSYVEIPVLVALREKSSKIDKPLQRERNPSNGVRIKEVSQVAESEALTSDQLVDGTDDDRRYAILLQDEENRKSMQQPRKNSSSGSASNMFYIKINEDEIANDYPLPSYYKTSEEETDELILYDASYEVQSEHLPHRMLHNWALYNSDLRFISLELLPMKQCDDIDVNIFGSGVVTDDNGSWISLNDPDSGSQSHDPDGMCIFLSQIKEWMIEFGSDDIISISIRTDVAWYRLGKPSKLYAPWWKPVLKTARVGISILTFLRVESRVARLSFADVTKRLSGLQANDKAYISSDPLAVERYLVVHGQIILQLFAVYPDDNVKRCPFVVGLASKLEDRHHTKWIIKKKKISLKELNLNPRAGMAPVASKRKAMQATTTRLVNRIWGEFYSNYSPEDPLQATAAENGEDEVEEEGGNGEEEVEEEGENGLTEDTVPEPVEVQKPHTPKKIRGSSGKREIKWDGESLGKTSAGEPLYQQALVGGEMVAVGGAVTLEVDDPDEMPAIYFVEYMFESTDHCKMLHGRFLQRGSMTVLGNAANERELFLTNECMTTQLKDIKGVASFEIRSRPWGHQYRKKNITADKLDWARALERKVKDLPTEYYCKSLYSPERGGFFSLPLSDIGRSSGFCTSCKIREDEEKRSTIKLNVSKTGFFINGIEYSVEDFVYVNPDSIGGLKEGSKTSFKSGRNIGLRAYVVCQLLEIVPKESRKADLGSFDVKVRRFYRPEDVSAEKAYASDIQELYFSQDTVVLPPGALEGKCEVRKKSDMPLSREYPISDHIFFCDLFFDTSKGSLKQLPANMKPKFSTIKDDTLLRKKKGKGVESEIESEIVKPVEPPKEIRLATLDIFAGCGGLSHGLKKAGVSDAKWAIEYEEPAGQAFKQNHPESTVFVDNCNVILRAIMEKGGDQDDCVSTTEANELAAKLTEEQKSTLPLPGQVDFINGGPPCQGFSGMNRFNQSSWSKVQCEMILAFLSFADYFRPRYFLLENVRTFVSFNKGQTFQLTLASLLEMGYQVRFGILEAGAYGVSQSRKRAFIWAAAPEEVLPEWPEPMHVFGVPKLKISLSQGLHYAAVRSTALGAPFRPITVRDTIGDLPSVENGDSRTNKEYKEVAVSWFQKEIRGNTIALTDHICKAMNELNLIRCKLIPTRPGADWHDLPKRKVTLSDGRVEEMIPFCLPNTAERHNGWKGLYGRLDWQGNFPTSVTDPQPMGKVGMCFHPEQHRILTVRECARSQGFPDSYEFAGNINHKHRQIGNAVPPPLAFALGRKLKEALHLKKSPQHQP
1046	*Ascobolus* Masc2	MELTPELSGVSTDLGGGGSIFAHWRMKEESPAPTEILDDLNVLEWEKTTRDYSKEDLRIADQLFSIEDEHQSLPFETADAEDGTPTEEEEEKELPMRTLDNFVLYDASDLELAALDLIGTELNIHAVGTVGPIYTEGEEDEQEDEDEDVSPPVRTGTQATSASVTQMTVELYIRNIVQYEFCFNDDGTVETWIQTTNAHYKLLQPAKCYTSLYRPVNDCLNVITAIITLAPESTTMSLKDLLKVMDDKAQAVSYEEVERMSEFIVQHLDQWMETAPKKKSKLIEKSKVYIDLNNLAGIDMVSGVRPPPVRRVTGRSSAPKKRIVRNMNDAVLLHQNETTVTNWIHQLSAGMFGRALNVLGAETADVENLTCDPASAKFVVPQRRLHKRLKWETRGHIPVSEEEYKHIYQGKKYAKFFEAVRAVDESKLTIKLGDLVYVLDQDPKVTQTQFATAGREGRKKGAEKEKIQVRFGRVLSIRQPDSNSKDAQNVFIHVQWLVLGCDTILQEMASRRELFLTDSCDTVFADVIYGVAKLTPLGAKDIPTVEFHESMATMMGENEFFVRFKYNYQDGSFTDLKDVDAEQIGTLQPRVNTHRNPGYCSNCRIKYDNERTGDKWIYENDTEGEPRLFRSSKGWCIYAQEFVYLQPVEKQPGTTFRVGY

SEQ	Description	Sequence
		ISEINKSSVIVELLARVDDDDKSGHISYSDPRHLYFTGTDIKVTFDKIIRKCFVFHDSGDQKAKAPLMYGTLQRDLYYYRYEKRKGKAELVPVREIRSIHEQTLNDWESRTQIERHGAVSGKKLKGLDIFAGCGGLTLGLDLSGAVDTKWDIEFAPSAANTLALNFPDAQVFNQCANVLLSRAIQSEDEGSLDIEYDLQGRVLPDLPKKGEVDFIYGGPPCQGFSGVNRYKKGNDIKNSLVATFLSYVDHYKPRFVLLENVKGLITTKLGNSKNAEGKWEGGISNGVVKFIYRTLISMNYQCRIGLVQSGEYGVPQSRPRVIFLAARMGERLPDLPEPMHAFEVLDSQYALPHIKRYHTTQNGVAPLPRITIGEAVSDLPKFQYANPGVWPRHDPYSSAKAQPSDKTIEKFSVSKATSFVGYLLQPYHSRPQSEFQRRLRTKLVPSDEPAEKTSLLTTKLVTAHVTRLFNKETTQRIVCVPMWPGADHRSLPKEMRPWCLVDPNSQAEKHRFWPGLFGRLGMEDFFSTALTDVQPCGKQGKVLHPTQRRVYTVRELARAQGFPDWFAFTDGDADSGLGGVKKWHRNIGNAVPVPLGEQIGRCIGYSVWWKDDMIAQLREDGADEDEEMIDGNDQWVEELNTQMAADMPGLPLLVTHLLNLCVYRRLYGPNAKEFLPARVYDKKLEGGRRRLVWAML
1047	*Neurospora* Dim2	MDSPDRSHGGMFIDVPAETMGFQEDYLDMFASVLSQGLAKEGDYAHHQPLPAGKEECLEPIAVATTITPSPDDPQLQLQLELEQQFQTESGLNGVDPAPAPESEDEADLPDGFSDESPDDDFVVQRSKHITVDLPVSTLINPRSTFQRIDENDNLVPPPQSTPERVAVEDLLKAAKAAGKNKEDYIEFELHDFNFYVNYAYHPQEMRPIQLVATKVLHDKYYFDGVLKYGNTKHYVTGMQVLELPVGNYGASLHSVKGQIWVRSKHNAKKEIYYLLKKPAFEYQRYYQPFLWIADLGKHVVDYCTRMVERKREVTLGCFKSDFIQWASKAHGKSKAFQNWRAQHPSDDFRTSVAANIGYIWKEINGVAGAKRAAGDQLFRELMIVKPGQYFRQEVPPGPVVTEGDRTVAATIVTPYIKECFGHMILGKVLRLAGEDAEKEKEVKLAKRLKIENKNATKADTKDDMKNDTATESLPTPLRSLPVQVLEATPIESDIVSIVSSDLPPSENNPPPLTNGSVKPKAKANPKPKPSTQPLHAAHVKYLSQELVNKIKVGDVISTPRDDSSNTDTKWKPTDTDDHRWFGLVQRVHTAKTKSSGRGLNSKSFDVIWFYRPEDTPCCAMKYKWRNELFLSNHCTCQEGHHARVKGNEVLAVHPVDWFGTPESNKGEFFVRQLYESEQRRWITLQKDHLTCYHNQPPKPPTAPYKPGDTVLATLSPSDKFSDPYEVVEYFTQGEKETAFVRLRKLLRRRKVDRQDAPANELVYTEDLVDVRAERIVGKCIMRCFRPDERVPSPYDRGGTGNMFFITHRQDHGRCVPLDTLPPTLRQGFNPLGNLGKPKLRGMDLYCGGGNFGRGLEEGGVVEMRWANDIWDKAIHTYMANTPDPNKTNPFLGSVDDLLRLALEGKFSDNVPRPGEVDFIAAGSPCPGFSLLTQDKKVLNQVKNQSLVASFASFVDFYRPKYGVLENVSGIVQTFVNRKQDVLSQLFCALVGMGYQAQLILGDAWAHGAPQSRERVFLYFAAPGLPLPDPPLPSHSHYRVKNRNIGFLCNGESYVQRSFIPTAFKFVSAGEGTADLPKIGDGKPDACVRFPDHRLASGITPYIRAQYACIPTHPYGMNFIKAWNNGNGVMSKSDRDLFPSEGKTRTSDASVGWKRLNPKTLFPTVTTTSNPSDARMGPGLHWDEDRPYTVQEMRRAQGYLDEEVLVGRTTDQWKLVGNSVSRHMALAIGLKFREAWLGTLYDESAVVATATATATTAAAVGVTVPVMEEPGIGTTESSRPSRSPVHTAVDLDDSKSERSRSTTPATVLSTSSAAGDGSANAAGLEDDDNDDMEMMEVTRKRSSPAVDEEGMRPSKVQKVEVTVASPASRRSSRQASRNPTASPSSKASKATTHEAPAPEELESDAESYSETYDKEGFDGDYHSGHEDQYSEEDEEEEYAEPETMTVNGMTIVKL
1048	*Drosophila* dDnmt2	MVFRVLELFSGIGGMHYAFNYAQLDGQIVAALDVNTVANAVYAHNYGSNLVKTRNIQSLSVKEVTKLQANMLLMSPPCQPHTRQGLQRDTEDKRSDALTHLCGLIPECQELEYILMENVKGFESSQARNQFIESLERSGFHWREFILTPTQFNVPNTRYRYYCIARKGADFPFAGGKIWEEMPGAIAQNQGLSQIAEIVEENVSPDFLVPDDVLTKRVLVMDIIHPAQSRSMCFTKGYTHYTEGTGSAYTPLSEDESHRIFELVKEIDTSNQDASKSEKILQQRLDLLHQVRLRYFTPREVARLMSFPENFEFPPETTNRQKYRLLGNSINVKVVGELIKLLTIK
1049	*S. pombe* Pmt1	MLSTKRLRVLELYSGIGGMHYALNLANIPADIVCAIDINPQANEIYNLNHGKLAKHMDISTLTAKDFDAFDCKLWTMSPSCQPFTRIGNRKDILDPRSQAFLNILNVLPHVNNLPEYILIENVQGFEESKAAEECRKVLRNCGYNLIEGILSPNQFNIPNSRSRWYGLARLNFKGEWSIDDVFQFSEVAQKEGEVKRIRDYLEIERDWSSYMVLESVLNKWGHQFDIVKPDSSSCCCFTRGYTHLVQGAGSILQMSDHENTHEQFERNRMALQLRYFTAREVARLMGFPESLEWSKSNVTEKCMYRLLGNSINVKVVSYLISLLLEPLNF
1050	*Arabidopsis* DRM1	MVMSHIFLISQIQEVEHGDSDDVNWNTDDDELAIDNFQFSPSPVHISATSPNSIQNRISDETVASFVEMGFSTQMIARAIEETAGANMEPMMILETLFNYSASTEASSSKSKVINHFIAMGFPEEHVIKAMQEHGDEDVGEITNALLTYAEVDKLRESEDMNININDDDDDNLYSLSSDDEEDELNNSSNEDRILQALIKMGYLREDAAIAIERCGEDASMEEVVDFICAAQMARQFDEIYAEPDKKELMNNNKKRRTYTETPRKPNTDQLISLPKEMIGFGVPNHPGLMMHRPVPIPDIARGPPFFYYENVAMTPKGVWAKISSHLYDIVPEFVDSKHFCAAARKRGYIHNLPIQNRFQIQPPQHNTIQEAFPLTKRWWPSWDGRTKLNCLLTCIASSRLTEKIREALERYDGETPLDVQKWVMYECKKWNLVWVGKNKLAPLDADEMEKLLGFPRDHTRGGGISTTDRYKSLGNSFQVDTVAYHLSVLKPLFPNGINVLSLFTGIGGGEVALHRLQIKMNVVVSVEISDANRNILRSFWEQTNQKGILREFKDVQKLDDNTIERLMDEYGGFDLVIGGSPCNNLAGGNRHHRVGLGGEHSSLFFDYCRILEAVRRKARHMRR
1051	*Arabadopsis* DRM2	MVIWNNDDDDFLEIDNFQSSPRSSPIHAMQCRVENLAGVAVTTSSLSSPTETTDLVQMGFSDEVFATLFDMGFPVEMISRAIKETGPNVETSVIIDTISKYSSDCEAGSSKSKAIDHFLAMGFDEEKVVKAIQEHGEDNMEAIANALLSCPEAKKLPAAVEEEDGIDWSSSDDDTNYTDMLNSDDEKDPNSNENGSKIRSLVKMGFSELEASLAVERCGENVDIAELTDFLCAAQMAREFSEFYTEHEEQKPRHNIKKRRFESKGEPRSSVDDEPIRLPNPMIGFGVPNEPGLITHRSLPELARGPPFFYYENVALTPKGVWETISRHLFEIPPEFVDSKYFCVAARKRGYIHNLPINNRFQIQPPPKYTIHDAFPLSKRWWPEWDKRTKLNCILTCTGSAQLTNRIRVALEPYNEEPEP

SEQ	Description	Sequence
		PKHVQRYVIDQCKKWNLVWVGKNKAAPLEPDEMESILGFPKNHTRGGGMSRTERFKSLGNSFQVDTVAYHLSVLKPIFPHGINVLSLFTGIGGGEVALHRLQIKMKLVVSVEISKVNRNILKDFWEQTNQTGELIEFSDIQHLTNDTIEGLMEKYGGFDLVIGGSPCNNLAGGNRVSRVGLEGDQSSLFFEYCRILEVVRARMRGS
1052	Arabadopsis CMT1	MAARNKQKKRAEPESDLCFAGKPMSVVESTIRWPHRYQSKKTKLQAPTKKPANKGGKKEDEEIIKQAKCHFDKALVDGVLINLNDDVYVTGLPGKLKFIAKVIELFEADDGVPYCRFRWYYRPEDTLIERFSHLVQPKRVFLSNDENDNPLTCIWSKVNIAKVPLPKITSRIEQRVIPPCDYYYDMKYEVPYLNFTSADDGSDASSSLSSDSALNCFENLHKDEKFLLDLYSGCGAMSTGFCMGASISGVKLITKWSVDINKFACDSLKLNHPETEVRNEAAEDFLALLKEWKRLCEKFSLVSSTEPVESISELEDEEVEENDDIDEASTGAELEPGEFEVEKFLGIMFGDPQGTGEKTLQLMVRWKGYNSSYDTWEPYSGLGNCKEKLKEYVIDGFKSHLLPLPGTVYTVCGGPPCQGISGYNRYRNNEAPLEDQKNQQLLVFLDIIDFLKPNYVLMENVVDLLRFSKGFLARHAVASFVAMNYQTRLGMMAAGSYGLPQLRNRVFLWAAQPSEKLPPYPLPTHEVAKKFNTPKEFKDLQVGRIQMEFLKLDNALTLADAISDLPPVTNYVANDVMDYNDAAPKTEFENFISLKRSETLLPAFGGDPTRRLFDHQPLVLGDDDLERVSYIPKQKGANYRDMPGVLVHNNKAEINPRFRAKLKSGKNVVPAYAISFIKGKSKKPFGRLWGDEIVNTVVTRAEPHNQCVIHPMQNRVLSVRENARLQGFPDCYKLCGTIKEKYIQVGNAVAVPVGVALGYAFGMASQGLTDDEPVIKLPFKYPECMQAKDQI
1053	*Arabadopsis* CMT2	MLSPAKCESEEAQAPLDLHSSSRSEPECLSLVLWCPNPEEAAPSSTRELIKLPDNGEMSLRRSTTLNCNSPEENGGEGRVSQRKSSRGKSQPLLMLTNGCQLRRSPRFRALHANFDNVCSVPVTKGGVSQRKFSRGKSQPLLTLTNGCQLRRSPRFRAVDGNFDSVCSVPVTGKFGSRKRKSNSALDKKESSDSEGLTFKDIAVIAKSLEMEIISECQYKNNVAEGRSRLQDPAKRKVDSDTLLYSSINSSKQSLGSNKRMRRSQRFMKGTENEGEENLGKSKGKGMSLASCSFRRSTRLSGTVETGNTETLNRRKDCGPALCGAEQVRGTERLVQISKKDHCCEAMKKCEGDGLVSSKQELLVFPSGCIKKTVNGCRDRTLGKPRSSGLNTDDIHTSSLKISKNDTSNGLTMTTALVEQDAMESLLQGKTSACGAADKGKTREMHVNSTVIYLSDSDEPSSIEYLNGDNLTQVESGSALSSGGNEGIVSLDLNNPTKSTKRKGKRVTRTAVQEQNKRSICFFIGEPLSCEEAQERWRWRYELKERKSKSRGQQSEDDEDKIVANVECHYSQAKVDGHTFSLGDFAYIKGEEEETHVGQIVEFFKTTDGESYFRVQWFYRATDTIMERQATNHDKRRLFYSTVMNDNPVDCLISKVTVLQVSPRVGLKPNSIKSDYYFDMEYCVEYSTFQTLRNPKTSENKLECCADVVPTESTESILKKKSFSGELPVLDLYSGCGGMSTGLSLGAKISGVDVVTKWAVDQNTAACKSLKLNHPNTQVRNDAAGDFLQLLKEWDKLCKRYVFNNDQRTDTLRSVNSTKETSGSSSSSDDDSDSEEYEVEKLVDICFGDHDKTGKNGLKFKVHWKGYRSDEDTWELAEELSNCQDAIREFVTSGFKSKILPLPGRVGVICGGPPCQGISGYNRHRNVDSPLNDERNQQIIVFMDIVEYLKPSYVLMENVVDILRMDKGSLGRYALSRLVNMRYQARLGIMTAGCYGLSQFRSRVFMWGAVPNKNLPPFPLPTHDVIVRYGLPLEFERNVVAYAEGQPRKLEKALVLKDAISDLPHVSNDEDREKLPYESLPKTDFQRYIRSTKRDLTGSAIDNCNKRTMLLHDHRPFHINEDDYARVCQIPKRKGANFRDLPGLIVRNNTVCRDPSMEPVILPSGKPLVPGYVFTFQQGKSKRPFARLWWDETVPTVLTVPTCHSQALLHPEQDRVLTIRESARLQGFPDYFQFCGTIKERYCQIGNAVAVSVSRALGYSLGMAFRGLARDEHLIKLPQNFSHSTYPQLQETIPH
1054	*Arabadopsis* CMT3	MAPKRKRPATKDDTTKSIPKPKKRAPKRAKTVKEEPVTVVEEGEKHVARFLDEPIPESEAKSTWPDRYKPIEVQPPKASSRKKTKDDEKVEIIRARCHYRRAIVDERQIYELNDDAYVQSGEGKDPFICKIIEMFEGANGKLYFTARWFYRPSDTVMKEFEILIKKKRVFFSEIQDTNELGLLEKKLNILMIPLNENTKETIPATENCDFFCDMNYFLPYDTFEAIQQETMMAISESSTISSDTDIREGAAAISEIGECSQETEGHKKATLLDLYSGCGAMSTGLCMGAQLSGLNLVTKWAVDMNAHACKSLQHNHPETNVRNMTAEDFLFLLKEWEKLCIHFSLRNSPNSEEYANLHGLNNVEDNEDVSEESENEDDGEVFTVDKIVGISFGVPKKLLKRGLYLKVRWLNYDDSHDTWEPIEGLSNCRGKIEEFVKLGYKSGILPLPGGVDVVCGGPPCQGISGHNRFRNLLDPLEDQKNKQLLVYMNIVEYLKPKFVLMENVVDMLKMAKGYLARFAVGRLLQMNYQVRNGMMAAGAYGLAQFRLRFFLWGALPSEIIPQFPLPTHDLVHRGNIVKEFQGNIVAYDEGHTVKLADKLLLKDVISDLPAVANSEKRDEITYDKDPTTPFQKFIRLRKDEASGSQSKSKSKKHVLYDHHPLNLNINDYERVCQVPKRKGANFRDFPGVIVGPGNVVKLEEGKERVKLESGKTLVPDYALTYVDGKSCKPFGRLWWDEIVPTVVTRAEPHNQVIIHPEQNRVLSIRENARLQGFPDDYKLFGPPKQKYIQVGNAVAVPVAKALGYALGTAFQGLAVGKDPLLTLPEGFAFMKPTLPSELA
1055	*Neurospora* Rid	MAEQNPFVIDDEDDVIQIHDEEEVEEEVAEVIDITEDDIEPSELDRAFGSRPKEETLPSLLLRDQGFIVRPGMTVELKAPIGRFAISFVRVNSIVKVRQAHVNNVTIRGHGFTRAKEMNGMLPKQLNECCLVASIDTRDPRP
1056	*E. coli* strain 12 hsdM	MNNNDLVAKLWKLCDNLRDGGVSYQNYVNELASLLFLKMCKETGQEAEYLPEGYRWDDLKSRIGQEQLQFYRKMLVHLGEDDKKLVQAVFHNVSTTITEPKQITALVSNMDSLDWYNGAHGKSRDDFGDMYEGLLQKNANETKSGAGQYFTPRPLIKTIIHLLKPQPREVVQDPAAGTAGFLIEADRYVKSQTNDLDDLDGDTQDFQIHRAFIGLELVPGTRRLALMNCLLHDIEGNLDHGGAIRLGNTLGSDGENLPKAHIVATNPPFGSAAGTNITRTFVHPTSNKQLCFMQHIIETLHPGGRAAVVVPDNVLFEGGKGTDIRRDLMDKCHLHTILRLPTGIFYAQGVKTNVLFFTKGTVANPNQDKNCTDDVWVYDLRTNMPSFGKRTPFTDEHLQPFERVYGEDPHGLSPRTEGEWSFNAEETEVADSEENKNTDQHLATSRWRKFSREWIRTAKSDSLDISWLKDKDSIDADSLPEPDVLAAEAMGELVQALSELDALMRELGASDEADLQRQLLEEAFGGVKE

SEQ	Description	Sequence
1057	*E. coli* strain 12 hsdS	MSAGKLPEGWVIAPVSTVTTLIRGVTYKKEQAINYLKDDYLPLIRANNIQNGKFDTTDLVFVPKNLVKESQKISPEDIVIAMSSGSKSVVGKSAHQHLPFECSFGAFCGVLRPEKLIFSGFIAHFTKSSLYRNKISSLSAGANINNIKPASFDLINIPIPPLAEQKIIAEKLDTLLAQVDSTKARFEQIPQILKRFRQAVLGGAVNGKLTEKWRNFEPQHSVFKKLNFESILTELRNGLSSKPNESGVGHPILRISSVRAGHVDQNDIRFLECSESELNRHKLQDGDLLFTRYNGSLEFVGVCGLLKKLQHQNLLYPDKLIRARLTKDALPEYIEIFFSSPSARNAMMNCVKTTSGQKGISGKDIKSQVVLLPPVKEQAEIVRRVEQLFAYADTIEKQVNNALARVNNLTQSILAKAFRGELTAQWRAENPDLISGENSAAALLEKIKAERAASGGKKASRKKS
1058	*T. aquaticus* M TaqI	MGLPPLLSLPSNSAPRSLGRVETPPEVVDFMVSLAEAPRGGRVLEPACAHGPFLRAFREAHGTAYRFVGVEIDPKALDLPPWAEGILADFLLWEPGEAFDLILGNPPYGIVGEASKYPIHVFKAVKDLYKKAFSTWKGKYNLYGAFLEKAVRLLKPGGVLVFVVPATWLVLEDFALLREFLAREGKTSVYYLGEVFPQKKVSAVVIRFQKSGKGLSLWDTQESESGFTPILWAEYPHWEGEIIRFETEETRKLEISGMPLGDLFHIRFAARSPEFKKHPAVRKEPGPGLVPVLTGRNLKPGWVDYEKNHSGLWMPKERAKELRDFYATPHLVVAHTKGTRVVAAWDERAYPWREEFHLLPKEGVRLDPSSLVQWLNSEAMQKHVRTLYRDFVPHLTLRMLERLPVRREYGFHTSPESARNF
1059	*E. coli* M EcoDam	MKKNRAFLKWAGGKYPLLDDIKRHLPKGECLVEPFVGAGSVFLNTDFSRYILADINSDLISLYNIVKMRTDEYVQAARELFVPETNCAEVYYQFREEFNKSQDPFRRAVLFLYLNRYGYNGLCRYNLRGEFNVPFGRYKKPYFPEAELYHFAEKAQNAFFYCESYADSMARADDASVVYCDPPYAPLSATANFTAYHTNSFTLEQQAHLAEIAEGLVERHIPVLISNHDTMLTREWYQRAKLHVVKVRRSISSNGGTRKKVDELLALYKPGVVSPAKK
1060	*C. crescentus* M CcrMI	MKFGPETIIHGDCIEQMNALPEKSVDLIFADPPYNLQLGGDLLRPDNSKVDAVDDHWDQFESFAAYDKFTREWLKAARRVLKDDGAIWVIGSYHNIFRVGVAVQDLGFWILNDIVWRKSNPMPNFKGTRFANAHETLIWASKSQNAKRYTFNYDALKMANDEVQMRSDWTIPLCTGEERIKGADGQKAHPTQKPEALLYRVILSTTKPGDVILDPFFGVGTTGAAAKRLGRKFIGIEREAEYLEHAKARIAKVVPIAPEDLDVMGSKRAEPRVPFGTIVEAGLLSPGDTLYCSKGTHVAKVRPDGSITVGDLSGSIHKIGALVQSAPACNGWTYWHFKTDAGLAPIDVLRAQVRAGMN
1061	*C. difficile* CamA	MDDISQDNFLLSKEYENSLDVDTKKASGIYYTPKIIVDYIVKKTLKNHDIIKNPYPRILDISCGCGNFLLEVYDILYDLFEENIYELKKKYDENYWTVDNIHRHILNYCIYGADIDEKAISILKDSLTNKKVVNDLDESDIKINLFCCDSLKKKWRYKFDYIVGNPPYIGHKKLEKKYKKFLLEKYSEVYKDKADLYFCFYKKIIDILKQGGIGSVITPRYFLESLSGKDLREYIKSNVNVQEIVDFLGANIFKNIGVSSCILTFDKKKTKETYIDVFKIKNEDICINKFETLEELLKSSKFEHFNINQRLLSDEWILVNKDDETFYNKIQEKCKYSLEDIAISFQGIITGCDKAFILSKDDVKLNLVDDKFLKCWIKSKNINKYIVDKSEYRLIYSNDIDNENTNKRILDEIIGLYKTKLENRRECKSGIRKWYELQWGREKLFFERKKIMYPYKSNENRFAIDYDNNFSSADVYSFFIKEEYLDKFSYEYLVGILNSSVYDKYFKITAKKMSKNIYDYYPNKVMKIRIFRDNNYEEIENLSKQIISILLNKSIDKGKVEKLQIKMDNLIMDSLGI
1062	KAP1	MAASAAAASAAAASAASGSPGPGEGSAGGEKRSTAPSAAASASASAAASSPAGGGAEALELLEHCGVCRERLRPEREPRLLPCLHSACSACLGPAAPAAANSSGDGGAAGDGTVVDCPVCKQQCFSKDIVENYFMRDSGSKAATDAQDANQCCTSCEDNAPATSYCVECSEPLCETCVEAHQRVKYTKDHTVRSTGPAKSRDGERTVYCNVHKHEPLVLFCESCDTLTCRDCQLNAHKDHQYQFLEDAVRNQRKLLASLVKRLGDKHATLQKSTKEVRSSIRQVSDVQKRVQVDVKMAILQIMKELNKRGRVLVNDAQKVTEGQQERLERQHWTMTKIQKHQEHILRFASWALESDNNTALLLSKKLIYFQLHRALKMIVDPVEPHGEMKFQWDLNAWTKSAEAFGKIVAERPGTNSTGPAPMAPPRAPGPLSKQGSGSSQPMEVQEGYGFGSGDDPYSSAEPHVSGVKRSRSGEGEVSGLMRKVPRVSLERLDLDLTADSQPPVFKVFPGSTTEDYNLIVIERGAAAAATGQPGTAPAGTPGAPPLAGMAIVKEEETEAAIGAPPTATEGPETKPVLMALAEGPGAEGPRLASPSGSTSSGLEVVAPEGTSAPGGGPGTLDDSATICRVCQKPGDLVMCNQCEFCFHLDCHLPALQDVPGEEWSCSLCHVLPDLKEEDGSLSLDGADSTGVVAKLSPANQRKCERVLLALFCHEPCRPLHQLATDSTFSLDQPGGTLDLTLIRARLQEKLSPPYSSPQEFAQDVGRMFKQFNKLTEDKADVQSIIGLQRFFETRMNEAFGDTKFSAVLVEPPPMSLPGAGLSSQELSGGPGDGP
1063	MECP2	MVAGMLGLREEKSEDQDLQGLKDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEAGKAETSEGSGSAPAVPEASASPKQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRGSPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVKRVLEKSPGKLLVKMPFQTSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEAKKKAVKESSIRSVQETVLPIKKRKTRETVSIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRKSKESSPKGRSSSASSPPKKEHHHHHHHSESPKAPVPLLPPLPPPPPEPESSEDPTSPPEPQDLSSSVCKEEKMPRGGSLESDGCPKEPAKTQPAVATAATAAEKYKHRGEGERKDIVSSSMPRPNREEPVDSRTPVTERVS
1064	linker	SGSETPGTSESATPES
1065	linker	SGGS
1066	linker	SGGSSGSETPGTSESATPESSGGS
1067	linker	SGGSSGGSSGSETPGTSESATPESSGGSSGGS
1068	linker	GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS

SEQ	Description	Sequence
1069	XTEN linker (XTEN16)	SGSETPGTSESATPES
1070	XTEN linker	SGGSSGGSSGSETPGTSESATPES
1071	XTEN linker	SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS
1072	XTEN linker	SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS
1073	XTEN linker	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS
1074	NLS	PKKKRKV
1075	NLS	AVKRPAATKKAGQAKKKKLD
1076	NLS	MSRRRKANPTKLSENAKKLAKEVEN
1077	NLS	PAAKRVKLD
1078	NLS	KLKIKRPVK
1079	NLS	MDSLLMNRRKFLYQFKNVRWAKGRRETYLC
1092	XTEN linker (XTEN80)	GGPSSGAPPPSGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
1236	Plasmid for fusion protein with mRNA001	CGTCGATCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAAGGAGACCCAAGCTACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCC

SEQ	Description	Sequence
		ACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGTCCCGGCCAGGGGAACGGCCCTTCCAGTGTCGGATCTGCATGAGAAACTTTTCAAAGAAGTTCAATCTCCTTCAGCATACCCGGACCCACACTGGAGAGAAACCCTTTCAGTGCAGGATATGTATGCGGAATTTTTCCCGGCAAGATAATTTGAATTCCCATTTGAGAACACATACCGGGAGTCAGAAGCCTTTCCAATGCCGGATTTGCATGAGGAACTTCTCCCGAAGCCATAATTTGAAACTCCATACTAGAACACATACAGGCGAGAAGCCATTCCAGTGTAGGATCTGCATGCGCAATTTTAGCCAATCAACCACTCTTAAACGCCATCTGAGAACGCATACAGGTAGTCAGAAGCCTTTTCAGTGCAGGATCTGCATGAGGAATTTTAGTCGCAACACGAACTTGACTAGACACACAAGAACGCATACTGGAGAGAAGCCCTTTCAGTGTAGGATTTGTATGCGGAACTTCAGCATTAAACACAACCTGGCAAGGCATCTGAGGACTCATTTGCGCGGGTCTAGCCCCAAGAAGAAGAGAAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCCGGACCCTGGTGACATTCAAGGACGTGTTCGTGGACTTCACCCGGGAGGAGTGGAAGCTGCTGGACACAGCCCAGCAGATCGTGTACAGGAACGTGATGCTGGAGAACTATAAGAATCTGGTGTCTCTGGGCTACCAGCTGACAAAGCCAGATGTGATCCTGCGGCTGGAGAAGGGAGAGGAGCCCTGGCTGGTGTAGTCTAGAAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCCTGTTAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAGTGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAATCCAGCACAGTGGCGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCA

SEQ	Description	Sequence
		TGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
1237	Plasmid for fusion protein with mRNA002	CGTCGATCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGAGACCCAAGCTACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGTCCCGGCCAGGGGAACGGCCCTTCCAGTGTCGGATCTGCATGAGAAACTTTTCAAAGAAGTTCAATCTGCTTCAGCACACCCGGACCCACACTGGAGAGAAACCCTTTCAGTGCAGGATATGTATGCGGAATTTTTCCCGAAAAGATTACTTGATTAGCCACCTCCGAACACATACCGGGAGTCAGAAGCCTTTCCAATGCCGGATTTGCATGAGGAACTTCTCCAGGAGCCACAACCTTAAACTGCACACAAGAACACATACAGGCGAGAAGCCATTCCAGTGTAGGATCTGCATGCGCAATTTTAGCCAATCCACAACATTGAAAAGACATCTTCGGACGCATACAGGTAGTCAGAAGCCTTTTCAGTGCAGGATCTGCATGAGGAATTTTAGTCGACAAGATAATCTTGGCCGACATCTTCGAACGCATACTGGAGAGAAGCCCTTTCAGTGTAGGATTTGTATGCGGAACTTCAGCGTAGTAAACAACTTGAACAGACACTTGAAAACTCATTTGCGCGGGTCTAGCCCCAAGAAGAAGAGAAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCAC

SEQ	Description	Sequence
		CCCTGAGTCCCGGACCCTGGTGACATTCAAGGACGTGTTCGTGGACTTCACCCGGGAGGAGTGGAAGCTGCTGGACACAGCCCAGCAGATCGTGTACAGGAACGTGATGCTGGAGAACTATAAGAATCTGGTGTCTCTGGGCTACCAGCTGACAAAGCCAGATGTGATCCTGCGGCTGGAGAAGGGAGAGGAGCCCTGGCTGGTGTAGTCTAGAAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGCCGCCTCCCCGCCTGTTAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAGTGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAATCCAGCACAGTGGCGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
1238	Plasmid for fusion protein with mRNA0003	CGTCGATCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAG

SEQ	Description	Sequence
		AGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAAGGAGACCCAAGC TACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCG CAAGGTCAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGA GAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGT GCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTC TATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTC CGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCC CTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCG GCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAG ACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACAT CTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACA CAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGT GAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAA GGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCC CGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGG CTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCT GGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTT TGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGG ATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGA GGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTC TCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGA GGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGG AGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGA CGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAA TCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCAC CAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCG CAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAG GGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGT GAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGA GTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAA GGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACA CACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGC AAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCT GAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCC AGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCAT CAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAA GCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCC ACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTC CTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGA GGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGA GGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCAC AGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGTCCCG GCCAGGGGAACGGCCCTTCCAGTGTCGGATCTGCATGAGAAACTTTTCAAAAAAGTTTAA CCTTCTCCAACACACACGAACCCACACTGGAGAGAAACCCTTTCAGTGCAGGATATGTAT GCGGAATTTTTCCAGAAAAGATTATTTGATCAGTCATCTGCGAACACATACCGGGAGTCA GAAGCCTTTCCAATGCCGGATTTGCATGAGGAACTTCTCCAGGAGTCATAACCTCCGGTT GCACACACGCACACATACAGGCGAGAAGCCATTCCAGTGTAGGATCTGCATGCGCAATTT TAGCCAGAGTACGACCCTGAAGAGACATCTGCGGACGCATACAGGTAGTCAGAAGCCTTT TCAGTGCAGGATCTGCATGAGGAATTTTAGTCGGCAAGATAATTTGGGGAGACACTTGAG AACGCATACTGGAGAGAAGCCCTTTCAGTGTAGGATTTGTATGCGGAACTTCAGCGTTGT GAATAATTTGAATCGGCATCTCAAAACTCATTTGCGCGGGTCTAGCCCCAAGAAGAAGAG AAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCAC CCCTGAGTCCCGGACCCTGGTGACATTCAAGGACGTGTTCGTGGACTTCACCCGGGAGGA GTGGAAGCTGCTGGACACAGCCCAGCAGATCGTGTACAGGAACGTGATGCTGGAGAACTA TAAGAATCTGGTGTCTCTGGGCTACCAGCTGACAAAGCCAGATGTGATCCTGCGGCTGGA GAAGGGAGAGGAGCCCTGGCTGGTGTAGTCTAGAAATCAACCTCTGGATTACAAAATTTG TGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGC TTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTA TAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGT GGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCA GCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGC CTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTT GTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCG CGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGG CCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGAT CTCCCTTTGGGCCGCCTCCCCGCCTGTTAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAACTAGTGGCGCCTGATGCGGTATTTTCTCCTTACGCA TCTGTGCGGTATTTCACACCGCATAATCCAGCACAGTGGCGGCCCGTTTAAACCCGCTGA

SEQ	Description	Sequence
		TCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
1239	Plasmid for fusion protein with mRNA0004	CGTCGATCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAAGGAGACCCAAGCTACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCT

SEQ	Description	Sequence
		GGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGTCCCGGCCAGGGGAACGGCCCTTCCAGTGTCGGATCTGCATGAGAAACTTTTCACGACGCCACATTTTGGACAGACATACTCGGACCCACACTGGAGAGAAACCCTTTCAGTGCAGGATATGTATGCGGAATTTTTCCCGCCAGGACAACTTGGGGCGGCATCTGCGCACACATACCGGGAGTCAGAAGCCTTTCCAATGCCGGATTTGCATGAGGAACTTCTCCCAATCTACCACTCTTAAACGACACTTGCGCACACATACAGGCGAGAAGCCATTCCAGTGTAGGATCTGCATGCGCAATTTTAGCCGCCGGGACGGCCTGGCAGGGCACCTTAAGACGCATACAGGTAGTCAGAAGCCTTTTCAGTGCAGGATCTGCATGAGGAATTTTAGTGTTCATCATAACCTCGTTAGGCATCTGAGAACGCATACTGGAGAGAAGCCCTTTCAGTGTAGGATTTGTATGCGGAACTTCAGCATCAGTCACAATTTGGCGCGGCACCTTAAGACTCATTTGCGCGGGTCTAGCCCCAAGAAGAAGAGAAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCCGGACCCTGGTGACATTCAAGGACGTGTTCGTGGACTTCACCCGGGAGGAGTGGAAGCTGCTGGACACAGCCCAGCAGATCGTGTACAGGAACGTGATGCTGGAGAACTATAAGAATCTGGTGTCTCTGGGCTACCAGCTGACAAAGCCAGATGTGATCCTGCGGCTGGAGAAGGGAGAGGAGCCCTGGCTGGTGTAGTCTAGAAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGCCGCCTCCCCGCCTGTTAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAGTGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAATCCAGCACAGTGGCGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGC

SEQ	Description	Sequence
		TGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
1240	Plasmid for fusion protein with mRNA0005	CGTCGATCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAAGGAGACCCAAGCTACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGC

SEQ	Description	Sequence
		AAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCT GAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCC AGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCAT CAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAA GCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCC ACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTC CTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGA GGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGA GGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCAC AGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGTCCCG GCCAGGGGAACGGCCCTTCCAGTGTCGGATCTGCATGAGAAACTTTTCACGCCGGGAGGT ATTGGAAAACCATTTGCGAACCCACACTGGAGAGAAACCCTTTCAGTGCAGGATATGTAT GCGGAATTTTTCCCGGCGGGATAATCTCAATCGGCACTTGAAAACACATACCGGGAGTCA GAAGCCTTTCCAATGCCGGATTTGCATGAGGAACTTCTCCCAATCCACTACCCTCAAGCG ACATCTGCGGACACATACAGGCGAGAAGCCATTCCAGTGTAGGATCTGCATGCGCAATTT TAGCCGAAGGGATGGGCTGGCGGGCCATCTTAAGACGCATACAGGTAGTCAGAAGCCTTT TCAGTGCAGGATCTGCATGAGGAATTTTAGTGTCCATCACAACCTGGTCAGACACCTTAG GACGCATACTGGAGAGAAGCCCTTTCAGTGTAGGATTTGTATGCGGAACTTCAGCATATC ACATAACCTTGCCCGACACTTGAAGACTCATTTGCGCGGGTCTAGCCCCAAGAAGAAGAG AAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCAC CCCTGAGTCCCGGACCCTGGTGACATTCAAGGACGTGTTCGTGGACTTCACCCGGGAGGA GTGGAAGCTGCTGGACACAGCCCAGCAGATCGTGTACAGGAACGTGATGCTGGAGAACTA TAAGAATCTGGTGTCTCTGGGCTACCAGCTGACAAAGCCAGATGTGATCCTGCGGCTGGA GAAGGGAGAGGAGCCCTGGCTGGTGTAGTCTAGAAATCAACCTCTGGATTACAAAATTTG TGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGC TTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTA TAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGT GGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCA GCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGC CTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTT GTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCG CGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGG CCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGAT CTCCCTTTGGGCCGCCTCCCCGCCTGTTAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAACTAGTGGCGCCTGATGCGGTATTTTCTCCTTACGCA TCTGTGCGGTATTTCACACCGCATAATCCAGCACAGTGGCGGCCCGTTTAAACCCGCTGA TCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCT TCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCA TCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAG GGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCT GAGGCGGAAAGAACCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGC GTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGC GGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATA ACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCG CGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCT CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAA GCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTC TCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGT AGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCG CCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCT TGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGC TGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCG CTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAG AAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAG GGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAAT GAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCT TAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGAC TCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAA TGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCG GAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATT GTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCA TTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTT CCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCT TCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGG CAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTG AGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGG CGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA

SEQ	Description	Sequence
		AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
1241	Plasmid for fusion fusion protein with mRNA0006	CGTCGATCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAAGGAGACCCAAGCTACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGTCCCGGCCAGGGGAACGGCCCTTCCAGTGTCGGATCTGCATGAGAAACTTTTCACGCAGGGCAGTGTTGGATAGACATACCCGGACCCACACTGGAGAGAAACCCTTTCAGTGCAGGATATGTATGCGGAATTTTTCCCGACAAGATAATCTGGGGAGGCATCTGCGGACACATACCGGGAGTCAGAAGCCTTTCCAATGCCGGATTTGCATGAGGAACTTCTCCCAATCAACTACCCTGAAGCGACATCTGCGCACACATACAGGCGAGAAGCCATTCCAGTGTAGGATCTGCATGCGCAATTTTAGCCGCCGCGATGGGCTGGCTGGACACCTGAAGACGCATACAGGTAGTCAGAAGCCTTT

SEQ	Description	Sequence
		TCAGTGCAGGATCTGCATGAGGAATTTTAGTGTTCATCACAACTTGGTCCGACACCTTCGGACGCATACTGGAGAGAAGCCCTTTCAGTGTAGGATTTGTATGCGGAACTTCAGCATTTCACACAACCTCGCGCGCCACTTGAAAACTCATTTGCGCGGGTCTAGCCCCAAGAAGAAGAGAAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCCGGACCCTGGTGACATTCAAGGACGTGTTCGTGGACTTCACCCGGGAGGAGTGGAAGCTGCTGGACACAGCCCAGCAGATCGTGTACAGGAACGTGATGCTGGAGAACTATAAGAATCTGGTGTCTCTGGGCTACCAGCTGACAAAGCCAGATGTGATCCTGCGGCTGGAGAAGGGAGAGGAGCCCTGGCTGGTGTAGTCTAGAAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGCCGCCTCCCCGCCTGTTAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAGTGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAATCCAGCACAGTGGCGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
1242	Plasmid for fusion protein with mRNA0021	CGTCGATCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGT

SEQ	Description	Sequence
		ATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAAGGAGACCCAAGCTACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGTCCCGGCCAGGGGAACGGCCCTTCCAGTGTCGGATCTGCATGAGAAACTTTTCAAGAGCAGATAATCTGGGTCGGCACCTCCGCACCCACACTGGAGAGAAACCCTTTCAGTGCAGGATATGTATGCGGAATTTTTCCCGCAACACGCATCTCAGTTATCACCTTAAAACACATACCGGGAGTCAGAAGCCTTTCCAATGCCGGATTTGCATGAGGAACTTCTCCAGGGGCGACGGCTTGAGGCGGCATCTTCGCACACATACAGGCGAGAAGCCATTCCAGTGTAGGATCTGCATGCGCAATTTTAGCCGCAGAGACAATTTGAACAGACATCTCAAAACGCATACAGGTAGTCAGAAGCCTTTTCAGTGCAGGATCTGCATGAGGAATTTTAGTCGAGCAAGAAACTTGACGCTGCACACCCGGACGCATACTGGAGAGAAGCCCTTTCAGTGTAGGATTTGTATGCGGAACTTCAGCGACCCTTCATCTTTGAAGCGCCATCTTCGCACTCATTTGCGCGGGTCTAGCCCCAAGAAGAAGAGAAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCCGGACCCTGGTGACATTCAAGGACGTGTTCGTGGACTTCACCCGGGAGGAGTGGAAGCTGCTGGACACAGCCCAGCAGATCGTGTACAGGAACGTGATGCTGGAGAACTATAAGAATCTGGTGTCTCTGGGCTACCAGCTGACAAAGCCAGATGTGATCCTGCGGCTGGAGAAGGGAGAGGAGCCCTGGCTGGTGTAGTCTAGAAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGG

SEQ	Description	Sequence
		CCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCCTGTTAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAGTGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAATCCAGCACAGTGGCGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAGTGCCACCTGA
1243	Plasmid for fusion protein with mRNA0037	CGTCGATCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAAGGAGACCCAAGCTACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGT

SEQ	Description	Sequence
		GAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGTCCCGGCCAGGGGAACGGCCCTTCCAGTGTCGGATCTGCATGAGAAACTTTTCAAGAGTGGATCATCTCCATCGACACCTCCGGACCCACACTGGAGAGAAACCCTTTCAGTGCAGGATATGTATGCGGAATTTTTCCCGGAGGGAACATTTGTCCGGACATCTCAAGACACATACCGGGGGAGGCGGTAGTCAGAAGCCTTTCCAATGCCGGATTTGCATGAGGAACTTCTCCCAAAGTTCCAGCCTCGTCCGCCATCTTCGCACACATACAGGCGAGAAGCCATTCCAGTGTAGGATCTGCATGCGCAATTTTAGCCGCAAGGAGCGATTGGCAACCCACCTCAAGACGCATACAGGTAGTCAGAAGCCTTTTCAGTGCAGGATCTGCATGAGGAATTTTAGTGTCGCACATAACCTCACAAGGCATCTGCGCACGCATACTGGAGAGAAGCCCTTTCAGTGTAGGATTTGTATGCGGAACTTCAGCATTAGTCATAACCTGGCAAGGCATCTCAAAACTCATTTGCGCGGGTCTAGCCCCAAGAAGAAGAGAAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCCGGACCCTGGTGACATTCAAGGACGTGTTCGTGGACTTCACCCGGGAGGAGTGGAAGCTGCTGGACACAGCCCAGCAGATCGTGTACAGGAACGTGATGCTGGAGAACTATAAGAATCTGGTGTCTCTGGGCTACCAGCTGACAAAGCCAGATGTGATCCTGCGGCTGGAGAAGGGAGAGGAGCCCTGGCTGGTGTAGTCTAGAAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCCTGTTAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAGTGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAATCCAGCACAGTGGCGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCA

SEQ	Description	Sequence
		GTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
1244	Plasmid for fusion protein with mRNA0038	CGTCGATCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAAGGAGACCCAAGCTACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGT

SEQ	Description	Sequence
		GAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGTCCCGGCCAGGGGAACGGCCCTTCCAGTGTCGGATCTGCATGAGAAACTTTTCACGCAAGCACCACCTTGGGAGACATACCAGAACCCACACTGGAGAGAAACCCTTTCAGTGCAGGATATGTATGCGGAATTTTTCCCGACGGGAACACCTCACGATTCATTTGCGGACACATACCGGGGGAGGCGGTAGTCAGAAGCCTTTCCAATGCCGGATTTGCATGAGGAACTTCTCCCAGAGCTCATCTCTCGTGCGGCACCTGCGGACACATACAGGCGAGAAGCCATTCCAGTGTAGGATCTGCATGCGCAATTTTAGCCGGAAGGAGCGATTGGCGACGCACCTGAAAACGCATACAGGTAGTCAGAAGCCTTTTCAGTGCAGGATCTGCATGAGGAATTTTAGTGTAGCCCACAACCTGACTAGGCATTTGAGGACGCATACTGGAGAGAAGCCCTTTCAGTGTAGGATTTGTATGCGGAACTTCAGCATTTCTCACAATCTCGCGCGACATTTGAAAACTCATTTGCGCGGGTCTAGCCCCAAGAAGAAGAGAAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCCGGACCCTGGTGACATTCAAGGACGTGTTCGTGGACTTCACCCGGGAGGAGTGGAAGCTGCTGGACACAGCCCAGCAGATCGTGTACAGGAACGTGATGCTGGAGAACTATAAGAATCTGGTGTCTCTGGGCTACCAGCTGACAAAGCCAGATGTGATCCTGCGGCTGGAGAAGGGAGAGGAGCCCTGGCTGGTGTAGTCTAGAAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCCTGTTAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAGTGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAATCCAGCACAGTGGCGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGG

SEQ	Description	Sequence
		TTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA
1245	Plasmid for fusion protein with mRNA0039	CGTCGATCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAAGGAGACCCAAGCTACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCAATCACGATCAGGAGTTCGACCCCCCTAAGGTGTACCCACCAGTGCCTGCAGAGAAGAGGAAGCCAATCCGGGTGCTGAGCCTGTTTGATGGCATCGCCACCGGCCTGCTGGTGCTGAAGGATCTGGGCATCCAGGTGGACCGGTACATCGCCTCCGAGGTGTGCGAGGATTCTATCACCGTGGGCATGGTGCGCCACCAGGGCAAGATCATGTATGTGGGCGACGTGCGGTCCGTGACACAGAAGCACATCCAGGAGTGGGGCCCATTCGATCTGGTGATCGGCGGCAGCCCCTGTAATGACCTGTCCATCGTGAACCCTGCAAGGAAGGGACTGTACGAGGGAACCGGCCGGCTGTTCTTTGAGTTTTATAGACTGCTGCACGACGCCAGGCCTAAGGAGGGCGACGATAGACCATTCTTTTGGCTGTTCGAGAATGTGGTGGCTATGGGCGTGAGCGATAAGAGGGACATCTCCAGGTTTCTGGAGTCTAACCCCGTGATGATCGATGCAAAGGAGGTGTCCGCCGCACACAGAGCCAGGTATTTCTGGGGCAATCTGCCAGGAATGAACAGGCCACTGGCAAGCACCGTGAATGACAAGCTGGAGCTGCAGGAGTGCCTGGAGCACGGAAGGATCGCCAAGTTTTCCAAGGTGCGCACAATCACCACACGGAGCAATTCCATCAAGCAGGGCAAGGATCAGCACTTCCCCGTGTTCATGAACGAGAAGGAGGACATCCTGTGGTGTACCGAGATGGAGAGAGTGTTCGGCTTTCCAGTGCACTACACAGACGTGTCTAACATGAGCAGGCTGGCAAGGCAGCGGCTGCTGGGCAGATCTTGGAGCGTGCCCGTGATCAGGCACCTGTTCGCCCCTCTGAAGGAGTATTTTGCCTGCGTGAGCAGCGGCAACTCCAATGCCAACAGCCGGGGCCCCTCTTTCAGCTCCGGATTGGTGCCTCTGAGCCTGAGGGGCTCCCACATGGCAGCAATCCCCGCCCTGGACCCCGAGGCCGAGCCTAGCATGGACGTGATCCTGGTGGGCTCTAGCGAGCTGTCCTCTAGCGTGTCTCCAGGAACCGGAAGGGATCTGATCGCATACGAGGTGAAGGCCAATCAGCGGAACATCGAGGACATCTGTATCTGCTGTGGCAGCCTGCAGGTGCACACACAGCACCCACTGTTCGAGGGAGGAATCTGCGCACCCTGTAAGGATAAGTTCCTGGACGCCCTGTTTCTGTACGACGATGACGGCTACCAGTCCTATTGCTCTATCTGCTGTTCCGGCGAGACCCTGCTGATCTGCGGCAATCCAGATTGTACAAGGTGCTATTGTTTTGAGTGCGTGGACTCTCTGGTGGGACCAGGCACCAGCGGAAAGGTGCACGCCATGTCCAACTGGGTGTGCTACCTGTGCCTGCCATCCTCTCGCAGCGGACTGCTGCAGCGGAGAAGGAAGTGGAGATCCCAGCTGAAGGCCTTCTATGATAGGGAGTCTGAGAACCCCCTGGAGATGTTTGAGACCGTGCCAGTGTGGCGCCGGCAGCCCGTGAGGGTGCTGAGCCTGTTCGAGGATATCAAGAAGGAGCTGACATCCCTGGGCTTTCTGGAGTCCGGCTCTGACCCCGGACAGCTGAAGCACGTGGTGGATGTGACCGACACAGTGCGGAAGGATGTGGAGGAGTGGGGCCCTTTCGACCTGGTGTACGGAGCAACCCCTCCACTGGGACACACATGCGACAGACCCCCTTCTTGGTACCTGTTCCAGTTTCACCGCCTGCTGCAGTATGCAAGGCCAAAGCCAGGCAGCCCTAGACCATTCTTTTGGATGTTCGTGGATAATCTGGTGCTGAACAAGGAGGATCTGGACGTGGCCAGCAGGTTTCTGGAGATGGAGCCAGTGACCATCCCAGACGTGCACGGCGGCTCCCTGCAGAATGCCGTGCGCGTGTGGTCTAACATCCCTGCCATCAGAAGCAGGCACTGGGCACTGGTGAGCGAGGAGGAGCTGTCCCTGCTGGCCCAGAATAAGCAGAGCAGCAAGCTGGCCGCCAAGTGGCCTACAAAGCTGGTGAAGAACTGCTTCCTGCCACTGCGGGAGTACTTCAAGTATTTTTCCACCGAGCTGACATCTAGCCTGGGAGGACCCTCCTCTGGCGCCCCACCACCTAGCGGCGGCTCCCCTGCCGGCTCTCCAACCAGCACAGAGGAGGGCACCAGCGAGTCCGCCACACCAGAGTCTGGACCTGGCACCAGCACAGAGCCATCCGAGGGCTCTGCCCCAGGCTCTCCTGCAGGCAGCCCTACCTCCACCGAAGAGGGCACCAGCACAGAGCCTTCTGAGGGCAGCGCCCCAGGCACCTCTACAGAGCCAAGCGAGCTCGAGTCCCGGCCAGGGGAACGGCCCTTCCAGTGTCGGATCTGCATGAGAAACTTTTCACGAGTCGATCACCTCCACCGCCACCTGCGAACCCACACTGGAGAGAAACCCTTTCAGTGCAGGATATGTAT

SEQ	Description	Sequence
		GCGGAATTTTTCCAGGTCCGACCACCTCAGCTTGCACTTGAAGACACATACCGGGGGAGGCGGTAGTCAGAAGCCTTTCCAATGCCGGATTTGCATGAGGAACTTCTCCCAATCTAGTTCATTGGTACGACATCTTAGGACACATACAGGCGAGAAGCCATTCCAGTGTAGGATCTGCATGCGCAATTTTAGCCGAAAAGAGCGGCTGGCGACCCACTTGAAAACGCATACAGGTAGTCAGAAGCCTTTTCAGTGCAGGATCTGCATGAGGAATTTTAGTGTAGCGCATAACTTGACACGGCACTTGCGCACGCATACTGGAGAGAAGCCCTTTCAGTGTAGGATTTGTATGCGGAACTTCAGCATTTCCCATAATCTGGCGCGGCACCTGAAGACTCATTTGCGCGGGTCTAGCCCCAAGAAGAAGAGAAAGGTGGGAGTCGACGGATCCAGCGGCTCCGAGACCCCAGGCACATCTGAGAGCGCCACCCCTGAGTCCCGGACCCTGGTGACATTCAAGGACGTGTTCGTGGACTTCACCCGGGAGGAGTGGAAGCTGCTGGACACAGCCCAGCAGATCGTGTACAGGAACGTGATGCTGGAGAACTATAAGAATCTGGTGTCTCTGGGCTACCAGCTGACAAAGCCAGATGTGATCCTGCGGCTGGAGAAGGGAGAGGAGCCCTGGCTGGTGTAGTCTAGAAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCCTGTTAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAGTGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAATCCAGCACAGTGGCGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGA

CLAIMS

1. A method of modifying an epigenetic state of a hepatitis B virus (HBV) gene or genome, comprising contacting the HBV gene or genome with an epigenetic editing system,

   wherein the epigenetic editing system comprises

   a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or

   one or more nucleic acid molecules encoding thereof,

   wherein the first DNA binding domain binds a first target region of the HBV gene or genome, and

   wherein the contacting results in a reduction of:

   number of HBV viral episomes,

   replication of the HBV gene or genome, or

   expression of a protein product encoded by the HBV gene or genome,

   wherein said reduction is at least about 20% compared to contacting the HBV gene or genome with a suitable control.

2. A method of treating an HBV infection in a subject comprising administering an epigenetic editing system to the subject,

   wherein the epigenetic editing system comprises

   a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or

   one or more nucleic acid molecules encoding thereof,

   wherein the first DNA binding domain binds a first target region of a HBV gene or genome, and

   wherein the contacting results in a reduction of:

   number of HBV viral episomes,

   replication of the HBV gene or genome, or

   expression of a protein product encoded by the HBV gene or genome,

   wherein said reduction is at least about 20% compared to administering a suitable control.

3. A method of modulating expression of an HBV gene or genome comprising contacting the HBV gene or genome with an epigenetic editing system,

wherein the epigenetic editing system comprises

a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or

one or more nucleic acid molecules encoding thereof,

wherein the first DNA binding domain binds a first target region of the HBV gene or genome, and

wherein the contacting results in a reduction of expression of a gene product encoded by the HBV gene or genome, optionally, wherein the gene product is a nucleic acid or a protein,

wherein said reduction is at least about 20% compared to contacting the HBV genome with a suitable control.

4. A method of inhibiting viral replication in a cell infected with an HBV comprising administering an epigenetic editing system,

   wherein the epigenetic editing system comprises

   a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or

   one or more nucleic acid molecules encoding thereof,

   wherein the first DNA binding domain binds a first target region of a HBV gene or genome, and wherein the epigenetic editing system targets a target region of the HBV gene or genome, and

   wherein the contacting results in a reduction of number of HBV viral episomes or replication of the HBV gene or genome,

   wherein said reduction is at least about 20% compared to administering a suitable control.
5. The method of any one of claims 1-4, wherein the HBV genome is a covalently closed circular DNA (cccDNA) or an HBV integrated DNA.
6. The method of any one of claims 1-5, wherein the HBV genome comprises HBV genotype A, HBV genotype B, HBV genotype C, HBV genotype D, HBV genotype E, HBV genotype F, HBV genotype G or HBV genotype H.
7. The method of any one of claims 1-6, wherein the HBV genome comprises a sequence with at least 80% identity to an HBV genome sequence provided herein.
8. The method of claim 7, wherein the first target region is located in a region of the HBV genome within nucleotide 0-303, 1000-2448 or 2802-3182 of an HBV genome provided herein.

9. The method of any one of claims 1-6, wherein the first target region of the HBV genome is located in a CpG island.
10. The method of any one of claims 1-6, wherein the first target region of the HBV genome is located in a promotor.
11. The method of any one of claims 1-6, wherein the first target region of the HBV genome is located in a section of the HBV genome that encodes a transcript selected from the group consisting of a pgRNA, a precure mRNA, a preS mRNA, a S mRNA, and a X mRNA.
12. The method of any one of claims 1-11, wherein the first DNA binding domain comprises a CRISPR-Cas protein.
13. The method of claim 12, wherein the epigenetic editing system further comprises a first guide RNA (gRNA) that comprises a region complementary to a strand of the first target region.
14. The method of claim 13, wherein the gRNA comprises a sequence selected from a gRNA provided herein, e.g., in Table 14 and/or 15.
15. The method of any one of claims 1-11, wherein the first DNA binding domain comprises a zinc-finger protein.
16. The method of claim 15, wherein the zinc-finger protein comprises a zinc-finger motif with a sequence selected from any zinc finger or zinc finger motif provided herein, e.g., in Table 1.
17. The method of claim 15 or 16, wherein the zinc-finger protein comprises a sequence of any of the zinc finger epigenetic repressors provided herein.
18. The method of any one of claims 1-17, wherein the transcriptional repressor domain comprises ZIM3.
19. The method of any one of claims 1-18, wherein the first DNMT domain is a DNMT3A domain or a DNMT3L domain.
20. The method of claim 19, wherein the first DNMT domain comprises a sequence of a DNMT domain provided herein.
21. The method of any one of claims 1-20, wherein the epigenetic editing system further comprises a second DNMT domain or a nucleic acid encoding thereof.
22. The method of claims 21, wherein the second DNMT domain is a DNMT3A domain or a DNMT3L domain.
23. The method of claim 22, wherein the second DNMT domain comprises a sequence of a DNMT domain provided herein.
24. The method of any one of claims 21-23, wherein the epigenetic editing system comprises a fusion protein or a nucleic acid encoding thereof, and wherein the fusion protein comprises

the first DNA binding domain, the first DNMT domain, the repressor domain and the second DNMT domain.

25. The method of claim 24, wherein the fusion protein further comprises a nuclear localization sequence (NLS).
26. The method of claim 25, wherein the fusion protein comprises a sequence of a fusion protein provided herein.
27. The method of any one of claims 1-21, wherein the epigenetic editing system further comprises a second DNA binding domain or a nucleic acid encoding thereof, wherein the second DNA binding domain binds a second target region of the HBV genome.
28. The method of claim 27, wherein the second target region is located in a region of the HBV genome within nucleotide 0-303, 1000-2448 or 2802-3182.
29. The method of claim 27, wherein the second target region of the HBV genome is located in a CpG island.
30. The method of claim 27, wherein the second target region of the HBV genome is located in a promotor.
31. The method of claim 27, wherein the second target region of the HBV genome is located in a section of the HBV genome that encodes a transcript selected from the group consisting of a pgRNA, a precure mRNA, a preS mRNA, a S mRNA, and a X mRNA.
32. The method of any one of claims 27-31, wherein the second DNA binding domain comprises a CRISPR-Cas protein.
33. The method of claim 32, wherein the epigenetic editing system further comprises a second gRNA that comprises a region complementary to a strand of the second target region.
34. The method of claim 33, wherein the gRNA comprises a sequence selected from a gRNA sequence provided herein, e.g., a sequence provided in Table 14 and/or 15.
35. The method of any one of claims 27-31, wherein the second DNA binding domain comprises a zinc-finger protein.
36. The method of claim 35, wherein the zinc-finger protein comprises a zinc-finger motif with a sequence selected from a zinc finger motif sequence provided herein, e.g., a zinc finger motif provided in Table 1.
37. The method of claim 35 or 36, wherein the zinc-finger protein comprises a sequence of a zinc finger motif provided in Table 1.
38. The method of any one of claims 27-37, wherein the epigenetic editing system comprises a first fusion protein or a first nucleic acid encoding thereof and a second fusion protein or a second nucleic acid encoding thereof,

wherein the first fusion protein comprises the first DNA binding domain and the first DNMT domain, and

wherein the second fusion protein comprises the second DNA binding domain and the transcriptional repressor domain.

39. The method of claim 38, wherein the first fusion protein comprises a sequence of a fusion protein provided herein.
40. The method of claim 38, wherein the second fusion protein comprises a sequence of a fusion protein provided herein.
41. The method of any one of claims 38-40, wherein the epigenetic editing system further comprises a third DNA binding domain or a nucleic acid encoding thereof, wherein the third DNA binding domain binds to a third target region of the HBV genome.
42. The method of claim 41, wherein the third target region is located in a region of the HBV genome within nucleotide 0-303, 1000-2448 or 2802-3182.
43. The method of claim 41, wherein the third target region of the HBV genome is located in a CpG island.
44. The method of claim 41, wherein the third target region of the HBV genome is located in a promotor.
45. The method of claim 41, wherein the third target region of the HBV genome is located in a section of the HBV genome that encodes a transcript selected from the group consisting of a pgRNA, a precure mRNA, a preS mRNA, a S mRNA, and a X mRNA.
46. The method of any one of claims 41-45, wherein the third DNA binding domain comprises a CRISPR-Cas protein.
47. The method of claim 46, wherein the epigenetic editing system further comprises a third gRNA that comprises a region complementary to a strand of the third target region.
48. The method of claim 47, wherein the third gRNA comprises a sequence selected from a gRNA sequence provided herein, e.g., of a gRNA sequence provided in Table 14 and/or 15.
49. The method of any one of claims 41-45, wherein the third DNA binding domain comprises a zinc-finger protein.
50. The method of claim 49, wherein the zinc-finger protein comprises a zinc-finger motif with a sequence selected from a zinc finger motif provided herein.
51. The method of claim 49 or 50, wherein the zinc-finger protein comprises a sequence of a zinc finger motif provided in Table 1.
52. The method of any one of claims 41-51, wherein the epigenetic editing system further comprises a second DNMT domain or a nucleic acid encoding thereof.

53. The method of claim 52, wherein the second DNMT domain is a DNMT3A domain or a DNMT3L domain.
54. The method of claim 53, wherein the epigenetic editing system comprises a third fusion protein or a nucleic acid encoding thereof, wherein the third fusion protein comprises the third DNA binding domain and the second DNMT domain.
55. The method of claim 54, wherein the third fusion protein comprises a sequence of a fusion protein provided herein.
56. An epigenetic editing system comprising:

    a fusion protein or a nucleic acid encoding the fusion protein,

    wherein the fusion protein comprises:

    (a) a DNA-binding domain that binds a target region of a HBV gene or genome,

    (b) a first DNA methyltransferase (DNMT) domain, and

    (c) a transcriptional repressor domain.
57. The epigenetic system of claim 56, wherein the epigenetic editing system is capable of reducing a number of the HBV viral episome, replication of the HBV, or expression of a gene product encoded by the HBV gene or genome, wherein said reduction is at least about 20% compared to contacting the HBV gene or genome with a suitable control.
58. The epigenetic system of claim 56 or 57, wherein the HBV genome is a covalently closed circular DNA (cccDNA) or an HBV integrated DNA.
59. The epigenetic system of any one of claims 56-58, wherein the HBV genome comprises HBV genotype A, HBV genotype B, HBV genotype C, HBV genotype D, HBV genotype E, HBV genotype F, HBV genotype G or HBV genotype H.
60. The epigenetic system of any one of claims 56-59, wherein the HBV genome comprises a sequence with at least 80% identity to an HBV genome sequence provided herein.
61. The epigenetic system of any one of claims 56-60, wherein the target region is located in a region of the HBV genome within nucleotide 0-303, 1000-2448 or 2802-3182 of an HBV genome sequence provided herein.
62. The epigenetic system of any one of claims 56-60, wherein the target region of the HBV genome is located in a CpG island.
63. The epigenetic system of any one of claims 56-60, wherein the target region of the HBV genome is located in a promotor.
64. The epigenetic system of any one of claims 56-60, wherein the target region of the HBV genome is located in a section of the HBV genome that encodes a transcript selected from

the group consisting of a pgRNA, a precure mRNA, a preS mRNA, a S mRNA, and a X mRNA.

65. The epigenetic system of claims 56-64, wherein the DNA binding domain comprises a CRISPR-Cas protein.
66. The epigenetic system of claim 65, wherein the epigenetic editing system further comprises a gRNA that comprises a region complementary to a strand of the target region.
67. The epigenetic system of claim 66, wherein the gRNA comprises a sequence selected from a gRNA sequence provided herein, e.g., in Table 14 and/or 15.
68. The epigenetic system of any one of claims 56-64, wherein the DNA binding domain comprises a zinc-finger protein.
69. The epigenetic system of claim 68, wherein the zinc-finger protein comprises a zinc-finger motif with a sequence selected from a zinc finger motif provided herein.
70. The epigenetic system of claim 68 or 69, wherein the zinc-finger protein comprises a sequence of a zinc finger motif provided in Table 1.
71. The epigenetic system of any one of claims 56-70, wherein the transcriptional repressor domain comprises a sequence of a transcriptional repressor provided herein.
72. The epigenetic system of any one of claims 56-71, wherein the first DNMT domain is a DNMT3A domain or a DNMT3L domain.
73. The epigenetic system of claim 72, wherein the DNMT domain comprises a sequence of a DNMT domain provided herein.
74. The epigenetic system of any one of claims 56-73, wherein the fusion protein further comprises a second DNMT domain.
75. The epigenetic system of claim 74, wherein the second DNMT domain is a DNMT3A domain or a DNMT3L domain.
76. The epigenetic system of any one of claims 56-75, wherein the fusion protein further comprises a nuclear localization sequence (NLS).
77. The epigenetic system of claim 76, wherein the fusion protein comprises a sequence of a fusion protein provided herein.
78. An epigenetic editing system comprising:

    a first fusion protein or a nucleic acid encoding the first fusion protein, wherein the first fusion protein comprises a first DNA binding domain and a first DNMT domain, wherein the first DNA binding domain binds a first target region of a HBV genome, and

    a second fusion protein or a nucleic acid encoding the second fusion protein, wherein the second fusion protein comprises a second DNA binding domain and a transcriptional repressor domain, wherein the second DNA binding domain binds a second target region of the HBV genome.

79. The epigenetic system of claim 78, wherein the epigenetic editing system is capable of reducing a number of the HBV viral episome, replication of the HBV, or expression of a gene product encoded by the HBV genome, wherein said reduction is at least about 20% compared to contacting the HBV genome with a suitable control.
80. The epigenetic system of claim 78 or 79, wherein the HBV genome is a covalently closed circular DNA (cccDNA) or an HBV integrated DNA.
81. The epigenetic system of any one of claims 78-80, wherein the HBV genome comprises HBV genotype A, HBV genotype B, HBV genotype C, HBV genotype D, HBV genotype E, HBV genotype F, HBV genotype G or HBV genotype H.
82. The epigenetic system of any one of claims 78-81, wherein the HBV genome comprises a sequence with at least 80% identity to an HBV genome provided herein.
83. The epigenetic system of any one of claims 78-81, further comprising a third fusion protein or a nucleic acid encoding the third fusion protein, wherein the third fusion protein comprises a third DNA binding domain and a second DNMT domain, wherein the third DNA binding domain binds a third target region of the HBV genome.
84. The epigenetic system of claim 83, wherein the first target region, the second target region or the third target region is located in a region of the HBV genome within nucleotide 0-303, 1000-2448 or 2802-3182 of an HBV genome provided herein.
85. The epigenetic system of claim 83, wherein the first target region, the second target region or the third target region of the HBV genome is located in a CpG island.
86. The epigenetic system of claim 83, wherein the first target region, the second target region or the third target region of the HBV genome is located in a promotor.
87. The epigenetic system of claim 83, wherein the first target region, the second target region or the third target region of the HBV genome is located in a section of the HBV genome that encodes a transcript selected from the group consisting of a pgRNA, a precure mRNA, a preS mRNA, a S mRNA, and a X mRNA.
88. The epigenetic system of claim 83, wherein the first DNA binding domain, the second DNA binding domain or the third DNA binding domain comprises a CRISPR-Cas protein.
89. The epigenetic system of claim 88, wherein the epigenetic editing system further comprises a first gRNA that comprises a region complementary to a strand of the first target region, a second gRNA that comprises a region complementary to a strand of the second target region or a third RNA that comprises a region complementary to a strand of the third target region.

90. The epigenetic system of claim 89, wherein the first gRNA comprises a sequence selected from a gRNA sequence provided herein, e.g., provided in Table 14 and/or 15, the second gRNA comprises a sequence selected from a gRNA sequence provided herein, e.g., provided in Table 14 and/or 15, and/or the third gRNA comprises a sequence selected from a gRNA sequence provided herein, e.g., provided in Table 14 and/or 15.
91. The epigenetic system of claim 83, wherein the first DNA binding domain, the second DNA binding domain or the third DNA binding domain comprises a zinc-finger protein.
92. The epigenetic system of claim 91, wherein the zinc-finger protein comprises a zinc-finger motif with a sequence selected from a zinc finger motif provided herein.
93. The epigenetic system of claim 91 or 92, wherein the zinc-finger protein comprises a sequence of a zinc finger motif provided in Table 1.
94. The epigenetic system of any one of claims 78-93, wherein the transcriptional repressor domain comprises ZIM3.
95. The epigenetic system of any one of claims 78-94, wherein the first DNMT domain is a DNMT3A domain or a DNMT3L domain.
96. The epigenetic system of claim 95, wherein the first DNMT domain comprises a sequence of a DNMT provided herein.
97. The epigenetic system of claim 83, wherein the second DNMT domain is a DNMT3A domain or a DNMT3L domain.
98. The epigenetic system of claim 97, wherein the second DNMT domain comprises a sequence of a DNMT domain provided herein.
99. The epigenetic system of any one of claim 78-98, wherein the first fusion protein comprises a sequence of a fusion protein provided herein.
100. The epigenetic system of any one of claims 78-99, wherein the second fusion protein comprises a sequence of a fusion protein provided herein.
101. The epigenetic system of any one of claims 83-99, wherein the third fusion protein comprises a sequence of a fusion protein provided herein.
102. The method of any one of claims 1-55, wherein the epigenetic editing system comprises a nucleic acid sequence provided in Table 20.
103. A method of treating an HDV infection in a subject comprising administering an epigenetic editing system to the subject,

wherein the epigenetic editing system comprises

a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or one or more nucleic acid molecules encoding thereof,

wherein the first DNA binding domain binds a first target region of a HBV gene or genome, and

wherein the contacting results in a reduction of:

number of HDV viral episomes,

replication of the HDV gene or genome, or

expression of a protein product encoded by the HDV gene or genome,

wherein said reduction is at least about 20% compared to administering a suitable control.

104. A method of inhibiting viral replication in a cell infected with an HDV comprising administering an epigenetic editing system,

wherein the epigenetic editing system comprises

a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or

one or more nucleic acid molecules encoding thereof,

wherein the first DNA binding domain binds a first target region of a HBV gene or genome, and wherein the epigenetic editing system targets a target region of the HBV gene or genome, and

wherein the contacting results in a reduction of number of HDV viral episomes or replication of the HDV gene or genome,

wherein said reduction is at least about 20% compared to administering a suitable control.

105. The method of claim 103 or 104, wherein the first DNA binding domain comprises a CRISPR-Cas protein.

106. The method of claim 105, wherein the epigenetic editing system further comprises a first guide RNA (gRNA) that comprises a region complementary to a strand of the first target region.

107. The method of claim 106, wherein the gRNA comprises a sequence selected from a gRNA provided herein, e.g., in Table 14 and/or 15.

108. The method of claim 103 or 104, wherein the first DNA binding domain comprises a zinc-finger protein.

109. The method of claim 108, wherein the zinc-finger protein comprises a zinc-finger motif with a sequence selected from any zinc finger or zinc finger motif provided herein, e.g., in Table 1.

110. The method of claim 108 or 109, wherein the zinc-finger protein comprises a sequence of any of the zinc finger epigenetic repressors provided herein.

111. The method of any one of claims 103-110, wherein the transcriptional repressor domain comprises ZIM3.

112. The method of any one of claims 103-111, wherein the first DNMT domain is a DNMT3A domain or a DNMT3L domain.

113. The method of claim 112, wherein the first DNMT domain comprises a sequence of a DNMT domain provided herein.

114. The method of any one of claims 103-113, wherein the epigenetic editing system further comprises a second DNMT domain or a nucleic acid encoding thereof.

115. The method of claim 114, wherein the second DNMT domain is a DNMT3A domain or a DNMT3L domain.

116. The method of claim 115, wherein the second DNMT domain comprises a sequence of a DNMT domain provided herein.

117. The method of any one of claims 114-116, wherein the epigenetic editing system comprises a fusion protein or a nucleic acid encoding thereof, and wherein the fusion protein comprises the first DNA binding domain, the first DNMT domain, the repressor domain and the second DNMT domain.

118. The method of claim 117, wherein the fusion protein further comprises a nuclear localization sequence (NLS).

119. The method of claim 118, wherein the fusion protein comprises a sequence of a fusion protein provided herein.

120. The method of any one of claims 103-119, wherein the first DNA binding domain binds a target region of a HBV gene or genome encoding or controlling expression of an S-antigen.

121. A method comprising administering an epigenetic editing system to a subject in need thereof,

wherein the epigenetic editing system comprises

a first DNA binding domain, a first DNMT domain, and a transcriptional repressor domain or

one or more nucleic acid molecules encoding thereof, wherein the first DNA binding domain binds a first target region of a HBV gene or genome, and

wherein the contacting results in a reduction of:

number of HBV viral episomes,

replication of the HBV gene or genome, or

expression of a protein product encoded by the HBV gene or genome,

wherein said reduction is at least about 20% compared to administering a suitable control.

122. The method of any one of claims 103-121 wherein the epigenetic editing system comprises a nucleic acid sequence provided in Table 20.

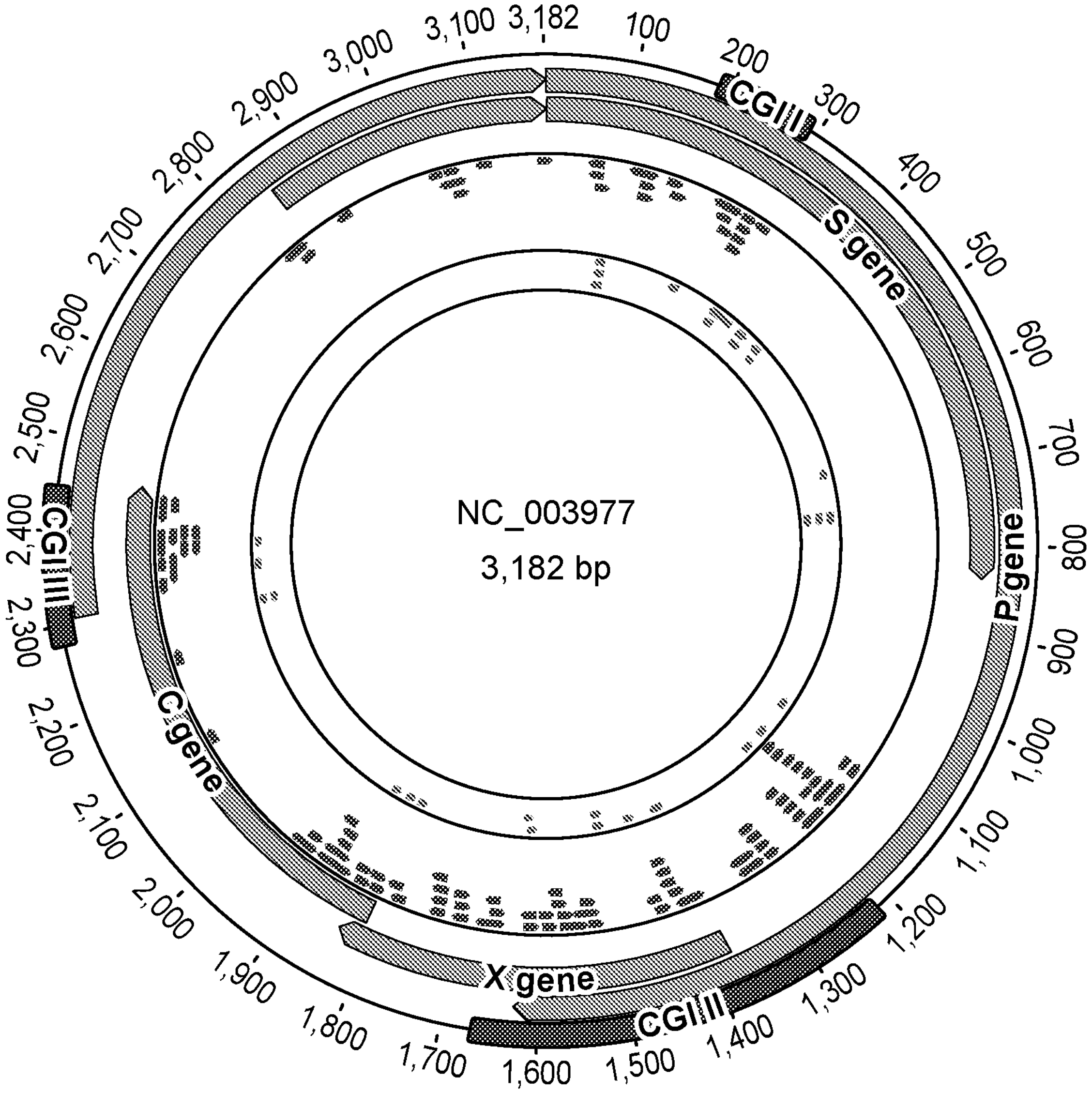

NC_003977
3,182 bp
3,182
100
200
300
400
500
600
700
800
900
1,000
1,100
1,200
1,300
1,400
1,500
1,600
1,700
1,800
1,900
2,000
2,100
2,200
2,300
2,400
2,500
2,600
2,700
2,800
2,900
3,000
3,100
CGI I
S gene
P gene
CGI II
X gene
C gene
CGI III

FIGURE 1

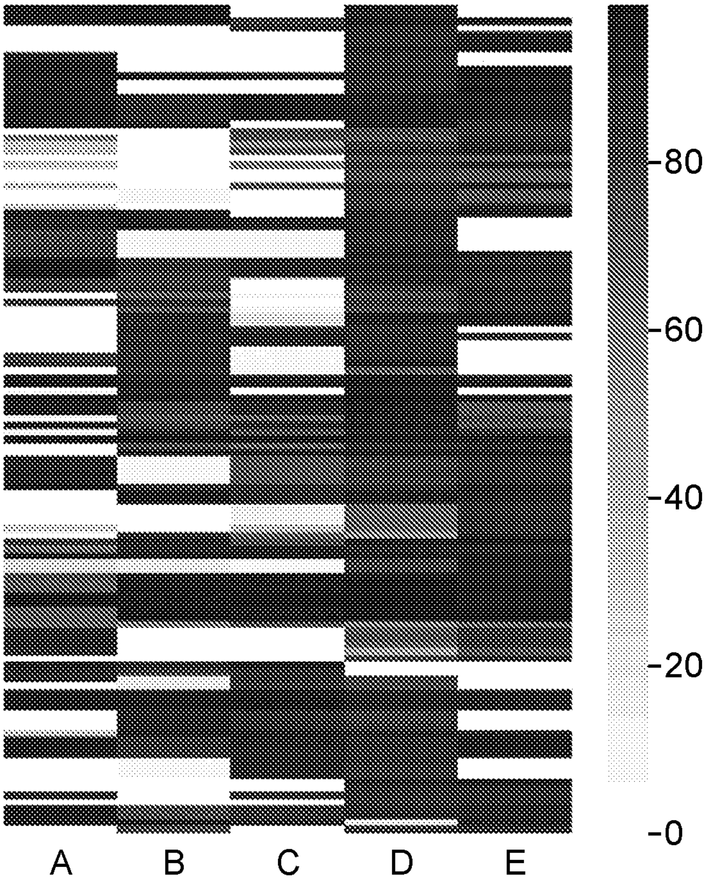


FIGURE 2

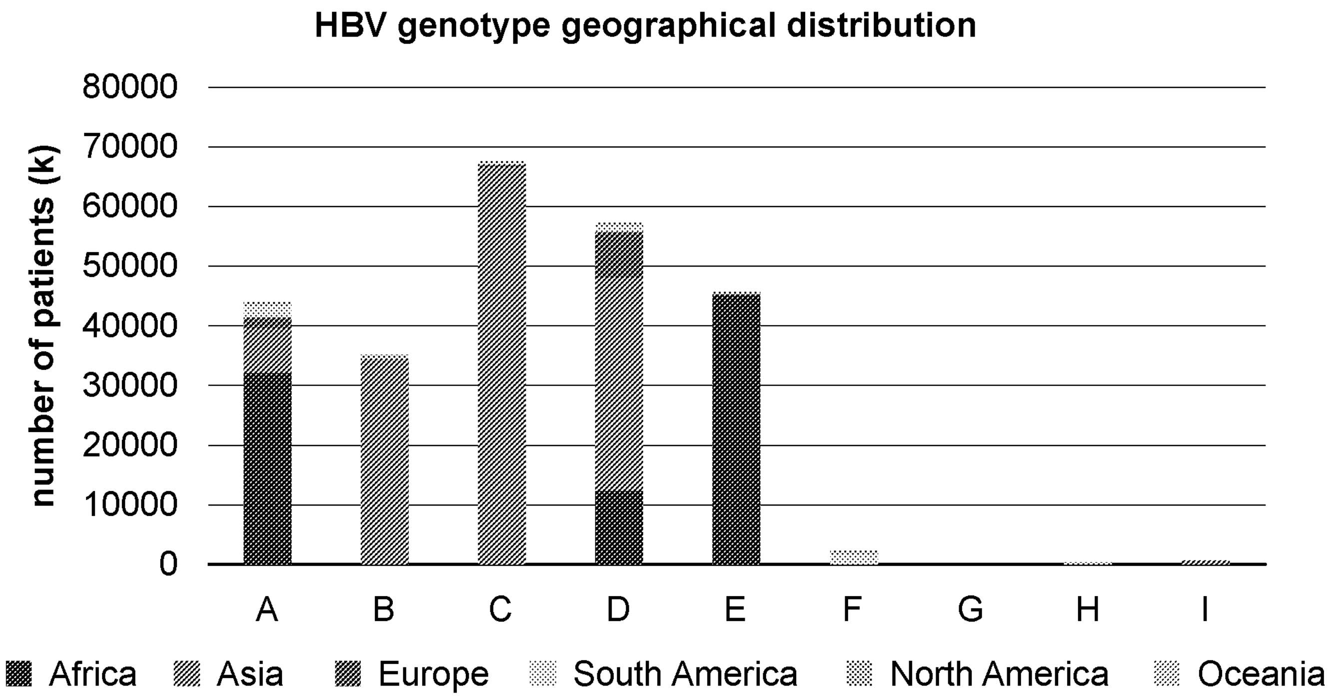

HBV genotype geographical distribution
number of patients (k)
80000
70000
60000
50000
40000
30000
20000
10000
0
A
B
C
D
E
F
G
H
I
Africa
Asia
Europe
South America
North America
Oceania

FIGURE 3

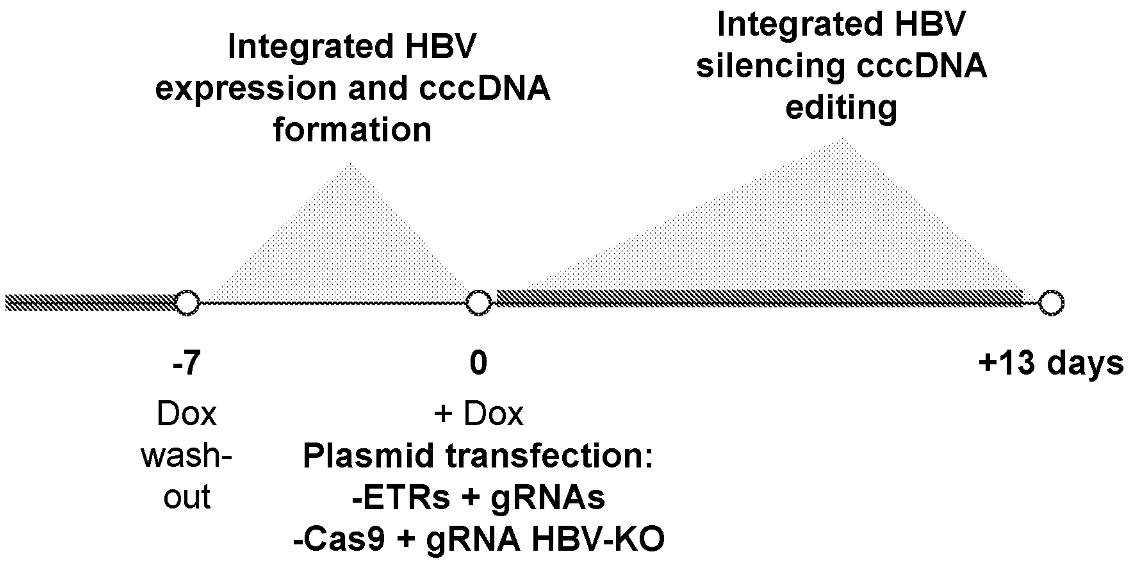


FIGURE 4A

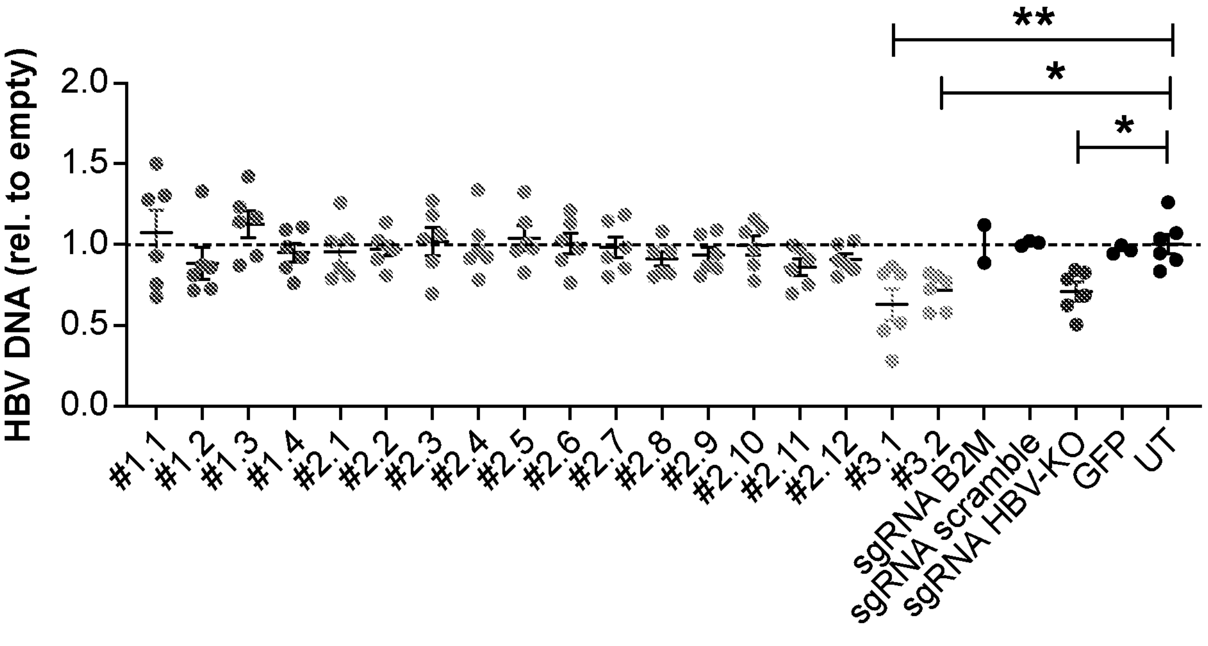


FIGURE 4B

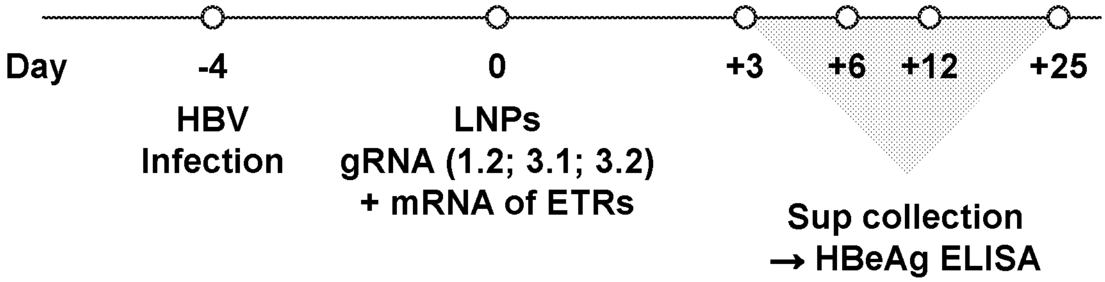


FIGURE 5A

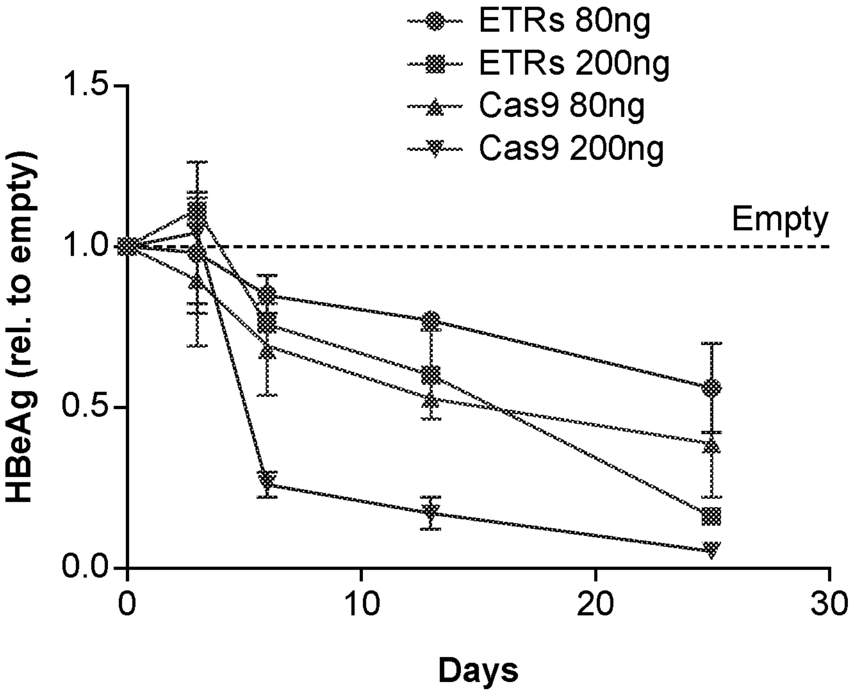


FIGURE 5B

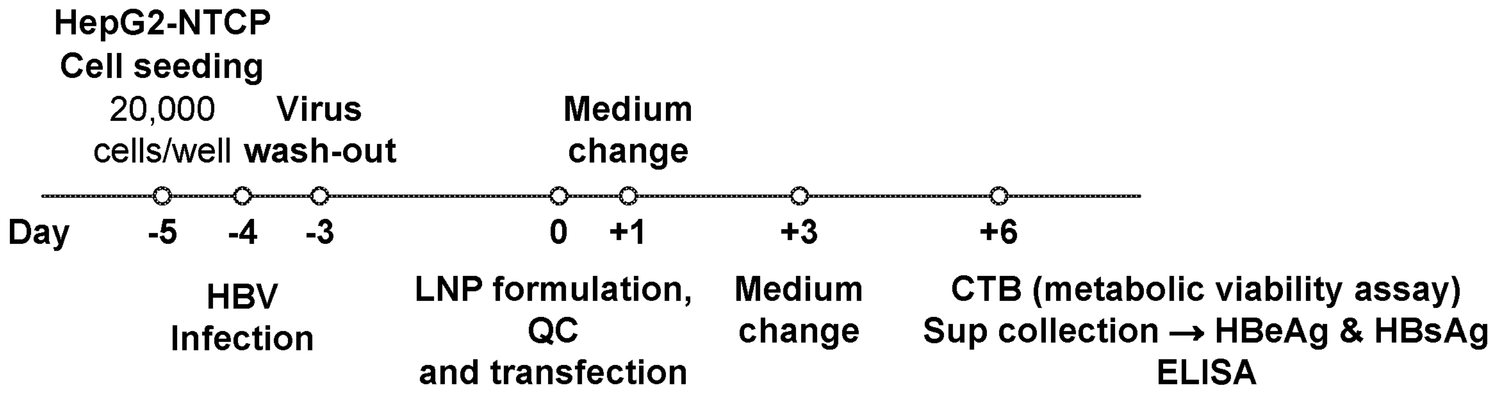

HepG2-NTCP
Cell seeding
20,000
cells/well
Virus
wash-out
Medium
change
Day
-5
-4
-3
0
+1
+3
+6
HBV
Infection
LNP formulation,
QC
and transfection
Medium
change
CTB (metabolic viability assay)
Sup collection → HBeAg & HBsAg
ELISA

FIGURE 6

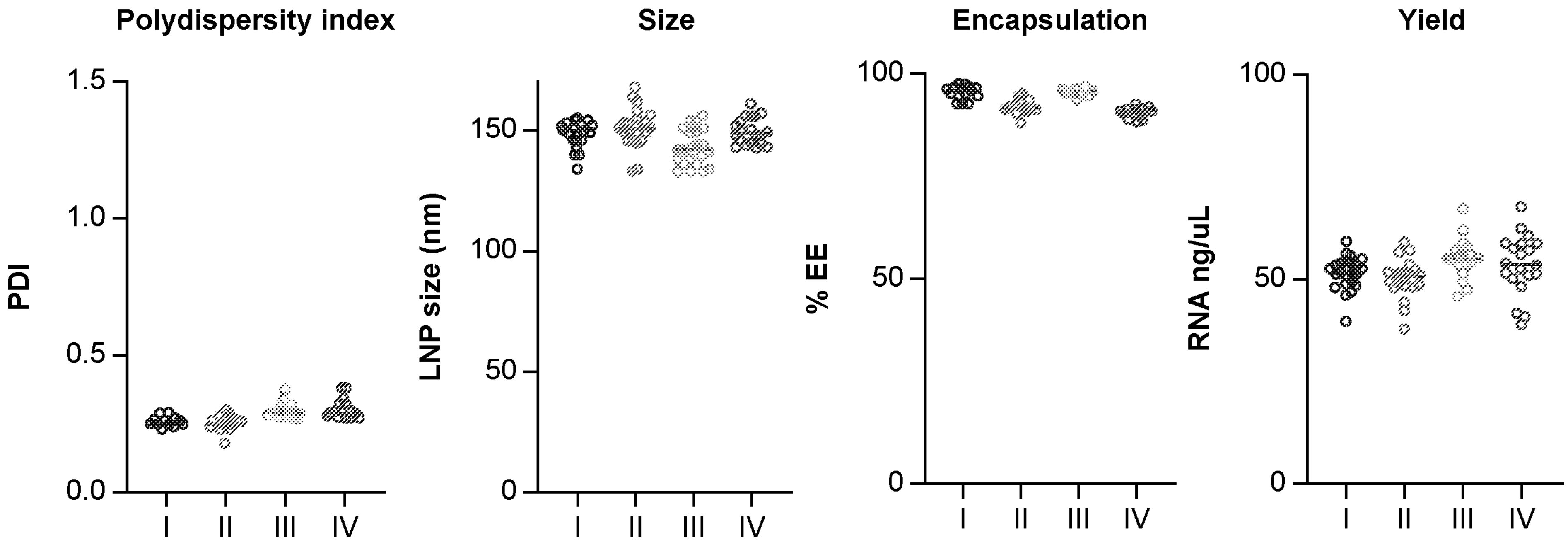

Polydispersity index
PDI
1.5
1.0
0.5
0.0
I
II
III
IV
Size
LNP size (nm)
150
100
50
0
I
II
III
IV
Encapsulation
% EE
100
50
0
I
II
III
IV
Yield
RNA ng/uL
100
50
0
I
II
III
IV

FIGURE 7

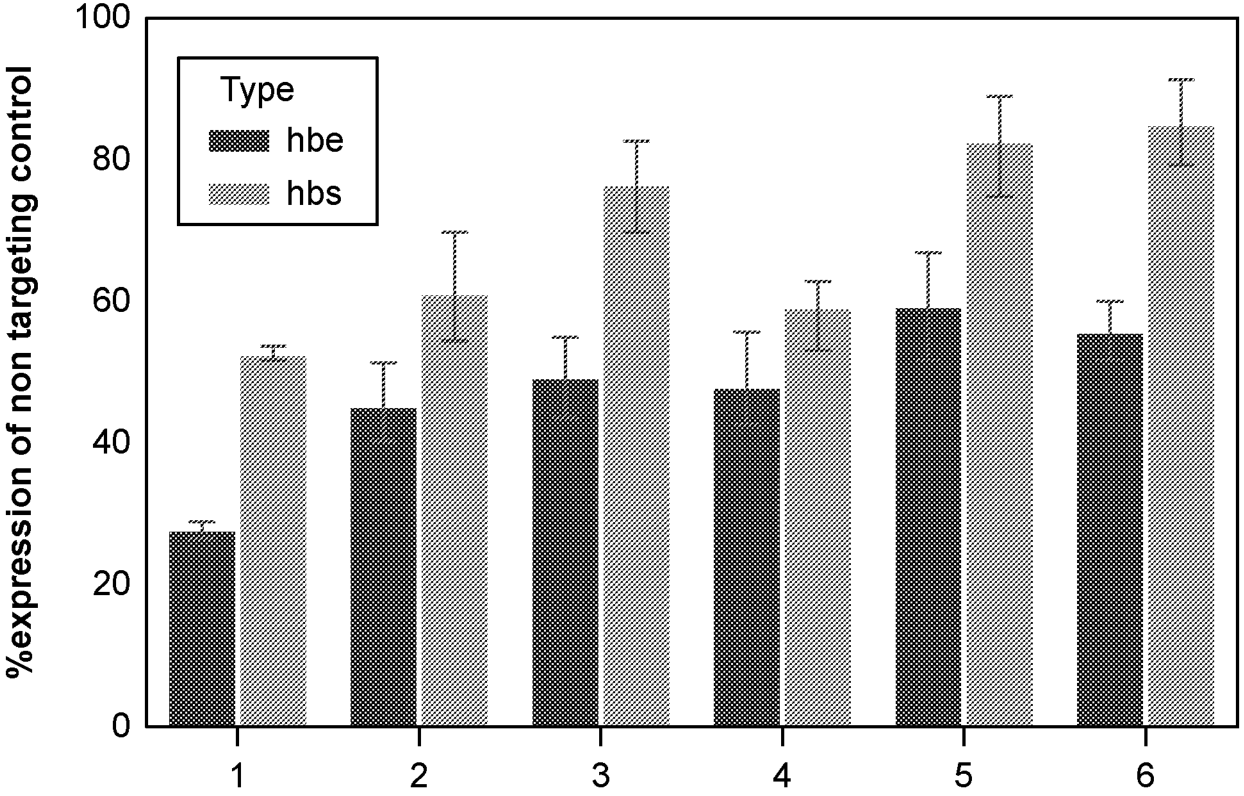

100
80
60
40
20
0
%expression of non targeting control
Type
hbe
hbs
1
2
3
4
5
6

FIGURE 8

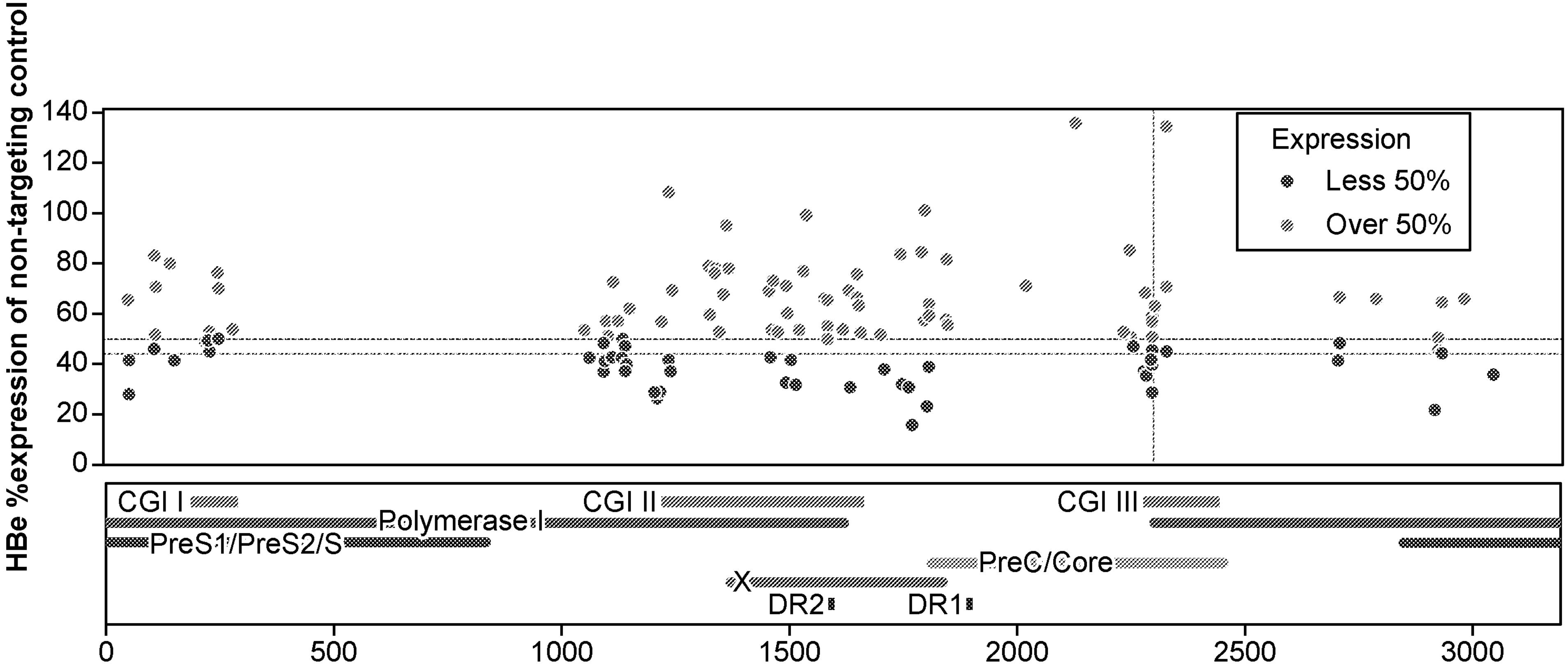

HBe %expression of non-targeting control
0
20
40
60
80
100
120
140
Expression
Less 50%
Over 50%
CGI I
CGI II
CGI III
Polymerase
PreS1/PreS2/S
PreC/Core
X
DR2
DR1
0
500
1000
1500
2000
2500
3000

FIGURE 9

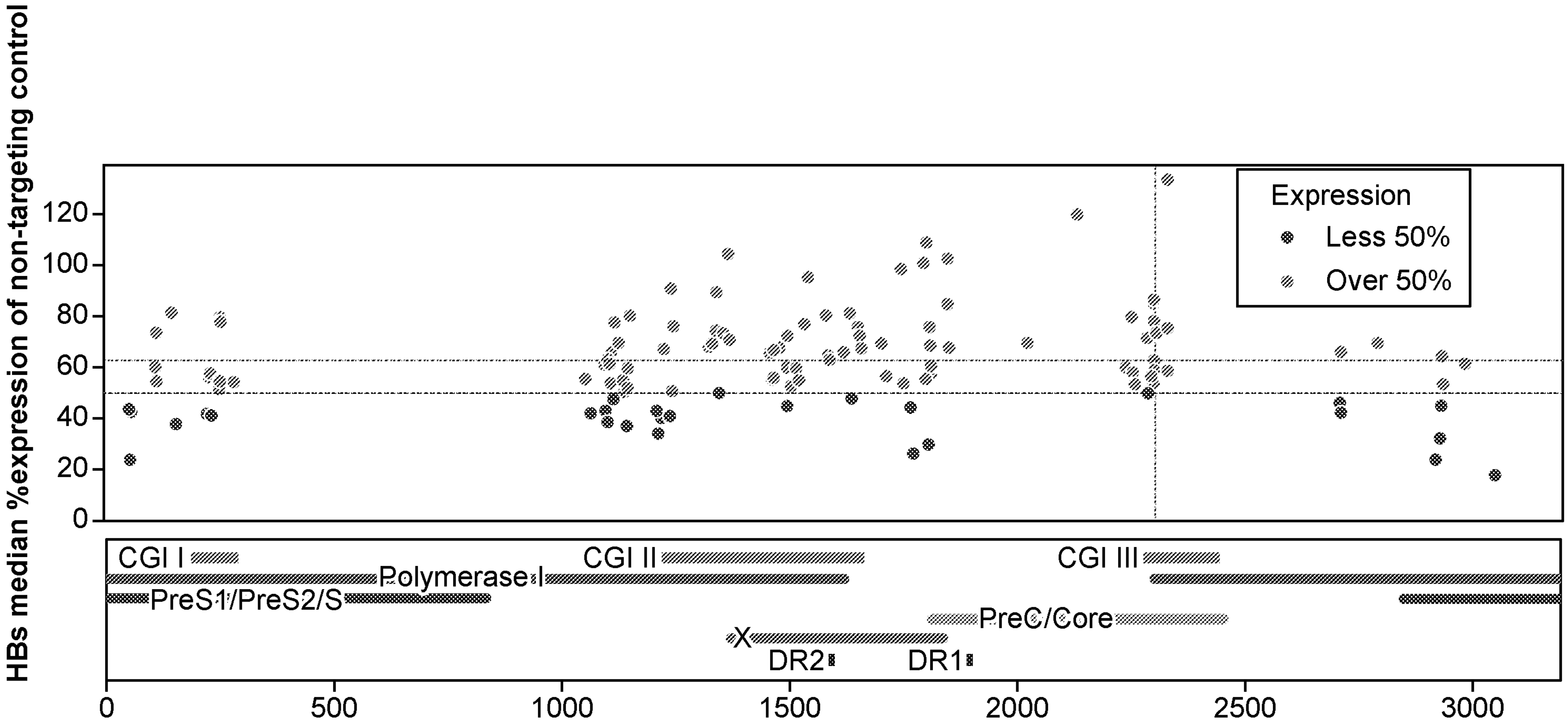

HBs median %expression of non-targeting control
0
20
40
60
80
100
120
Expression
● Less 50%
● Over 50%
CGI I
CGI II
CGI III
Polymerase
PreS1/PreS2/S
PreC/Core
X
DR2
DR1
0
500
1000
1500
2000
2500
3000

FIGURE 10

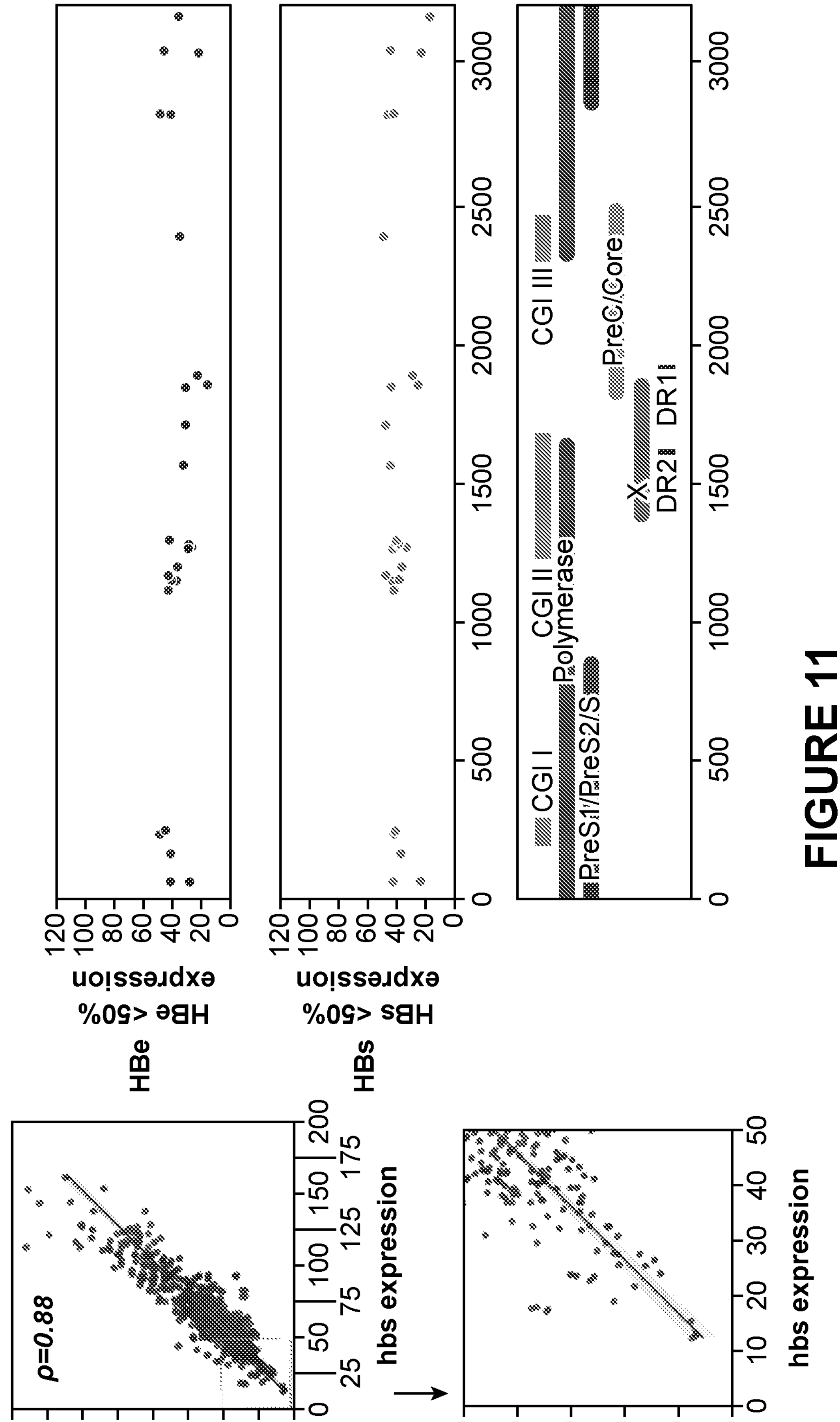

ρ=0.88
hbe expression
0
25
50
75
100
125
150
175
200
hbs expression
0
25
50
75
100
125
150
175
200
hbe expression
0
10
20
30
40
50
hbs expression
0
10
20
30
40
50
HBe
HBs
HBe <50% expression
HBs <50% expression
0
20
40
60
80
100
120
0
500
1000
1500
2000
2500
3000
CGI I
CGI II
CGI III
Polymerase
PreS1/PreS2/S
X
DR2
DR1
PreC/Core

FIGURE 11

]

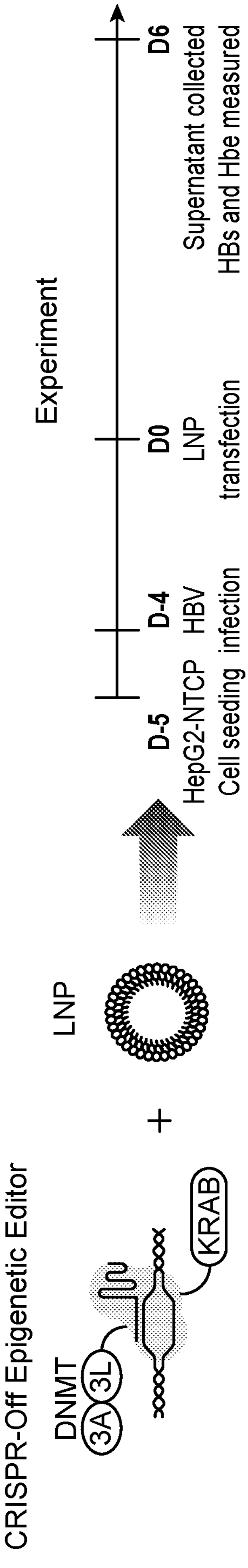


FIGURE 12A

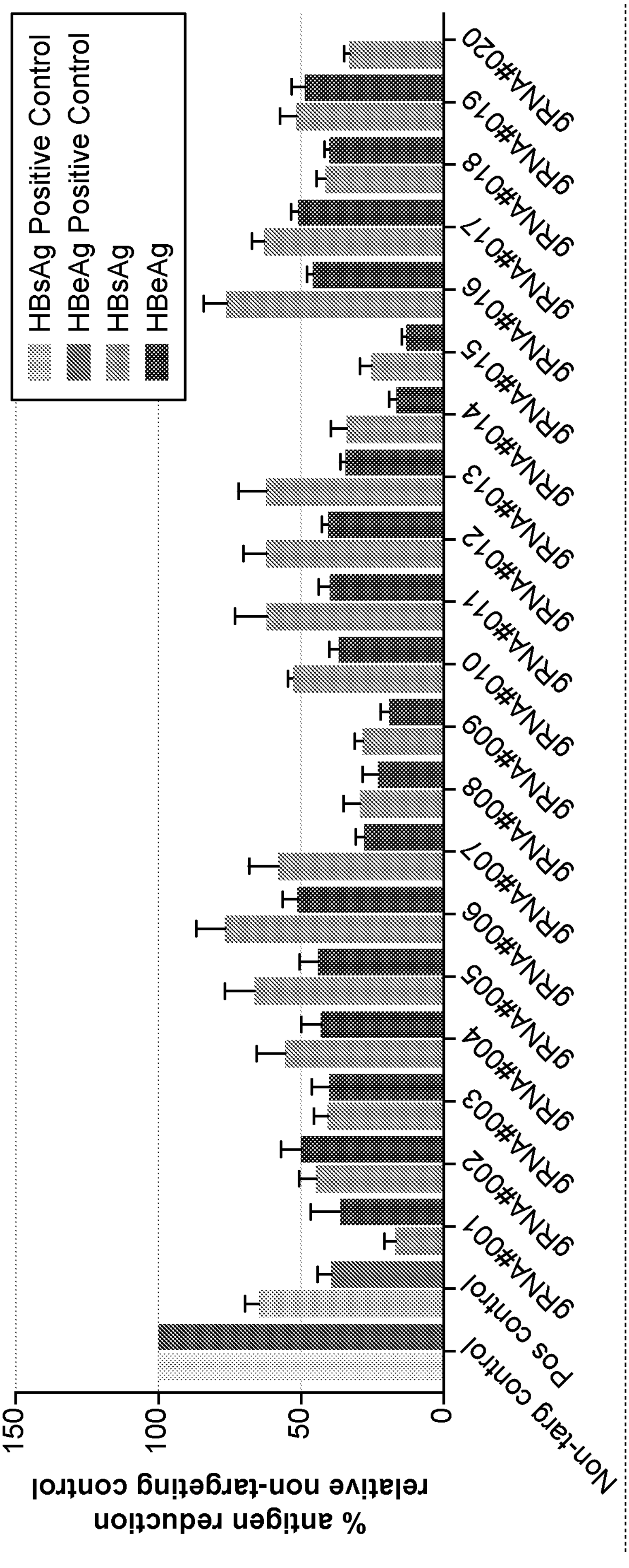


FIGURE 12B

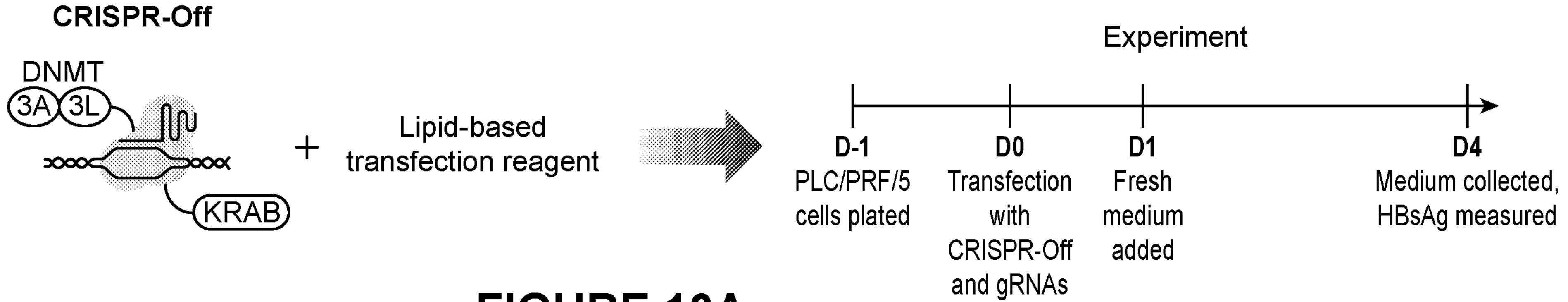
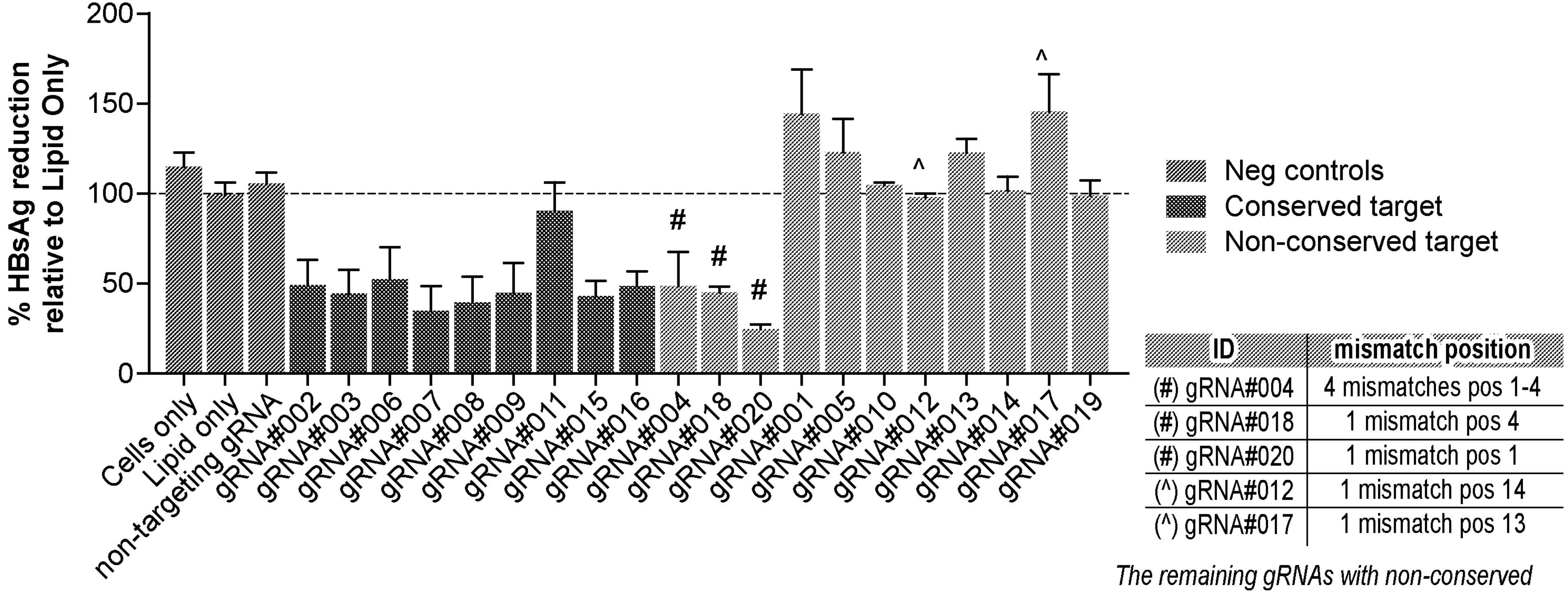


FIGURE 13A



ID	mismatch position
(#) gRNA#004	4 mismatches pos 1-4
(#) gRNA#018	1 mismatch pos 4
(#) gRNA#020	1 mismatch pos 1
(^) gRNA#012	1 mismatch pos 14
(^) gRNA#017	1 mismatch pos 13

The remaining gRNAs with non-conserved target do not have protospacer adjacent motif (PAM)

FIGURE 13B

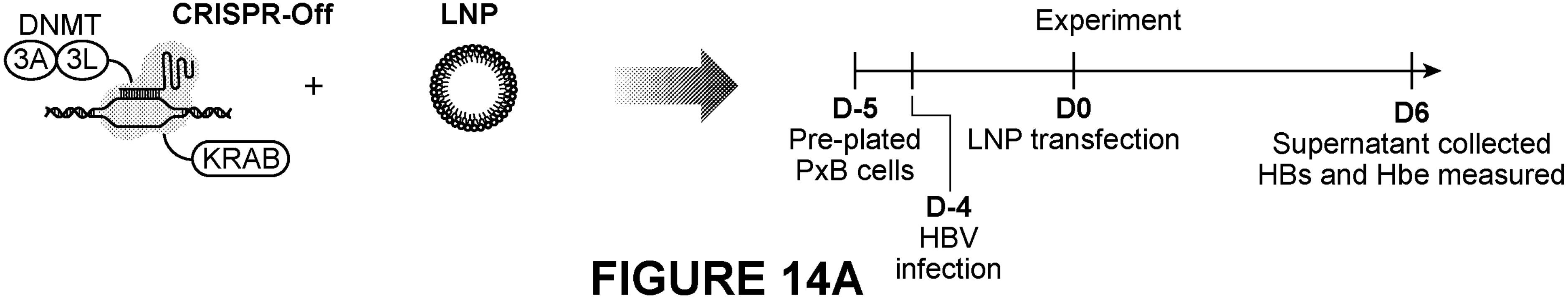


FIGURE 14A

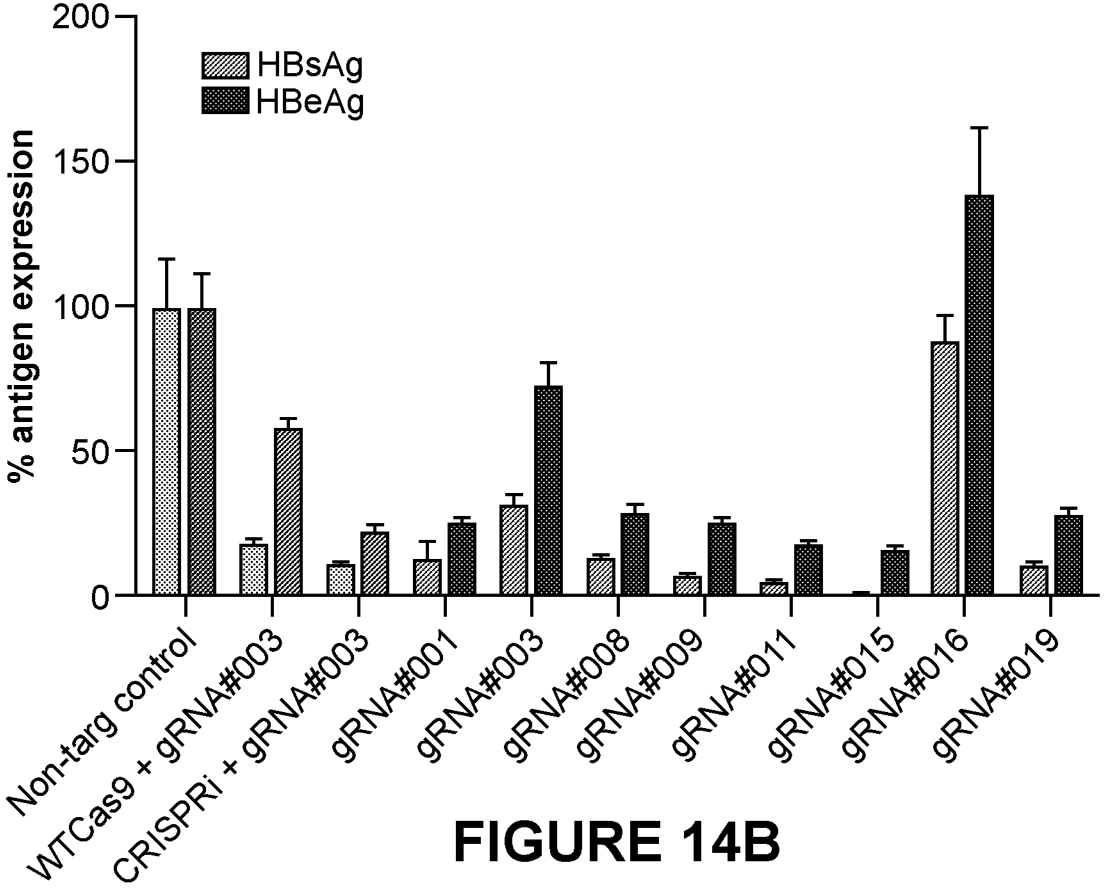


FIGURE 14B

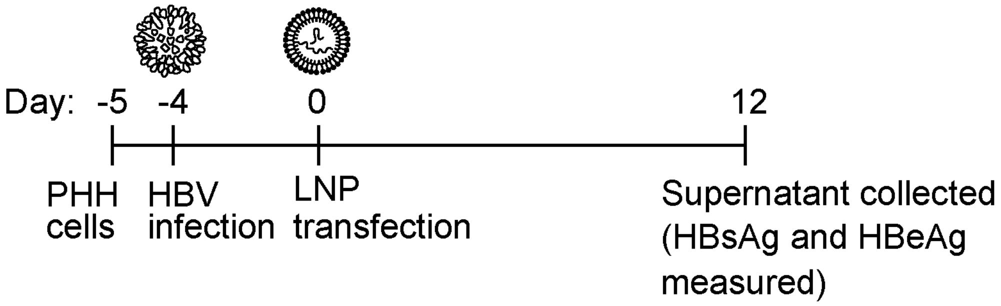


FIGURE 14C

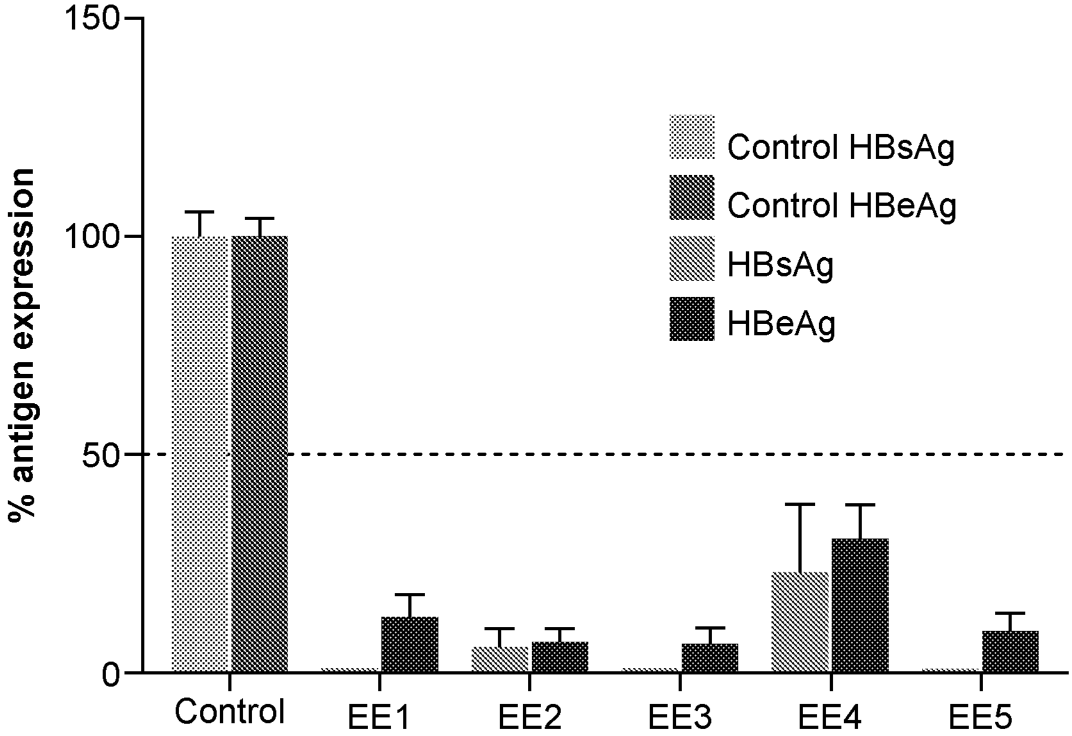


FIGURE 14D

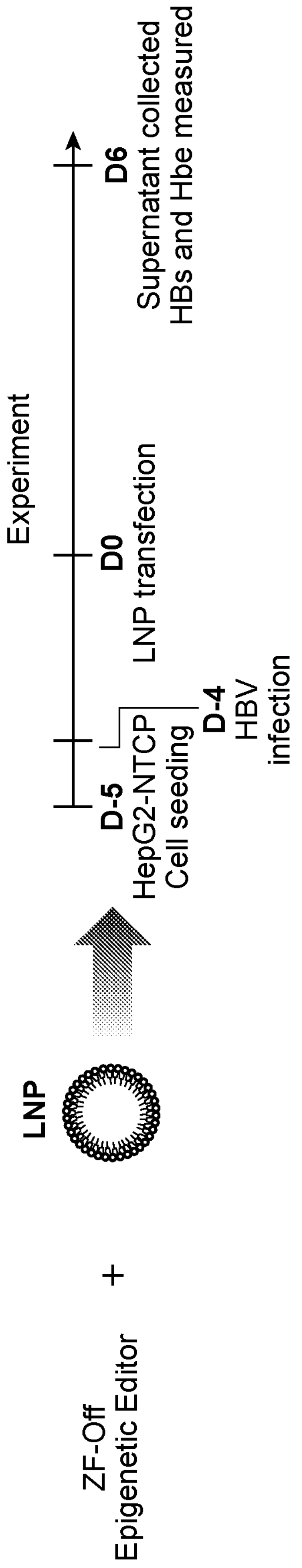


FIGURE 15A

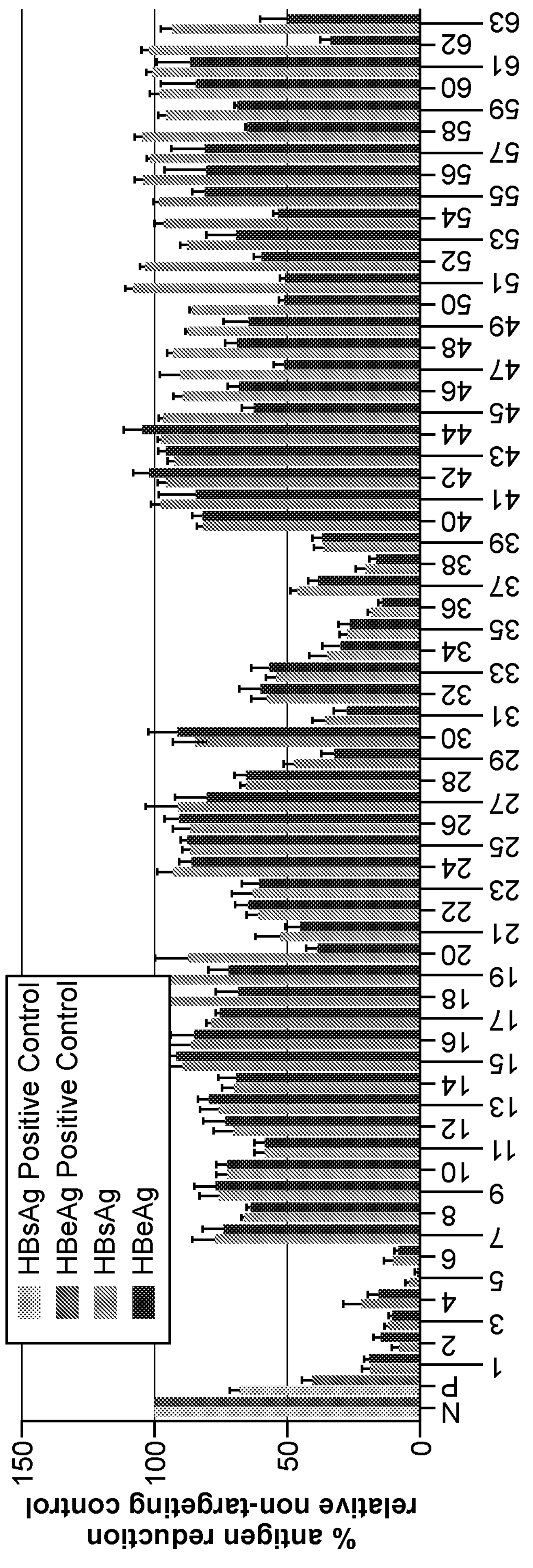


FIGURE 15B

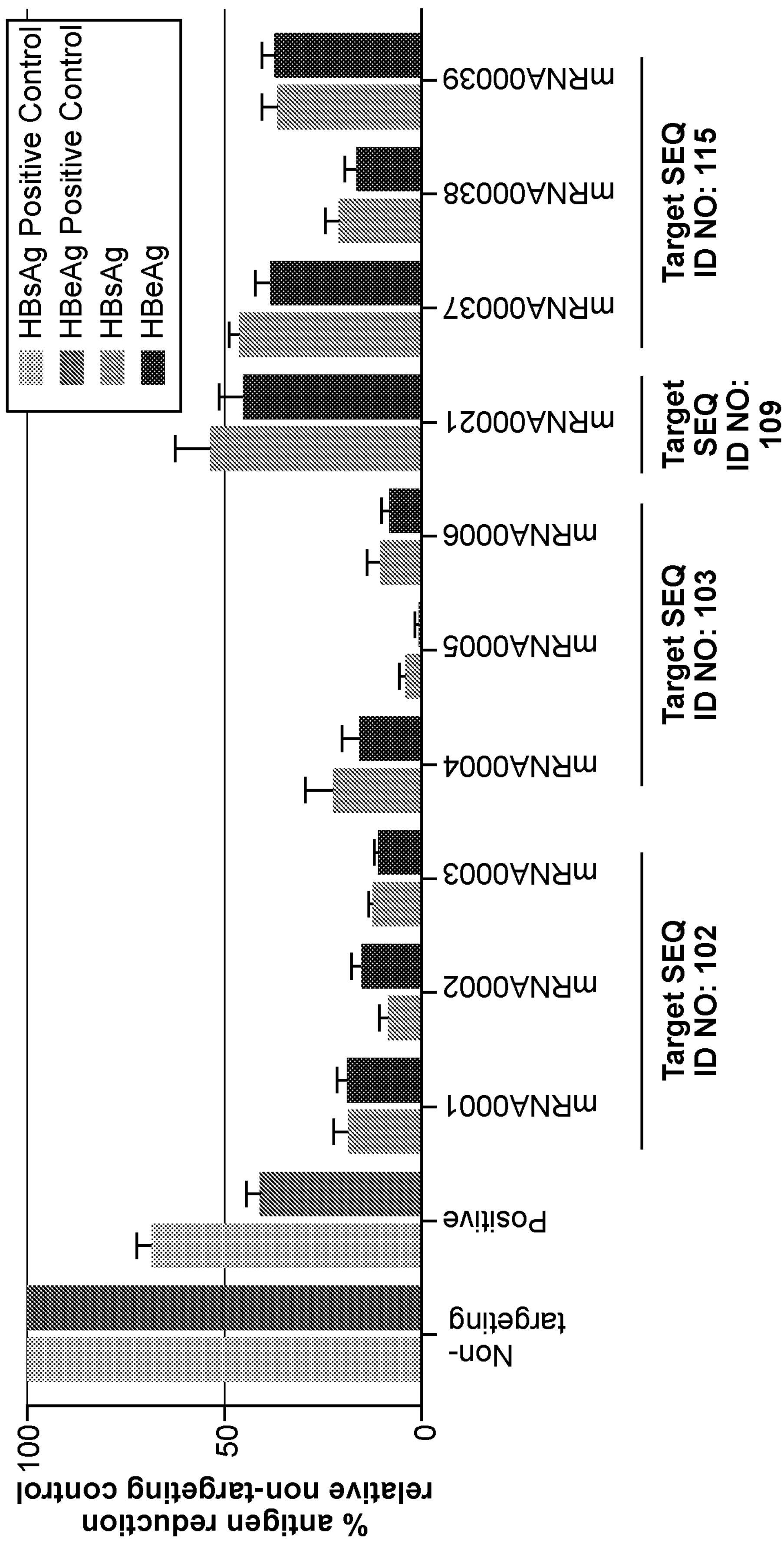

% antigen reduction
relative non-targeting control
100
50
0
HBsAg Positive Control
HBeAg Positive Control
HBsAg
HBeAg
Non-targeting
Positive
mRNA0001
mRNA0002
mRNA0003
mRNA0004
mRNA0005
mRNA0006
mRNA00021
mRNA00037
mRNA00038
mRNA00039
Target SEQ ID NO: 102
Target SEQ ID NO: 103
Target SEQ ID NO: 109
Target SEQ ID NO: 115

FIGURE 16A

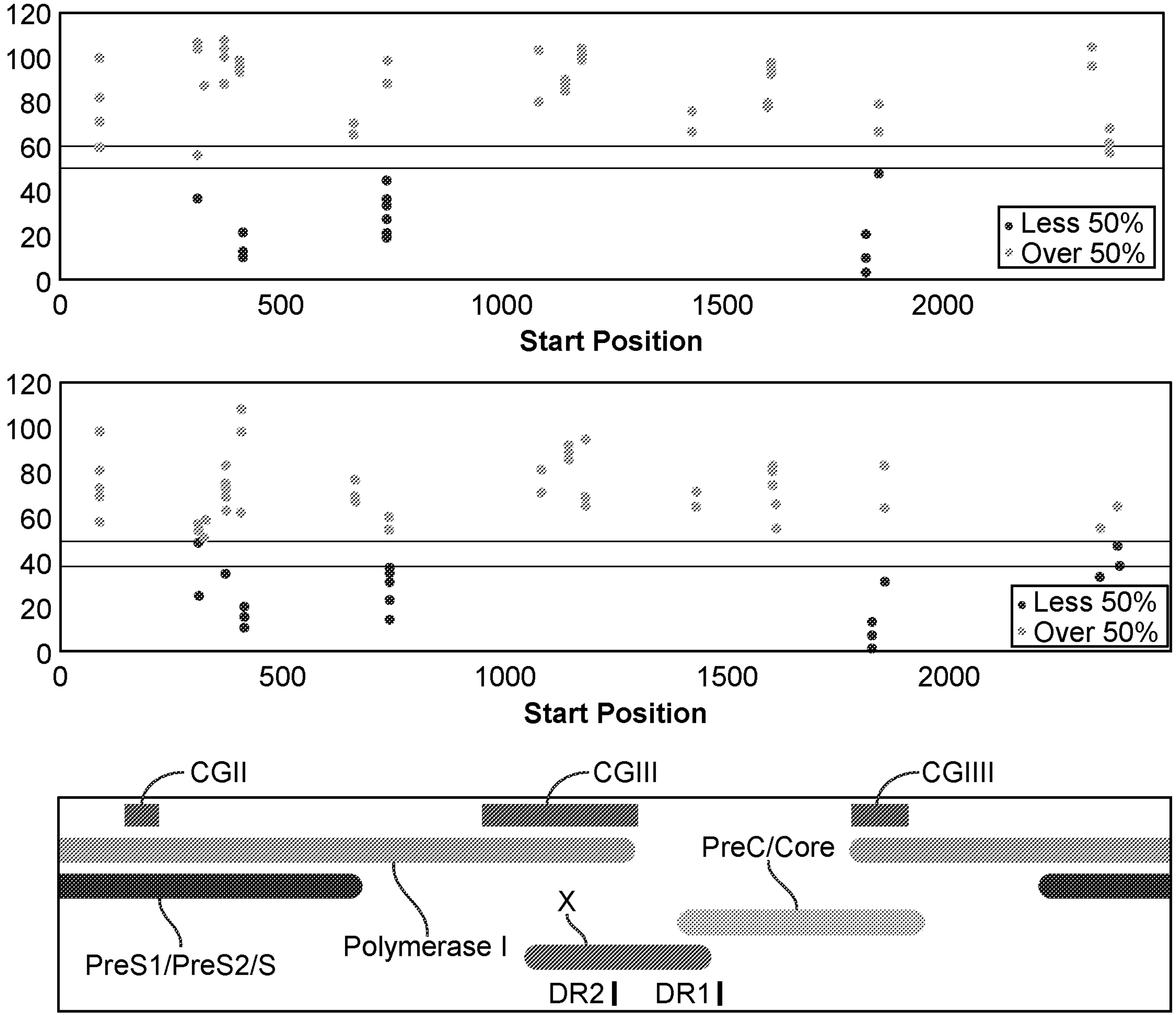

120
100
80
60
40
20
0
0
500
1000
1500
2000
Start Position
Less 50%
Over 50%
120
100
80
60
40
20
0
0
500
1000
1500
2000
Start Position
Less 50%
Over 50%
CGII
CGIII
CGIIII
PreC/Core
X
PreS1/PreS2/S
Polymerase I
DR2
DR1

FIGURE 16B

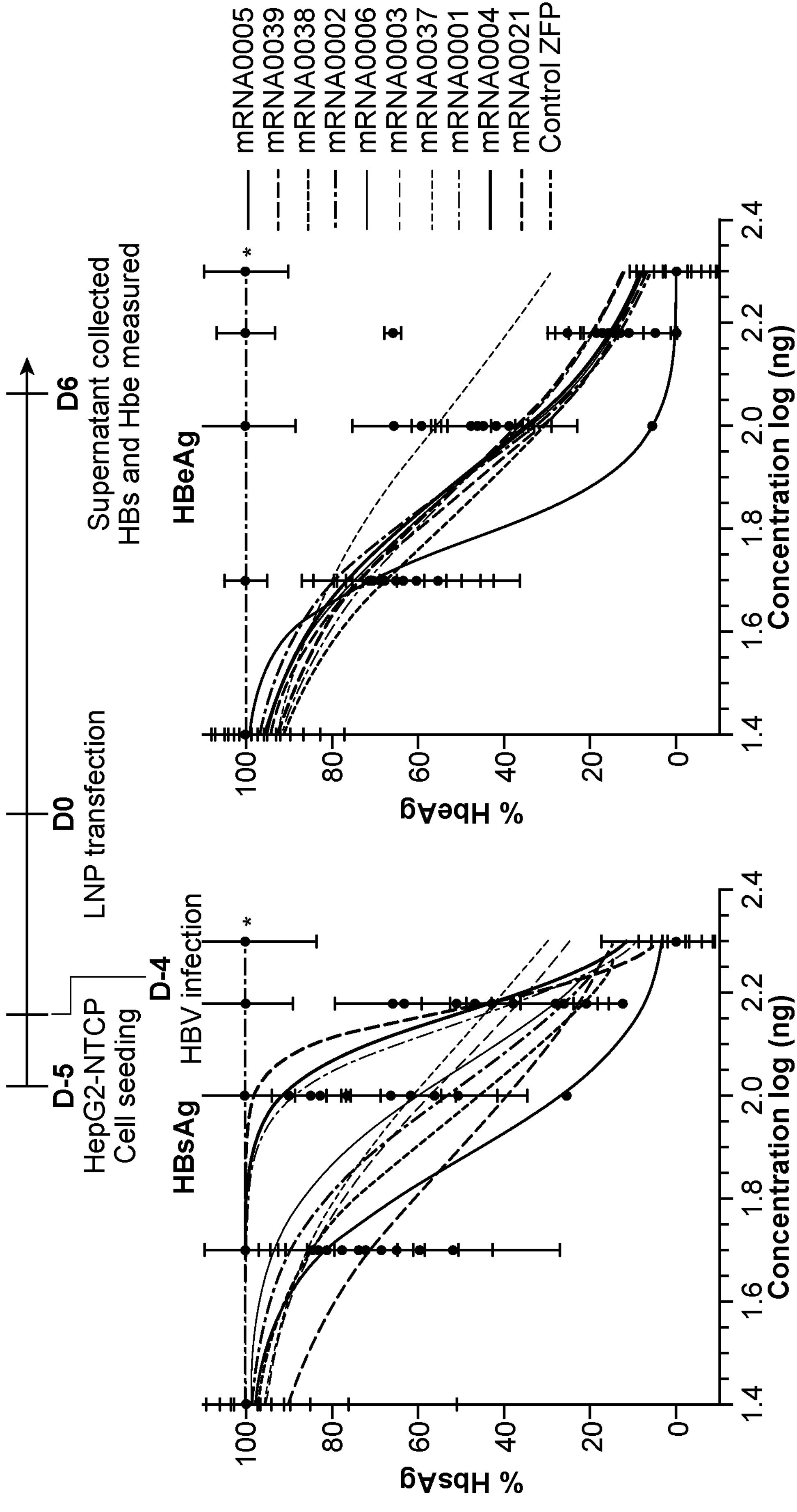

D-5
HepG2-NTCP
Cell seeding
D-4
HBV infection
D0
LNP transfection
D6
Supernatant collected
HBs and Hbe measured
HBsAg
% HbsAg
100
80
60
40
20
0
1.4
1.6
1.8
2.0
2.2
2.4
Concentration log (ng)
*
HBeAg
% HbeAg
100
80
60
40
20
0
1.4
1.6
1.8
2.0
2.2
2.4
Concentration log (ng)
*
mRNA0005
mRNA0039
mRNA0038
mRNA0002
mRNA0006
mRNA0003
mRNA0037
mRNA0001
mRNA0004
mRNA0021
Control ZFP*

FIGURE 17

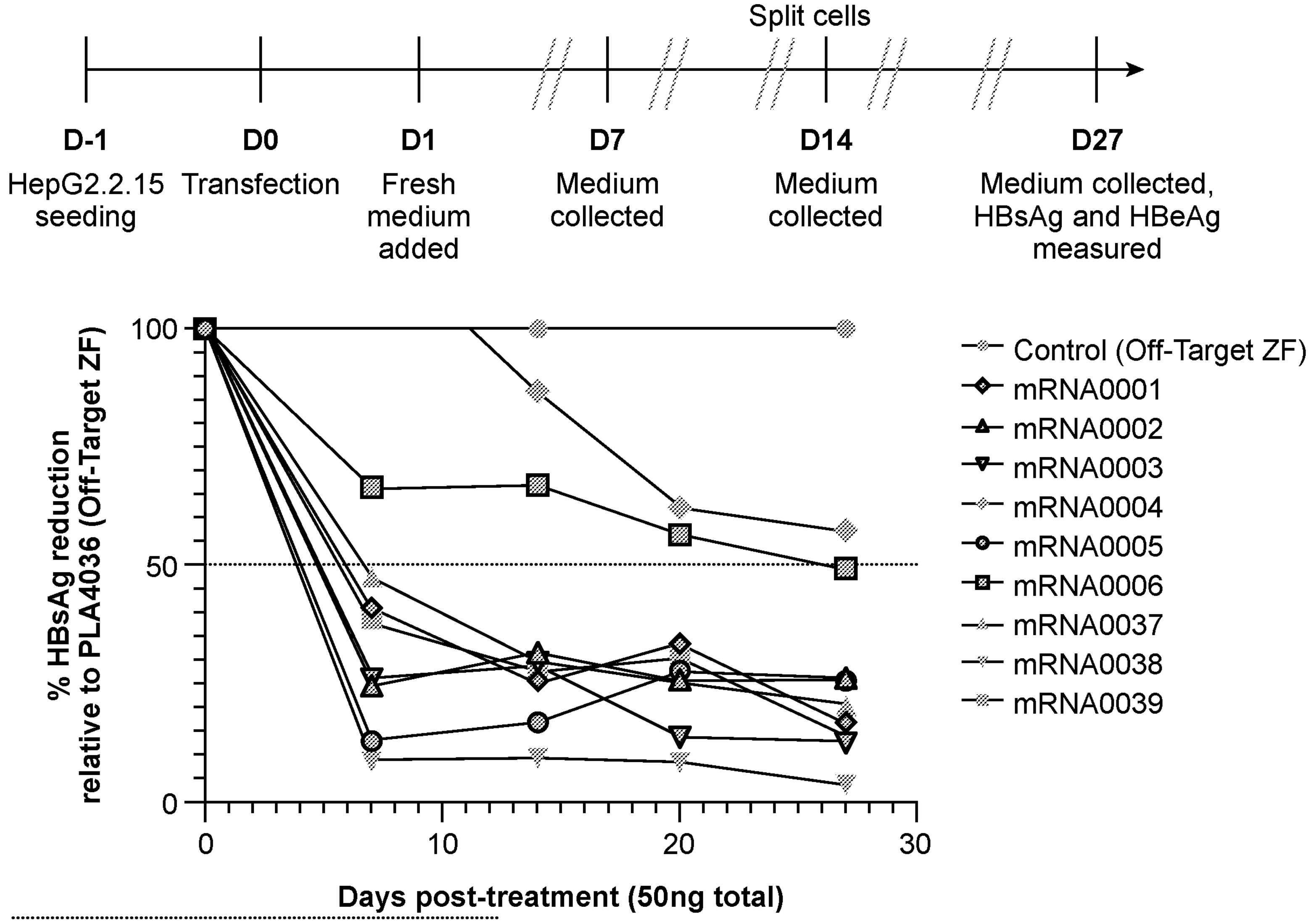

Split cells
D-1
HepG2.2.15 seeding
D0
Transfection
D1
Fresh medium added
D7
Medium collected
D14
Medium collected
D27
Medium collected, HBsAg and HBeAg measured
% HBsAg reduction relative to PLA4036 (Off-Target ZF)
100
50
0
0
10
20
30
Days post-treatment (50ng total)
Control (Off-Target ZF)
mRNA0001
mRNA0002
mRNA0003
mRNA0004
mRNA0005
mRNA0006
mRNA0037
mRNA0038
mRNA0039

FIGURE 18

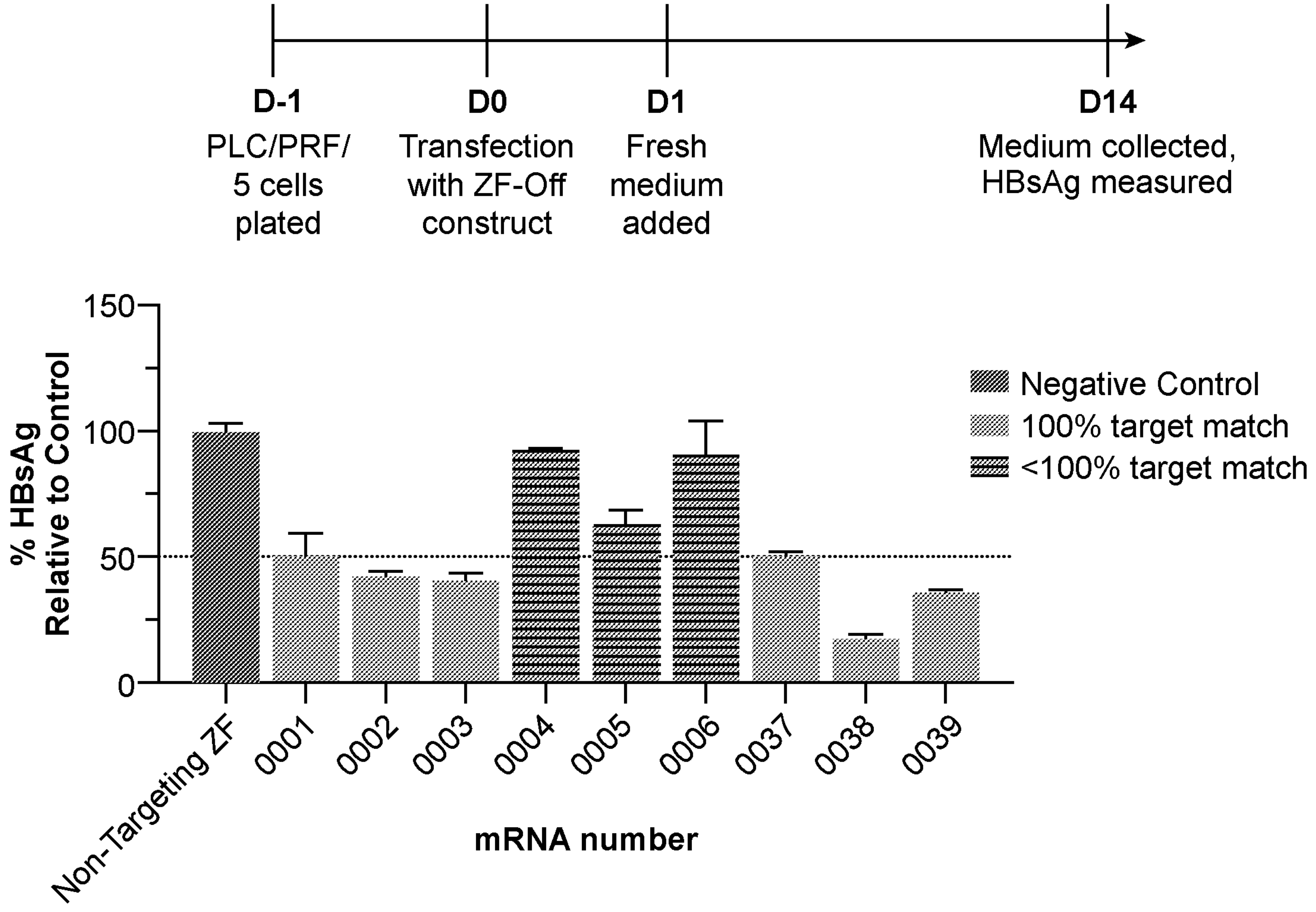

D-1
PLC/PRF/ 5 cells plated
D0
Transfection with ZF-Off construct
D1
Fresh medium added
D14
Medium collected, HBsAg measured
% HBsAg Relative to Control
150
100
50
0
Negative Control
100% target match
<100% target match
Non-Targeting ZF
0001
0002
0003
0004
0005
0006
0037
0038
0039
mRNA number

FIGURE 19

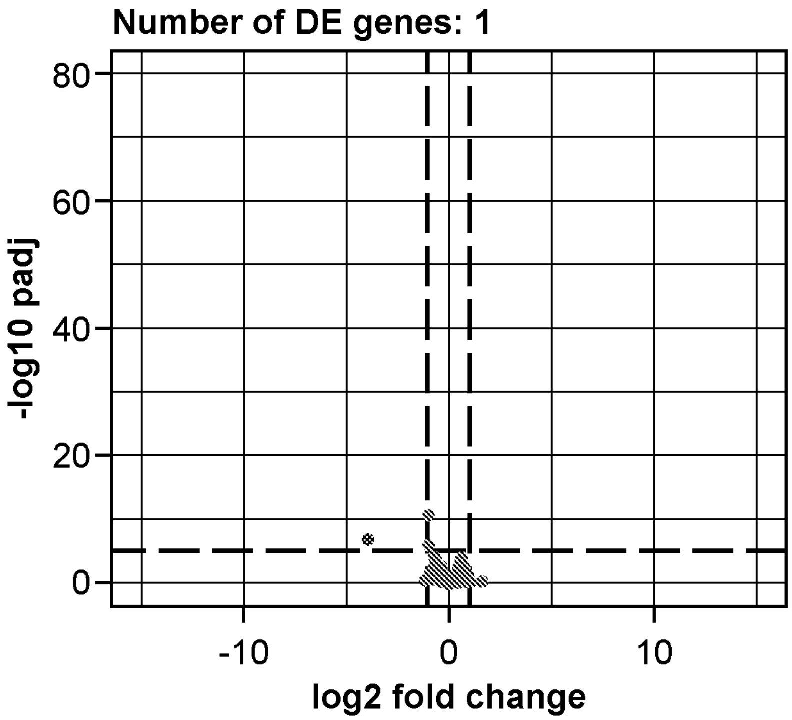

Number of DE genes: 1
80
60
40
20
0
-log10 padj
-10
0
10
log2 fold change

FIGURE 20A

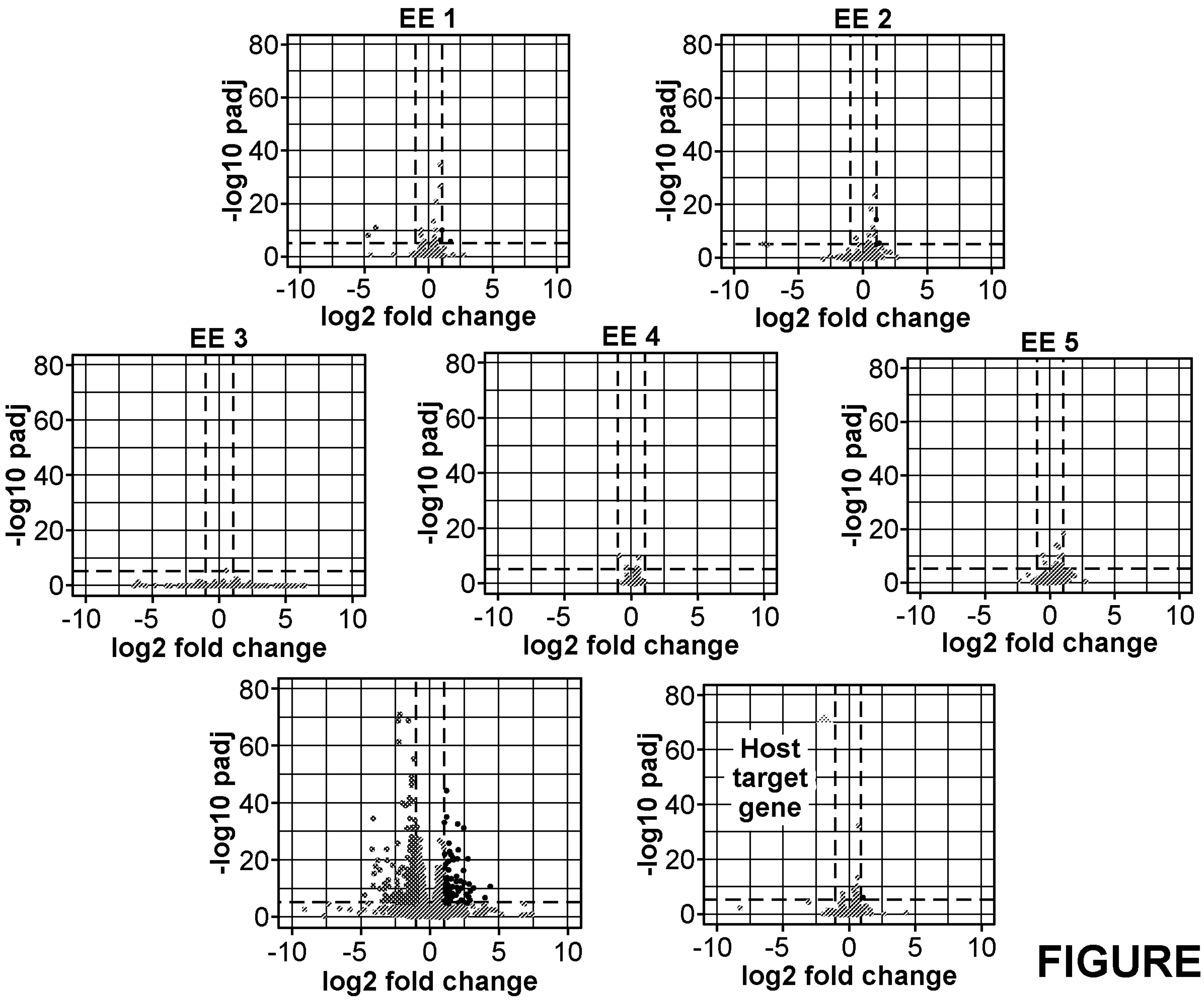

EE 1
EE 2
EE 3
EE 4
EE 5
-log10 padj
log2 fold change
80
60
40
20
0
-10
-5
0
5
10
Host target gene

FIGURE 20B

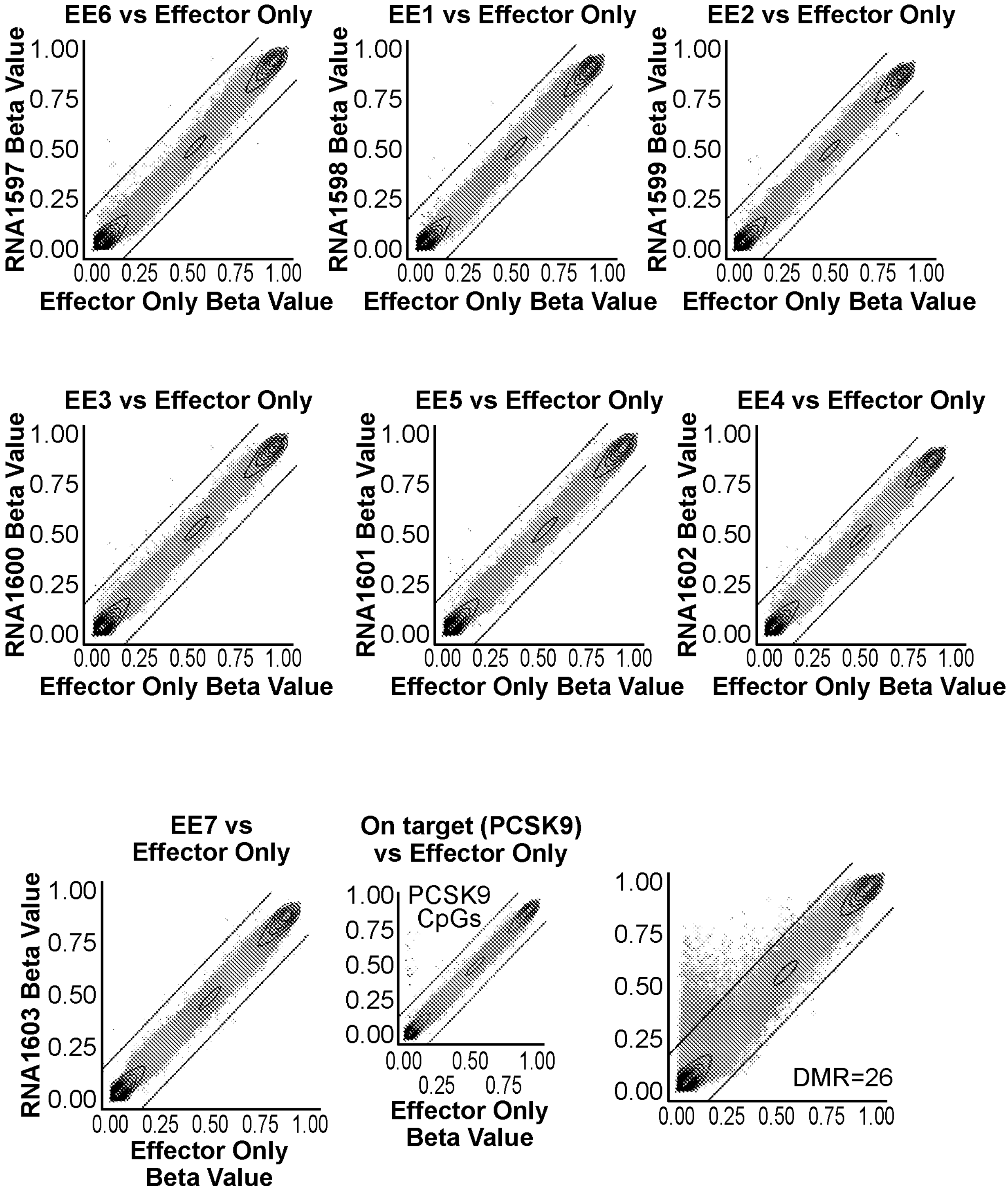

EE6 vs Effector Only
RNA1597 Beta Value
Effector Only Beta Value
EE1 vs Effector Only
RNA1598 Beta Value
Effector Only Beta Value
EE2 vs Effector Only
RNA1599 Beta Value
Effector Only Beta Value
EE3 vs Effector Only
RNA1600 Beta Value
Effector Only Beta Value
EE5 vs Effector Only
RNA1601 Beta Value
Effector Only Beta Value
EE4 vs Effector Only
RNA1602 Beta Value
Effector Only Beta Value
EE7 vs Effector Only
RNA1603 Beta Value
Effector Only Beta Value
On target (PCSK9) vs Effector Only
PCSK9 CpGs
Effector Only Beta Value
DMR=26
0.00
0.25
0.50
0.75
1.00

FIGURE 20C

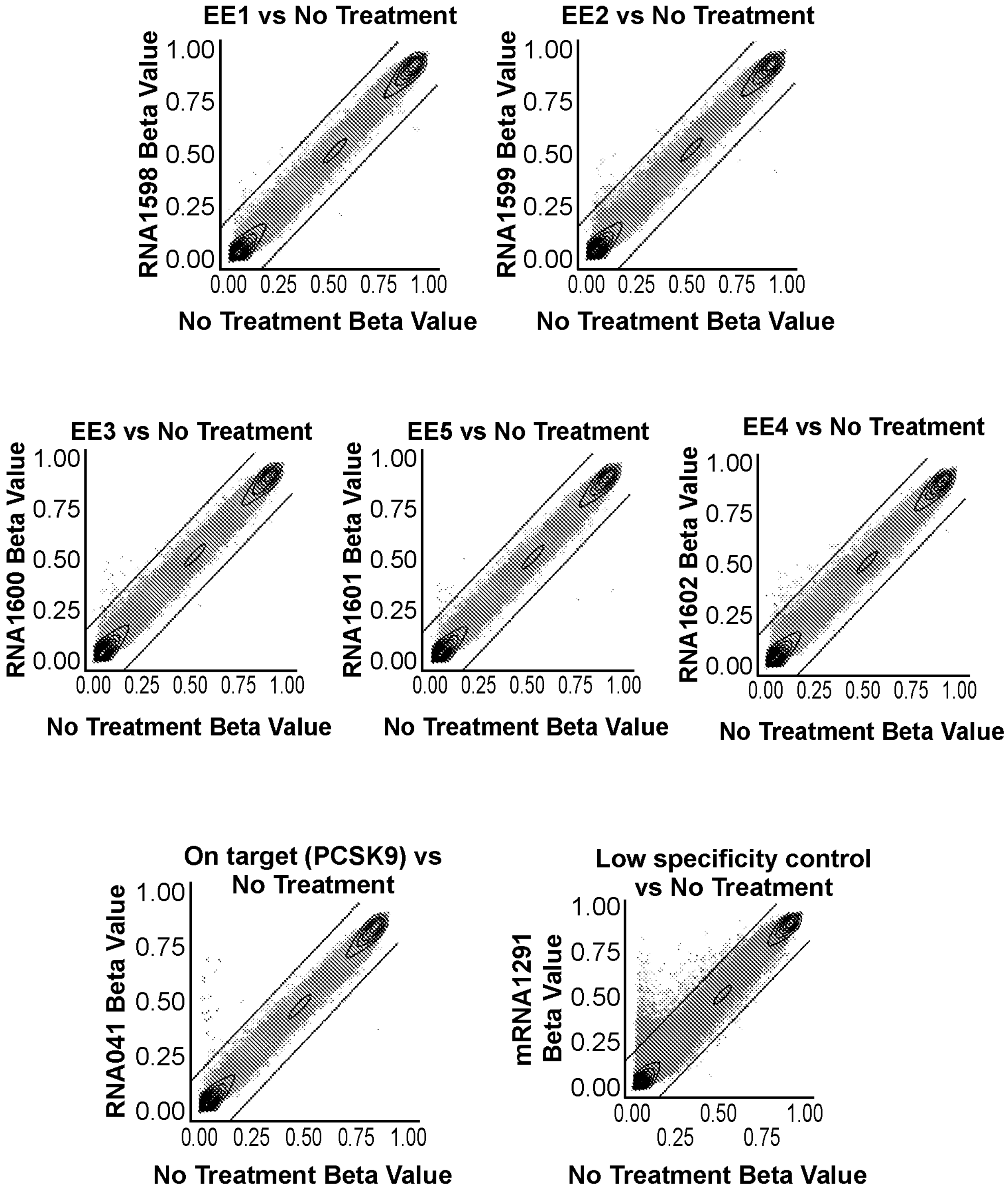

EE1 vs No Treatment
RNA1598 Beta Value
1.00
0.75
0.50
0.25
0.00
0.00 0.25 0.50 0.75 1.00
No Treatment Beta Value
EE2 vs No Treatment
RNA1599 Beta Value
1.00
0.75
0.50
0.25
0.00
0.00 0.25 0.50 0.75 1.00
No Treatment Beta Value
EE3 vs No Treatment
RNA1600 Beta Value
1.00
0.75
0.50
0.25
0.00
0.00 0.25 0.50 0.75 1.00
No Treatment Beta Value
EE5 vs No Treatment
RNA1601 Beta Value
1.00
0.75
0.50
0.25
0.00
0.00 0.25 0.50 0.75 1.00
No Treatment Beta Value
EE4 vs No Treatment
RNA1602 Beta Value
1.00
0.75
0.50
0.25
0.00
0.00 0.25 0.50 0.75 1.00
No Treatment Beta Value
On target (PCSK9) vs No Treatment
RNA041 Beta Value
1.00
0.75
0.50
0.25
0.00
0.00 0.25 0.50 0.75 1.00
No Treatment Beta Value
Low specificity control vs No Treatment
mRNA1291 Beta Value
1.00
0.75
0.50
0.25
0.00
0.00 0.50 1.00
0.25 0.75
No Treatment Beta Value

FIGURE 20D

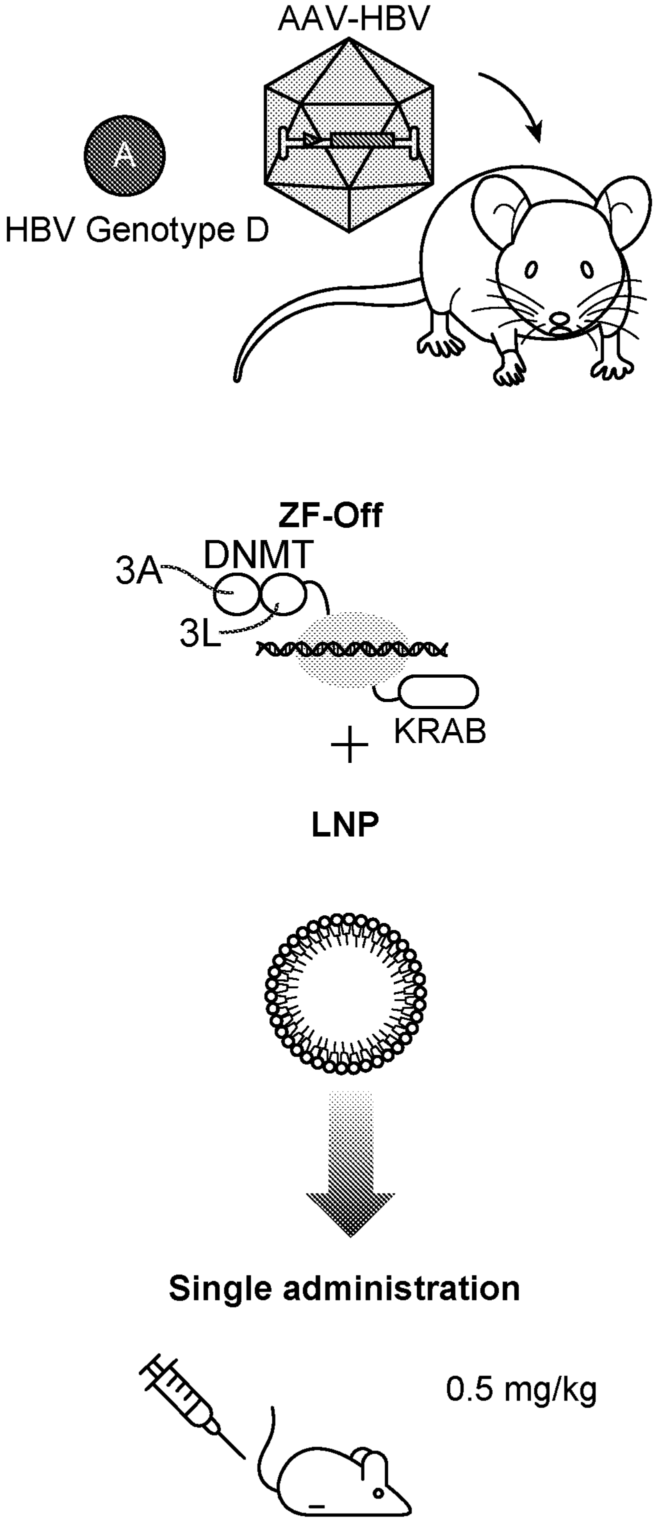

AAV-HBV
A
HBV Genotype D
ZF-Off
DNMT
3A
3L
KRAB
+
LNP
Single administration
0.5 mg/kg

FIGURE 21