Supporting Information

for

Bisulfite-free and single-nucleotide resolution sequencing of DNA epigenetic modification of 5-hydroxymethylcytosine by engineered deaminase

Neng-Bin Xie,[1,2,†] Min Wang,[1,†] Tong-Tong Ji,[1] Xia Guo,[1] Jiang-Hui Ding,[1]

Bi-Feng Yuan,[1,2,3,*] Yu-Qi Feng[1,2]

[1] Sauvage Center for Molecular Sciences, Department of Chemistry, Wuhan University, Wuhan 430072, China.

[2] School of Public Health, Wuhan University, Wuhan 430071, China.

[3] Department of Radiation and Medical Oncology, Zhongnan Hospital of Wuhan University, Wuhan 430071, China.

[†] These authors contributed equally to this work.

* Corresponding author:

Bi-Feng Yuan. E-mail: bfyuan@whu.edu.cn

Table of Contents

Colony sequencing

The colony sequencing was carried out according to the previous reports.[1, 2] Briefly, to quantitatively determine the ratio of C/(C+T) at given sites after the treatment of eA3A proteins, the deaminase-treated DNA was amplified and the PCR products were ligated into pClone007 Versatile Simple vector (Tsingke, Beijing, China). The resulting vector was then transformed into competent *E. coli* cells. Individual clone was randomly picked, lysed in TE buffer, amplified by PCR using M13 forward and reverse primers and the PCR products were then sequenced using an ABI3700 (Applied Biosystems, Inc.). Fifty positive clones per sample were picked and subjected to sequencing. The percentage of C/(C+T) at given sites could be calculated from the statistical analysis of fifty sequences, which reflect the percentage of 5hmC at given sites.

Table S1. Sequences of oligonucleotides.

Name	Sequence (5’ to 3’)
GC-C	GTATGATGCGAATGAGATGTATTG
AC-C	GTATGATACGAATGAGATGTATTG
TC-C	GTATGATTCGAATGAGATGTATTG
CC-C	GTATGATCCGAATGAGATGTATTG
GC-5mC	GTATGATG5mCGAATGAGATGTATTG
AC-5mC	GTATGATA5mCGAATGAGATGTATTG
TC-5mC	GTATGATT5mCGAATGAGATGTATTG
CC-5mC	GTATGATC5mCGAATGAGATGTATTG
GC-5hmC	GTATGATG5hmCGAATGAGATGTATTG
AC-5hmC	GTATGATA5hmCGAATGAGATGTATTG
TC-5hmC	GTATGATT5hmCGAATGAGATGTATTG
CC-5hmC	GTATGATC5hmCGAATGAGATGTATTG

Note: the “C” represented 5’-aza-2’-deoxycytidine.

Table S2. Sequences of double-stranded DNA.

Name	Sequence (5’ to 3’)
DNA-C	GAGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGA TTATGAGTATGTATAGTGTTAGGAAGAGTGTAGTAATAG GATGAAGATGATTATATGATCGATGGTCCGTATGCGTAG AATACGTTGTTGTAGTGATTATAATGGAGTGAGAATGTA GATGAGTGGAGTAGGTAGTAAGATGTAGTGGTGAAGAG AGTAATTGTTAGTGGAATGTTGGAGAGGAT
DNA-5mC	GAGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGA TTATGAGTATGTATAGTGTTAGGAAGAGTGTAGTAATAG GATGAAGATGATTATATGAT5mCGATGGT5mC5mCGTAT G5mCGTAGAATA5mCGTTGTTGTAGTGATTATAATGGAG TGAGAATGTAGATGAGTGGAGTAGGTAGTAAGATGTAG TGGTGAAGAGAGTAATTGTTAGTGGAATGTTGGAGAGG AT
DNA-5hmC	GAGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGA TTATGAGTATGTATAGTGTTAGGAAGAGTGTAGTAATAG GATGAAGATGATTATATGAT5hmCGATGGT5hmC5hmCGT ATG5hmCGTAGAATA5hmCGTTGTTGTAGTGATTATAATG GAGTGAGAATGTAGATGAGTGGAGTAGGTAGTAAGATG TAGTGGTGAAGAGAGTAATTGTTAGTGGAATGTTGGAG AGGAT
DNA-C5mC	GAGTGACGCTGAGCTTGACGTCGCGCGTC5mCGAGT5mC 5mCGATGAGAGGTGATTATGAGTATGTATAGTGTTAGGA AGAGTGTAGTAATAGGATGAAGATGATTATATGATTGAT GGTTTGTATGTGTAGAATATGTTGTTGTAGTGATTATAA TGGAGTGAGAATGTAGATGAGTGGAGTAGGTAGTAAGA TGTAGTGGTGACAGAGAGTAATTGTTAGTGGAATGTTGG AGAGGAT
DNA-C5hmC	GAGTGACGCTGAGCTTGACGTCGCGCGTC5hmCGAGT5h mC5hmCGATGAGAGGTGATTATGAGTATGTATAGTGTTA GGAAGAGTGTAGTAATAGGATGAAGATGATTATATGAT TGATGGTTTGTATGTGTAGAATATGTTGTTGTAGTGATT ATAATGGAGTGAGAATGTAGATGAGTGGAGTAGGTAGT AAGATGTAGTGGTGACAGAGAGTAATTGTTAGTGGAAT GTTGGAGAGGAT

Table S3. Information of detected 5hmC sites in genomic DNA of human lung tissue.

Genome location	Site type	Primers for detection (from 5’ to 3’)
chr3:29918016, CRCh37	GC	S: TTTAAATATTTTTTTAGGTGAATTATTTATG
		R: ACTATTACCTTTATCCTCCTCCCAT
chr5:14180809, CRCh37	AC	S: TGGGAAGATTGTTTGAGGTTAGG
		R: CATAAAACCACTCACTCTATCACCC
chr4:169198493, CRCh37	TC	S: GGTATTTATAGTAGTTTTGGAAGAATTTAGTG
		R: ACAAATATATAACACATCCATAAACACCT
chr1:68672111, CRCh37	CC	S: TAAATGTGGAGGAGGAATTAGATTTAT
		R: AAAATCACCTTTCCATCCCTACC

Figure S1. SDS-PAGE analysis of the purified wtA3A, eA3A-5 and eA3A-9. "M" represents protein marker. (A) SDS-PAGE analysis of the purified wtA3A. (B) SDS-PAGE analysis of the purified eA3A-5. (C) SDS-PAGE analysis of the purified eA3A-9. Line 1, without addition of IPTG; line 2, with addition of IPTG; line 3, purified proteins.

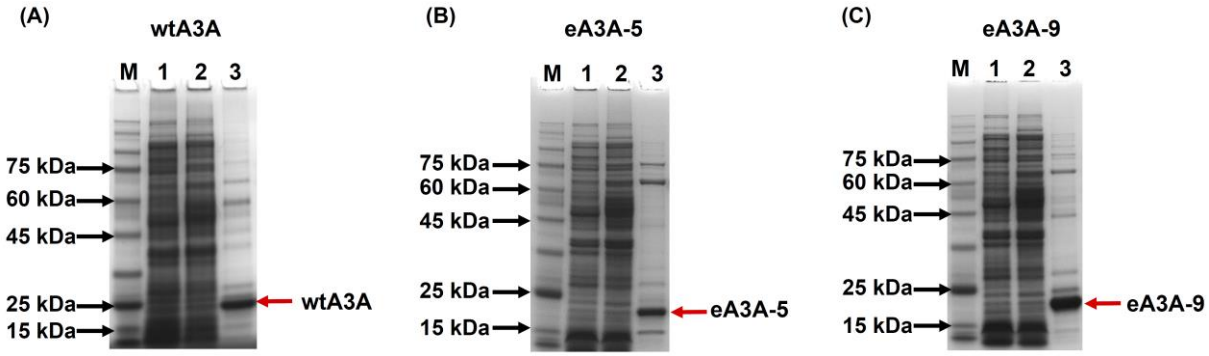


Figure S2. Principle of the AMD-seq method. (A) C, 5mC and 5hmC can be deaminated by wtA3A to form U, T and 5-hydroxymethyluracil (5hmU), respectively. All these deaminated products base pair with adenine. (B) In AMD-seq, DNA is firstly treated with *β*-GT that can transfer a glucosyl moiety to 5hmC to form 5gmC. Subsequent wtA3A treatment will lead to the deamination of C and 5mC to form U and T, both of which are read as T during sequencing. However, 5gmC is resistant to the deamination by wtA3A and then is still read as C during sequencing. Therefore, the remaining C manifests the original 5hmC site in DNA.

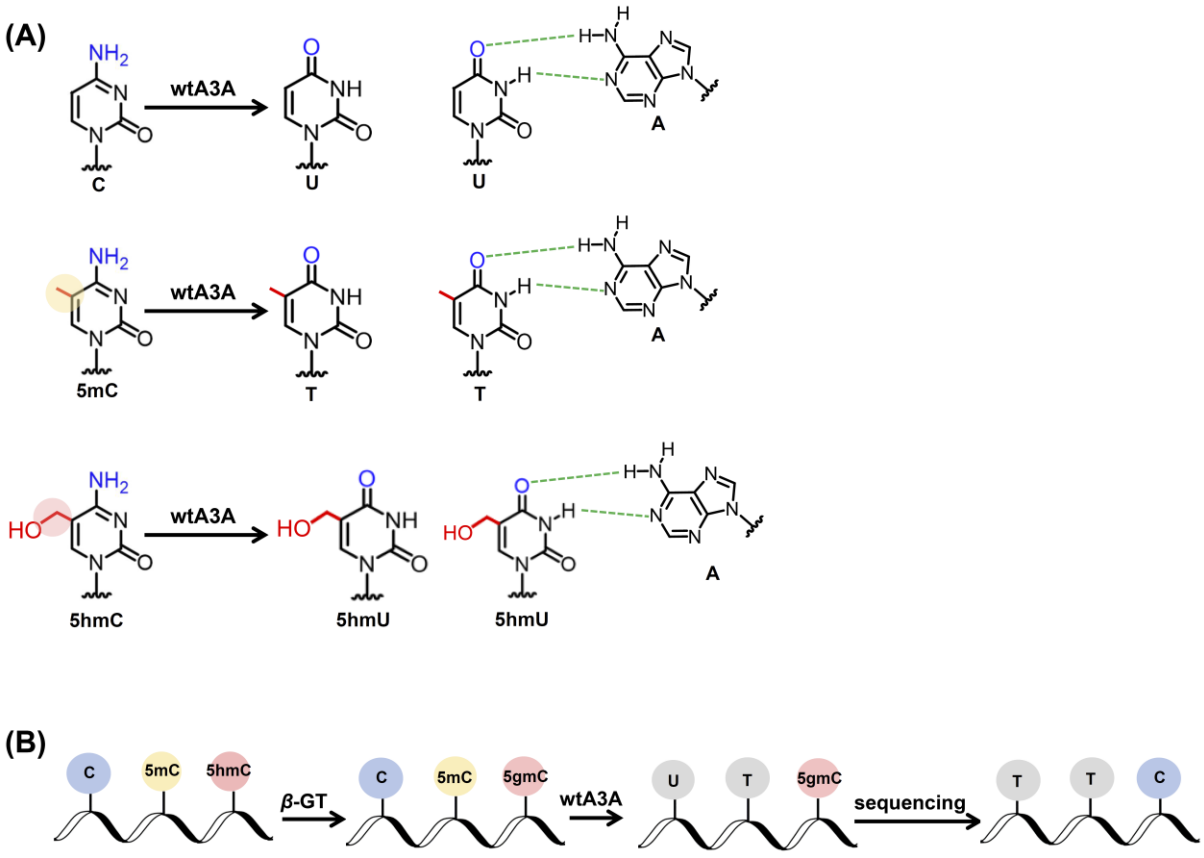


Figure S3. Characterization of the deaminase selectivity of wtA3A toward C, 5mC and 5hmC in different sequence context by Sanger sequencing. The 224-bp DNA-C, DNA-5mC and DNA-5hmC were used for the evaluation. All the C and 5mC were deaminated and then read as T. 5hmC in TC and CC sites were also deaminated and read as T. However, 5hmC in GC and AC sites were partially deaminated and therefore were partially read as C and partially read as T.

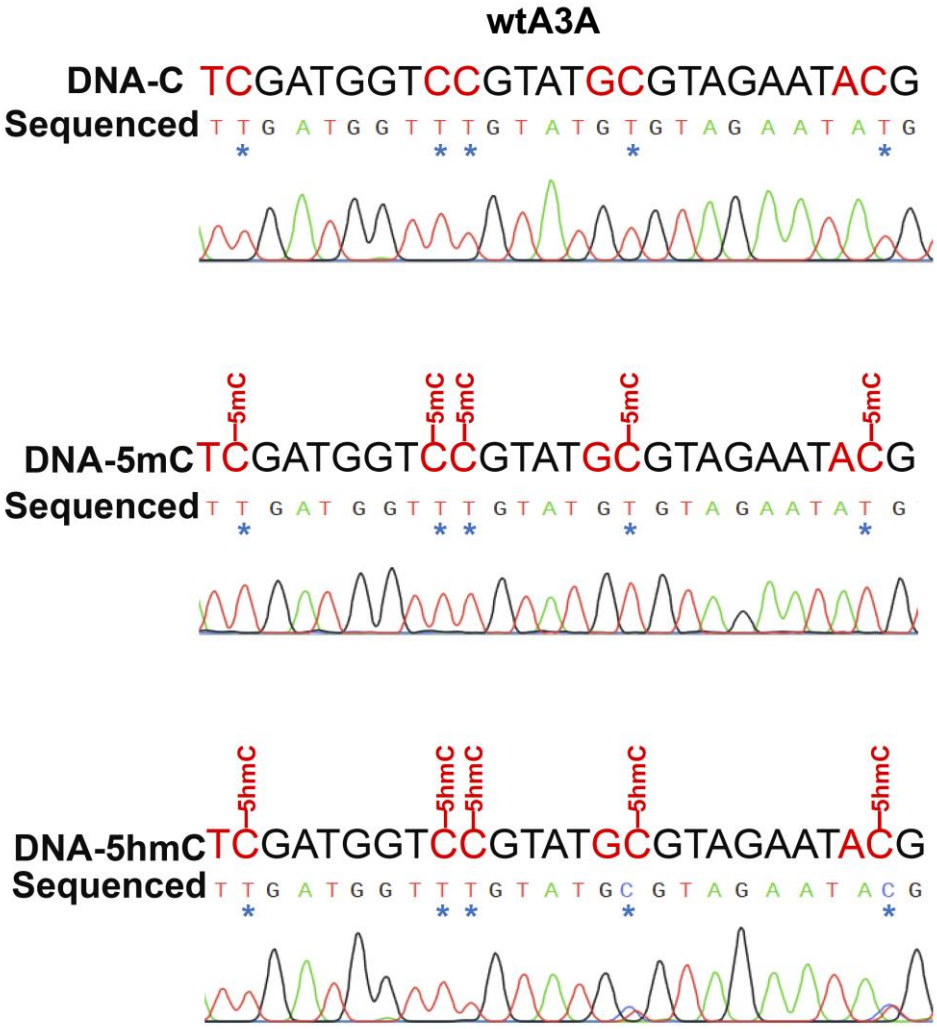


Figure S4. Characterization of the deaminase selectivity of eA3A-1 toward C, 5mC and 5hmC in different sequence context by Sanger sequencing. The 224-bp DNA-C, DNA-5mC and DNA-5hmC were used for the evaluation. (A) The amino acid composition of wtA3A and eA3A proteins. (B) The sequencing results of DNA-C, DNA-5mC, and DNA-5hmC after eA3A-1 treatment. All the C were deaminated and read as T; but 5mC and 5hmC were partially deaminated and were partially read as C and partially read as T.

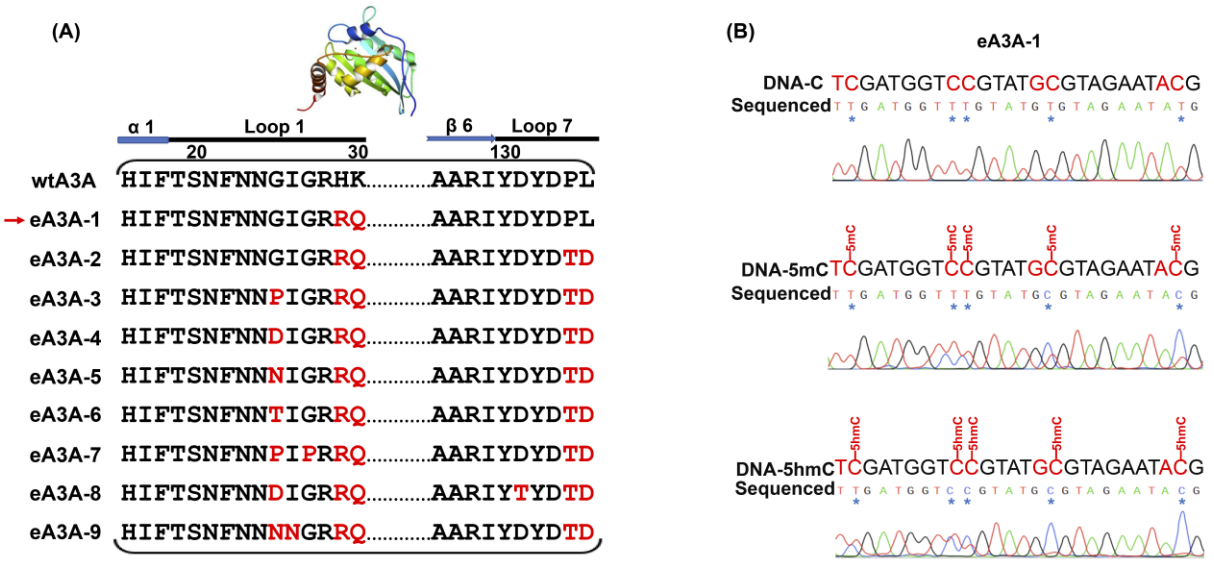


Figure S5. Characterization of the deaminase selectivity of eA3A-2 toward C, 5mC and 5hmC in different sequence context by Sanger sequencing. The 224-bp DNA-C, DNA-5mC and DNA-5hmC were used for the evaluation. (A) The amino acid composition of wtA3A and eA3A proteins. (B) The sequencing results of DNA-C, DNA-5mC, and DNA-5hmC after eA3A-2 treatment. In GC and AC sites, all the C and 5mC were deaminated and read as T; but all the 5hmC were resistant to deamination and were still read as C. In TC and CC sites, all the C and 5mC were deaminated and read as T; but 5hmC were partially deaminated and were partially read as C and partially read as T.

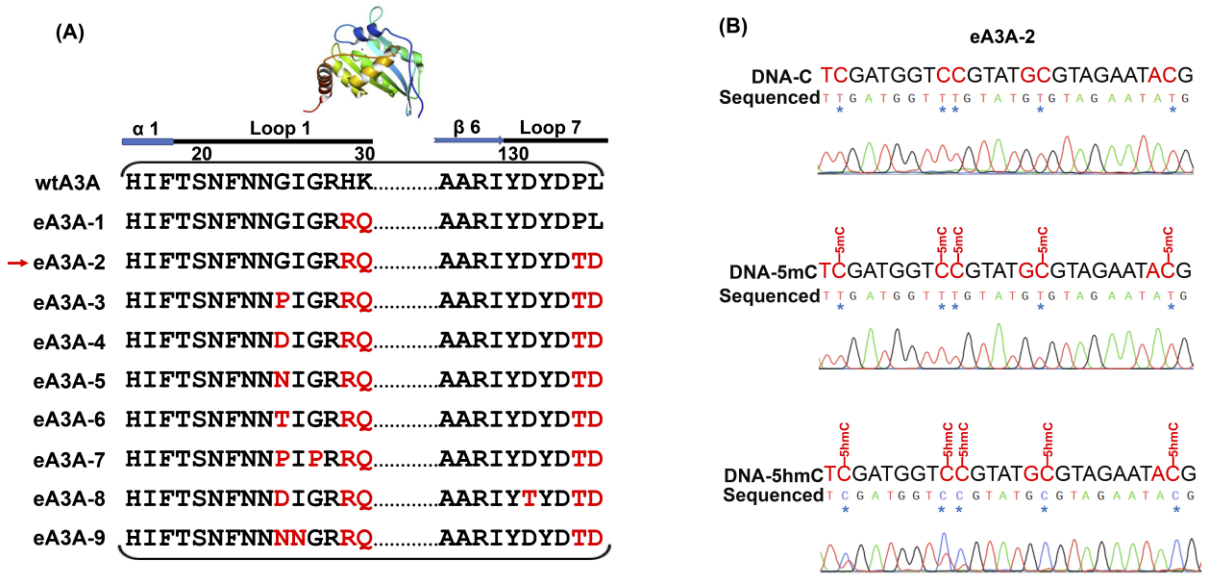


Figure S6. Characterization of the deaminase selectivity of eA3A-3 toward C, 5mC and 5hmC in different sequence context by Sanger sequencing. The 224-bp DNA-C, DNA-5mC and DNA-5hmC were used for the evaluation. (A) The amino acid composition of wtA3A and eA3A proteins. (B) The sequencing results of DNA-C, DNA-5mC, and DNA-5hmC after eA3A-3 treatment. All the C were deaminated and read as T; but 5mC were partially deaminated and were partially read as C and partially read as T; while 5hmC were resistant to deamination and were still read as C.

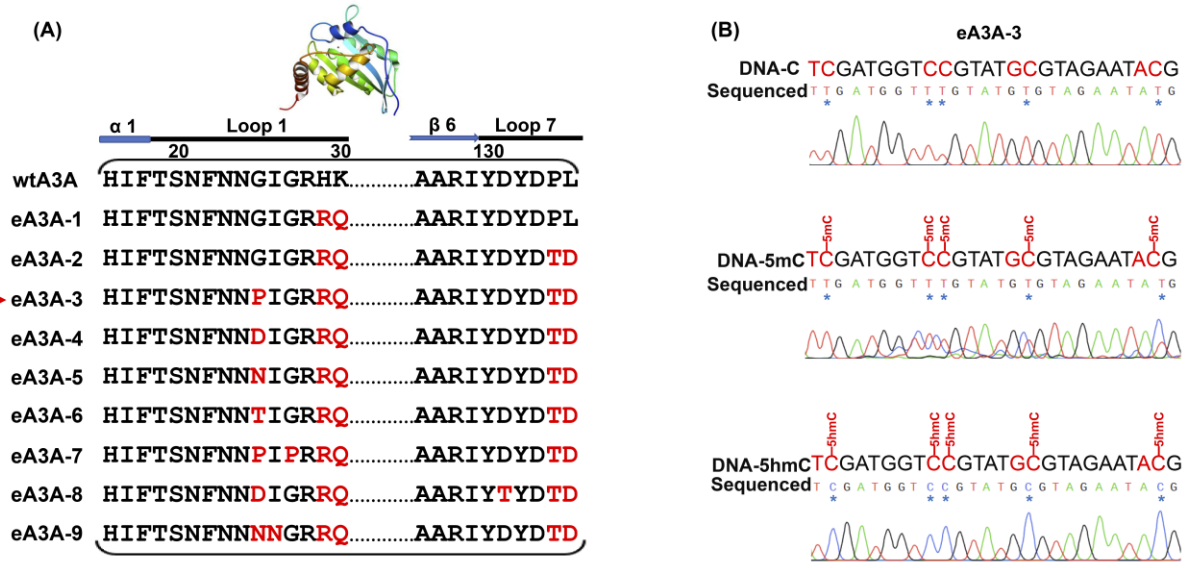


Figure S7. Characterization of the deaminase selectivity of eA3A-4 toward C, 5mC and 5hmC in different sequence context by Sanger sequencing. The 224-bp DNA-C, DNA-5mC and DNA-5hmC were used for the evaluation. (A) The amino acid composition of wtA3A and eA3A proteins. (B) The sequencing results of DNA-C, DNA-5mC, and DNA-5hmC after eA3A-4 treatment. All the C were deaminated and read as T; but 5mC were partially deaminated and were partially read as C and partially read as T; while 5hmC were resistant to deamination and were still read as C.

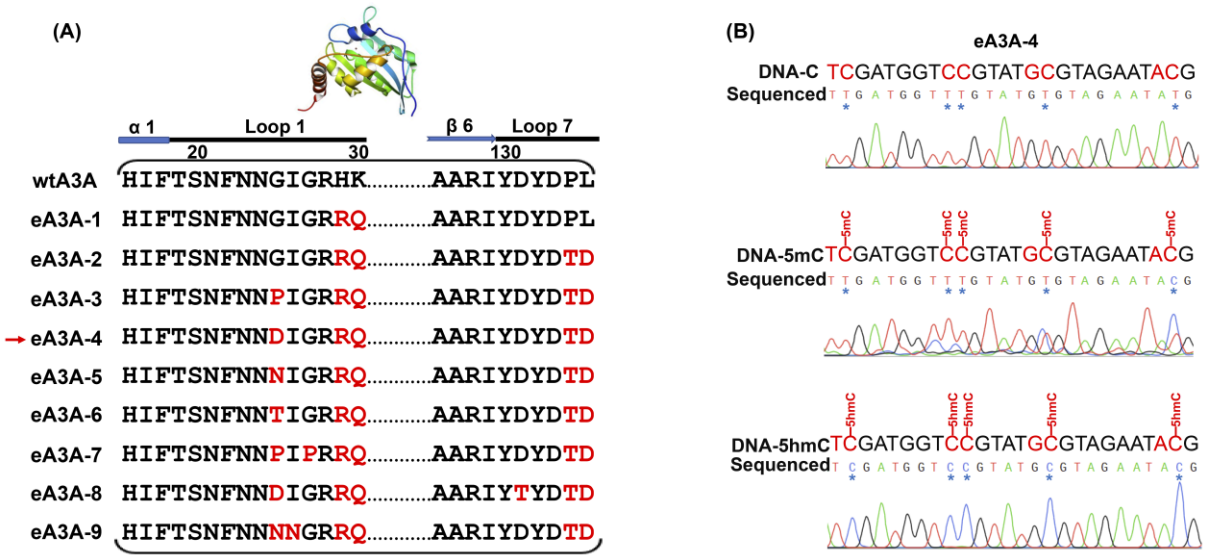


Figure S8. Characterization of the deaminase selectivity of eA3A-6 toward C, 5mC and 5hmC in different sequence context by Sanger sequencing. The 224-bp DNA-C, DNA-5mC and DNA-5hmC were used for the evaluation. (A) The amino acid composition of wtA3A and eA3A proteins. (B) The sequencing results of DNA-C, DNA-5mC, and DNA-5hmC after eA3A-6 treatment. In GC and AC sites, all the C and 5mC were deaminated and read as T; but all the 5hmC were resistant to deamination and were still read as C. In TC and CC sites, all the C and 5mC were deaminated and read as T; but 5hmC were partially deaminated and were partially read as C and partially read as T.

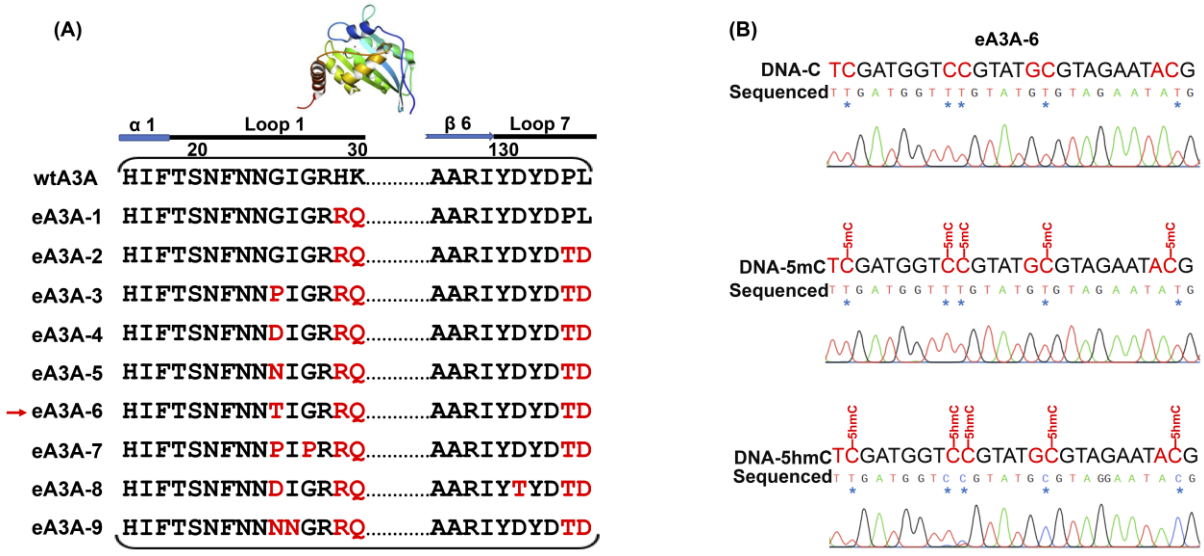


Figure S9. Characterization of the deaminase selectivity of eA3A-7 toward C, 5mC and 5hmC in different sequence context by Sanger sequencing. The 224-bp DNA-C, DNA-5mC and DNA-5hmC were used for the evaluation. (A) The amino acid composition of wtA3A and eA3A proteins. (B) The sequencing results of DNA-C, DNA-5mC, and DNA-5hmC after eA3A-7 treatment. In GC and AC sites, all the C were deaminated and read as T; but 5mC were partially deaminated and were partially read as C and partially read as T; while all the 5hmC were resistant to deamination and were still read as C. In TC and CC sites, all the C and 5mC were deaminated and then read as T; but all the 5hmC were resistant to deamination and were still read as C.

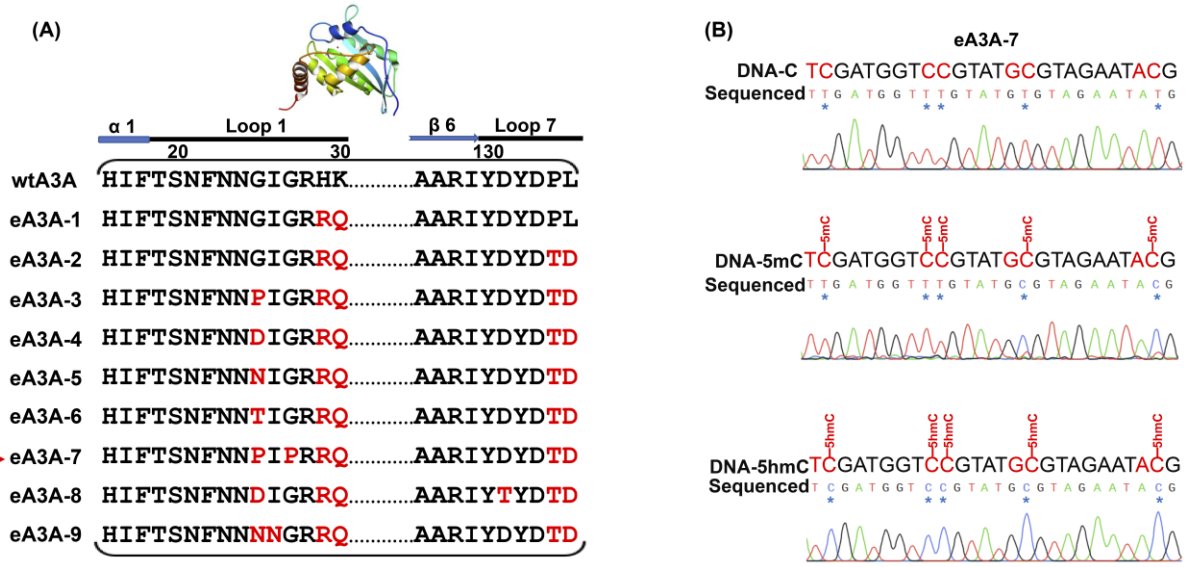


Figure S10. Characterization of the deaminase selectivity of eA3A-8 toward C, 5mC and 5hmC in different sequence context by Sanger sequencing. The 224-bp DNA-C, DNA-5mC and DNA-5hmC were used for the evaluation. (A) The amino acid composition of wtA3A and eA3A proteins. (B) The sequencing results of DNA-C, DNA-5mC, and DNA-5hmC after eA3A-8 treatment. In GC and AC sites, all the C were deaminated and read as T; but 5mC were partially deaminated and were partially read as C and partially read as T; while all the 5hmC were resistant to deamination and were still read as C. In TC and CC sites, all the C and 5mC were deaminated and then read as T; but all the 5hmC were resistant to deamination and were still read as C.

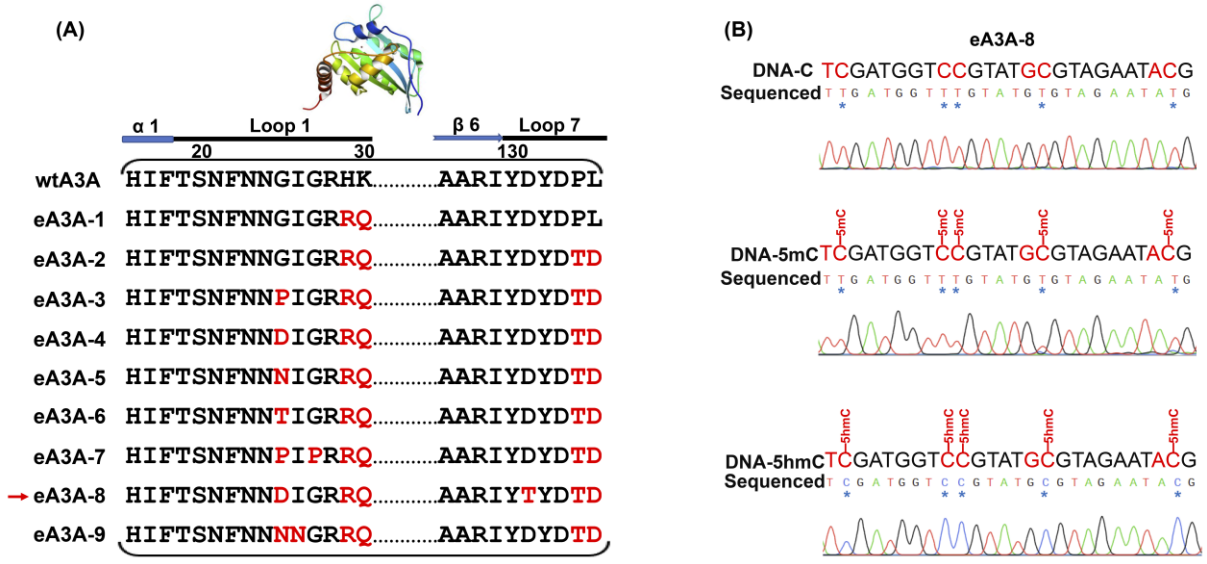


Figure S11. Characterization of 5mC and 5hmC in CC site by EDM-seq analysis with Sanger sequencing. DNA-C5mC and DNA-C5hmC were used for the evaluation. After eA3A-9 treatment, 5mC sites in both C5mC and 5mC5mC were deaminated and read as T; while 5hmC sites in both C5hmC and 5hmC5hmC were not deaminated and still read as C.

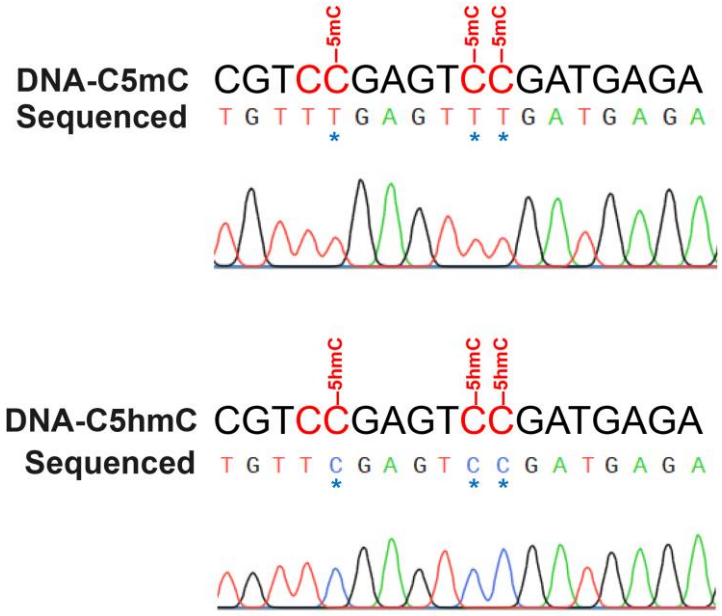


Figure S12. Extracted-ion chromatograms of dA, dG and dT from eA3A-5 or eA3A-9 treated DNA by LC-MS/MS analysis. The mixture of GC-C and AC-C was treated with eA3A-5. Another mixture of TC-C and CC-C was treated with eA3A-9. (A) Extracted-ion chromatograms of dA, dG and dT from GC-C and AC-C mixture without or with eA3A-5 treatment. (B) Extracted-ion chromatograms of dA, dG and dT from TC-C and CC-C mixture without or with eA3A-9 treatment.

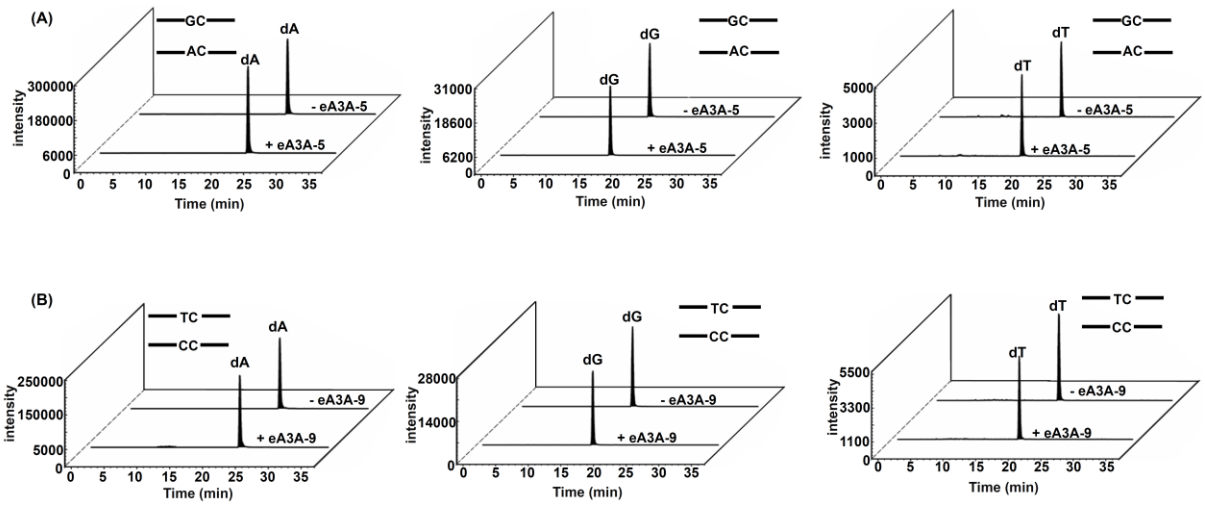


Figure S13. Evaluation of the deaminase activity of eA3A-5 to C, 5mC and 5hmC in GC and AC sites by colony sequencing. The 224-bp DNA-C, DNA-5mC and DNA-5hmC were first denatured to ssDNA and then separately treated with eA3A-5 followed by colony sequencing. Fifty clones for each sample were randomly picked up and sequenced. (A) C in GC and AC sites from DNA-C were all deaminated and read as T. (B) 5mC in GC and AC sites from DNA-5mC were all deaminated and read as T. (C) 5hmC in GC and AC sites from DNA-5hmC were resistant to deamination and were still read as C. Blue stick represents T; Orange stick represents C.

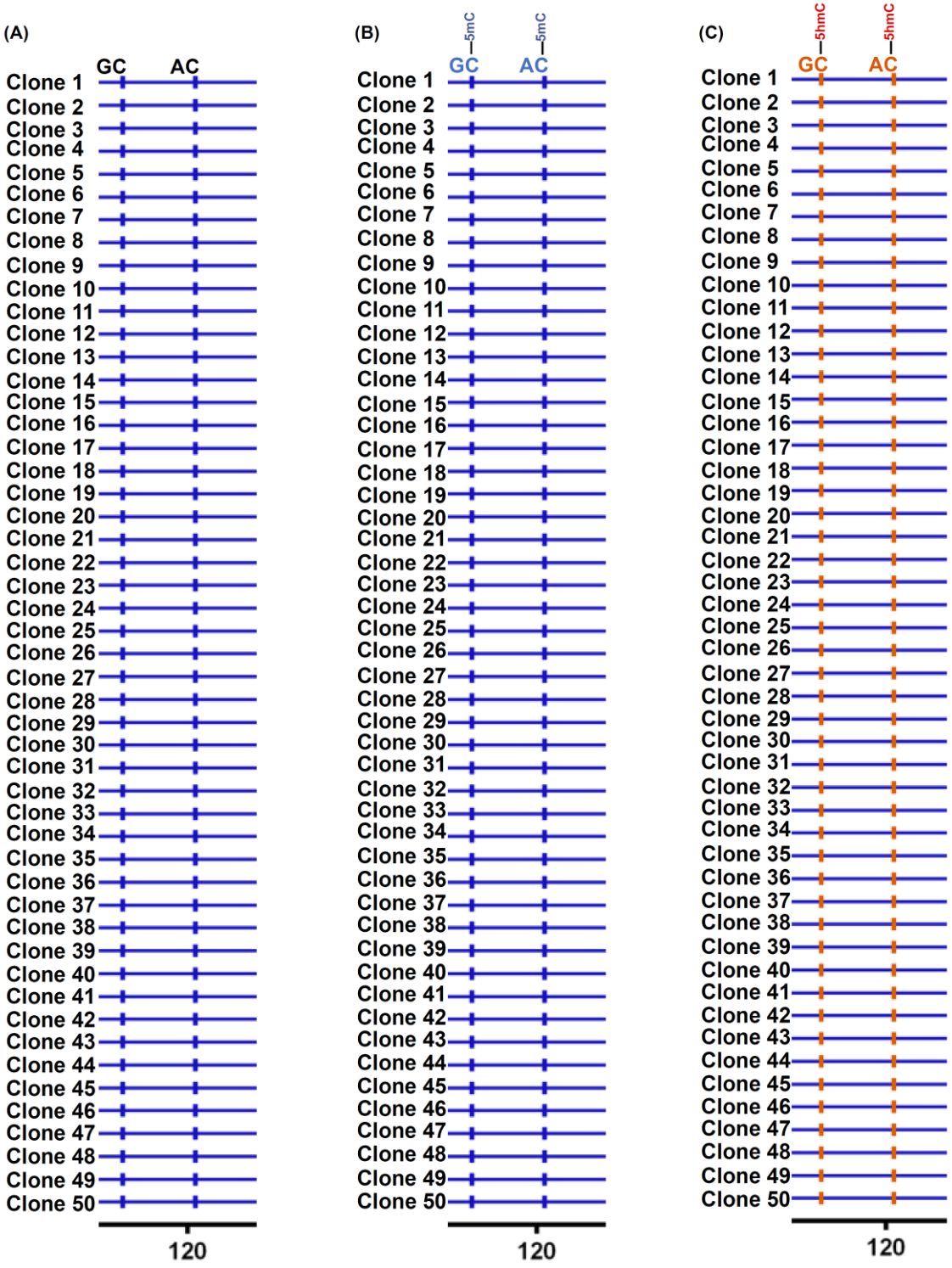


Figure S14. Evaluation of the deaminase activity of eA3A-9 to C, 5mC and 5hmC in TC and CC sites by colony sequencing. The 224-bp DNA-C, DNA-5mC and DNA-5hmC were first denatured to ssDNA and then separately treated with eA3A-9 followed by colony sequencing. Fifty clones for each sample were randomly picked up and sequenced. (A) C in TC and CC sites from DNA-C were all deaminated and read as T. (B) 5mC in TC and CC sites from DNA-5mC were all deaminated and read as T. (C) 5hmC in TC and CC sites from DNA-5hmC were resistant to deamination and were still read as C. Blue stick represents T; Orange stick represents C.

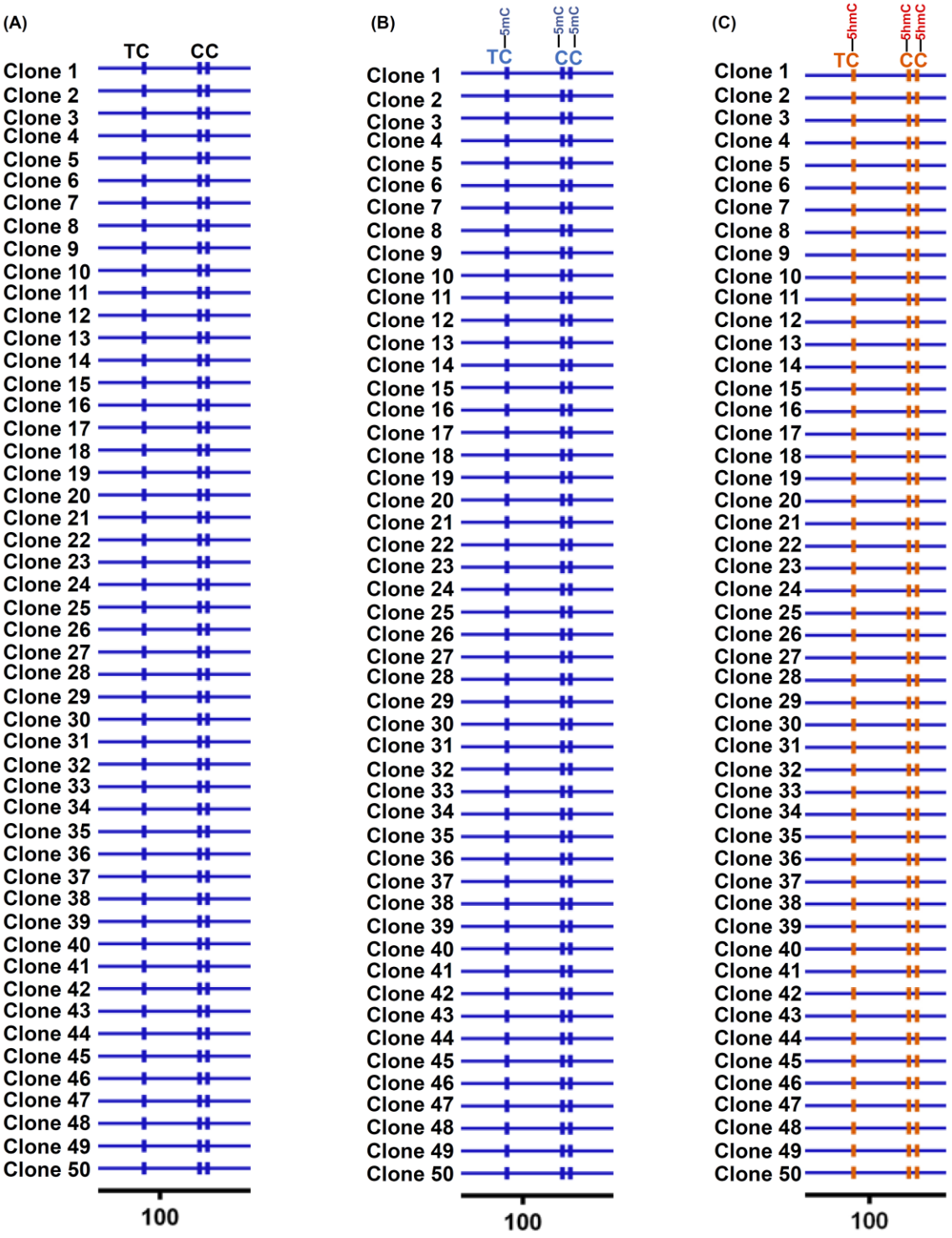


Figure S15. Evaluation of the deaminase activity of wtA3A to C, 5mC and 5hmC in GC and AC sites by colony sequencing. The 224-bp DNA-C, DNA-5mC and DNA-5hmC were first denatured to ssDNA and then separately treated with wtA3A followed by colony sequencing. Fifty clones for each sample were randomly picked up and sequenced. (A) C in GC and AC sites from DNA-C were all deaminated and read as T. (B) 5mC in GC and AC sites from DNA-5mC were all deaminated and read as T. (C) 5hmC in GC and AC sites from DNA-5hmC were partially deaminated, and therefore were partially read as T and partially read as C. Blue stick represents T; Orange stick represents C.

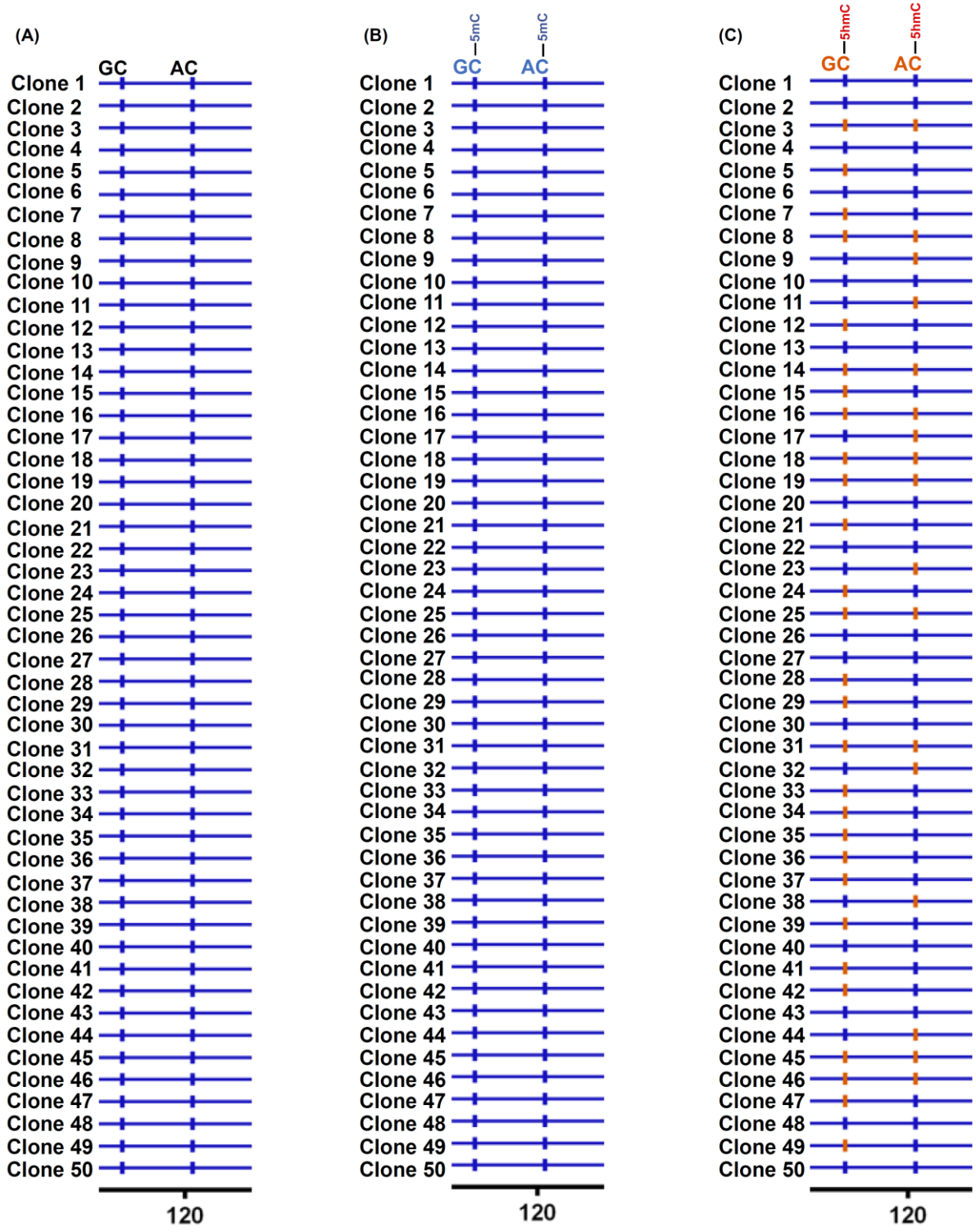


Figure S16. Evaluation of the deaminase activity of wtA3A to C, 5mC and 5hmC in TC and CC sites by colony sequencing. The 224-bp DNA-C, DNA-5mC and DNA-5hmC were first denatured to ssDNA and then separately treated with wtA3A followed by colony sequencing. Fifty clones for each sample were randomly picked up and sequenced. (A) C in TC and CC sites from DNA-C were all deaminated and read as T. (B) 5mC in TC and CC sites from DNA-5mC were all deaminated and read as T. (C) 5hmC in TC and CC sites from DNA-5hmC were all deaminated and read as T. Blue stick represents T; Orange stick represents C.

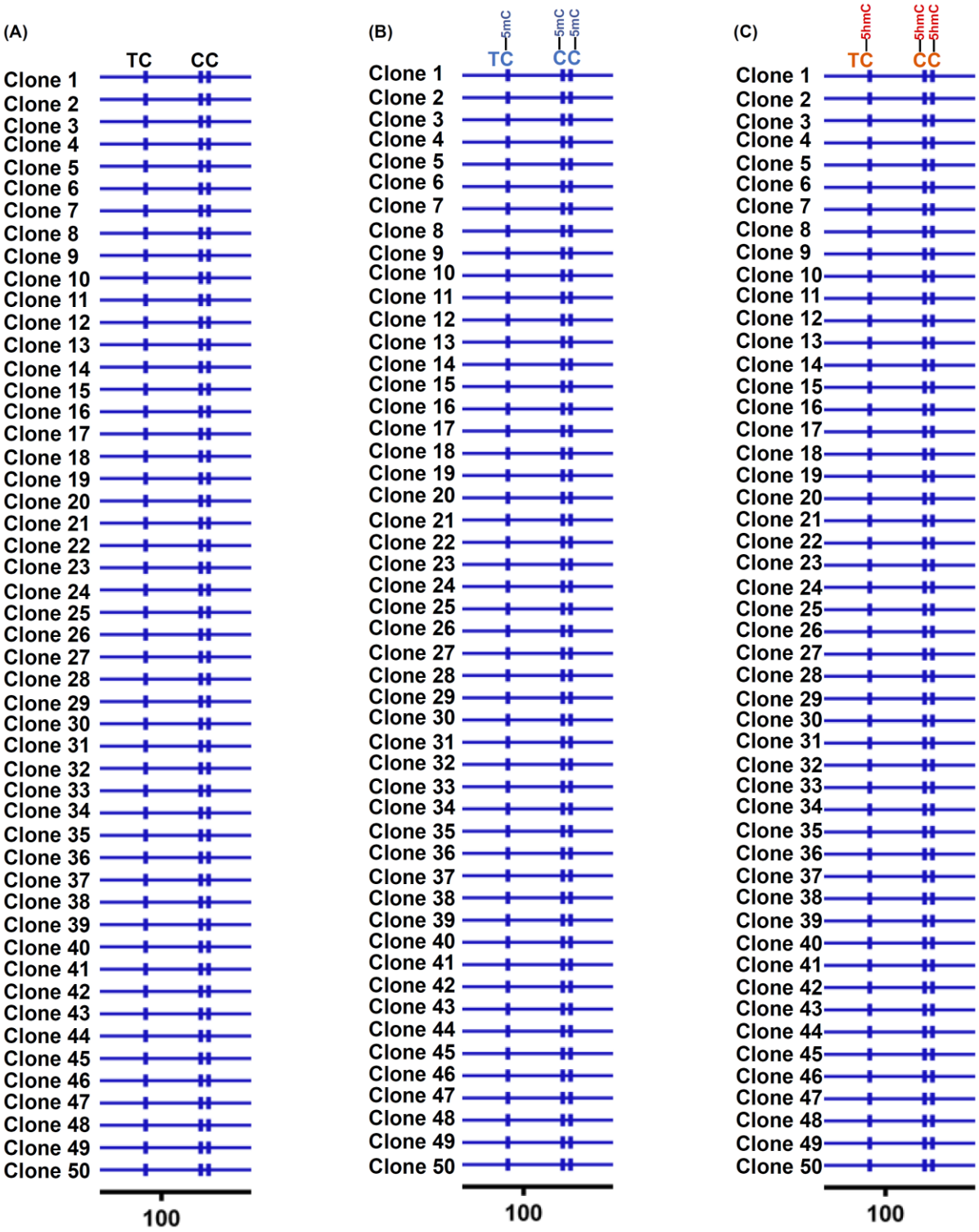


Figure S17. Quantitative evaluation of the level of 5hmC at AC and TC sites of DNA using EDM-seq with Sanger sequencing. (A) DNA-C and DNA-5hmC were mixed at different ratios with DNA-5hmC ranging from 0% to 100%. The mixtures were treated with eA3A-5 followed by Sanger sequencing. (B) Linear regression of the measured percentage versus theoretical percentage of 5hmC at AC site by EDM-seq with Sanger sequencing. (C) DNA-C and DNA-5hmC were mixed at different ratios with DNA-5hmC ranging from 0% to 100%. The mixtures were treated with eA3A-9 followed by Sanger sequencing. (D) Linear regression of the measured percentage versus theoretical percentage of 5hmC at TC site by EDM-seq with Sanger sequencing.

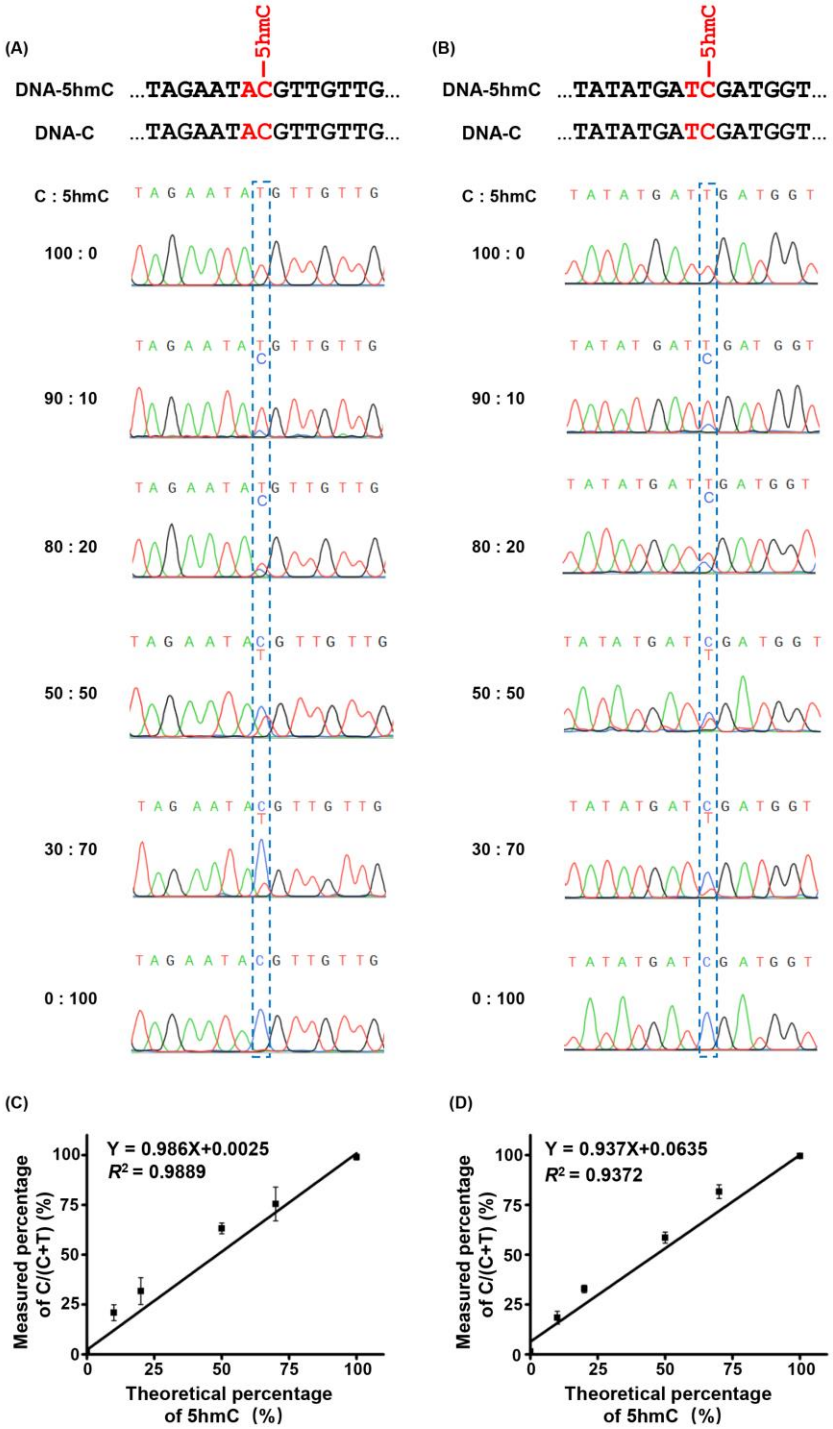
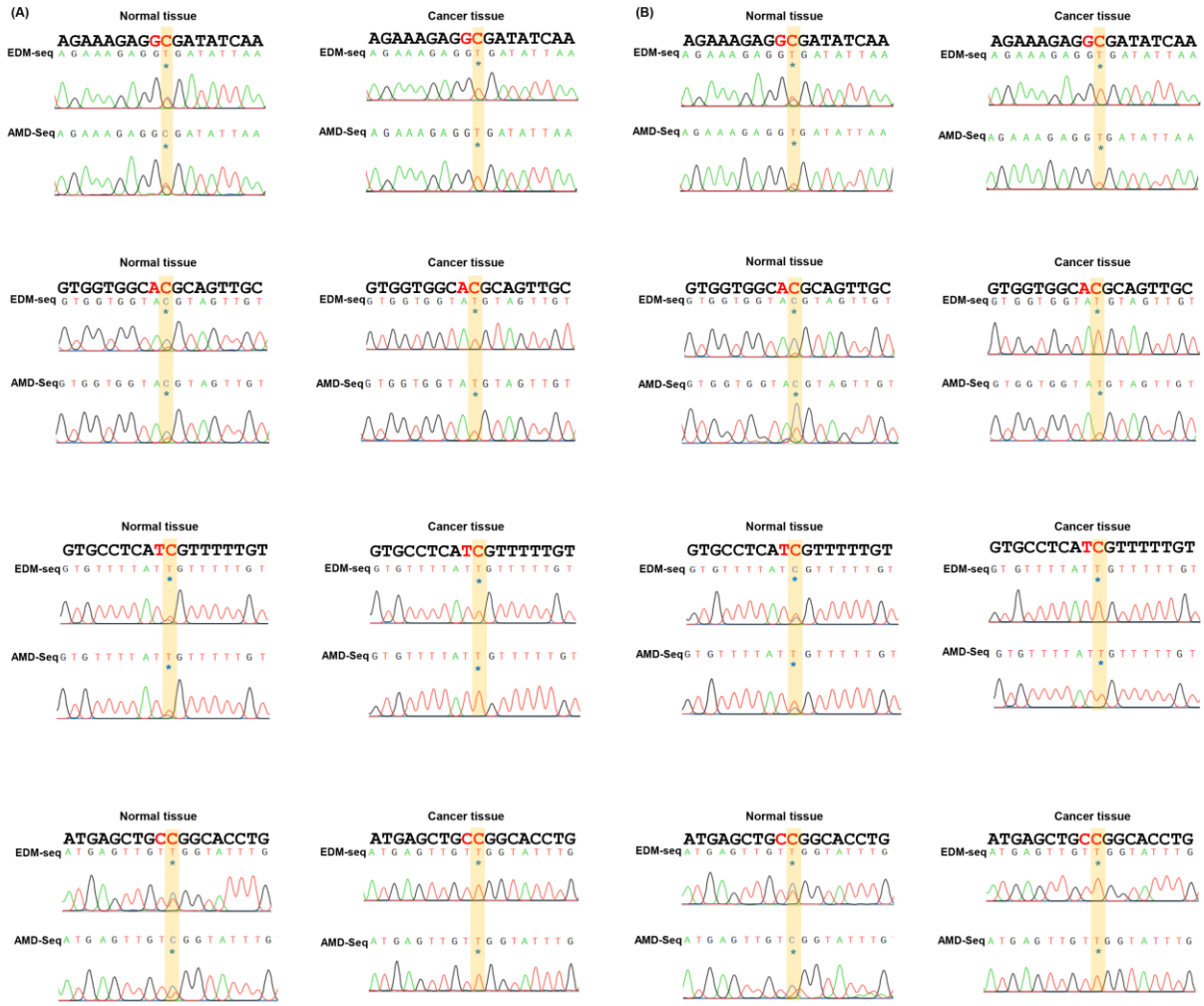


Figure S18. Site-specific detection of 5hmC in genomic DNA of lung cancer tissue and adjacent normal tissue by EDM-seq and AMD-seq. Shown in (A) and (B) are two replicates related to Figure 7.



References

1. F. Tang, S. Liu, Q. Y. Li, J. Yuan, L. Li, Y. Wang, B. F. Yuan and Y. Q. Feng, *Chem Sci*, 2019, **10**, 4272–4281.
2. E. K. Schutsky, J. E. DeNizio, P. Hu, M. Y. Liu, C. S. Nabel, E. B. Fabyanic, Y. Hwang, F. D. Bushman, H. Wu and R. M. Kohli, *Nat Biotechnol*, 2018, **36**, 1083-1090.