Supplementary Information (SI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2025

Supporting Information

for

Orthologous Mammalian A3A-Mediated Single-Nucleotide Resolution Sequencing of DNA Epigenetic Modification 5-Hydroxymethylcytosine

Xia Guo,[1,2,3,#] Jianyuan Wu,[4,#] Tong-Tong Ji,[3] Min Wang,[1] Shan Zhang,[3] Jun Xiong,[1] Fang-Yin Gang,[1] Wei Liu,[1] Yao-Hua Gu,[1,5] Yu Liu,[1,6,*] Neng-Bin Xie,[1,*] and Bi-Feng Yuan[1,2,3,*]

[1] Department of Occupational and Environmental Health, School of Public Health, Wuhan University, Department of Radiation and Medical Oncology, Zhongnan Hospital of Wuhan University, Wuhan 430071, China.

[2] Research Center of Public Health, Renmin Hospital of Wuhan University, Wuhan 430060, China.

[3] College of Chemistry and Molecular Sciences, Hubei Key Laboratory of Biomass Resource Chemistry and Environmental Biotechnology, Wuhan University, Wuhan 430072, China.

[4] Clinical Trial Center, Zhongnan Hospital of Wuhan University, Wuhan 430071, China.

[5] School of Nursing, Wuhan University, Wuhan 430071, China.

[6] Hubei Key Laboratory of Tumor Biological Behaviors, Cancer Clinical Study Center, 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

Neng-Bin Xie. E-mail: nengbinxie@whu.edu.cn

Yu Liu. E-mail: liuyu97@whu.edu.cn

Table of Contents

Page S3-S4	*In vitro* expression and purification of orthologous mammalian A3A proteins; Colony sequencing; Enzymatic digestion of DNA; LC-MS/MS analysis.
Page S5	Table S1. Sequences of oligonucleotides.
Page S6	Table S2. Information of orthologous mammalian A3A proteins.
Page S7	Table S3. Sequences of dsDNA.
Page S8	Table S4. The MRM parameters for analysis of nucleosides by LC-MS/MS.
Page S9	Table S5. Information of detected 5hmC sites in genomic DNA of human lung tissue and the adjacent normal tissue and the PCR primers.
Page S10	Figure S1. Multiple sequence alignment of mammalian A3A proteins using CLUSTALW.
Page S11	Figure S2. Expression and purification of hA3A and evaluation of the deaminase activity of hA3A.
Page S12	Figure S3. Expression and purification of cowA3A and evaluation of the deaminase activity of cowA3A.
Page S13	Figure S4. Expression and purification of porpoiseA3A and evaluation of the deaminase activity of porpoiseA3A.
Page S14	Figure S5. Expression and purification of pandaA3A and evaluation of the deaminase activity of pandaA3A.
Page S15	Figure S6. Expression and purification of gmA3A.
Page S16	Figure S7. Expression and purification of dogA3A.
Page S17	Figure S8. LC-MS/MS analysis of dA, dG, and T from gmA3A or dogA3A treated DNA.
Page S18	Figure S9. Evaluation of the deaminase activity of gmA3A toward 5mC and 5hmC at TC and CC sites by colony sequencing.
Page S19	Figure S10. Evaluation of the deaminase activity of dogA3A toward 5mC and 5hmC at GC and AC sites by colony sequencing.
Page S20	Figure S11. Schematic overview of the ACE-seq method.
Page S21	Figure S12. Site-specific and quantitative detection of 5hmC in genomic DNA of lung cancer tissue and the matched adjacent normal tissue by OMA-seq and ACE-seq at chr4:169198493 (TC site).
Page S22	Figure S13. Site-specific and quantitative detection of 5hmC in genomic DNA of lung cancer tissue and the matched adjacent normal tissue by OMA-seq and ACE-seq at chr2:101493058 (CC site).
Page S23	Figure S14. Site-specific and quantitative detection of 5hmC in genomic DNA of lung cancer tissue and the matched adjacent normal tissue by OMA-seq and ACE-seq at chr1:211984122 (GC site).
Page S24	Figure S15. Site-specific and quantitative detection of 5hmC in genomic DNA of lung cancer tissue and the matched adjacent normal tissue by OMA-seq and ACE-seq at chr3:46967292 (AC site).
Page S25	References

Methods

In vitro expression and purification of orthologous mammalian A3A proteins

The sequences for orthologous mammalian A3A proteins are obtained from the NCBI database and are listed in Table S2. For the expression and purification of the human A3A (hA3A) protein, the coding sequence was incorporated into the pET-41a(+) plasmid, which was synthesized *de novo* by TsingKe Co., Ltd. (Beijing, China). This plasmid includes a human rhinovirus 3C protease (HRV 3C) site between the glutathione S-transferase (GST) tag and the hA3A protein sequence. The resulting plasmid, pET-41a(+)-hA3A, was transformed into the *E. coli* BL21(DE3) *pLysS* strain (Sangon). The transformed *E. coli* cells were cultured in LB medium at 37°C, supplemented with kanamycin (30 μg/mL) and chloramphenicol (10 μg/mL). When the optical density at 600 nm (OD600) reached 0.6 to 0.8, hA3A expression was induced by adding isopropyl-*β*-D-thiogalactoside (IPTG) to a final concentration of 1 mM and incubating at 25°C for 20 h. The *E. coli* cells were then harvested by centrifugation at 8,000 rpm for 3 min. For purification of the hA3A protein, the cell pellets were resuspended in PBS buffer and sonicated using an Ultrasonic Homogenizer JY92-IIN (Scientz). The lysate was centrifuged at 8,000 rpm for 20 min to remove cell debris. The resulting supernatant was filtered through a 0.22-μm membrane and incubated with BeyoGold™ GST-tag Purification Resin (Beyotime) for 2 h. After digestion with HRV 3C protease (Sangon), the supernatant containing hA3A protein was concentrated using a 10-kDa ultrafiltration spin column (Millipore) and equilibrated with a storage solution containing 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 0.01 mM EDTA, 0.5 mM DTT, and 0.01% Tween-20. The purity of the protein was assessed by SDS-PAGE, and was stored at -80°C after the addition of 25% glycerol. The concentration of the purified protein was quantified using the BCA protein assay kit (Beyotime).

The expression and purification of other mammalian A3A proteins, including cowA3A, pandaA3A, porpoiseA3A, gmA3A, and dogA3A, were conducted in a similar manner to that of hA3A.

Colony sequencing

Colony sequencing was performed as previously described.[1, 2] Briefly, DNA treated with gmA3A or dogA3A was amplified by PCR and purified using KAPA Pure beads (Roche) according to the manufacturer's protocol. The purified products (50 ng) were then cloned into the pCE2 TA/Blunt-Zero vector (Vazyme Biotech Co., Ltd, Nanjing, China) and transformed into competent *E. coli* cells. Fifty clones were randomly picked up and sequenced, and the percentage of T/(C+T) at specific sites was calculated from the sequence data, providing a measure of gmA3A or dogA3A deamination activity at those sites.

Enzymatic digestion of DNA.

DNA digestion was performed as previously described.[3] Briefly, DNA was digested in a 50-μL solution containing 5 μL of neutral buffer (500 mM Tris–HCl, 10 mM $MgCl_2$, 100 mM NaCl, 10 mM $ZnSO_4$, pH 7.0), 2.5 U of DNase I, 90 U of S1 nuclease, 7.5 U of alkaline phosphatase, and 0.125 U of venom phosphodiesterase I at 37°C for 6 h. The reaction mixture was then diluted with 250 μL of H_2O and extracted with 300 μL of chloroform three times to remove proteins. The aqueous layer was lyophilized and reconstituted in 50 μL of H_2O, followed by LC-MS/MS analysis.

LC-MS/MS analysis

The analysis of nucleosides was conducted using an LC-MS/MS system, which included a Shimadzu 8045 mass spectrometer (Kyoto, Japan) with a Turbo Ionspray electrospray ionization source, coupled with a Shimadzu LC-30AD UPLC system. Separation of the digested nucleosides was achieved on a Shim-pack GIST C18 column (2.1 mm × 100 mm, 2.0 μm, Shimadzu, Japan) at a flow rate of 0.3 mL/min and a temperature of 40°C. Solvent A (0.05% formic acid) and solvent B (methanol) were used as the mobile phases, with a 15-min gradient for separation as follows: 0-3 min at 5% B, 3-7 min ramping from 5% to 80% B, 7-10 min at 80% B, 10-12 min ramping from 80% to 5% B, and 12-15 min at 5% B. The nucleosides were monitored using multiple reaction monitoring (MRM) in positive-ion mode. The optimized MRM parameters are provided in Table S4.

Table S1. Sequences of oligonucleotides.

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

Note: the "C" represented 5′-aza-2′-deoxycytidine.

Table S2. Information of orthologous mammalian A3A proteins.

Name	Source	NCBI Reference Sequence	Amino acid composition
hA3A	*Homo sapiens*	NP_001180218.1	MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLC YEVERLDNGTSVKMDQHRGFLHNQAKNLLCGF YGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPC FSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPL YKEALQMLRDAGAQVSIMTYDEFKHCWDTFVD HQGCPFQPWDGLDEHSQALSGRLRAILQNQGN
gmA3A	*Chlorocebus aethiops*	ADO85881.1 (GenBank)	MDGSPASRPRHLMDPDTFTFNFNNDLSILGRRQT YLCYEVERLDNGTWVPMDERRGFLHNKAKNLP HGDYGCHAELCFLGEVPSWQLDPAQTYRVTWFI SWSPCFSRGCAGQVRAFLQENTHMKLRIFAARIY DSDFLYEKALRTLRDAGAQVSIMTYEEFKHCWD TFVDHQGRPFQPWDGLDEHSQALSGRLRAILQN QGN
cowA3A	*Bos taurus*	NP_001157408.1	MDEYTFTENFNNQGRPSKTYLCYEMERLDGNAT IPLDEYKGFVRNKGLDQPEKPCHAELYFLGKIRS WNLDRNQHYRLTCFISWSPCYDCAQKLTTFLKE NHHISLHILASRIYTRNHFGCHQSGLCELQAAGA RITIMTFEDFKHCWETFVDHKGKPFQPWEGLNV KSQALCAELQAILKTQQN
porpoiseA3A	*Neophocaena asiaeorientalis asiaeorientalis*	XP_024617343.1	MEASTAPWTSCLLDENTFTENFMNRLRPRKTYL CYKVEILDGDARVPLDEKKGFVRNKVTDPACPQ QAGPPYCGTLRVEGCQLHTGCPTSSLTPGPCRCY RCSAWNQGANEPGMPRHAECYFLDRIRSWNLD RGLHYRLTCFISWTPCHSCAQELATFLGENSHVS LHIFASRIYRRPGYEAGILTLRAAGAQIAIMTSKE FQHCWENFVDHQERPFRPWVGLEVESQHQCNEL QAILQTQAN
pandaA3A	*Ailuropoda melanoleuca*	XP_002914629.1	MDAGAEAWDRHLLDEDTFTENFRNDDWPSRTY LCYKVEGPDQGSGVPLGQDKGILHNKPAQGPEP SRHAECYLLEQIQSWNLDPKLHYGVTCFLSWSPC AKCAQKMARFLQENSHVSLKLFASRLYTRERWD EDYKEGLRTLKRAGASIAIMTYREFEHCWKTFVL HDQEGSCFQPWPFLHKESQKFSEKLQAILQGA
dogA3A	*Canis lupus familiaris*	NP_001333061.1	MEAGPEDWDRHLLDENTFTQNFRNDHNPSKTYL CYQVELSDGSSGVLLDQDKDIVQNEGGGGQHAE WFLLEHIRSRNLDQKLSYKVTCFLSWTPCEKCAE EIIRFLAKNRHVSLSILASRIYTMGPYVKGLRELY DAGVHISIMTFRDFEYCWQTFVDHQDSPFQPWA DLDRRSQQLSQQLRAILQKEPEGWTSVCL

Table S3. Sequences of dsDNA.

Name	Sequence (5′ to 3′)
DNA-C	AGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGATTATG AGTATGTATAGTGTTAGGAAGAGTGTAGTAATAGGATGAAGA TGATTATATGATCGATGGTCCGTATGCGTAGAATACGTTGTTG TAGTGATTATAATGGAGTGAGAATGTAGATGAGTGGAGTAGG TAGTAAGATGTAGTGGTGATAGAGAGTAATTGTTAGTGGAAT GTTGG
DNA-5mC	AGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGATTATG AGTATGTATAGTGTTAGGAAGAGTGTAGTAATAGGATGAAGA TGATTATATGAT5mCGATGGT5mC5mCGTATG5mCGTAGAATA 5mCGTTGTTGTAGTGATTATAATGGAGTGAGAATGTAGATGA GTGGAGTAGGTAGTAAGATGTAGTGGTGATAGAGAGTAATTG TTAGTGGAATGTTGG
DNA-5hmC	AGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTGATTATG AGTATGTATAGTGTTAGGAAGAGTGTAGTAATAGGATGAAGA TGATTATATGAT5hmCGATGGT5hmC5hmCGTATG5hmCGTAGA ATA5hmCGTTGTTGTAGTGATTATAATGGAGTGAGAATGTAGA TGAGTGGAGTAGGTAGTAAGATGTAGTGGTGATAGAGAGTAA TTGTTAGTGGAATGTTGG
228-bp DNA	ACTAGTAGTGACGCTGAGCTTGACGTCGCGCGATGAGAGGTG ATTATGAGTATGTATAGTGTTAGGAAGAGTGTAGTAATAGGA TGAAGATGATTATATGATCGATGGTCCGTATGCGTAGAATAC GTTGTTGTAGTGATTATAATGGAGTGAGAATGTAGATGAGTG GAGTAGGTAGTAAGATGTAGTGGTGATAGAGAGTAATTGTTA GTGGAATGTTGGCTCGAG

Table S4. The MRM parameters for analysis of nucleosides by LC-MS/MS.

Nucleosides	Precursor ion (*m/z*)	Product ion (*m/z*)	Q1 Prerod (V)	CE (V)	Q2 Prerod (V)
dG	268.2	152.1	-22	-50	-29
dA	252.2	136.1	-20	-50	-20
dC	228.2	112.1	-11	-50	-20
T	243.2	127.0	-12	-40	-22
5mC	242.2	126.1	-18	-10	-25
5hmC	258.2	142.1	-19	-9	-24

Table S5. Information of detected 5hmC sites in genomic DNA of human lung tissue and the adjacent normal tissue and the PCR primers. The GRCh37 version of human genomic DNA (from NCBI) was used for the genome location.

Genome location	Site type	Primers		Sequence (5′ to 3′)
chr4:169198493	TC	Without gmA3A treatment	Forward	GGTATCCACAGCAGCTTTGGAAGAACTC
			Reverse	GCAAATATATAGCACATCCATAAGCACCT
		With gmA3A treatment	Forward	GGTATTTATAGTAGTTTTGGAAGAATTTAGTG
			Reverse	ACAAATATATAACACATCCATAAACACCT
chr2:101493058	CC	Without gmA3A treatment	Forward	GGCCTTTCCATTTGGCCATATGTCA
			Reverse	AAACAGATGGGATTAGGCCTGAGCC
		With gmA3A treatment	Forward	GTTGTTTAGTTGTATTTGGAGATTATAG
			Reverse	AAACCACCATAACATACACACTCCT
chr1:211984122	GC	Without dogA3A treatment	Forward	CACTTATTGGGAATAAAAAACAAAC
			Reverse	CCTGGCCTAAAGTCTAGTATCTGAC
		With dogA3A treatment	Forward	GAAGTTAGGAAAAGAAGTTATTGTGAGTTG
			Reverse	AACCTACTTCCCTCTATTATCACCTATATAATAAT
chr3:46967292	AC	Without dogA3A treatment	Forward	GGGACTGGCTGCCCTCCAGAGGC
			Reverse	CCTTGAGCTAAGATGTGCAGCCACTGCTGTGT
		With dogA3A treatment	Forward	GGAGGAGAAGTATAGAATGGAGTGTTATTAGGTAGG
			Reverse	TCTCTCTAACCTCCAATCCCCTTATCTATCAAAT

Figure S1. Multiple sequence alignment of mammalian A3A proteins using CLUSTALW. Conserved regions are highlighted in blue, with the intensity of blue indicating the degree of sequence conservation.

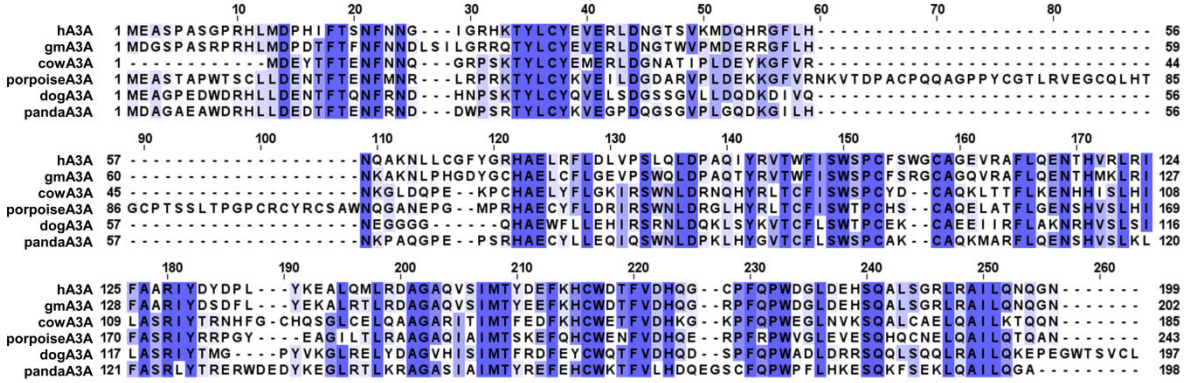


Figure S2. Expression and purification of hA3A and evaluation of the deaminase activity of hA3A. (A) Schematic illustration of the pET-41a(+)-hA3A plasmid. (B) SDS-PAGE analysis of the purified hA3A. (C) Characterization of the deaminase selectivity of hA3A towards C, 5mC and 5hmC in different sequence contexts by Sanger sequencing.

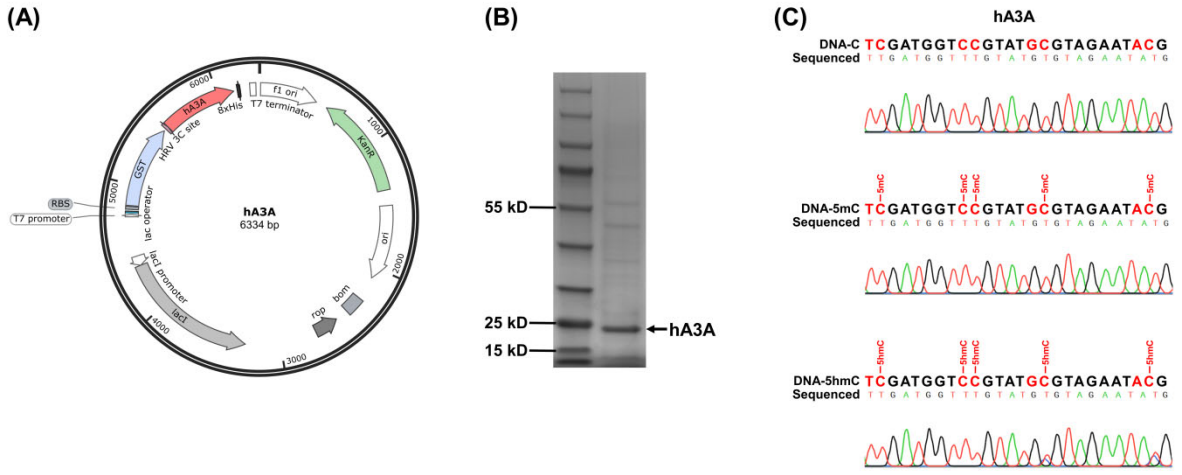


Figure S3. Expression and purification of cowA3A and evaluation of the deaminase activity of cowA3A. (A) Schematic illustration of the pET-41a(+)-cowA3A plasmid. (B) SDS-PAGE analysis of the purified cowA3A. (C) Characterization of the deaminase selectivity of cowA3A towards C, 5mC and 5hmC in different sequence contexts by Sanger sequencing.

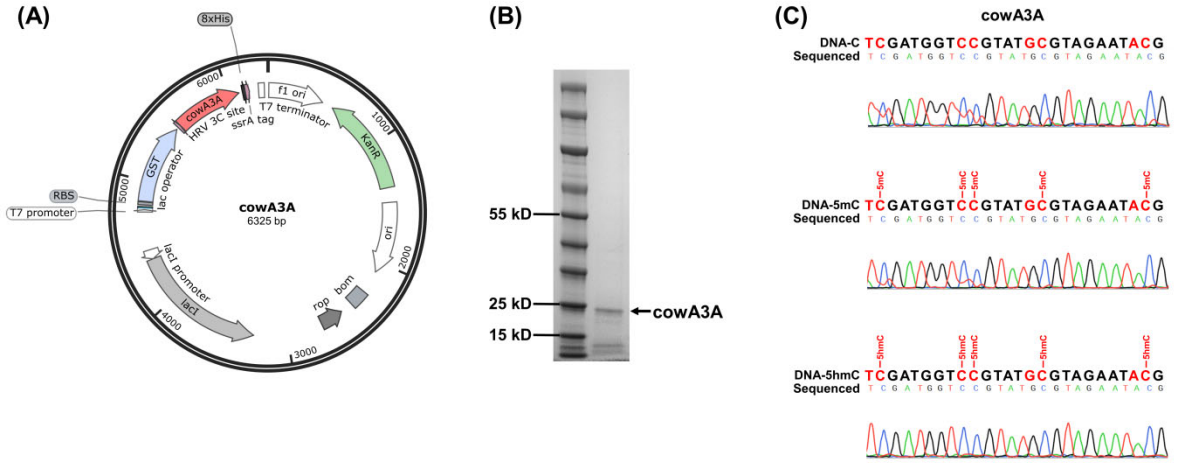


Figure S4. Expression and purification of porpoiseA3A and evaluation of the deaminase activity of porpoiseA3A. (A) Schematic illustration of the pET-41a(+)-porpoiseA3A plasmid. (B) SDS-PAGE analysis of the purified porpoiseA3A. (C) Characterization of the deaminase selectivity of porpoiseA3A towards C, 5mC and 5hmC in different sequence contexts by Sanger sequencing.

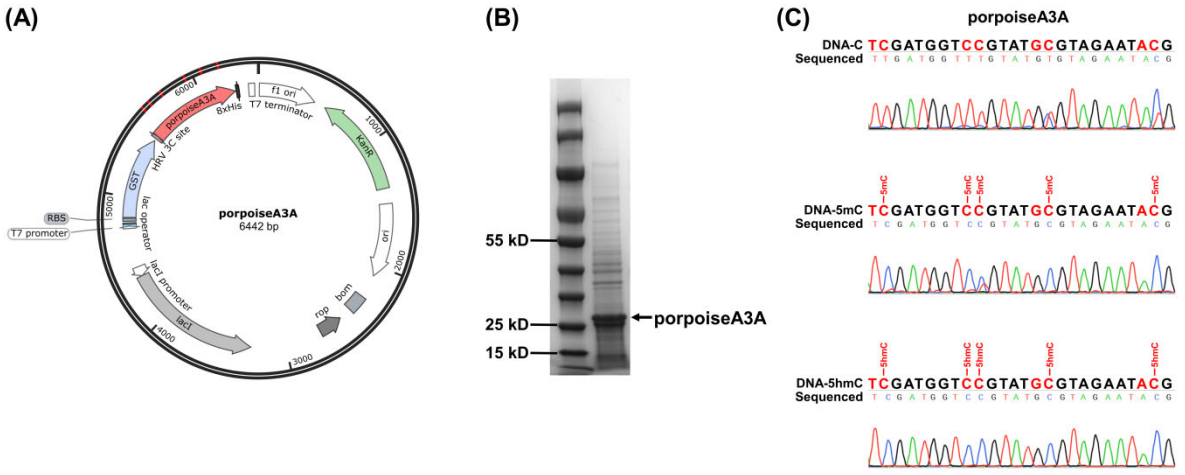


Figure S5. Expression and purification of pandaA3A and evaluation of the deaminase activity of pandaA3A. (A) Schematic illustration of the pET-41a(+)-pandaA3A plasmid. (B) SDS-PAGE analysis of the purified pandaA3A. (C) Characterization of the deaminase selectivity of pandaA3A towards C, 5mC and 5hmC in different sequence contexts by Sanger sequencing.

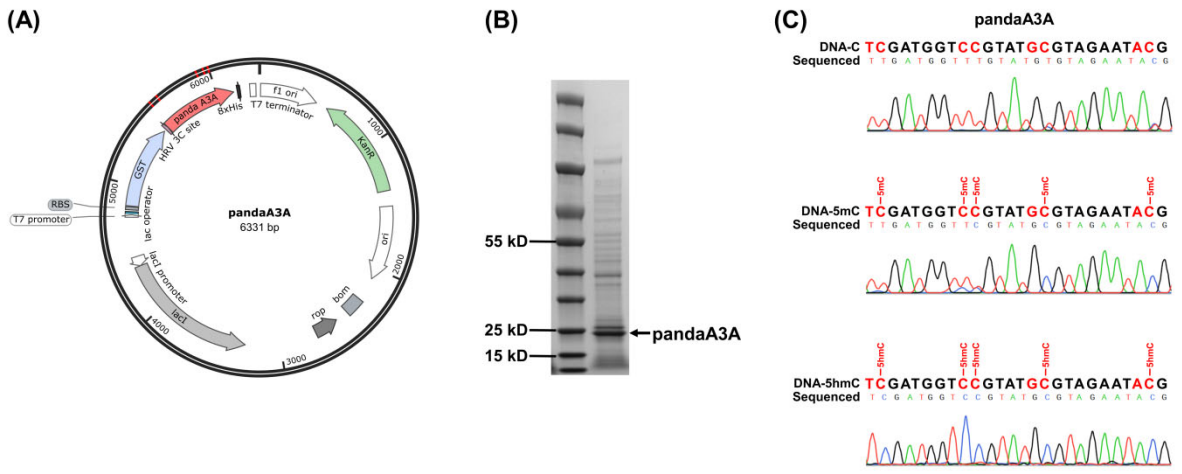


Figure S6. Expression and purification of gmA3A. (A) Schematic illustration of the pET-41a(+)-gmA3A plasmid. (B) SDS-PAGE analysis of the purified gmA3A.

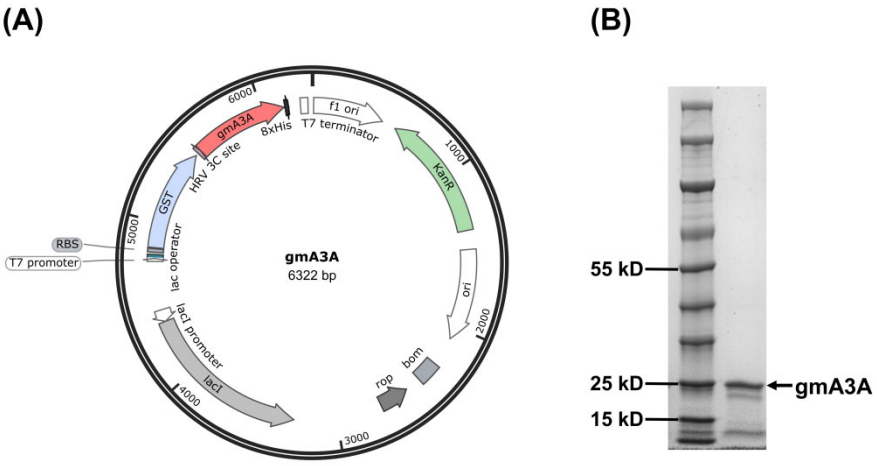


Figure S7. Expression and purification of dogA3A. (A) Schematic illustration of the pET-41a(+)-dogA3A plasmid. (B) SDS-PAGE analysis of the purified dogA3A.

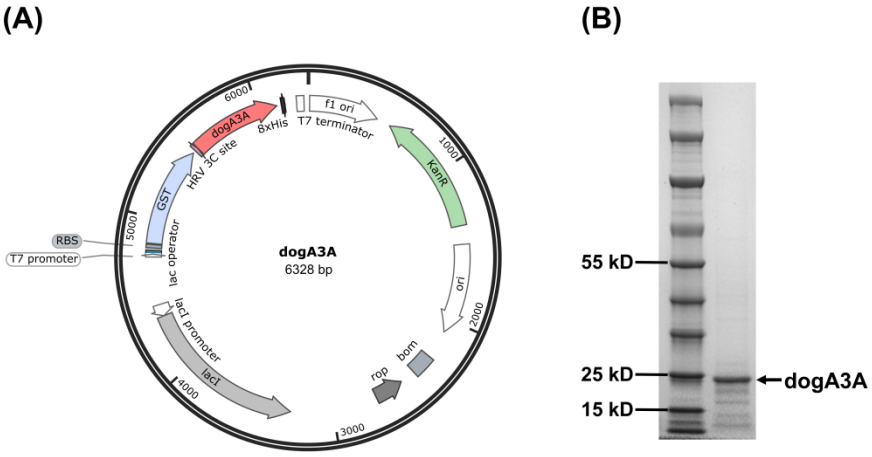


Figure S8. LC-MS/MS analysis of dA, dG, and T from gmA3A or dogA3A treated DNA. (A) Extracted-ion chromatograms of dA, dG, and T in DNA with gmA3A treatment. (B) Extracted-ion chromatograms of dA, dG, and T in DNA with dogA3A treatment.

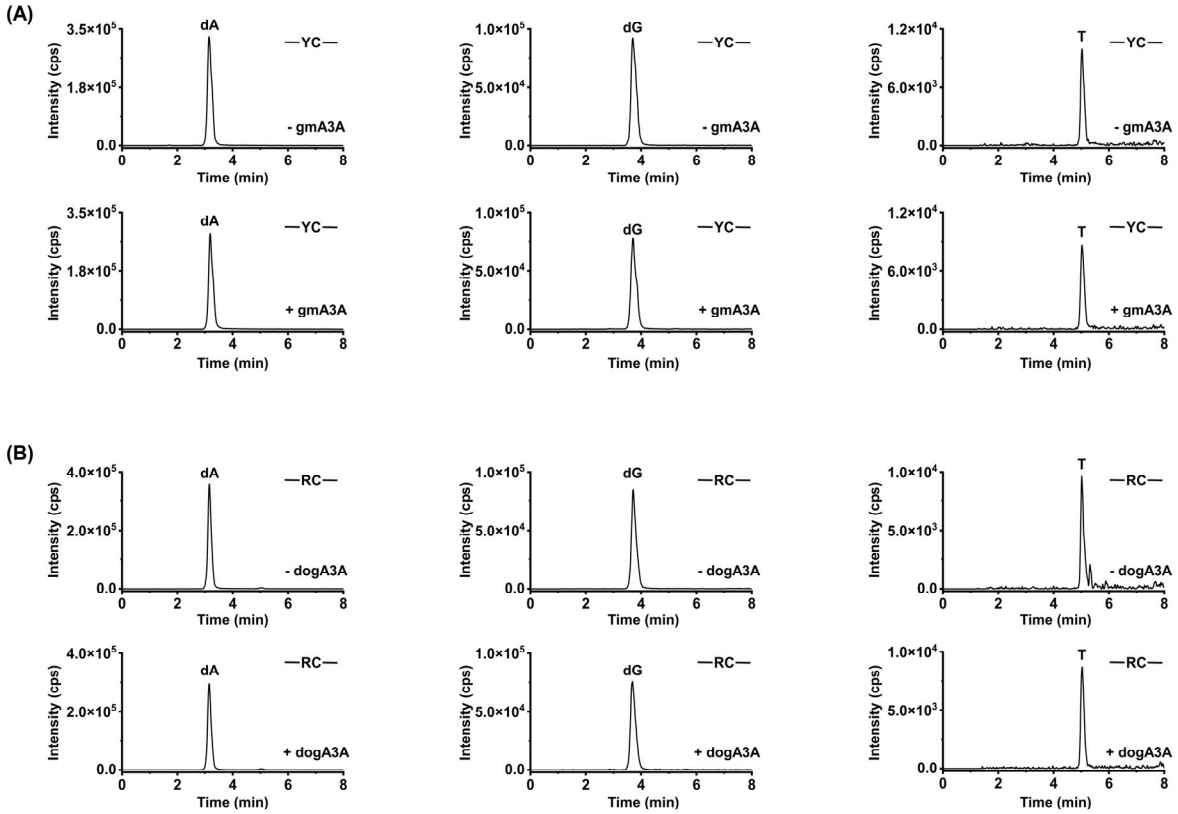


Figure S9. Evaluation of the deaminase activity of gmA3A toward 5mC and 5hmC at TC and CC sites by colony sequencing. (A) 5mC at TC and CC sites from DNA-5mC were all deaminated and read as T. (B) 5hmC at TC and CC sites from DNA-5hmC were resistant to deamination and were still read as C (only one was deaminated and read as T at CC sites).

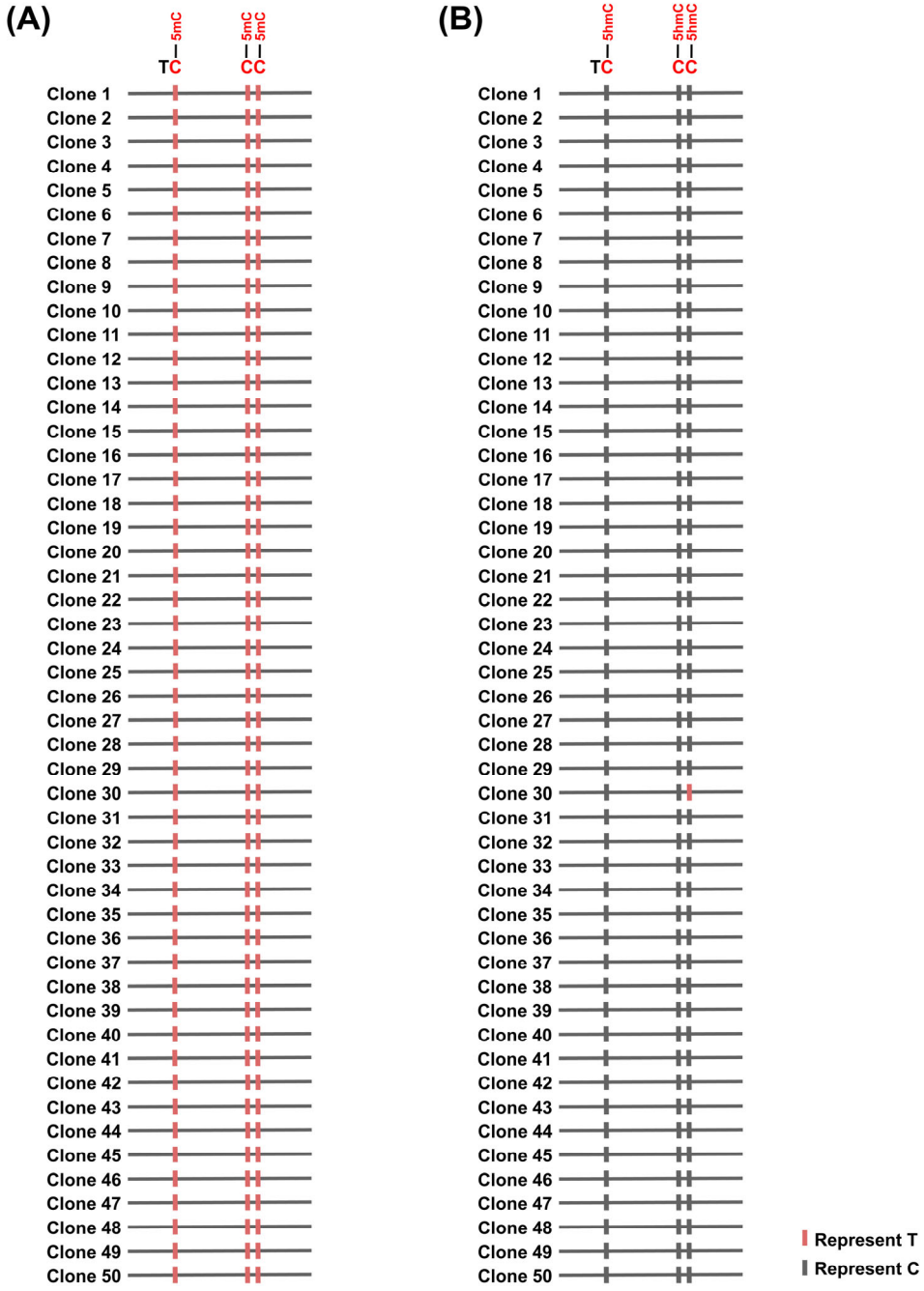


Figure S10. Evaluation of the deaminase activity of dogA3A toward 5mC and 5hmC at GC and AC sites by colony sequencing. (A) 5mC at GC and AC sites from DNA-5mC were all deaminated and read as T. (B) 5hmC at GC and AC sites from DNA-5hmC were resistant to deamination and were still read as C (only one was deaminated and read as T at AC sites).

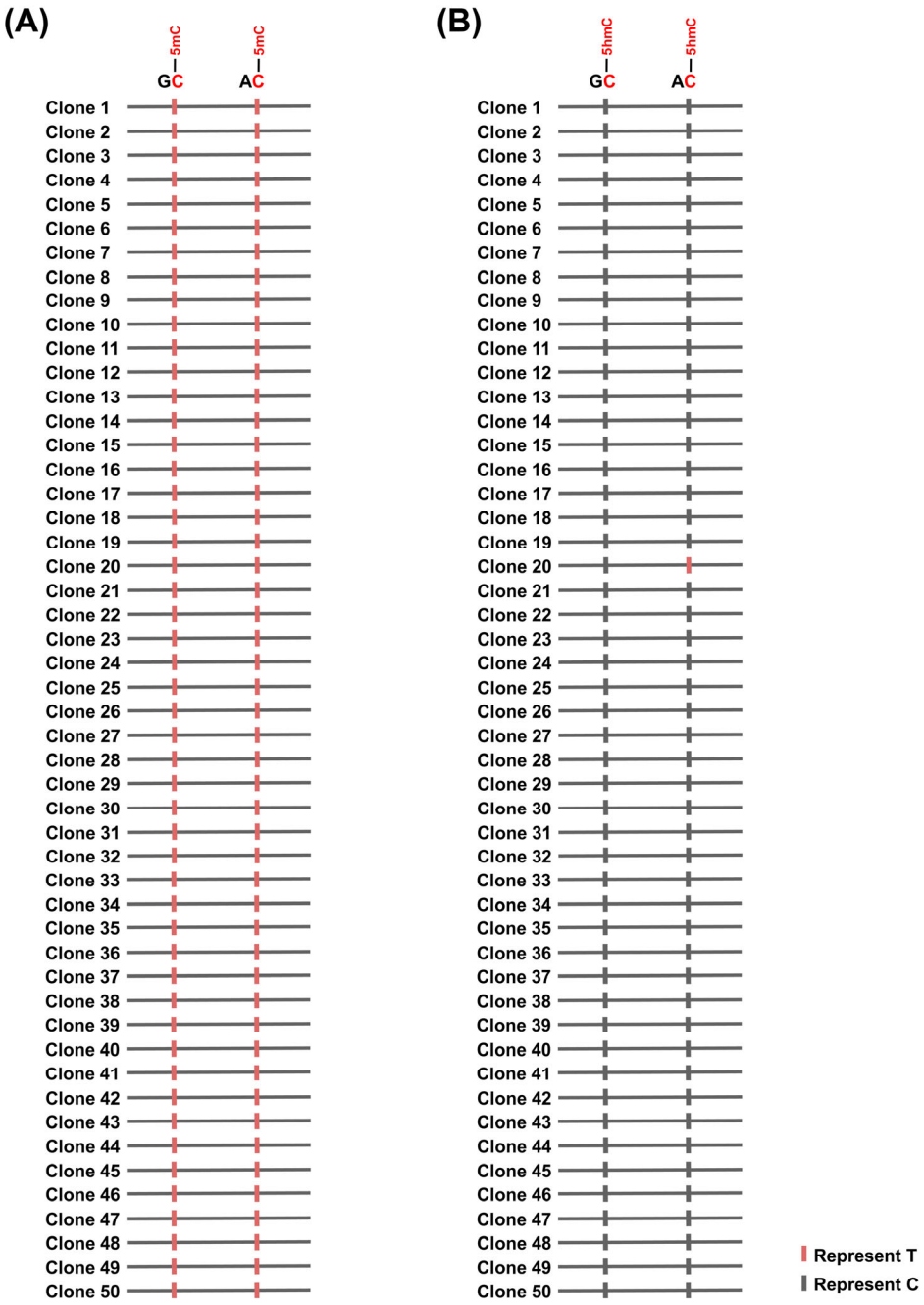


Figure S11. Schematic overview of the ACE-seq method. (A) Deamination of C, 5mC, and 5hmC by hA3A yields U, T, and 5-hydroxymethyluracil (5hmU), respectively, all of which pair with A. Glycosylated 5hmC (*β*-glucosyl-5-hydroxymethyl-2′-deoxycytidine, 5gmC) is resistant to deamination by hA3A and still pairs with G. (B) hA3A completely deaminates C and 5mC, resulting in T reads. In contrast, 5hmC is protected from deamination by glycosylation, allowing it to be read as C.

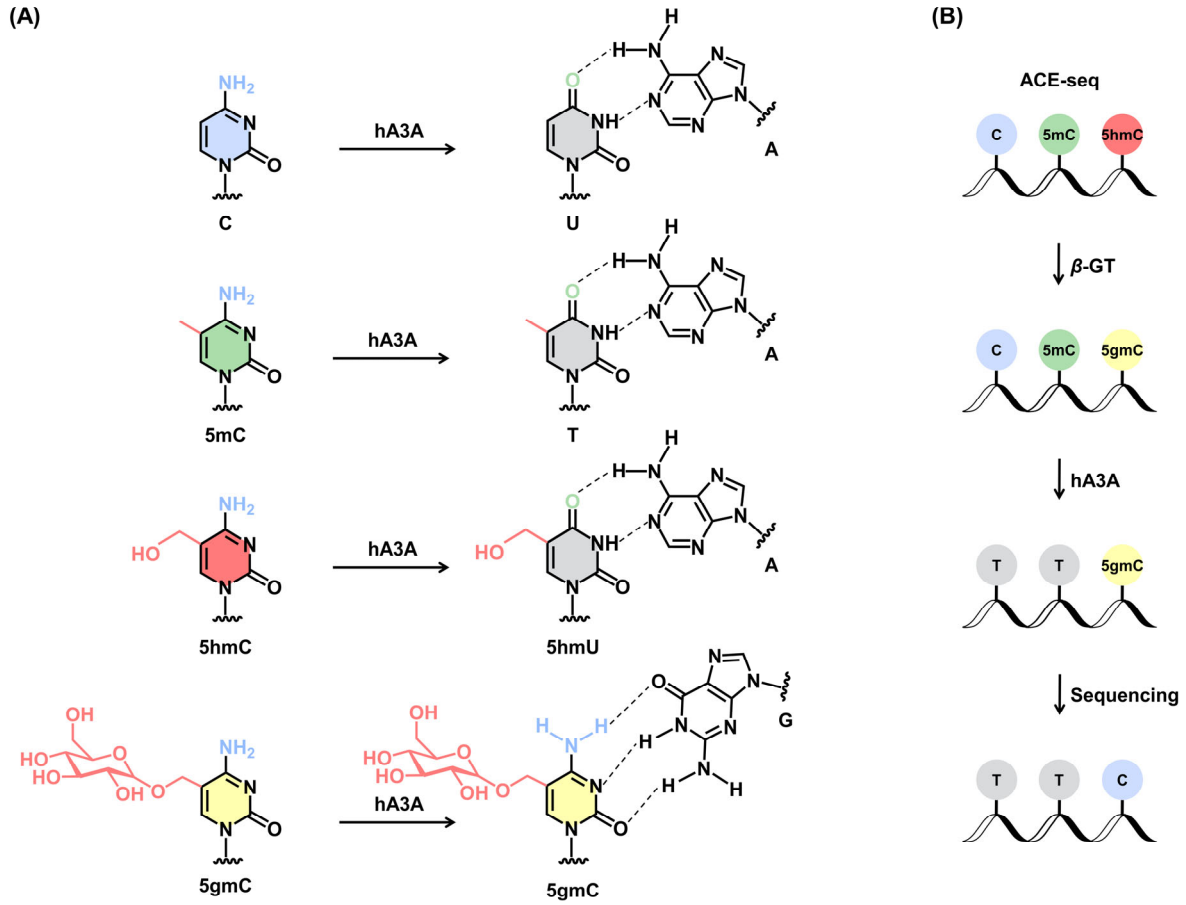


Figure S12. Site-specific and quantitative detection of 5hmC in genomic DNA of lung cancer tissue and the matched adjacent normal tissue by OMA-seq and ACE-seq at chr4:169198493 (TC site). (A) Normal tissue. (B) Cancer tissue.

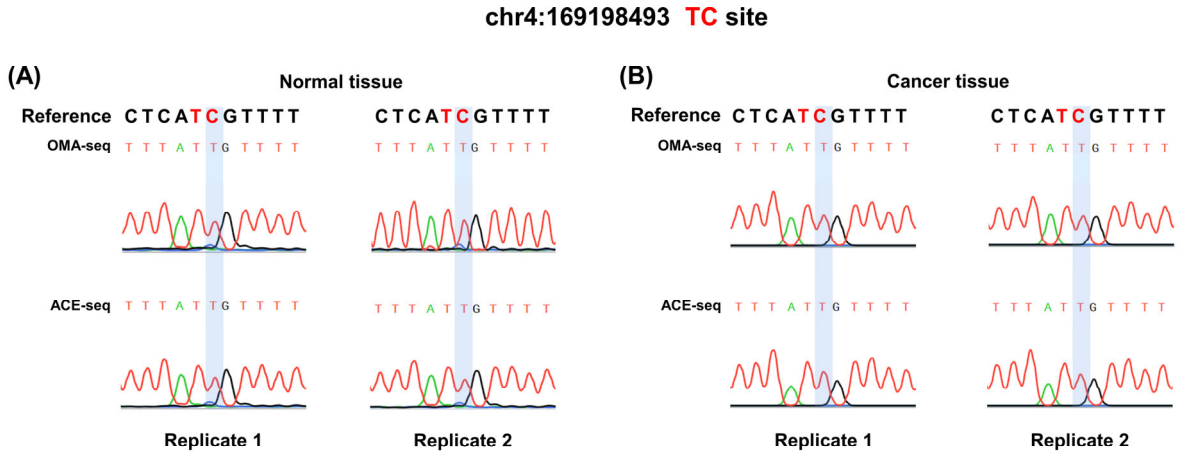


Figure S13. Site-specific and quantitative detection of 5hmC in genomic DNA of lung cancer tissue and the matched adjacent normal tissue by OMA-seq and ACE-seq at chr2:101493058 (CC site). (A) Normal tissue. (B) Cancer tissue.

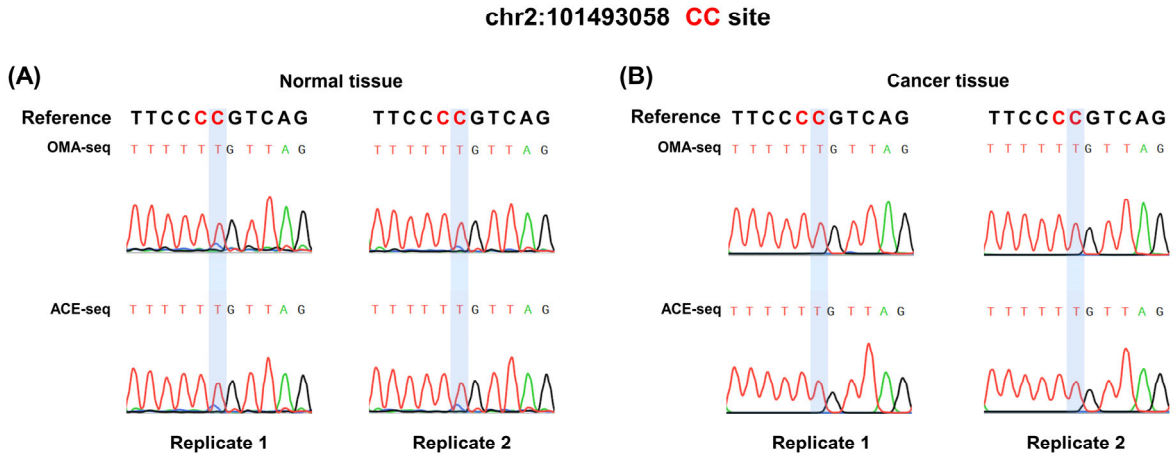


Figure S14. Site-specific and quantitative detection of 5hmC in genomic DNA of lung cancer tissue and the matched adjacent normal tissue by OMA-seq and ACE-seq at chr1:211984122 (GC site). (A) Normal tissue. (B) Cancer tissue.

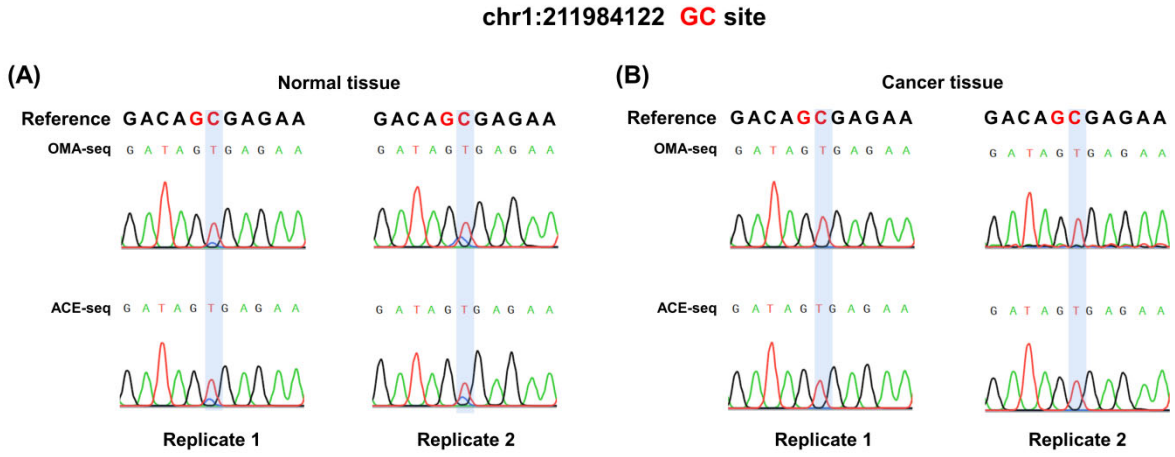
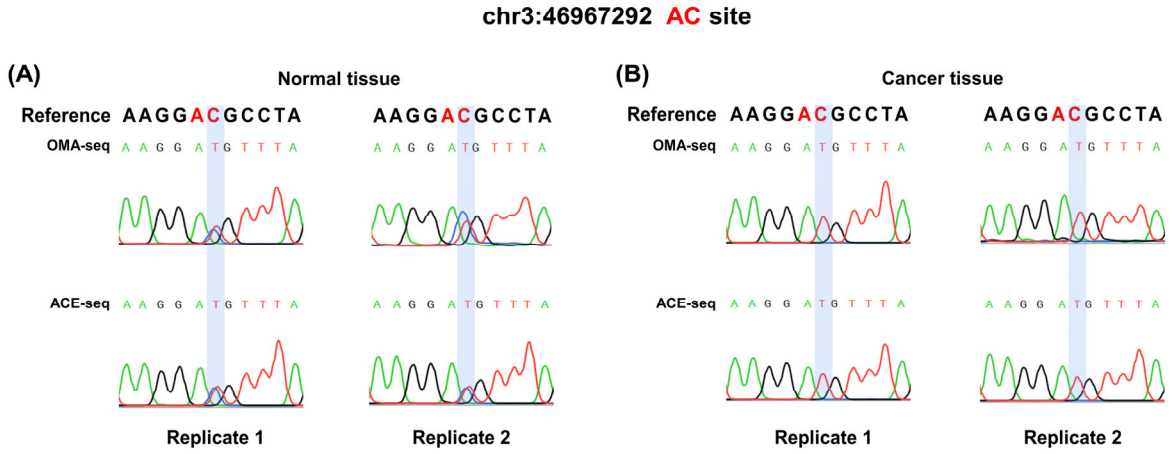


Figure S15. Site-specific and quantitative detection of 5hmC in genomic DNA of lung cancer tissue and the matched adjacent normal tissue by OMA-seq and ACE-seq at chr3:46967292 (AC site). (A) Normal tissue. (B) Cancer tissue.



References

1. Xiong, J.; Wang, P.; Shao, W. X.; Li, G. J.; Ding, J. H.; Xie, N. B.; Wang, M.; Cheng, Q. Y.; Xie, C. H.; Feng, Y. Q.; Ci, W. M.; Yuan, B. F., Genome-wide mapping of N-4-methylcytosine at single-base resolution by APOBEC3A-mediated deamination sequencing. *Chem Sci* **2022,** *13* (34), 9960-9972.

2. Tang, F.; Liu, S.; Li, Q. Y.; Yuan, J.; Li, L.; Wang, Y.; Yuan, B. F.; Feng, Y. Q., Location analysis of 8-oxo-7,8-dihydroguanine in DNA by polymerase-mediated differential coding. *Chem Sci* **2019,** *10*, 4272–4281.

3. Ma, C. J.; Li, G.; Shao, W. X.; Min, Y. H.; Wang, P.; Ding, J. H.; Xie, N. B.; Wang, M.; Tang, F.; Feng, Y. Q.; Ci, W.; Wang, Y.; Yuan, B. F., Single-Nucleotide Resolution Mapping of N(6)-Methyladenine in Genomic DNA. *ACS Cent Sci* **2023,** *9* (9), 1799-1809.