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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 September 2013 (19.09.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/138644 A2

(51) International Patent Classification:
C12N 9/78 (2006.01)

(21) International Application Number:
PCT/US2013/031620

(22) International Filing Date:
14 March 2013 (14.03.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/611,295 15 March 2012 (15.03.2012) US

(71) Applicant: **NEW ENGLAND BIOLABS, INC.** [US/US]; 240 County Road, Ipswich, MA 01938 (US).

(72) Inventors; and
(71) Applicants : **VAISVILA, Romualdas** [LT/US]; 100 Central Street, Unit 8, Ipswich, MA 01938 (US). **JOHNSON, Heidi, Erika** [US/US]; 15 Goldfinch Way, Ipswich, MA 01938 (US). **VAINAUSKAS, Saulius** [—/US]; 4 Johnson Street, Newburyport, MA 01950 (US).

(72) Inventor: **DAVIS, Theodore B.**; 139 Depot Road, Boxford, MA 01921 (US).

(74) Agent: **STRIMPEL, Harriet, M.**; New England Biolabs, Inc., 240 County Road, Ipswich, MA 01938 (US).

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

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

Published:

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

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

(54) **Title**: METHODS AND COMPOSITIONS FOR DISCRIMINATION BETWEEN CYTOSINE AND MODIFICATIONS THEREOF, AND FOR METHYLOME ANALYSIS

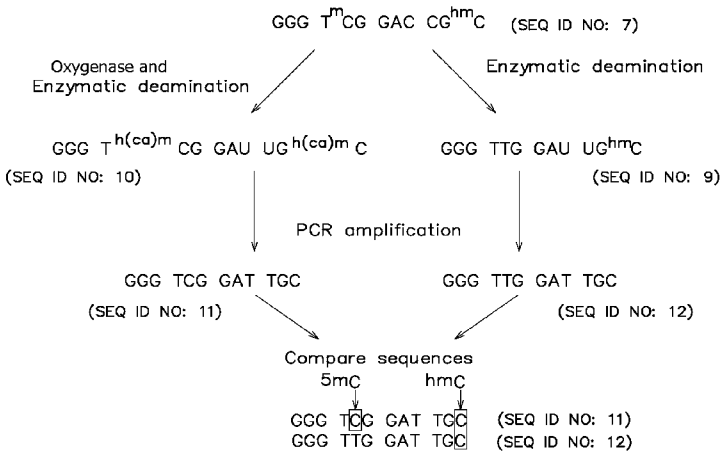


FIG. 13

(57) **Abstract**: Compositions and methods are provided for discrimination between cytosine and modifications thereof using cytidine deaminases and/or oxygenases. Variants of wild type cytidine deaminases are described which show reduced bias with respect to adjacent nucleotides upstream of the cytosine. The methods provide a rapid and convenient use of enzymes to obtain methylomes.

WO 2013/138644 A2

METHODS AND COMPOSITIONS FOR DISCRIMINATION BETWEEN CYTOSINE AND MODIFICATIONS THEREOF, AND FOR METHYLOME ANALYSIS

REFERENCE TO RELATED APPLICATIONS

The entire disclosure of each of the following patent applications is hereby incorporated by reference into the present application: U.S. 61/611,295, filed March 15, 2012; U.S. Application No. 61/722,968, filed November 6, 2012; U.S. Application No. 61/723,427, filed November 7, 2012; U.S. Application No. 61/724,041, filed November 8, 2012; U.S. Application No. 13/804,804, filed March 14, 2013; U.S. Application No. 13/826,395, filed March 14, 2013; and U.S. Application No. 13/827,087, filed March 14, 2013.

BACKGROUND

Cytidine deaminases include activation induced cytidine deaminase (AID) and apolipoprotein B mRNA editing enzymes, catalytic polypeptide-like (APOBEC). These enzymes are DNA mutators capable of inserting mutations into DNA and RNA by deaminating cytidine to form uridine. These enzymes play a role in antigen-driven antibody diversification processes and in an innate defense system against retroviruses. The human APOBEC family consists of 11 members: APOBEC-1 (Apo1), APOBEC-2 (Apo2), AID, APOBEC-3A, -3B, -3C, -3DE, -3F, -3G, -3H and APOBEC-4 (Apo4). Members of the APOBEC-3A family contain either single (A3A, A3C, A3H) or double (A3B, A3DE, A3F, and A3G) zinc-dependent deaminase domain (ZDD).

Attempts have been made to replace sodium bisulfite methylome

sequencing (Frommer, et al., *Proceedings of the National Academy of Sciences,* 89.5:1827-1831 (1992)) by using AID (Larijani, et al., *Molecular Immunology,* 42.5:599-604 (2005)). However problems were identified with the use of native AID including: difficulties in obtaining purified active enzyme due to toxicity in non-natural host cells, incomplete conversion of cytosine to uracil arising from low activity of enzymes and substrate bias, and a lack of *in vitro* assays suitable for selecting AID suitable for methylome sequencing. Hence, methylome sequencing continues to be performed by sodium bisulfite sequencing despite problems associated with this method that include the use of multiple biochemical steps, an inability to distinguish methyl from hydroxymethylcytosine, the requirement for heat to denature the DNA, additional shearing of DNA into small fragments by the chemical treatment, and a limitation on the length of a DNA that can be sequenced.

SUMMARY

In general, in one aspect, a protein variant, for example, a variant of a native cytidine deaminase is provided that includes a protein having at least 90% sequence homology or identity to APOBEC-3A and has at least one mutation. In an embodiment, a protein is provided having at least 90% sequence homology or identity to: (a) SEQ ID NO:1 and comprising at least one mutation at a position corresponding to an amino acid position selected from the group consisting of 25, 45, 109, 123 and 180 of SEQ ID NO:1; or (b) SEQ ID NO:2 and comprising at least one mutation at an amino position corresponding to 23, 25, 29, 45, 69, 104 or 123 of SEQ ID NO:2 or (c) SEQ ID NO:3 (AID) and having at least one mutation corresponding to a deletion at 117 of SEQ ID NO:4.

Embodiments include combining with the protein variant at least one of a purified oxygenase, a polymerase, a polynucleotide and/or at least one primer and dNTPs.

In general, in another aspect, an *in vitro* mixture is provided that includes a cytidine deaminase and a purified oxygenase.

Embodiments of the mixture may include one or more of the following features: the cytidine deaminase may include AID or mutant thereof; APOBEC or a mutant thereof; APOBEC-3A or a mutant thereof; or the compositions described above; and/or the oxygenase may be a methylpyrimidine oxygenase, for example, mYOX (eg mYOX1) or TET (eg TET1 or TET2 or TET3), a DNA polymerase; and/or at least one primer and dNTPs.

In general in one aspect, an *in vitro* mixture is provided that includes a cytidine deaminase and a DNA polymerase. The *in vitro* mixture may further include a polynucleotide and at least one primer and dNTPs.

In general, in one aspect, a method is provided for determining, for a cytidine deaminase, a cytosine preference according to an adjacent nucleotide, which includes: (a) reacting a polynucleotide (optionally single-stranded) containing a cytosine with a cytidine deaminase to convert the cytosine to uracil where the cytosine may be adjacent to any of adenine (A), guanine (G), thymine (T), uracil (U) or cytosine (C); (b) reacting the product of (a) with a glycosylase and an AP endonuclease so as to cleave the polynucleotide at the uracil; and

(c) detecting the cleavage product from (b) to determine how the activity of the cytidine deaminase for the cytosine adjacent to any of A, G, T, U or C.

An embodiment may include one or more features for example, the cytidine deaminase may include AID or mutant thereof, APOBEC or a mutant thereof, APOBEC-3A or a mutant thereof or the compositions described above, and/or the polynucleotide may be single-stranded and may be labeled at one end.

In general in one aspect, a method for differentiating an unmethylated cytosine (C) or a 5-methylcytosine (5-mC) from a 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC), 5-carboxycytosine (5-CaC) or 5-glycosylated hmC (5-ghmC) that includes reacting a polynucleotide optionally containing C, 5-mC, 5-hmC, 5-fC, 5-CaC and/or 5-ghmC, with a cytidine deaminase wherein the 5-mC is converted to a T and only the C is converted to a U; and (b) amplifying or cleaving the polynucleotide to identify the location of at least one converted nucleotide in the polynucleotide.

For aspects of the invention described herein, the cytidine deaminase may be a protein variant as described above; the polynucleotide may be single- stranded, the oxygenase maybe a methylpyrimidine oxygenase such as mYOX1 or a 5-mC oxygenase such as TET1, TET2 and TET3, for example, TET1.

In an embodiment of the method, 5-mC may be differentiated from C, by additionally reacting the polynucleotide prior to (a) with an oxygenase so as to generate a sequence wherein only C is altered to

uracil (U).

A further embodiment includes sequencing the polynucleotide in which C is converted to U only and sequencing the polynucleotide obtained in (a) where C is converted to U and 5-mC is converted to T and comparing the sequences to characterize 5-mC in the polynucleotide or alternatively, comparing the sequence of the deaminated polynucleotide from (a) with the sequence of the untreated polynucleotide.

Additional steps may include sequencing a first sample of the polynucleotide after the reaction with the oxygenase and cytidine deaminase to generate a first sequence; sequencing a second sample of the polynucleotide after a reaction with cytidine deaminase but not with the oxygenase to generate a second sequence; and optionally sequencing a third sample of the polynucleotide absent a reaction with the cytidine deaminase or the oxygenase; and comparing the first and at least one of the second and third sample sequences to detect cytosine and 5-mC or comparing the second sample sequence with at least one of the first and third sample sequences.

In another embodiment, cleaving the polynucleotide further includes cleaving the polynucleotide with a glycosylase and endonuclease at a U or cleaving the polynucleotide after DNA amplification, with a restriction endonuclease that recognizes a site after conversion of C to T in the polynucleotide.

Because mYOX1 acts as an oxygenase on single-stranded polynucleotide substrates, cytidine deaminases which also act on

single-stranded polynucleotides can be added after the oxygenase reaction without the need to manipulate the substrate and optionally without changing the buffer or reaction vessel. It may be desirable to form the *in vitro* mixture described above during a polynucleotide analysis for the presence of 5-mC or it may be desirable to form an *in vitro* mixture of the oxygenase and cytidine deaminase prior to the reaction where the conditions are modulated so that the oxygenase acts on the polynucleotide substrate before the cytidine deaminase acts.

In an embodiment of the method, cytidine deaminase having the characteristics of the protein described above is added to the reaction mixture containing the oxygenase after completion of the oxidation reaction.

In general in one aspect, a method for differentiating a methyl 5-mC from a C is described that includes: reacting a first polynucleotide with sodium bisulfite reaction reagents followed by a cytidine deaminase in the absence of an oxygenase, thereby converting 5-mC to T and converting C to U; and reacting a second polynucleotide, comprising an identical nucleotide sequence, with sodium bisulfite sequencing reagents without subsequent exposure to a cytidine deaminase, thereby converting C to U only.

An embodiment may include amplifying or cleaving the first and second polynucleotides to identify the location of at least one converted nucleotide in the polynucleotides.

An embodiment may include one or of the following features: the

cytidine deaminase is APOBEC-3A or variant thereof; the cytidine deaminase is a protein having at least 90% sequence homology with (a) SEQ ID NO:1 and comprising at least one mutation at a position corresponding to an amino acid position selected from the group consisting of 25, 45 109, 123 and 180 of SEQ ID NO:1; (b) SEQ ID NO:2 and comprising at least one mutation at an amino position corresponding to 23, 25, 29, 45, 69, 104 or 123 of SEQ ID NO:2 or (c) SEQ ID NO:3 (AID) and having at least one mutation corresponding to a deletion at 117 of SEQ ID NO:4; the 5-methylcytosine oxygenase is TET1, TET2 or TET3; the methylpyrimidine oxygenase is mYOX1 and/or the polynucleotide has a length of greater than 1Kb.

An embodiment may include one or more of the following features: adding the cytidine deaminase to the reaction mixture containing the oxygenase after completion of the oxidation reaction; and/or performing the method at a temperature less than 60°C.

The embodiments described above may be used to construct a methylome map.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows that APOBEC-3A expresses well in a cell-free transcription translation system (PURExpress®, New England Biolabs, Ipswich, MA) compared with *E.coli* transformed with DNA encoding APOBEC-3A which did not produce a detectable amount of protein as determined by gel electrophoresis.

Lane 1: Protein ladder (New England Biolabs, Ipswich, MA) arrow

approximates to a 25 kd band on the gel.

Lanes 2-7: Human APOBEC-3A was subcloned under a T7 phage promoter (2 clones, A3A-1 and A3A-2) and expressed in T7 Express Competent *E.coli* (New England Biolabs, Ipswich, MA). U = uninduced, T = after induction with 0.5 mM IPTG and cultivating 3 hours at 30°C, S = soluble fraction after induction with 0.5 mM IPTG and cultivating 3 hours at 30°C. No distinct bands corresponding to APOBEC-3A were observed.

Lanes 8-9: The same clones (A3A-1 and A3A-2) were expressed using PURExpress; the black arrow points to APOBEC-3A protein.

Lane 10 contains a control sample with no plasmid (A3A-1 or A3A-2) added in PURExpress. No band corresponding to AID or APOBEC-3A was observed.

Lane 11: A gene for human AID was subcloned under a T7 phage promoter, and expressed using PURExpress producing a band having a slightly larger molecular weight compared to APOBEC-3A than expected

Figure 2A-F shows an assay for determining the conversion of C to U using APOBEC-3A.

Figure 2A shows 44 bp single-stranded DNA (ssDNA) substrates having an internal unmodified C adjacent to a T and a 3'-FAM marker. The DNA is reacted with APOBEC-3A that converts a C to a U and with USER™ (New England Biolabs, Ipswich, MA). USER removes then

cleaves at the U thus generating two fragments where the labeled fragment is 26 bp.

Figure 2B shows a urea gel which reveals the presence of the 26 bp fragment in all samples except the control which was not reacted with A3A and hence only an uncleaved 44 bp band can be seen.

Figures 2C-2F show the cleavage pattern for an APOBEC-3A obtained by *in vitro* transcription-translation in the PureExpress system. This rapid easy assay shows the extent of deamination of APOBEC-3A for AC (Figure 2C), CC (Figure 2D), GC (Figure 2E) and TC (Figure 2F) substrates at various concentrations.

Figure 3 shows that primate APOBEC-3A family members vary in target sequence preferences: left to right: Human A3A, *Macaca mulatta* A3A, *Pan troglodytes* A3A, and *Pan paniscus* A3A.

Figure 4 shows that primate APOBEC-3A family members vary in target sequence preferences: left to right: *Pango pygmaeus* A3A, *Actus trivigatus* A3A, *Hylobates lar* A3A, and *Saimiri sciureus* A3A.

Figure 5A-B shows that mutations in human APOBEC-3A polypeptide change target sequence preferences.

Figure 5A shows mutations in the human APOBEC-3A polypeptide sequences that alter target sequence preferences.

Figure 5B shows activity of mutant enzymes where the mutation is described on the left and the assay is as described in Figure 2A.

Figure 5C shows mutations in *Pan troglodytes* APOBEC-3A that alter sequence preferences.

Figure 5D shows the activity of mutant enzymes where the mutations are listed on the left and cleavage is shown for oligonucleotides containing single AC, CC<GC and TC from left to right using the assay shown in Figure 2A.

Figure 6A shows the wild type sequence (SEQ ID NO:3) and a mutant sequence (SEQ ID NO:4).

Figure 6B shows the AID wild type and an AID mutant described in Figure 6A and the effect of this mutation on altering target cleavage preferences using urea gel electrophoresis and a substrate an oligonucleotide having a single C immediately preceded by an A.

Figure 7A and 7B shows an assay for detection of 5-mC in a polynucleotide.

Figure 7A shows a 71 base ssDNA substrate having an internal $T^{m}CTAA$ that is reacted with APOBEC-3A. Conversion of ^{m}C to T by APOBEC-3A and subsequent PCR amplification results in conversion of the sequence to a double-stranded product of PCR containing a TTTAA sequence which can be cleaved by the restriction endonuclease MseI (New England Biolabs, Ipswich, MA) (which recognizes TTAA) into two fragments, a 39 base pair fragment and a 77 base pair fragment. The bands can be readily identified using gel electrophoresis.

Figure 7B shows a gel demonstrating the presence of cleavage products in the form of two reduced molecular weight bands on the gel.

Figure 7C shows that the ability of a deaminase to convert 5m to T is unaltered in the presence of an adjacent U where the U might arise after conversion from a C to U by a bisulfite sequencing reaction.

Figure 7D shows a two-fold serial dilution of APOBEC-3A on the reaction described in Figure 7C. Lanes 3-9 show 77 bp and 39 bp fragments and no 116 bp DNA demonstrating complete conversion of 5-mC to T.

Figure 8A and 8B shows that 5-hmC is recognized as a C in the presence of APOBEC-3A.

Figure 8A provides the labeled single-stranded polynucleotide synthesized with a 5-hmC or hydroxymethyl uracil (5-hmU).

Figure 8B is a gel which shows that in the presence of USER, only the control polynucleotide containing 5-hmU was cleaved into two fragments. The polynucleotide with 5-hmC remained intact for all enzyme concentrations tested.

Figure 9A-B shows prior art methods for identifying 5-mC and 5-hmC residues in a substrate DNA.

Figure 9A shows a method called OxBS-Seq (Booth, et al. *Science*, 336.6083:934-937 (2012)). This method requires two bisulfite sequencing steps. In the first sequencing step, 5-hmC in

genomic DNA is oxidized to formylcytosine (5-fC) by KRuO4, a reagent that is destructive for DNA. 5-fC is subsequently converted into T by sodium bisulfite treatment. In the second bisulfite sequencing step, genomic DNA is subjected to bisulfite treatment without KRuO4 treatment. The first sequencing step reveals sites of 5-mC; this information is subtracted from the 5-mC plus 5-hmC sites provided by the second traditional bisulfite sequencing step.

Figure 9B shows a method called TAB-Seq (Yu, et al., *Cell,* 149:1368-1380 (2012)). This method can distinguish 5-hmC from 5-mC after one bisulfite sequencing step. However, multiple sequential enzyme reactions are required. 5-hmC is selectively protected from TET-mediated oxidation and bisulfite conversion by βGT-catalyzed glucosylation of 5-hmC to glycosylated 5-hmC. Next, 5-mC is oxidized by TET to 5-caC, and subsequently converted into T after bisulfite treatment and PCR. Therefore, TAB-Seq reveals sites of 5-hmC in genomic DNA.

Figure 10A-C shows an overview of locus specific differentiation of 5-mCs from 5-hmCs in bisulfite converted DNA.

Figure 10A shows a substrate with C, 5-mC and 5-hmC. Sodium bisulfite sequencing converts the C into U but does not affect 5-mC or 5-hmC.

Figure 10B shows the same initial substrate as in Figure 10A except the sample is deaminated using APOBEC-3A or AID so that the 5-mC is converted to a T, the C is converted to a U and the 5-hmC is unchanged.

Figure 10C shows the same initial substrate as in Figure 10A but this time, oxidation of the 5-mC and 5-hmC is achieved using TET1 or a methyl pyrimidine oxygenase (mYOX1) to convert the 5-mC and 5-hmC to 5-caC. Sequencing identifies these bases as C.

Figure 11 show a method to detect 5-hmC and 5-mC in a two-step method that first utilizes sodium bisulfite after which the sample is split into two aliquots where one aliquot is not further treated while the other aliquot is deaminated. A comparison of sequences reveals that those Cs in the non-deaminated aliquot that are absent in the deaminated aliquot are 5-mC. Because enzymatic deamination achieves the conversion of C to U, the sodium bisulfite step is optional if it is followed by an enzymatic deamination step.

Figure 12A-D shows examples of sequences and their fate as they are utilized in the method described in Figure 10.

Figure 12A shows a 389 bp substrate in which the methyl CpG are underlined.

Figure 12B shows a 389 bp substrate where after sodium bisulfite conversion, all Cs converted to Us except methyl in CpG sites.

Figure 12C shows a 389 bp substrate where after treating with APOBEC-3A enzyme. 5-mCs at CpG sites were converted to Ts.

Figure 12D shows that after PCR amplification all Us become Ts.

Figure 13 shows a method for detecting C, 5-mC and 5-hmC in the absence of a bisulfite sequencing step. An oligonucleotide is either reacted with an oxygenase (for example, TET1 or mYOXI) and a deaminase (APOBEC-3A or AID) or with a deaminase alone and then amplified and sequenced. APOBEC-3A converts a C to U but 5-mC and 5-hmC are oxidized to 5-caC which will be identified as C during sequencing. Thus the combination of APOBEC-3A or AID with an oxygenase has the same effect as sodium bisulfite sequencing. If the same oligonucleotide is treated with APOBEC-3A or an AID only then 5-mC will be converted to a T. If the amount of C and 5-mC is known than the sum total of 5-hmC+5-fC+5-caC can be calculated.

DETAILED DESCRIPTION OF THE EMBODIMENTS

The term "cytidine deaminase" is intended to encompass a naturally occurring enzyme, or an enzyme which is not known to be naturally occurring having at least 85%, 90%, 95% and 99% sequence similarity or identity to the wild type enzyme, where the enzyme may be mutated, the mutations optionally including one or more artificially introduced point mutations or deletions.

A wide range of cytidine deaminases have been identified from humans and other species by means of sequence homology. Eleven members of the human APOBEC-3A family are listed in Table 1. The table also lists preferences in recognition of nucleotide motifs, DNA substrate preference for single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA), and biological functions of the enzymes. The cytidine deaminase family of proteins can be identified in amino acid similarity searches through the occurrence of the ZDD (H/C)-x-E-x25-30P-C-x-x-C. The sequence for human APOBEC-3A is provided in

Figure 5 (SEQ ID NO:1) and human AID is provided in Figure 6 (SEQ ID NO:3).

The term "oxygenase" includes enzymes that catalyze the conversion of 5-mC and 5-hmC to 5-fC to 5-caC. The family that includes mYOXI is variously referred to as methylpyrimidine oxygenase, a cytosine oxygenase, and a 5-methyl oxygenase. Examples of mYOX include the following:

Name	Accession #	SEQ ID NO:	SEQUENCE
mYOX1	XP_002667965.1	24	MTTFKQQTIKEKETKRKYCIKGTTANLTQT HPNGPVCVNRGEEVANTTTLLDSGGGINK KSLLQNLLSKCKTTFQQSFTNANITLKDEK WLKNVRTAYFVCDHDGSVELAYLPNVLPK ELVEEFTEKFESIQTGRKKDTGYSGILDNS MPFNYVTADLSQELGQYLSEIVNPQINYYIS KLLTCVSSRTINYLVSLNDSYYALNNCLYPS TAFNSLKPSNDGHRIRKPHKDNLDITPSSL FYFGNFQNTEGYLELTDKNCKVFVQPGDVL FFKGNEYKHVVANITSGWRIGLVYFAHKG SKTKPYYEDTQKNSLKIHKETK
mYOX6	XP_002674105.1	25	MPMNYITSDLKTQLGEYLIGIVNPMLDETIT AALEILSPRTINYLTSLPHPYHILNNCIYPST AFNYLEPQIEKHRIKNAHKDTRDATPSVLF YLGDYDEKEGYLEFPEQNCKVFVKPGDLLL FKGNKYKHQVAPITSGTRLGLVYFAHKACK VMDFYDDYQKESLNKHKQQNQ
mYOX4	XP_002676528.1	26	MSINTTFNQKTTQSGEPPMMMRMTNSSTP PLTPKNCLPIFVYNDYGKLIREEQQQPTDII TNNNNSMMRSMPTTNRWETNPQTPLSVS PFQPLLPIPNFSHAFIVGNLPPSVSVRRKNR KMSEKPKNNSAPSKIMHQLELSVLNNQRR IAPKGPLADISNIQLPQQESTNKSNNTTPK KPRIRQLMLTTPLRESLQSNQSARSKYIDE EANNYSINDSPETTIIKTSNTKDSEHKAAM ATNLGLSTDDFECKPFETTTLPSVIDKNYLV VDKEGCTQLALLPNHIPTSVCKLIEVKCRK VSNLRHALKIQKASFYVNWWTKSQPMGY MCKDNESEIGKVVNEIAELLSDHCRNLLR

			MCNERVYKKISELKEDKFFAPCICFNILEHD LESRITKFHHDKMDYGVSVLFYFGDYSRG NLNVLDAGSSSTIVTRPGDAVILRGNYYKH SVQNIEPGNNKARYSIVFFAHSTHFLKKKY ELSPAAAKKAFLVDNPDFVSIKKRKQASSS SDVSVKKSKKSTEDNVEFIQTHTYLGNGY KSGHKNYQYYVKFNNSDQKEWKSYESLPK QAVASYWVKFKKLKSLSNQ
mYOX7	XP_002668594.1	27	MLEAQHHKLTIYTGMWGHMKPCVFIAADN CNKSGETIVENLLFKLGKIGSKLMEILSPFT MNFLSSLDPEIFLNHDLFPISATNFMIPGNK HRILKPHKDNQDVGLCIIFYFGNYNAPLEF VNKGSVFNTERGDVLLMRGSHFRHVVKPV DNGLLEHVHDPMRISVVLFAHKSLKMNPS YFLNAGSALKAHDEDFPEKAKKRKKKRK
mYOX8	XP_002676954.1	28	MFLRNILPENTTTEVTNILDKINQRRSKENY YIGSWGKSSSFLFKTNDTIFNELSSQFIKII NLLKNYVLEILKFGNNKMRKFLEKYNSSDF LSIYPTVCFNFLDKSVDENRILHIHPDKEDT GTSLIFYFGKFKGGAISFPELNFKLMVQSA DVLLFDGKNNLHAVESLHGKDDVRYSVVF FAHKADLGKTSYPMNRGEVMKGIKNKINN
mYOX5	XP_002668409.1	29	MDIGIDWRGTHFRHKNHLVKEEVCDRTN WIVLCPNGQVDIAFFPNAIPEELCLEMETV VANSDVDILSCKKAIIDGSWTRYGNGIYPV KTITTNQSILLHELNDKCGPFVLDKLKHINK NMFNKLDNINEDIKNYKIFAKYPTLALNVS HNENYNISKKPYRKHTDGNDIGLGVLTYFG SEIIEGGNLIIHIENLKVFNFPIQRRDLVFLN SKFYAHQVTKVTSGIRFGLVYFAGEAHFRV RNNDDFLPALPFNANDKELREERSKKGRK SMNEYKKRFLKKYLREKKKINKKRVKCKNK LK
mYOX2	XP_002682154.1	30	MGPLHVSQHDKKKPKHRRRKKQFLKAQAL TRVCWENEKSIDESGKTRVYKMIKEWEFL KGNNIQSNEPILSVYGVNDTIPKEISSNTII VTKEGMVEMALLKSVLPPSLLEECTQLCRE MSEWLATEKDIDKGSFFSGWWTMNMPM GYKCADSFRFELVDTKVKQIQALLHDTFQH ILELANPKLFAKLSKLTERGQTPVVCFNMIP TRNESVKEKFQGSYKSTDKVNRPKTNHRD RNDMGISAMFYMGKFGGGSLQLIRVNEHT PKTLVHIQAGDVVLLRANKYRHAVSPTRPQ SFPLANSSQTEVDDVKICENSSPTLNNPQA

			DDNTPTLINTCPKQEPTDGDNPVQSSKEP SNDYEQKRFSFIFFAHRSHFKHSKVYCGM GQRQALNAFKADHPYYQSQRMKKKLGDD CLDQSLILTEKRKPIKRNYALFNECGDDKQ EESDEEEYQQYEPKPTTEEYTIKVIVDHEKV FKGSDQSRKSYLYHIQWLGYPDETWEPYE HLDDCQVFEDYLKHHNISLFDEEEEDRKV DDSMLLPAWMHEDESLFEALLPIICCSTDN PRHHLDDVPPFDFNY
mYOX3	XP_002668005.1	31	MTEIVELSNIEPKDQKQAIIGGTWNRYGNS IEIVAGISDENNTLLDNLTNCCESFVLDKL WHLNRSMYNKLDTIEEKIKNFKTYAKYPSL ALNLLCKENYNGKVKPYRKHIDPNNNGMD VLMFFGKTFEGGNLIVSYHYTNIDFRMFTLP IQSGDLVFLNSRIYHHKVTKVTSGVRCGLV FFAGLDHFSVRKANYKKVKKEEYQKNMDD KLLALPFQQKDKDLRIERTKTGRKEIKQFH KNLQNNLPNKKRKK

The term "polynucleotide" includes a DNA or RNA having ranging in size from a few nucleotides (for example, 10 nucleotides) to a genome length. A polynucleotide as used herein may also be 1 kb or 2 kb or 5 kb or 10 kb in length unless otherwise specified.

The term "DNA polymerase" includes any polymerase suitable for amplifying DNA by isothermal or temperature cycling amplification methods.

In general "detecting", "determining" and "comparing" refer to standard gel based techniques such as SDS gels or TBE-urea gels described in the examples and equivalent methods well known in the art. These terms may be applied to sequencing, where DNA sequences are compared. There are a number of sequencing platforms that are commercially available and any of these may be used to determine or compare the sequences of polynucleotides.

The term "sodium bisulfite sequencing reagents" refers to a standard method for detecting 5-mC as is described in Frommer, et al., *Proceedings of the National Academy of Sciences,* 89.5:1827-1831 (1992).

A number of problems had to be solved to achieve the present embodiments. These include:

(a) a means to generate sufficient quantities of purified APOBEC-3A/AID to test reproducibly for deamination specificity;

(b) a simple, reliable assay was developed to determine whether the activity and specificity of an cytidine deaminase or its homolog or derivative might be suitable for the desired purpose; and

(c) determination of whether and to what extent any purified cytidine deaminase obtained from a natural source could deaminate a C adjacent to an A, T, G or C.

Means for generating sufficient quantities of purified APOBEC-3A

An *in vitro* transcription and translation system was successfully utilized to generate purified active APOBEC-3A and AID from a synthetic gene sequence. The product of synthesis could then be tested on synthetic oligonucleotide substrates containing a modified C in varying sequence contexts. The sequences of the oligonucleotides can be essentially any sequence containing a C or modified C although in embodiments of the assays, the oligonucleotide preferably has a single C unless otherwise specified with an alternative base immediately preceding the C. Figures 8-11 show that APOBEC-3A and

AID express well in *in vitro* transcription translations systems.

An assay for cytidine deaminases

Synthetic oligonucleotide substrates were prepared which contain a single internal C and a 3′ label (Integrated DNA Technologies, Coralville, IA). A synthetic oligonucleotide with multiple internal Cs may also be used as a substrate, to generate multiple fragments in the assay described in Figure 2. A synthetic oligonucleotide can be substituted by a naturally occurring DNA fragment which is denatured so as to provide single-stranded fragments suitable for reacting with APOBEC-3A and AID. Using substrates designed as described above, assays were designed to: (a) analyze the specificity of an APOBEC-3A or AID or variant thereof; (b) differentiate C from 5-mC; and (c) differentiate 5-hmC from 5-mC.

(a) Use of a cleavage agent for breaking ssDNA into 2 fragments at a cytidine deaminase modified nucleotide.

This assay for determining the activity of a cytidine deaminase relies on a change in a ssDNA that results in selective cleavage using a reagent. Although the detectable marker illustrated in Figure 2A and Figure 2B and in Example 2 was a fluorescent label, any label known in the art capable of attaching to the 3′ end or the 5′ end of an oligonucleotide and being detectable on a gel or other separation device, can be used. Examples of detection labels and capture tags for oligonucleotides are described in US Patent No. 6,368,801.

The conversion by APOBEC-3A or AID of a C into a U can be detected rapidly and simply by reacting the oligonucleotide with a

glycosylase/endonuclease mixture such as the commercially available USER or an equivalent which removes the U from the DNA thereby generating two fragments from the oligonucleotide, one of which retains the fluorescent label. Since the labeled cleavage product is significantly smaller than the full-length oligonucleotide, it can readily be detected by size separation such as shown using gel electrophoresis. Figures 3 and 4 show assay results for 8 human and simian APOBEC-3A enzymes using the rapid assay shown in Figure 2 on PURExpress generated samples. An advantage of this assay and others that utilize cytidine deaminase is that the temperature of the reaction may be less than 60°C for example, less than 50°C, or less than 40°C.

(b) Restriction endonuclease cleavage of dsDNA only after cytidine deaminase induced modification of a C to a T.

A change of a C to a T in a reaction with APOBEC-3A or AID was designed to result in the formation of a specific restriction endonuclease cleavage site that was not present previously. When this modified oligonucleotide was amplified, the resulting dsDNA could be cleaved by the restriction endonuclease to generate two fragments where previously only one was present. Example 3 and Figure 7A-D describe one instance of this method. PCR amplification is described here though any form of amplification can be used including various isothermal amplification methods such as transcription mediated amplification, nucleic acid sequence-based amplification, signal mediated amplification of RNA technology, strand displacement amplification, rolling circle amplification, loop-mediated isothermal amplification of DNA, isothermal multiple displacement amplification, helicase-dependent amplification, single primer isothermal

amplification, and circular helicase-dependent amplification.

(c) Conversion of 5-hmC to 5-hmU.

The oligonucleotides were reacted with a cytidine deaminase so that the 5-hmC was converted to 5-hmU. 5-hmU could be cleaved with a glycosylase and an endonuclease. In the example, SMUG (New England Biolabs, Ipswich, MA) was used with EndoVIII (New England Biolabs, Ipswich, MA) to cleave the synthetic oligonucleotide. A cleavage fragment could be detected by a label either at the 3' end or the 5' end of the oligonucleotide although a same, similar or different label could be used at either end of the oligonucleotide substrate to facilitate detection of the cleavage product. Example 4 and Figure 8 show results obtained using a 3'FAM label on a synthetic oligonucleotide with a single internal 5-hmC located in a sequence TCT in a substrate oligonucleotide and treated with APOBEC-3A.

The ability to readily make protein variants and the robust assay both described herein, provides a simple means to generate and evaluate mutants of cytidine deaminases for improved catalytic conversion of C and 5-mC in polynucleotide substrates that may be ssDNA fragments or genomes. An advantage of the methods described herein are that they are suited for analyzing large pieces or even entire genomes without the difficulties of shearing that arise when a temperature or alkali denaturation step is required or where the chemical also cleaves the DNA, as in bisulfite sequencing.

Mutants of APOBEC-3A and AID were generated that had enzymes with reduced or no bias to any particular sequence context for converting a C to U or a 5-mC to T. For example, using a human

APOBEC-3A sequence as a starting point, mutations were introduced into the gene and the mutants tested in the assay described above. The results for 12 mutants of two wild type APOBEC-3A are shown in Figure 5A-D and for 1 mutant of AID in Figure 6A-B. These mutations are intended to be representative and are not intended to be comprehensive. It would be straightforward based on the compositions and methods described herein to select a sequence of a cytidine deaminase such as APOBEC-3A from a sequence database, introduce mutated amino acids into the protein sequence and assay for altered substrate specificity using embodiments of the methods described here.

Various combinations of the 5 mutants of human APOBEC-3A and 7 mutants of *Pan troglodytes* APOBEC-3A are shown in the examples and in Figure 5A-D that have altered cleavage bias of substrates. For example, for human APOBEC-3A, E109 could be combined with one or more of any of G25, S45, R123 and D180, for example E109A could be combined with one or more of any of G25V, S45W, R123H and D180E. Similarly, R123 could be combined with one or more of any of E109 G25, S45, and D180 for example, R123H could be combined with one or more of any of E109Q G25V, S45W, and D180E. Similarly, D180 could be combined with one or more of any of G25, S45, R123 and E109 for example D180E could be combined with one or more of any of G25V, S45W, R123H and E109Q. Similarly, S45 could be combined with one or more of any of G25, E109, R123 and D180 for example S45W could be combined with one or more of any of G25V, E109Q, R123H and D180E. Similarly G25 could be combined with one or more of any of E109, S45, R123 and D180 for example G25V could be combined with one or more of any of

E109Q, S45W, R123H and D180E. In one example, human APOBEC-3A with an E190 mutation for example E109Q was used in embodiments of the method. In another example human APOBEC-3A with G25, S45, E109, R123 and D180 mutations for example G25V, S45W, E109Q, R123H and D180E mutations were used.

For *Pan troglodytes* APOBEC-3A, one or more of mutations at positions corresponding to 23, 25, 29, 45, 69, 104 and 123 can be introduced into (SEQ ID NO:2) to alter the sequence preference for the enzyme. Examples of specific mutations correspond to 23N, 25V, 29H, 45W, 69R, 104W and 123H. Examples of combinations of mutations include combining a mutation at positions corresponding to 23 with one or more or two or more mutations at positions corresponding to 25, 29, 45, 69, 104 and 123; or 25 with 29, 45, 69 104 and 123, or 29 with 45, 69, 104 or 123, or 45 with 69, 104 or 123, or 69 with 104 or 123, or 104 with 123. Three or more or four mutations selected from positions corresponding to 23, 25, 29, 45, 69, 104 and 123 can be selected or a mutant may be constructed that includes 5 mutations at positions corresponding to 23, 25, 29, 45, 69, 104 and 123 for example 23N, 25V, 29H, 45W, 69R, 104W and 123H.

With the disclosed assay, it is possible to mutate any cytidine deaminase at any site and test the mutant for altered site preference. In one embodiment of the invention, a cytidine deaminase with a site preference is selected to determine the locations of a subset of 5-mC residues present in a target nucleic acid.

In another embodiment of the invention, a cytidine deaminase with little or no site preference is preferred. Accordingly, a mutated

APOBEC-3A having a mutation corresponding to E109Q or S45W from SEQ ID NO:1 or C69R, T23N, or G25V in SEQ ID NO:2 may be selected with these features.

Mutants of wild type AID (SEQ ID NO:3) were created and can be used in embodiments of the methods described herein. For example a mutation at position 117 may be introduced, for example a deletion as shown in Figure 6A (SEQ ID NO:4).

In one embodiment, a method is provided which shows a time course leading to complete conversion to U of all 5-mCs in a substrate containing multiple 5-mCs which when amplified resulted in a T in place of a U (see for example Figure 8A-8B). APOBEC-3A is demonstrated to convert all of the 5-mC into U in a time that is greater than 2 hours although even after 1 hour about 95% is converted. It is expected that manipulating conditions such as one or more of pH, concentration and buffer and selected APOBEC-3A or AID variants results in substantially 100% conversion in a time frame of less than 2 hours.

Methylome sequencing

Sodium bisulfite sequencing has become an established method for mapping 5-mC in a genome as part of an epigenetic study. Unfortunately, sodium bisulfite sequencing cannot differentiate between 5-mC and intermediates of demethylation such as 5-hmC, 5-fC and 5-caC. In brain tissue, there is a significant amount of 5-hmC as well as small amounts of 5-fC and 5-caC while in all tissues, there are at least some of these modified bases. Another problem

associated with sodium bisulfite sequencing is that the method damages the target DNA causing extensive fragmentation. This complicates assembly of maps for a methylome. Another problem of sodium bisulfite sequencing is that it involves multiple chemical steps and therefore is intrinsically inefficient and costly to perform. Nonetheless, an embodiment of a method is provided that facilitates sodium bisulfite sequencing and ameliorates one or more of the above limitations. Accordingly, a one-enzyme step enables mC to be differentiated from a 5-hmC (see Figure 11). The sodium bisulfite reaction which precedes the deamination reaction was shown not to interfere with this reaction (see Figure 11).

Embodiments of the invention include methods for methylome construction that may utilize sodium bisulfite sequencing while reducing the number of steps to determine not just the occurrence of modified bases but the occurrence of methyl bases and not hydroxymethyl bases. Other embodiments do not utilize sodium bisulfite sequencing at all but rather utilize two enzyme reactions, in particular, a demethylase such as methyl-pyrimidine oxygenase or TET or analog thereof and an AID. A comparison between these oxygenases is given below. An example of the class of enzymes referred to as methyl pyrimidine oxygenases is an enzyme identified as mYOXI which is described in U.S. Application No. 13/827,087.

Table 1: Properties of oxygenases

Name	Methyl-pyrimidine oxygenase	5-methylcytosine oxygenase (TET)
Length	~300AA	~1600AA
Reaction temp.	34C	37C
Cofactors	2-oxoglutarate and Fe^{2+}	2-oxoglutarate and Fe^{2+}
Optimal pH	6-6.5	8

Substrate forms	DS-DNA, SS-DNA	DS-DNA
Substrate	5-mC (and depending on enzyme , T)	5-mC
Products	5-mC->5-hmC/5-fC/5-caC, T->5-hmU/5fU/5-caU	5-mC->5-hmC/5-fC/5-caC
Substrate specificity	Converts mCG to >90% 5-caC, coverts mCWG to a mix of 5-hmC/5-fC/5-caC	Similar to mYOX1
ATP effect	Inhibition	stimulation
Conserved Sequence feature	Contains characteristic 2OGFE-domains, presumably for binding 2OG and Fe^{2+}	In addition to 2OGFE-domains, there are long extra sequences.

Figure 10A-C shows a comparison of: sodium bisulfite sequencing in which an unmodified C is deaminated to form a U leaving 5-mCs detected as a C (Figure 10A); enzymatic deamination only (Figure 10B) in which an unmethylated C is converted to U, a 5-mC is converted to a T and no change is observed with 5-hmC which is read as a C in the sequencing reaction; and with an oxygenase reaction step (Figure 10C) in which a 5-mC and 5-hmC are converted to 5-caC which is recognized as C and thereby replicates the sodium bisulfite sequencing reaction. When parallel reactions are analyzed for samples +/- enzymatic deamination, after sodium bisulfite sequencing step, the results reveal which C residues were methyl and which were hydroxymethyl by subtraction (Figure 11).

The reaction pathways in Figure 13 remove the need for a sodium bisulfite sequencing reaction. This method contrasts with TAB-seq (Yu, et al. (2012)) which requires two separate enzyme reactions in addition to sodium bisulfite sequencing. In TAB-seq, 5-hmC is first labeled with a glucosyl transferase, and 5-mC is oxidized prior to sodium bisulfite sequencing to 5-caC which is then converted to 5-caU and hence to T (see Figure 9A-B).

Table 2. A summary of the human cytidine deaminase family of enzymes.

APOBEC-3A	HotSpot motif	Substrate	Function
AID	WRC	ssDNA	Somatic hypermutation Class switch recombination
Apo1	G/CTC	ssDNA	Lipid metabolism
A3A	T/CCA	ssDNA	Inhibits parvovirus, HPV, retroelements
A3B	C/GTC	ssDNA	Inhibits HIV, HBV, retroelements
A3C	TC/TC	ssDNA	Inhibits HBV, HPV, retroelements
A3DE	WWC	ssDNA	Inhibits HIV
A3F	TTC	ssDNA	Inhibits HIV, HBV, retroelement
A3G	CCC	ssDNA	Inhibits HIV, HBV, retroelement
A3H	unknown	unknown	Inhibits HPV
Apo4	unknown	unknown	unknown

EXAMPLES

Example 1: Synthesis of APOBEC-3A, AID and mutants thereof using PURExpress

The AID and APOBEC-3A DNA sequences were codon optimized and reverse synthesized to form the gene sequence. The synthesized DNA was subcloned by TOPO® TA Cloning® (Life Technologies, Carlsbad, CA) under a T7 phage promoter and inserted into NEB 5-alpha F'I^q Competent *E. coli* (New England Biolabs, Ipswich, MA). Mutant proteins were produced with the Q5® Site-Directed

Mutagenesis Kit (New England Biolabs, Ipswich, MA) before being subcloned as above. 200 ng of each plasmid was used as template for *in vitro* synthesis of AID, APOBEC-3A, or their mutants using PURExpress. The *in vitro* reaction was incubated at 37°C for 5 hours, and the synthesized proteins were boiled for 3 minutes, and after precipitation by centrifugation, a sample of the supernatant was loaded on an SDS gel. The results are shown in Figure 1.

Example 2: Biochemical assay for cytosine deamination (see Figure 2)

Table 3: The following components were combined in a 1X solution to convert C to U:

Reaction Component	**Volume, µl**	**Stock**	**Final Concentration**
Nuclease-free water	15.8		Total volume: 20 µl
Oligonucleotide	1	1 µM	50 nM
RNase A	0.2	10 mg/ml	1 µg
NEBuffer 1 (New England Biolabs, Ipswich, MA)	2	10x	1x
PURE extract (APOBEC-3A/AID	1		50 ng

The oligonucleotide was synthesized with a FAM label (Integrated DNA Technologies, Coralville, IA). The reaction mixture was incubated at 37°C for 1 hour, and 10 µl of 1x NEBuffer 1 were added, containing 0.5 µl (0.5 U) of USER, and incubated additional 30 minutes at 37°C.

The products were separated on a 15% TBE-Urea gel (Invitrogen, Grand Island, NY) using the XCell SureLock® Mini-Cell (Invitrogen, Grand Island, NY).

Serial dilutions were performed on samples in order to determine the activity of AID, APOBEC-3A, or their mutants in a fixed ratio (such as 1:1, 1:2, 1:4, 1:8, etc.) as indicated below:

Table 4: Components were combined in a 1X solution or a titration assay of the selected deaminase

Reaction Component w/o deaminase	**Volume, µl**	**Stock**	**Final Concentration**
Nuclease-free water	16.8		Total volume: 20 µl
Oligonucleotide	1	1 µM	50 nM
RNase A	0.2	10 mg/ml	1 µg
NEBuffer 1	2	10x	1x

1 µl of APOBEC-3A or AID (wild type or mutant) after *in vitro* transcription/translation using PURExpress (containing about 50 ng of DNA deaminase enzyme) was added to the first tube in the serial dilution (1:1) which then contained 40 µl of the reaction mixture with deaminase (1x solution). 20 µl from the first tube was placed into 20 µl of reaction components without enzyme in a second tube (1:2) and so forth for the desired numbers of dilutions resulting in 2x, 4x, 8x, 16x and 32x dilutions (1x) in a reaction volume of 20 µl. The reaction mixture was incubated at 37°C for 1 hour, and 10 µl of 1x NEBuffer 1 were added, containing 0.5 µl (0.5 U) of USER, and incubate additional 30 min at 37°C. The products were separated on a 15% TBE-Urea gel using the XCell SureLock Mini-Cell.

Generation of mutants

The APOBEC3A or AID mutants were generated by random mutagenesis or site-specific mutagenesis.

For random mutagenesis the error-prone PCR method according to Cirino, et al., *Methods in Molecular Biology*, 231:3-10 (2003), was

used; although the $MgCl_2$ concentration was increased to 7 mM concentration in the reaction.

For site-specific mutagenesis, the Q5® Site-Directed Mutagenesis Kit (New England Biolabs, Ipswich, MA) was used following the manufacturer's instructions. Target residues were selected according to sequence conservation, or predicted location in loop domains (Kohli, et al., *Journal of Biological Chemistry,* 284.34:22898-22904 (2009)) The mutant proteins were manufactured in cell-free PURExpress® system (see Example 1) and tested for activity as described above in Examples 2 and 3.

Example 3: APOBEC-3A activity assay on 5-mC containing substrate

A. Biochemical assay for 5-mC deamination

The following components were combined in a 1X solution. The following components were combined in a 1X solution in order to convert 5-mC to T (and C to U) in the oligonucleotide.

Table 5: Reaction components for 5-mC deamination

Reaction Component	**1x**
Nuclease-free water	16.5
Oligonucleotide * (1 µM)	1
RNase A (100 µg/ml)	0.2
NEBuffer 1 (10x)	2
PURE extract	0.3

* TATGGGGAAGGTTAGGGAAGATAAGAATAGAATGAAT/iMe-dC/GAAGGATGAATATGAGGTGAGGAGTAGGATGGG (SEQ ID NO:17)
iMe-dC=5-methylcytosine

The reaction mixture was incubated at 37°C for 12-16 hours, or overnight, and PCR reaction was performed followed by MseI

digestion.

Table 6: PCR reaction components

Component	**50µl 1x**
5X Epimark® HS *Taq* Reaction Buffer (New England Biolabs, Ipswich, MA)	10 µl
10 mM dNTPs	1 µl
10 µM Forward Primer	1 µl
10 µM Reverse Primer	1 µl
Deaminated Oligonucleotide (1 µM)	1 µl
Epimark HS *Taq* DNA Polymerase	0.25 µl
Nuclease-free water	35.75

Table 7: PCR cycling protocol

Cycling Step	**Temp**	**Time**
Initial denaturation	95°C	1 minute
Denaturation	95°C	30 seconds
Annealing	56°C	30 seconds
Extension	68°C	20 seconds
Final extension	68°C	5 minutes
Number of cycles	25 cycles	

Table 8: MseI digestion

Reaction Component	**20 µl 1x**
PCR product	8
Nuclease-free water	9.7
NEBuffer 4	2
MseI 10 u/µl	0.3

The reaction mixture was incubated at 37°C for 0.5 hours and the DNA digestion products were analyzed on 2% agarose gel (see Figure 7A-D). The 5-mC was converted to T and the PCR product formed after this conversion could be cleaved by the restriction endonuclease MseI.

B. APOBEC-3A activity assay on $U^{5m}C$ containing substrate

1 µl of human APOBEC-3A from PURExpress system extract was serially diluted 1:1, 1:2, 1:4, 1:8, 1:16, 1:32, 1:64 and reacted with 1pM ssDNA (TATGGGGAAGGTTAGGGAAGATAAGAATAGAATGATUmCTAAGGATGAATATGAGGTGAGGAGTAGGATGGG (SEQ ID NO:17) in NEBuffer 1 (10 mM Bis-Tris-Propane-HCl, 10 mM $MgCl_2$ 1 mM Dithiothreitol pH 7.0) in the presence of RNaseA (1 µg). Reactions were incubated for 14 hours (overnight) at 37°C. Deaminations were detected through PCR reaction (FW primer: 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGTATGGGGAAGGTTAGGGAAG (SEQ ID NO:18), REV primer: 5'-CCTCTCTATGGGCAGTCGGTGATCCCATCCTACTCCTCACCTC (SEQ ID NO:19)) and digestion with MseI restriction endonuclease as described in Example 3 and shown in Figure 7C and 7D.

Example 4: APOBEC-3A activity assay on 5-hmC containing substrate

1 µl of APOBEC-3A from PURExpress extract was reacted with 1 pM fluorescein (F)-labeled ssDNA in NEBuffer 1 (10mM Bis-Tris-Propane-HCl, 10mM $MgCl_2$ 1mM Dithiothreitol pH 7.0) in the presence of RNaseA (1 µg). Reactions were incubated for 60 minutes at 37°C. Deaminations were detected through breakage of DNA at abasic sites generated by 0.5 µl of each of SMUG1 (5u/µl) and EndoVIII (10u/µl) enzyme. The results are shown on a urea gel and show no conversion of 5-hmC to U except using when using the control oligonucleotide containing 5-hmU and SMUG1/EndoVIII which produces the second band. The results are shown in Figure 8A-B in which none of the 5-hmC is observed to be converted to 5-hmU.

Example 5: APOBEC-3A deaminates 5m-C but not 5-hmC in bisulfite converted DNA

A DNA substrate of 489 bp long containing four CpG sequences at positions 80, 174, 239 and 356 was methyl *in vitro* with SssI methylase (New England Biolabs, Ipswich, MA). Methyl substrate was bisulfite converted using EpiMark Bisulfite Conversion Kit, amplified using FW GAGGAGGAAAAGAGAAATGGAGTGTGG (SEQ ID NO:20) and REV CTCACACCTCTCCTCTCACTCAC (SEQ ID NO:21) primers and EpiMark HS *Taq* DNA Polymerase. PCR fragment was inserted into TOPO TA Cloning vector and transformed into NEB 5-alpha F'I^q Competent *E. coli*. When tested by sequencing, it was found that 100% of 5-mCs were identified as C.

An aliquot of the bisulfite converted DNA sample was treated with cytidine deaminase, or a mixture of cytidine deaminases for 4 hours to convert all 5-mCs to Ts (described in Example 1). After the reaction was completed, the deaminated DNA was amplified using FW GAGGAGGAAAAGAGAAATGGAGTGTGG (SEQ ID NO:20) and REV CTCACACCTCTCCTCTCACTCAC (SEQ ID NO:21) primers and Epimark HS *Taq* DNA Polymerase. PCR fragment was inserted into TOPO TA cloning vector and transformed into NEB 5-alpha F'I^q Competent *E.coli*. When tested by sequencing, it was found that 100% of 5-mCs were identified as Ts. The results are shown in Figures 11 and 12.

Example 6: Discrimination between 5-mC and 5-hmC using cytidine deaminase and DNA oxygenase enzymes

Bisulfite sequencing in Example 5 was replaced here by sequential oxygenase and deaminase treatment.

A. Use of TET as the oxygenase

The oxygenase reaction was performed as follows: 100 ng of NIH 3T3 Mouse Genomic DNA (New England Biolabs, Ipswich, MA) was incubated with TET1 (WiseGene, Chicago, IL) in 50 µl of reaction mixture for 3 hours. DNA was cleaned up using Zymo Clean & Concentrator™ kit (Zymo Research, Irvine, CA).

TET is active on dsDNA. 50 µl of TET treated and untreated NIH 3T3 Mouse Genomic DNA in two separate tubes were heated up to 99°C for 5 minutes, spun down briefly and put on ice.

B. Use of mYOX1 instead of TET as the oxygenase

In this example, a double-stranded substrate, containing 5-mC was oxidized with mYOX1. mYOX1 is active on ssDNA.

The sequence of the substrate was: CGGCGTTTCCGGGTTCCATAGGCTCCGCCCXGGACTCTGATGACCAGGGCA TCACA (X=5-mC) (SEQ ID NO:22)

Table 9: Reaction components for reaction of oligonucleotide with mYOX1

Reaction Component	**Volume, µl**	**Stock**	**Final Concentration**
ddH_2O	3		to 20 µL
Bis-Tris pH 6.0	1	1 M	50 mM
NaCl	1	1 M	50 mM
DTT	1	20 mM	1 mM
Ascorbic acid	2	20 mM	2 mM
α-ketoglutarate	2	10 mM	1 mM
$FeSO_4$	1	2 mM	100 µM
Oligonucleotide	4	10 µM	2 µM
MYOX1	5	16 µM	4 µM

The sequence for mYOX1 is as follows

MTTFKQQTIKEKETKRKYCIKGTTANLTQTHPNGPVCVNRGEEVANTTTLLDS
GGGINKKSLLQNLLSKCKTTFQQSFTNANITLKDEKWLKNVRTAYFVCDHDGS
VELAYLPNVLPKELVEEFTEKFESIQTGRKKDTGYSGILDNSMPFNYVTADLSQ
ELGQYLSEIVNPQINYYISKLLTCVSSRTINYLVSLNDSYYALNNCLYPSTAFNSL
KPSNDGHRIRKPHKDNLDITPSSLFYFGNFQNTEGYLELTDKNCKVFVQPGDV
LFFKGNEYKHVVANITSGWRIGLVYFAHKGSKTKPYYEDTQKNSLKIHKETK
(SEQ ID NO:23)

The reaction mixture was incubated at 34°C for 1 hour. 1 µL of Proteinase K (20 mg/mL^{-1}) (New England Biolabs, Ipswich, MA) was added to the reaction mixture and incubated for 1 hour at 50°C. DNA was isolated using the QIAquick® Nucleotide Removal Kit (Qiagen, Germany) and samples were analyzed by LC/MS. These conditions cause 5-mC to be oxidized to 5-hmC.

Oxidation with mYOX1 of Mouse NIH 3T3 genomic DNA

Table 10: Reaction components for oxygenase treatment of genomic DNA

Reaction Component	**Volume, µl**	**Stock**	**Final Concentration**
ddH_2O	15		to 50 µl
Bis-Tris pH 6.0	2.5	1 M	50 mM
NaCl	2.5	1 M	50 mM
DTT	2.5	20 mM	1 mM
Ascorbic acid	5	20 mM	2 mM
α-ketoglutarate	5	10 mM	1 mM
$FeSO_4$	2.5	2 mM	100 µM
NIH 3T3 DNA	10	200 ng/µL	2 µg
mYOX1	5	200 µM	20 µM

Human APOBEC-3A (100 ng from example 1), RNaseA (1 µg) were combined with ssDNA substrate (oxidated and non-oxidated NIH 3T3 Mouse Genomic DNA in separate tubes) and incubated in the reaction buffer (NEBuffer 1, (10mM Bis-Tris-Propane-HCl, 10 mM MgCl2 1 mM Dithiothreitol pH 7.0)) at 37°C for 12-16 hours. Reactions were terminated by incubating at 80°C for 10 minutes.

The oxidized and deaminated DNA or deaminated DNA only were PCR amplified and cloned. 2 µl from each sample (TET1+Deaminase and Deaminase only) were used for the PCR reaction using Epimark HS Taq DNA Polymerase. Primers were designed for deaminated DNA (all Cs become Ts, except oxidized 5m-C or 5hm-C). PCR products were visualized on 1.5% agarose gel and cloned into the TOPO TA Cloning vector and transformed into NEB 5-alpha F'I^q Competent *E.coli*. DNA was isolated from individual colonies and sequenced using nested primers.

The sequence results were interpreted as follows:

(a) Original sequence: GGGTmCGGACCGhmC (SEQ ID NO:7)

(b) After TET and deaminase treatment:GGGTCGGATTGC (SEQ ID NO:11)

(c) After deaminase only treatment: GGGTTGGATTGC (SEQ ID NO:12)

After alignment of sequences, TET and deaminase treatment (b) and deaminase only treatment (c) show Cs that corresponds to 5-hmC but not C. After TET and deaminase treatment only, C also corresponds to 5-mC but after deaminase treatment only, 5-mC corresponds to T.

What is claimed is:

1. An *in vitro* mixture comprising a cytidine deaminase and a purified oxygenase suitable for distinguishing between cytosine and 5-methylcytosine in a polynucleotide.

2. An *in vitro* mixture according to claim 1, wherein the oxygenase is a methylpyrimidine oxygenase or a 5-methylcytosine oxygenase.

3. An *in vitro* mixture according to claim 2, wherein the methylpyrimidine oxygenase is a mYOX.

4. An *in vitro* mixture according to claim 3, wherein the 5-methylcytosine oxygenase is TET1.

5. An *in vitro* mixture according to any of claims 1 through 4, further comprising a DNA polymerase.

6. An *in vitro* mixture comprising a cytidine deaminase and a DNA polymerase.

7. An *in vitro* mixture according to claims 5 or 6, further comprising a polynucleotide and at least one primer and dNTPs.

8. A composition comprising a variant cytidine deaminase having at least 90% sequence homology with (a) SEQ ID NO:1 and comprising at least one mutation at a position corresponding to an amino acid position selected from the group consisting of 25, 45, 109, 123 and 180 of SEQ ID NO:1; and/or (b) SEQ ID NO:2 and comprising at least one mutation at an amino position

corresponding to 23, 25, 29, 45, 69, 104 or 123 of SEQ ID NO:2 and/or (c) SEQ ID NO:3 having at least one mutation corresponding to a deletion at 117 of SEQ ID NO:4.

9. A composition according to claim 8, further comprising at least one of a purified oxygenase, a polymerase, a polynucleotide and/or at least one primer and dNTPs.

10. A method for determining for a cytidine deaminase, a cytosine preference according to an adjacent nucleotide, comprising:

(a) reacting a polynucleotide containing a cytosine with the cytidine deaminase to convert the cytosine to uracil where the cytosine may be adjacent to any of adenine, guanine, thymine, uracil or cytosine;

(b) reacting the product of (a) with a glycosylase and an AP endonuclease so as to cleave the polynucleotide at the uracil;

(c) detecting the cleavage product from (b) to determine the activity of the cytidine deaminase for the cytosine adjacent to any of adenine, guanine, thymine, uracil or cytosine.

11. A method according to claim 10, wherein the cytidine deaminase is characterized according to claim 8.

12. A method according to claims 10 or 11, wherein the polynucleotide is single stranded.

13. A method according to any of claims 10 to 12, wherein the polynucleotide is labeled at one end.

14. A method for differentiating a 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC), 5-carboxycytosine (5-CaC) or 5-glycosylated hydroxymethylcytosine (5-ghmC) from an unmethylated cytosine (C) or a 5-methylcytosine (5-mC), comprising:

(a) reacting a polynucleotide optionally containing C, 5-mC , 5-hmC, 5-fC, 5-CaC and/or 5-ghmC, with a cytidine deaminase wherein 5-mC is converted to a thymine (T) and C is converted to a uracil (U); and

(b) amplifying or cleaving the polynucleotide to identify the location of at least one converted nucleotide in the polynucleotide.

15. A method according to claim 14, wherein the cytidine deaminase is characterized according to claim 8.

16. A method according to claim 14 or 15, wherein cleaving the polynucleotide further comprises cleaving the polynucleotide with a glycosylase and endonuclease at a uracil (U) or cleaving the polynucleotide after DNA amplification, with a restriction endonuclease that recognizes a site after conversion of unmethylated cytosine (C) to thymine (T) in the polynucleotide.

17. A method according to any of claims 14 to 16, wherein the polynucleotide is a synthetic oligonucleotide.

18. A method according to claim 17, wherein the oligonucleotide is labeled at one end.

19. A method according to any of claims 14-18, further comprising differentiating a 5-methylcytosine (5-mC) from unmethylated cytosine (C), 5-hydroxymethylcytosine (5-hmC), 5-glycosylated hydroxymethylcytosine (5-ghmC), 5-carboxycytosine (5-CaC) and 5-formylcytosine (5-fC) by reacting the polynucleotide prior to (a) with an oxygenase so as to generate a sequence wherein only C is altered to uracil (U).

20. A method according to claim 19, further comprising comparing the polynucleotide sequence in which unmethylated cytosine (C) is converted to uracil (U) with the polynucleotide sequence obtained in (a) where C is converted to U and 5-methylcytosine (5-mC) is converted to thymine (T).

21. A method according to any of claims 14-20, wherein the polynucleotide in (a) is single-stranded.

22. A method according to any of claims 14-21, further comprising, comparing the sequence of the deaminated polynucleotide from (a) with the sequence of the untreated polynucleotide.

23. A method according to any of claims 14-22, further comprising: constructing a methylome map.

24. A method for differentiating unmethylated cytosine (C) from 5-methylcytosine (5-mC), 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC), 5-carboxycytosine (5-CaC) and/or 5-glycosylated hydroxymethylcytosine (5-ghmC) in a polynucleotide,

comprising:

(a) reacting a first sample of the polynucleotide optionally containing C, 5-mC, 5-hmC, 5-fC, 5-CaC and/or 5-ghmC , with an oxygenase followed by a cytidine deaminase, thereby converting C to uracil (U) and converting 5-mC to 5-hmC and 5-CaC; and

(b) amplifying or cleaving the polynucleotide to identify the location of the U in the polynucleotide.

25. A method according to claim 24, wherein the polynucleotide from step (a) is single-stranded.

26. A method according to claim 24 or 25, further comprising sequencing the first sample of the polynucleotide after the oxidation and deamination reactions to generate a first sequence; sequencing the second sample of the polynucleotide after the deamination reaction only to generate a second sequence; and optionally sequencing a third sample of the polynucleotide absent the oxidation or deamination reaction; and comparing the first sample sequence with the second sample sequence and optionally the third sample sequences to detect and differentiate unmethylated cytosine (C) and 5-methylcytosine (5-mC).

27. A method according to any of claim 24 through 26, wherein the oxygenase is a methylpyrimidine oxygenase or a 5-methylcytosine oxygenase.

28. A method according to claim 27, wherein the 5-methylcytosine oxygenase is TET1.

29. A method according to claim 27, wherein the 5-methylpyrimidine oxygenase is a mYOX.

30. A method for differentiating a 5-methylcytosine (5-mC) from an unmethylated cytosine (C), comprising:

reacting a first sample of a polynucleotide with sodium bisulfite sequencing reagents followed by a cytidine deaminase in the absence of an oxygenase, thereby converting 5-mC to thymine (T) and converting C to uracil (U); and

reacting a second sample of the polynucleotide, with sodium bisulfite sequencing reagents without subsequent exposure to a cytidine deaminase, thereby converting C to U while retaining 5-mC as 5-mC.

31. A method according to claim 30, further comprising amplifying the first and second samples of the polynucleotide for sequencing to identify the presence and associated location of at least one converted nucleotide in the polynucleotide.

32. A method according to claim 30 or 31, wherein the cytidine deaminase is APOBEC-3A or variant thereof.

33. A method according to any of claims 30-32, wherein the cytidine deaminase is characterized according to claim 8.

34. A method according to any of claims 30-33 wherein the oxygenase is 5-methylcytosine oxygenase or a 5-methylpyrimidine oxygenase.

35. A method according to claim 34, wherein the 5-methylcytosine oxygenase is Tet1.

36. A method according to claim 34, wherein the 5-methylpyrimidine oxygenase is a mYOX.

37. A method according to any of claims 30-36, wherein the cytidine deaminase is added to the reaction mixture containing the oxygenase after completion of the oxidation reaction.

38. A method according to any of claims 30-37, wherein the method is performed at a temperature less than 60°C.

39. A method according to any of claims 30-38, wherein the polynucleotide has a length of greater than 1Kb.

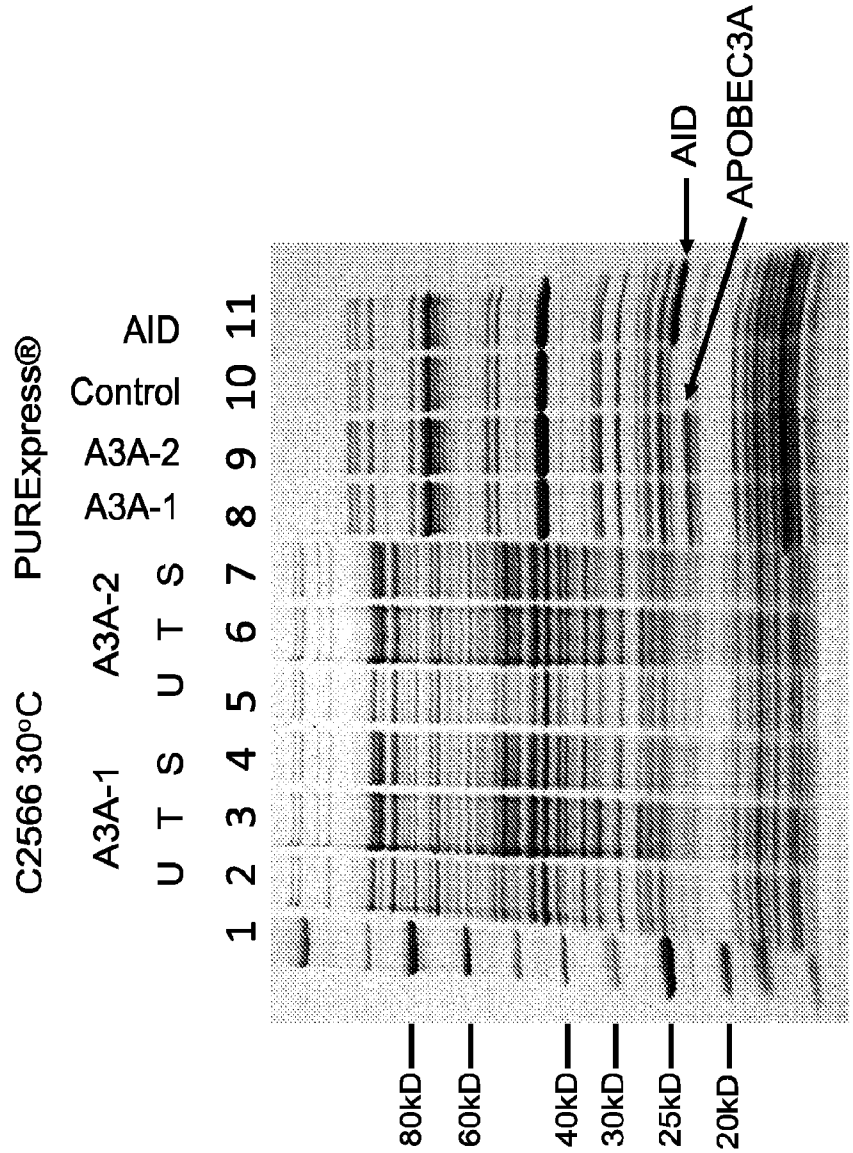

C2566 30°C
PURExpress®
A3A-1
A3A-2
AID
Control
A3A-2
A3A-1
U T S U T S
1 2 3 4 5 6 7 8 9 10 11
AID
APOBEC3A
80kD
60kD
40kD
30kD
25kD
20kD

FIG. 1

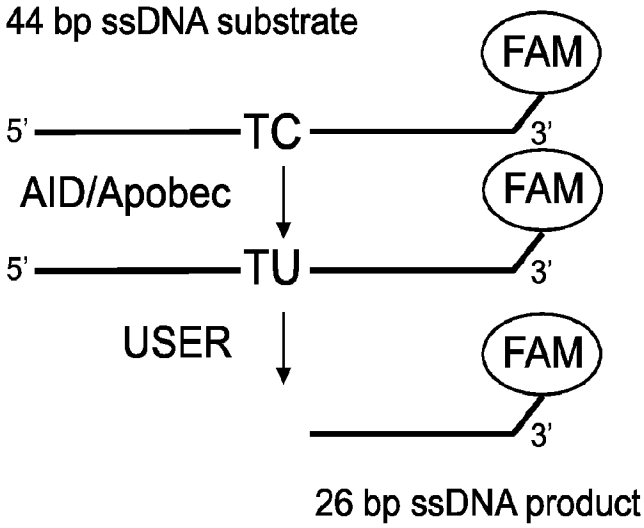

44 bp ssDNA substrate
FAM
5'
TC
3'
AID/Apobec
FAM
5'
TU
3'
USER
FAM
3'
26 bp ssDNA product

FIG. 2A

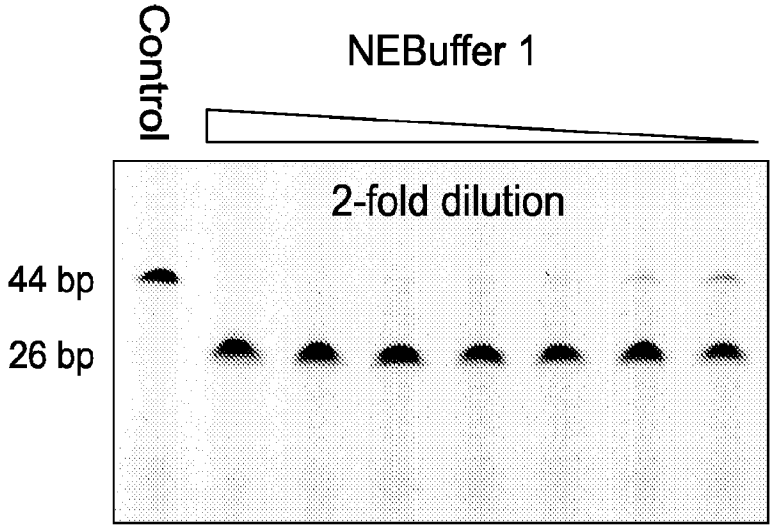
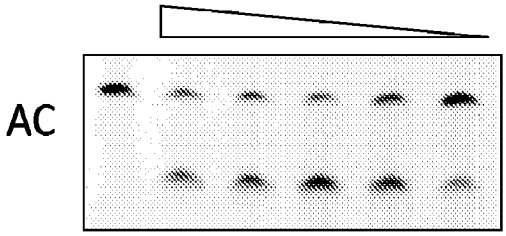
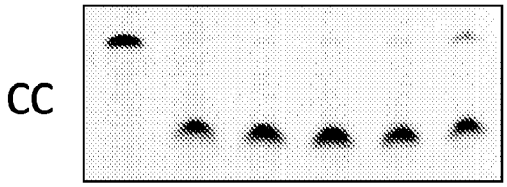

Control
NEBuffer 1
2-fold dilution
44 bp
26 bp

FIG. 2B



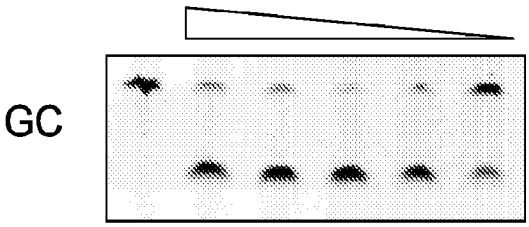
A3A wt PURE extract

FIG. 2C



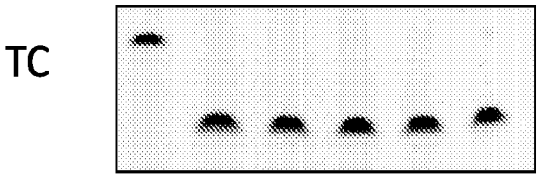
A3A wt PURE extract

FIG. 2D



A3A wt PURE extract

FIG. 2E



A3A wt PURE extract

FIG. 2F

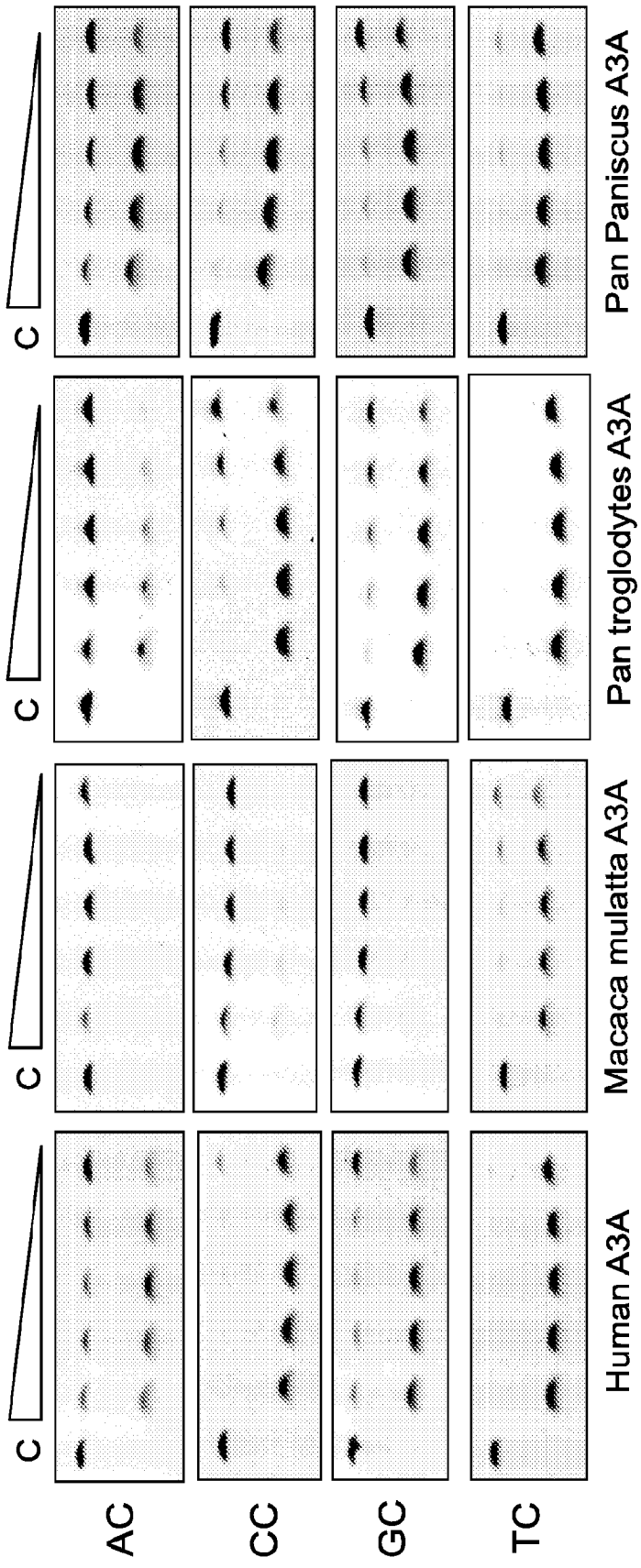

C
C
C
C
Human A3A
Macaca mulatta A3A
Pan troglodytes A3A
Pan Paniscus A3A
AC
CC
GC
TC

FIG. 3

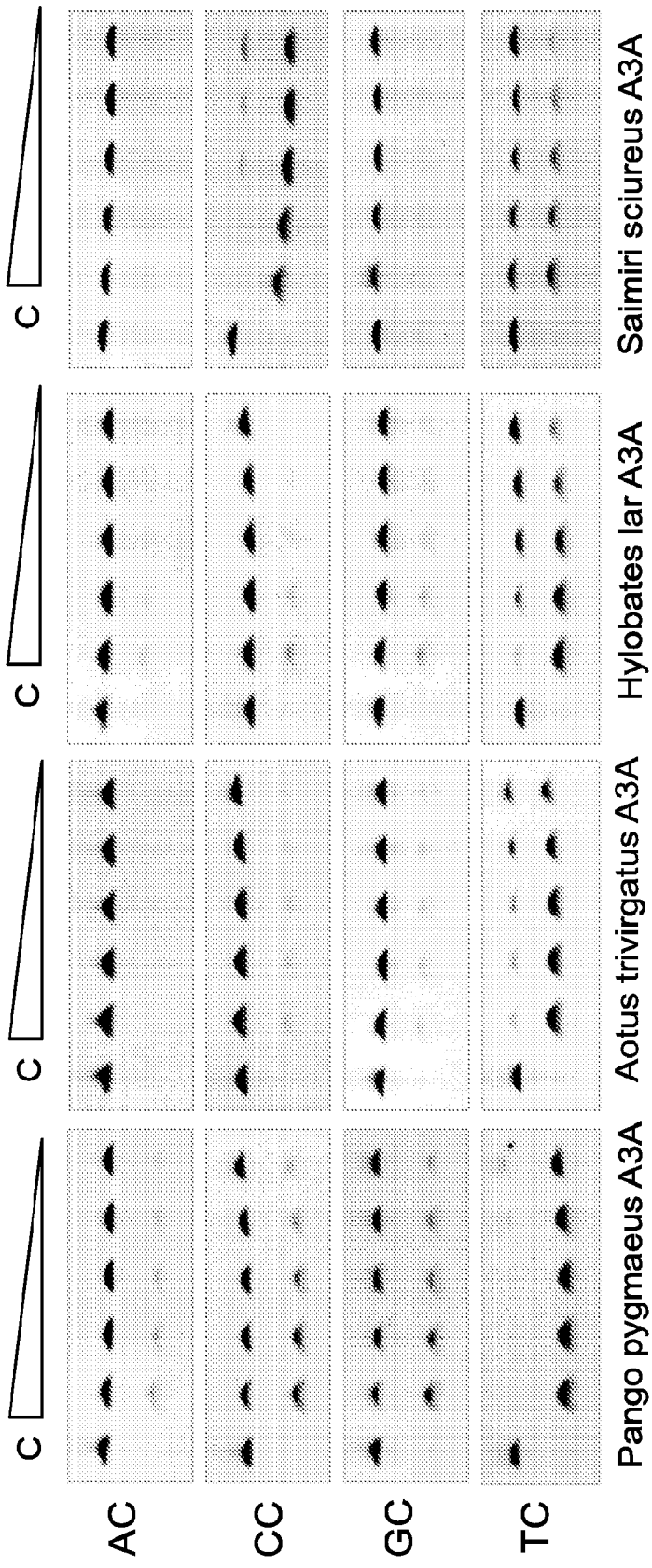

C
C
C
C
Pango pygmaeus A3A
Aotus trivirgatus A3A
Hylobates lar A3A
Saimiri sciureus A3A
AC
CC
GC
TC

FIG. 4

```
                                                                    60
Human A3A    MEASPASGPRHLMDPHIFTSNFNNG[V]IGRHKTYLCYEVERLDNGTS[W]VKMDQHRGFLHNQAK
                                                                   120
Human A3A    NLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGE[Q]VRAFLQENTHV
                                                                   180
Human A3A    RLR[H]IFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLD[E]
                                                  199
Human A3A    EHSQALSGRLRAILQNQGN      (SEQ ID NO:1)
```

FIG. 5A

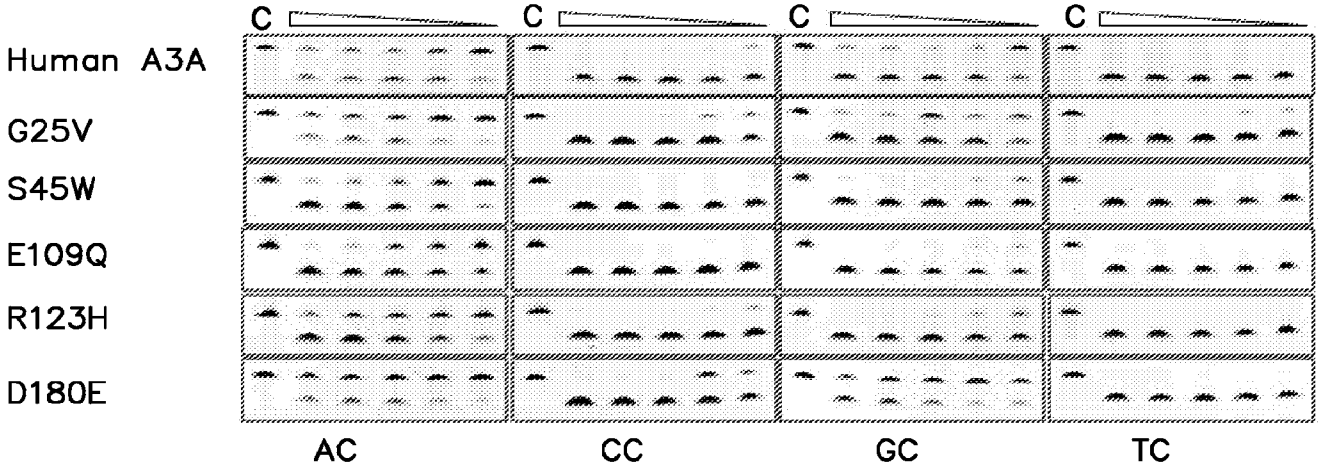


FIG. 5B

Pan_troglodytes MEASPASGPRHLMDPHIFTSNFTNGIGRRKTYLCYEVERLDNGTLVKMDQHRGFLHNQAK 60
N V H W
Pan_troglodytes NLLCGFYGCHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSRGCAGQVRAFLQENTHV 120
R W
Pan_troglodytes RLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLE 180
H
Pan_troglodytes EHSQALSGRLRAILQNQGN 199 (SEQ ID NO: 2)

FIG. 5C

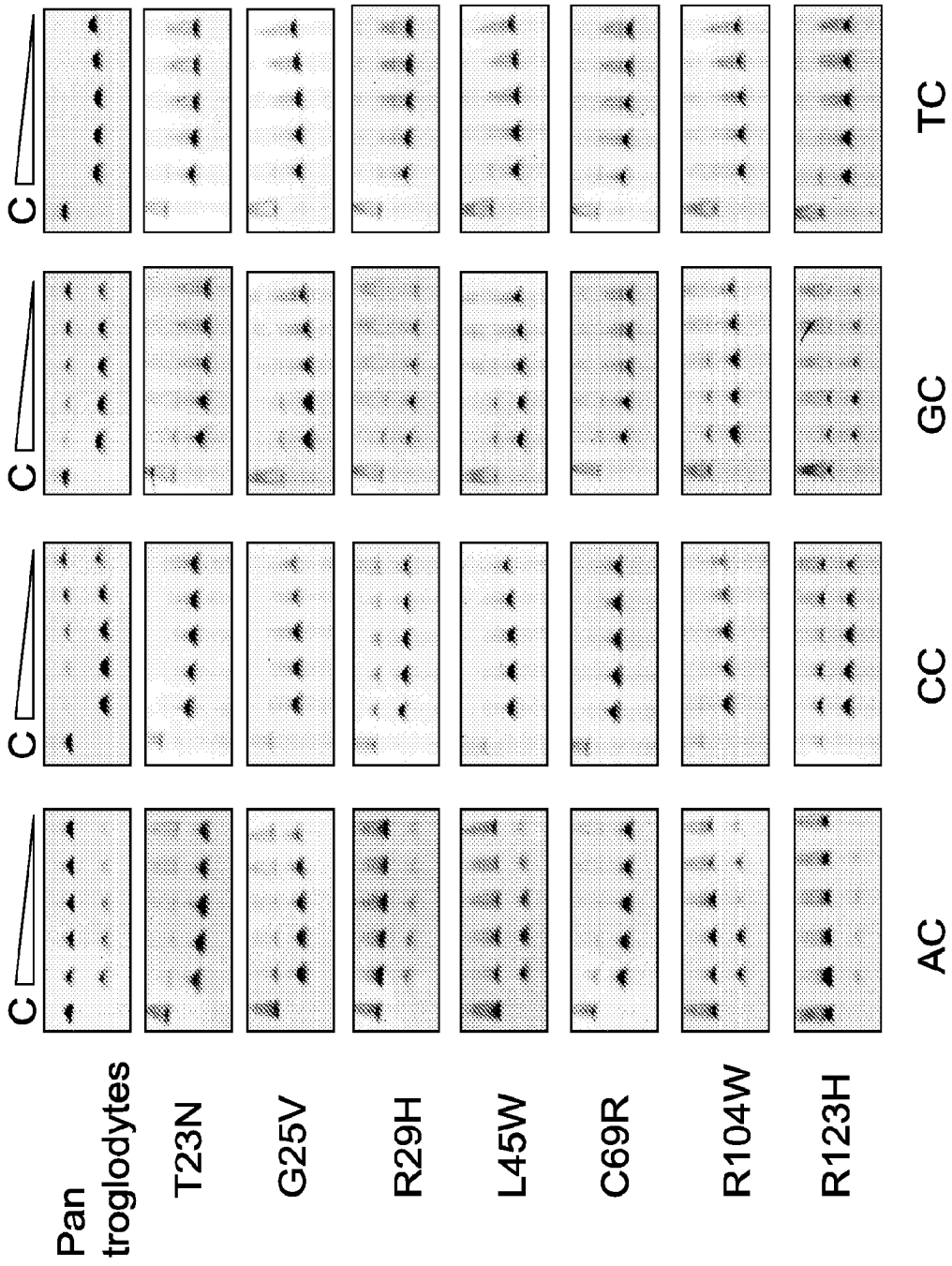

C
AC
C
CC
C
GC
C
TC
Pan troglodytes
T23N
G25V
R29H
L45W
C69R
R104W
R123H

FIG. 5D

```
                                                                60
AID wt      MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELL
AID_E117Δ   MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELL

                                                               120
AID wt      FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFC-DRK
AID_E117Δ   FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFC-DRK

                                                               180
AID wt      AEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRIL
AID_E117Δ   AEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRIL

                                                               198
AID wt      LPLYEVDDLRDAFRTLGL    (SEQ ID NO: 3)
AID_E117Δ   LPLYEVDDLRDAFRTLGL    (SEQ ID NO: 4)
```

FIG. 6A

FIG. 6B

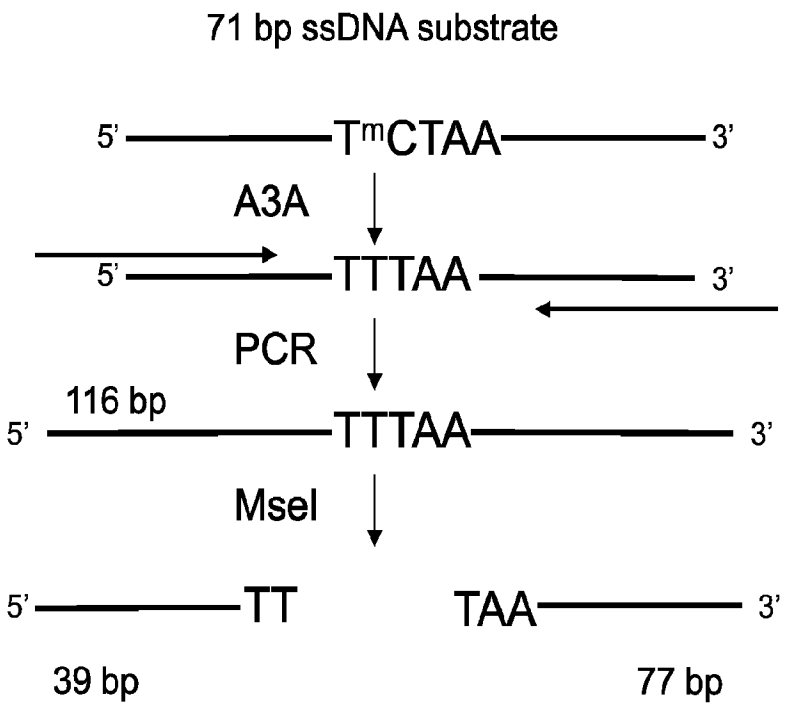

71 bp ssDNA substrate
5'
T^mCTAA
3'
A3A
5'
TTTAA
3'
PCR
116 bp
5'
TTTAA
3'
MseI
5'
TT
TAA
3'
39 bp
77 bp

FIG. 7A

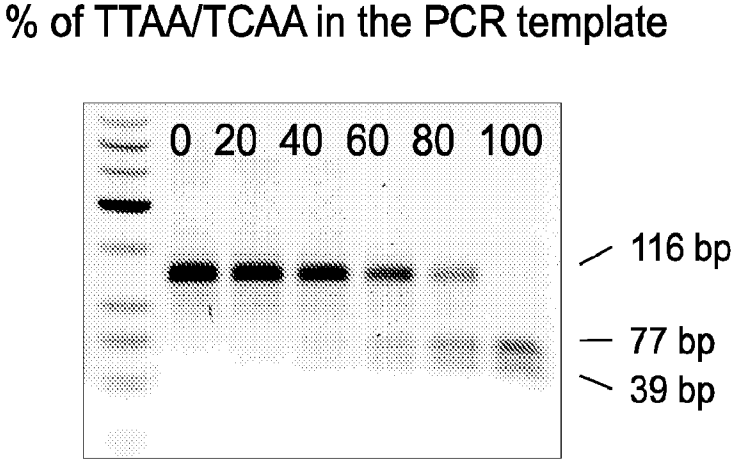

% of TTAA/TCAA in the PCR template
0 20 40 60 80 100
116 bp
77 bp
39 bp

FIG. 7B

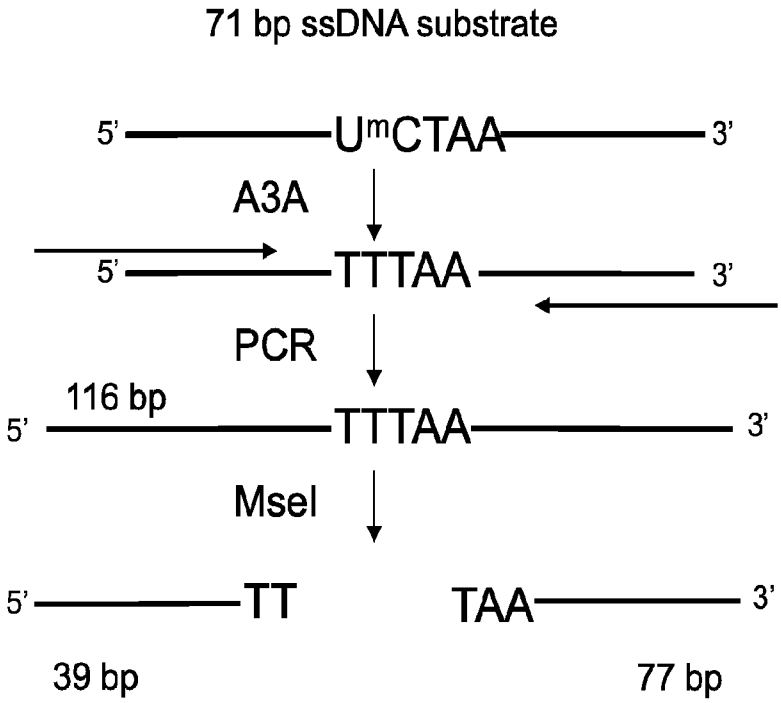
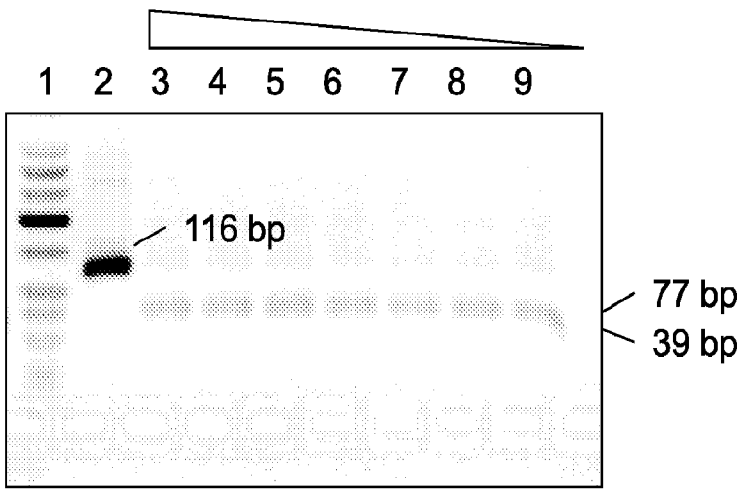


FIG. 7C



1) 2-Log DNA Ladder (NEB)
2) Control, no APOBEC3A
3-9) Serial 2x dilutions of APOBEC3A

FIG. 7D

5hmC oligo ATAAGAATAGAATGAAT5hmCTAAAAAATGAATATGAAATGAATAGTA/36-FAM/(SEQ ID NO:5)

5hmU oligo ATAAGAATAGAATGAAT5hmUTAAAAATGAATATGAAATGAATAGTA/36-FAM/(SEQ ID NO:6)

FIG. 8A

FIG. 8B

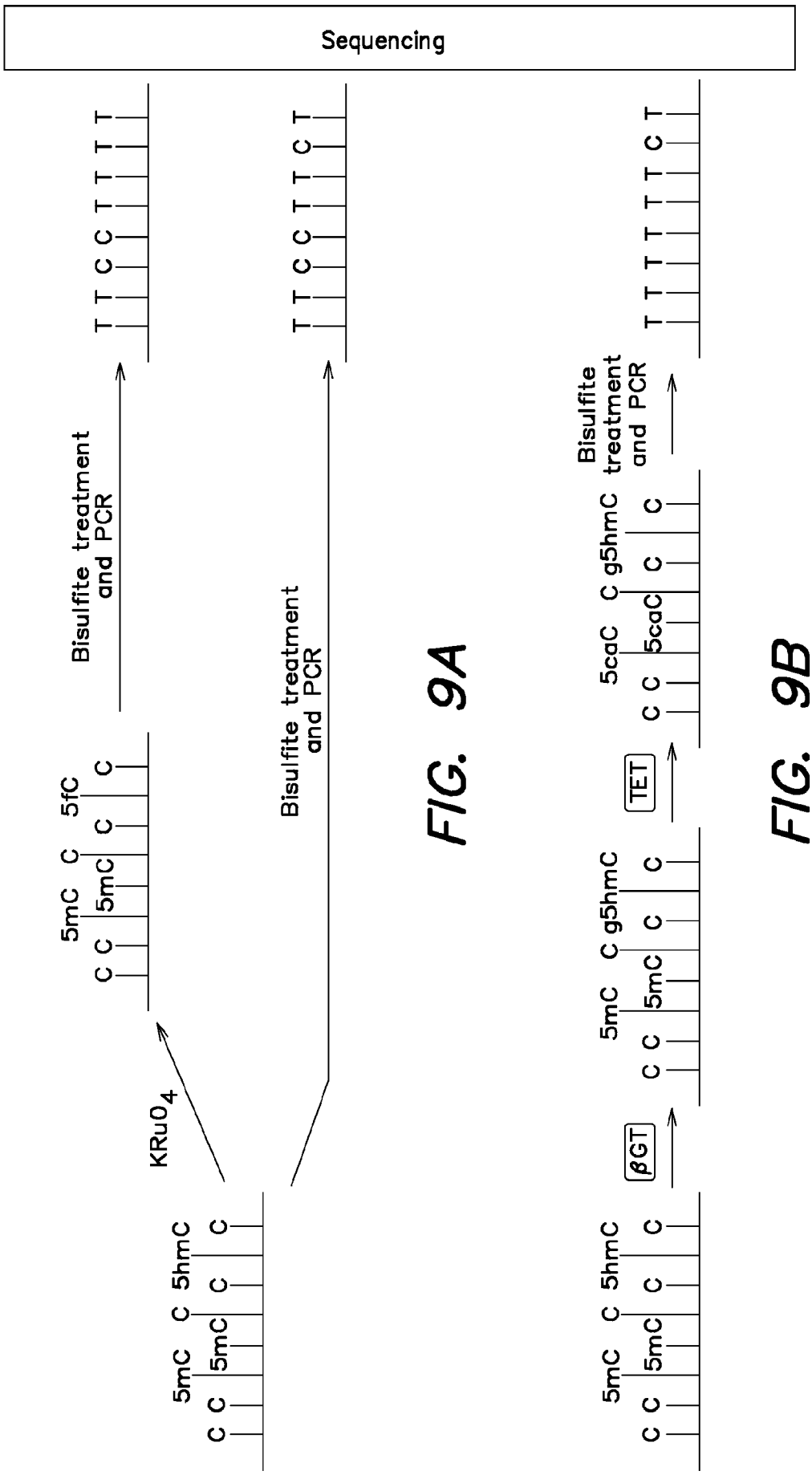

Sequencing
Bisulfite treatment and PCR
Bisulfite treatment and PCR
KRuO4
βGT
TET
Bisulfite treatment and PCR

FIG. 9A

FIG. 9B

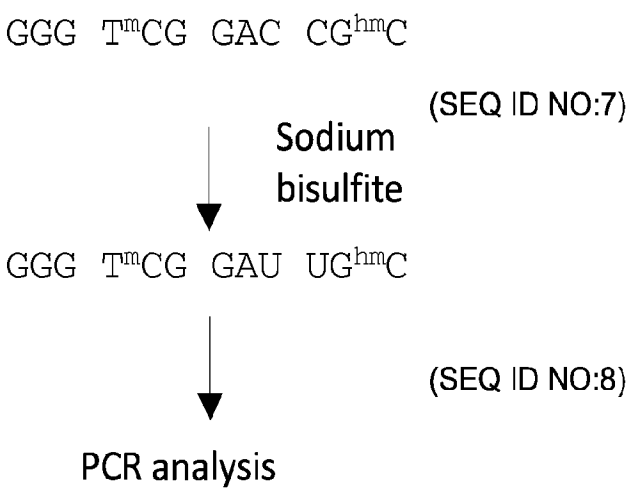


FIG. 10A

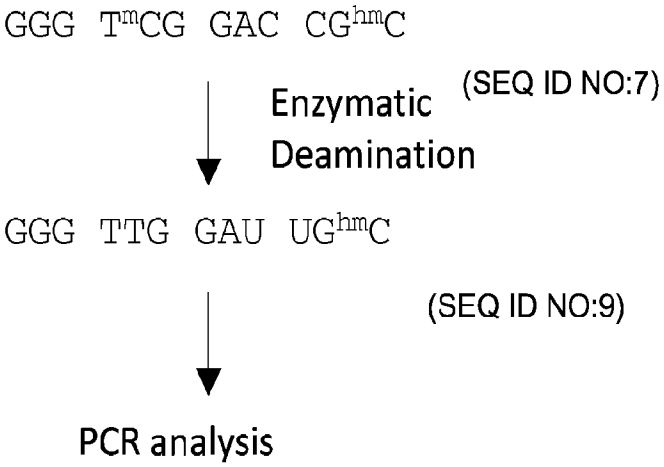


FIG. 10B

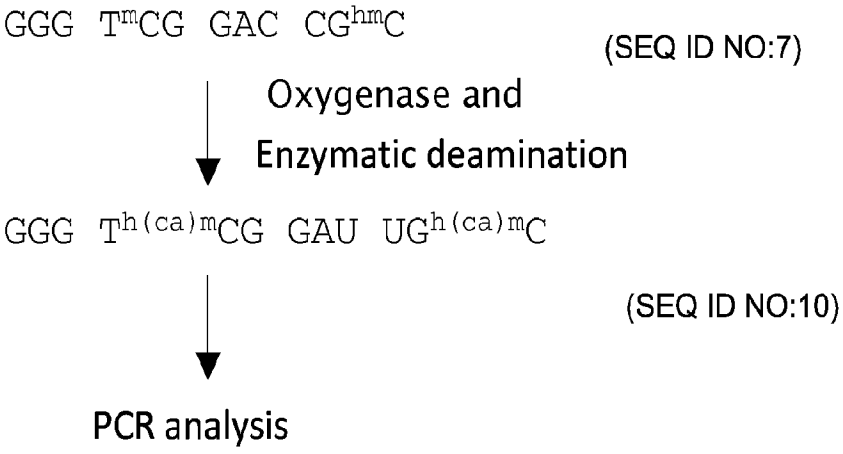


FIG. 10C

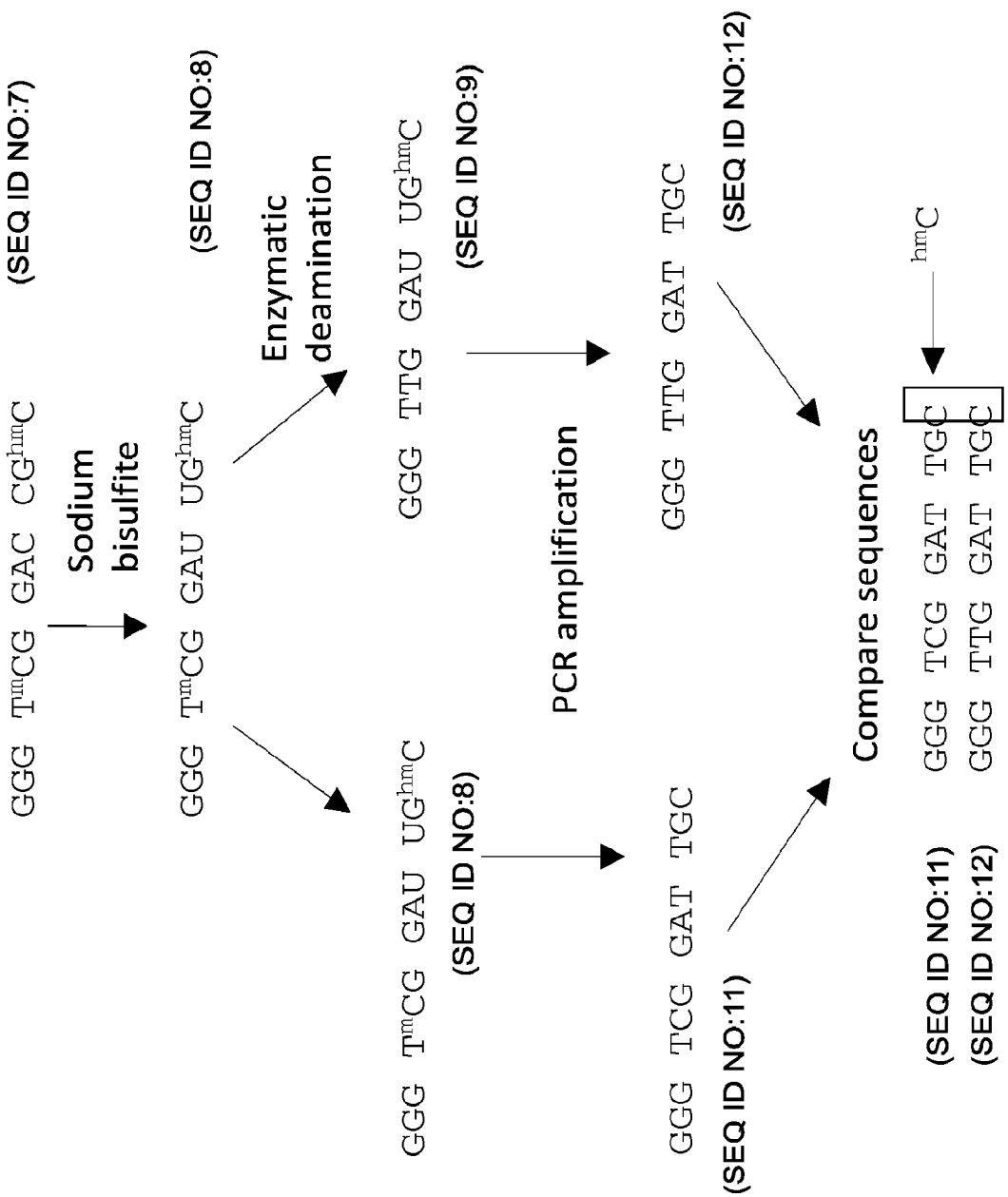

GGG T^{hm}CG GAC CG^{hm}C
(SEQ ID NO:7)
Sodium bisulfite
GGG T^{hm}CG GAU UG^{hm}C
(SEQ ID NO:8)
Enzymatic deamination
GGG TTG GAU UG^{hm}C
(SEQ ID NO:9)
PCR amplification
GGG T^{hm}CG GAU UG^{hm}C
(SEQ ID NO:8)
GGG TCG GAT TGC
(SEQ ID NO:11)
GGG TTG GAT TGC
(SEQ ID NO:12)
Compare sequences
^{hm}C
GGG TCG GAT TGC
(SEQ ID NO:11)
GGG TTG GAT TGC
(SEQ ID NO:12)

FIG. 11

A.
GTTAACGAGGAGGAAAAGAGAAATGGAGTGTGGGATGAGGAAAAGGAACAGAGAACCTGTCCCCATTGGAGACCACACAACAGAAGCCGGCACAGCTTTGTCCCTGTTTCCTTTCC
CTGAATGACTCCAGTTCCTTCTCAGTGAAGTTGGCAGACTGAGCCAGGTCCCACAGATCTATCACCCGGGGCTCTTCAAACTCTGCAGGAGAGCAAGGGCTGTCTATAGGTGGCA
AGTCAGGACACACCACCGGTGCTGGGGTCCAAGTGTCTTGTCATCCCAAGGGAATAGTGTGGGTTGCAACTTGGGAGCACTAAGTACCTGGTCTGTGCTCACAGCTCCAGCTTAG
ATGCAGCTTGATTGAATCCGGCTGCACCAGTGGAGGGGGTCACATGCTCTAGACCCAGAATCAACTGCTACCCCTGCCCAGGGCCATCAGCCTGCTCCTGGCTCTCAATCAAAAG
TGAGTGAGAGGAGAGGTGTGAGAGGATATC
(SEQ ID NO: 13)
Sodium bisulfite
B.
GTTAAUGAGGAGGAAAAGAGAAATGGAGTGTGGGATGAGGAAAAGGAAUAGAGAAUUTGTUUUUATTGGAGAUUAUAAAUAGAAGUCGGUAUAGUTTTGTUUUTGTTTUUT
TTUUUTGAATGAUTUUAGTTUUTTUTAGTGAAGTTGGUAGAUTGAGUUAGGTUUUAUAGATUTATUAUUCGGGGUTUTTUAAAUTUTGUAGGAGAGUAAGGGUTGTUTA
TAGGTGGUAAGTUAGGAUAUAUUAUCGGTGUTGGGGTUUAAGTGTUTTGTUATUUUAAGGGAATAGTGTGGGTTGUAAUTTGGGAGUAUTAAGTAUUTGGTUTGTGUTUA
UAGUTUUAGUTTAGATGUAGUTTGATTGAATUCGGUTGUAUUAGTGGAGGGGGTUAUATGUTUTAGAUUUAGAATUAAUTGUTAUUUUTGUUUAGGGUUATUAGUUTGU
TUUTGGUTUTUAATUAAAAGTGAGTGAGAGGAGAGGTGTGAGAGGATATU
(SEQ ID NO: 14)
Enzymatic deamination
C.
GTTAAUGAGGAGGAAAAGAGAAATGGAGTGTGGGATGAGGAAAAGGAAUAGAGAAUUTGTUUUUATTGGAGAUUAUAAAUAGAAGUTGGUAUAGUTTTGTUUUTGTTTUUT
TTUUUTGAATGAUTUUAGTTUUTTUTAGTGAAGTTGGUAGAUTGAGUUAGGTUUUAUAGATUTATUAUUTGGGGUTUTTUAAAUTUTGUAGGAGAGUAAGGGUTGTUTA
TAGGTGGUAAGTUAGGAUAUAUUAUTGGTGUTGGGGTUUAAGTGTUTTGTUATUUUAAGGGAATAGTGTGGGTTGUAAUTTGGGAGUAUTAAGTAUUTGGTUTGTGUTUA
UAGUTUUAGUTTAGATGUAGUTTGATTGAATUTGGUTGUAUUAGTGGAGGGGGTUAUATGUTUTAGAUUUAGAATUAAUTGUTAUUUUTGUUUAGGGUUATUAGUUTGU
TUUTGGUTUTUAATUAAAAGTGAGTGAGAGGAGAGGTGTGAGAGGATATU
(SEQ ID NO: 15)
PCR
D.
GTTAATGAGGAGGAAAAGAGAAATGGAGTGTGGGATGAGGAAAAGGAATAGAGAATTTGTTTTTATTGGAGATTATAATAGAAGTTGGTATAGTTTTGTTTTTGTTTTTTTTTTT
GAATGATTTTAGTTTTTTTTTTAGTGAAGTTGGTAGATTGAGTTAGGTTTTATAGATTTATTATTTGGGGTTTTTTAAATTTTGTAGGAGAGTAAGGGTTGTTTATAGGTGGTAAGTTA
GGATATATTATTTGGTGTTGGGGTTTAAGTGTTTTGTTATTTTAAGGGAATAGTGTGGGTTGTAATTTGGGAGTATTAAGTATTTGGTTTGTGTTTATAGTTTTAGTTTTAGATGTAGTT
TGATTGAATTTGGTTGTATTAGTGGAGGGGGTTATATGTTTTAGATTTAGAATTAATTGTTATTTTTTGTTTAGGGTTATTAGTTTGTTTTTTGGTTTTTTAATTAAAAGTGAGTGAGAGG
AGAGGTGTGAGAGGATATT
(SEQ ID NO: 16)

FIG. 12

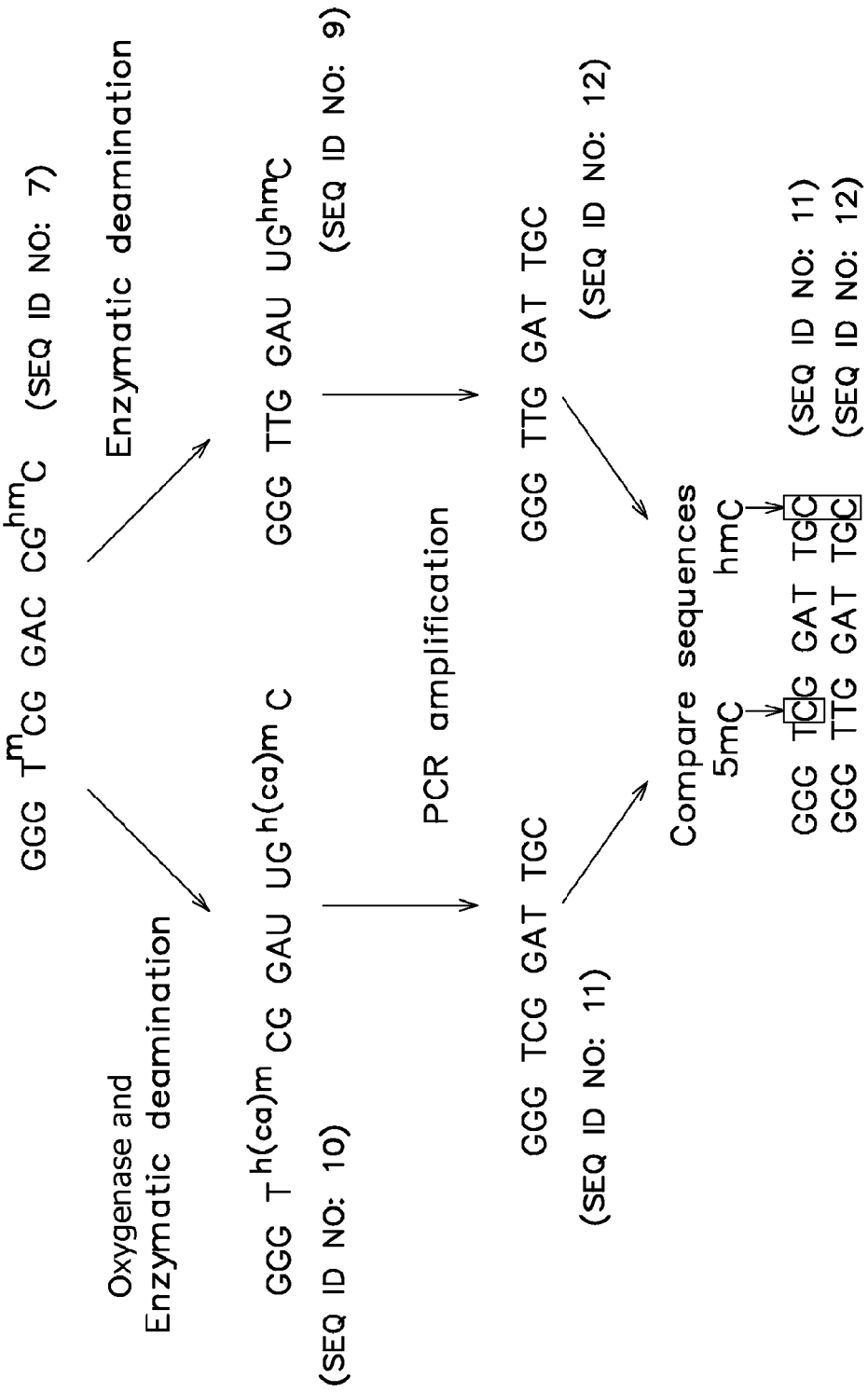

GGG T^mCG GAC CG^hmC (SEQ ID NO: 7)
Enzymatic deamination
GGG TTG GAU UG^hmC (SEQ ID NO: 9)
Oxygenase and Enzymatic deamination
GGG T^h(ca)m CG GAU UG^h(ca)m C (SEQ ID NO: 10)
PCR amplification
GGG TCG GAT TGC (SEQ ID NO: 11)
GGG TTG GAT TGC (SEQ ID NO: 12)
Compare sequences
5mC
hmC
GGG TCG GAT TGC (SEQ ID NO: 11)
GGG TTG GAT TGC (SEQ ID NO: 12)

FIG. 13