Supplementary Information

for

Genome-wide mapping of N^4-methylcytosine at single-base resolution by APOBEC3A-mediated deamination sequencing

Jun Xiong,[1, 2, 3 †] Ping Wang,[4, 5, †] Wen-Xuan Shao,[1, 2] Gaojie Li,[4, 5] Jiang-Hui Ding,[1, 2] Neng-Bin Xie,[1,2] Min Wang,[1, 2] Qing-Yun Cheng,[1] Conghua Xie,[1, 3] Weimin Ci,[4, 5*] Bi-Feng Yuan,[1,2,3,6 *] Yu-Qi Feng[1,2]

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

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

[3] Cancer Precision Diagnosis and Treatment and Translational Medicine Hubei Engineering Research Center, Zhongnan Hospital of Wuhan University, Wuhan, China.

[4] Key Laboratory of Genomics and Precision Medicine, and China National Center for Bioinformation, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China.

[5] University of Chinese Academy of Sciences, Beijing 100049, China.

[6] Wuhan Research Center for Infectious Diseases and Cancer, Chinese Academy of Medical Sciences.

[†] These authors contributed equally to this work.

* Corresponding author:

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

Weimin Ci. E-mail: ciwm@big.ac.cn

This file contains:

Supplementary Methods

Synthesis of 3-methyl-2′-deoxycytidine (3mdC)

3-Methyl-2′-deoxycytidine (3mdC) was synthesized according to the previous study with slight modifications.[1] Briefly, 1 mg of 2′-deoxycytidine in 400 μL of NaAc–HAc buffer (pH 4.3) was reacted with 100 μL of dimethyl sulfate at 37°C for 1 h. The 3mdC product was separated and purified by HPLC, and confirmed by high-resolution mass spectrometry analysis.

Expression and purification of recombinant A3A protein

The plasmid for A3A protein expression (pET-41a-A3A) was purchased from TsingKe Co., Ltd. pET-41a-A3A plasmid contains the full-length coding sequence of A3A in the pET-41a vector, which carries the glutathione S-transferase (GST) tag and the human rhinovirus 3C protease (HRV 3C) site. A3A protein was expressed and purified according to the previously described procedure with slight modification.[2] Briefly, the pET-41a-A3A plasmid was transformed into *E. coli* BL21 (DE3) competent cells. Protein expression in liquid culture was induced by the addition of 0.5 mM isopropyl-β-D-thiogalactoside (IPTG, Sangon) when the cell density reached an optical density at 600 nm of 0.4−0.7. The cells were grown for 15 h at 16°C in LB medium containing 100 μg/mL kanamycin (Sangon). Then the cells were pelleted and lysed by sonication in PBS buffer with 2 mM dithiothreitol and 50 μg/mL phenylmethylsulfonyl fluoride. Cellular debris was separated by centrifugation and filtered with a 0.22 μm membrane. The recombinant A3A protein was purified using glutathione sepharose 4B (GE Healthcare) following the manufacturer's protocol. After digestion with HRV 3C protease (Sangon), the protein was further purified with a size-exclusion column (Millipore, Darmstadt, Germany) equilibrated with a storage solution containing 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 40% glycerol, 0.01 mM EDTA, 0.5 mM dithiothreitol, and 0.01% Tween-20, and stored at −80°C freezer before use. The purity A3A protein was confirmed by SDS-polyacrylamide gel electrophoresis (Figure S1).

Enzymatic digestion of DNA

A3A protein-treated or untreated DNA (in 25 μL of H_2O) was digested in the solution with

360 U of S1 nuclease, 4 U of DNase I, 0.002 U of venom phosphodiesterase I, 30 U of alkaline phosphatase, and 3 μL of enzymatic buffer (500 mM Tris-HCl, 100 mM NaCl, 10 mM $MgCl_2$, 10 mM $ZnSO_4$, pH 7.0). The mixture (30 μL) was incubated at 37℃ for 3 h. After adding 270 μL H_2O, the resulting solution was extracted with chloroform twice. The resulting aqueous layer was collected and lyophilized to dryness and then reconstituted in water followed by analysis with LC-MS/MS. For the analysis of 4mC modification, 1 μg of *D. radiodurans* genome DNA was digested to nucleosides.

Confirmation of 4mC modification in *D. radiodurans* DNA by tandem mass spectrometry analysis and high-resolution mass spectrometry analysis

To unambiguously determine the existence of 4mC or/and 5mC in *D. radiodurans* DNA, we first explored the MS fragmentation behavior of 4mC and 5mC standards. The results showed that both 4mC and 5mC produced the ion *m/z* 126 under collision energy of 10 V (Figure S10). With higher collision energy of 40 V, the fragment ions of 4mC and 5mC were similar, but 4mC produced a characteristic ion at *m/z* 95 (Figure S10). After digestion into nucleosides, *D. radiodurans* DNA was analyzed by LC-MS/MS and the similar fragment ions to 4mC standard were observed, including the characteristic ion (*m/z* 95) (Figure S10).

We also applied high-resolution mass spectrometry to evaluate the fragmentation patterns of 4mC and 5mC. The results verified that a characteristic ion at *m/z* 95.0238 was produced only by 4mC standard and *D. radiodurans* DNA (Figure S11). We proposed that 4mC lose a ribose to form the fragment ion of *m/z* 126.0662, which can further lose a methylamino group on *N*4 position to form the fragment ion of *m/z* 95.0240 (Figure S12).

3-Methylcytosine (3mC), a damaged base in genomic DNA, possesses a similar fragmentation pattern to that of 4mC, including the characteristic ion of *m/z* 95. We optimized the separation conditions and 3mC was clearly separated with 4mC and 5mC. The LC-MS/MS analysis showed that 3mC was absent in *D. radiodurans* DNA (Figure S13).

Analysis of 4mC in specific loci by 4mC-AMD-seq in *D. radiodurans* DNA

100 ng of D. *radiodurans* DNA was denatured and treated with A3A protein (6 μM) in a final concentration of 20 mM MES (pH 6.5), 0.1% Triton X-100, and 10% DMSO at 37°C for

2 h. The deamination reaction was terminated by incubating the solution at 95°C for 10 min. Selected loci were amplified by PCR using EpiMark® Hot Start Taq DNA polymerase (New England Biolabs). The PCR products were purified and subjected to Sanger sequencing (TsingKe).

Distribution of 4mC in gene body and upstream and downstream of genes

The 100 bp upstream and downstream regions of genes were divided into 30 bins, and the gene body was divided into 60 bins. The average methylation level of each bin was calculated as follows:

$$\text{Methylation level} = \frac{\sum_{i=1}^{n} N_{Ci}}{\sum_{i=1}^{n} (N_{Ci} + N_{Ti})} \times 100\%$$

N_{Ci} refers to the number of ‘‘C’’ bases that are counted as 4mC sites. N_{Ti} refers to the number of ‘‘T’’ bases that are counted as unmethylated C sites.

Analysis of the methylation level distribution of 4mC sites

The sequence of 12 bp upstream and downstream of each 4mC site was extracted to obtain a 25 bp length motif. The ratio of 4mC/C in each motif was calculated based on the number of unmethylated C and methylated C in the motif. The numbers of various methylation levels were calculated based on the value counts of Pandas. Finally, the percentage of methylated C proportion was calculated.

Table S1. The sequences of primers and PCR amplificated dsDNA.

DNA	Sequence (5′ to 3′)	
Primers		
Primers for DNA-C, DNA-4mC, DNA-5mC, and 4mC spike-in	Original & A3A deamination	Forward: GAGTGAGTGAGGGAGGAAG Reverse: CCACTCACAATTCCACACAACATAC
Primers for 5mC spike-in	Original	Forward: TTCTGACAACGATGGGAGGA Reverse: CAACTTTATCCCCCTCCATC
	A3A deamination	Forward: ATAATGATTGGAGGATTGAAGGAGT Reverse: CCACCTCCATCCAATCTATTAAT
Primers for C spike-in	Original	Forward: TGCTGATAAATCTGGAGCCG Reverse: GAACGAAAACTCACGTTAAGGG
	A3A deamination	Forward: AGTGGGGAGAGATGGTAAG Reverse: TCTTCACCTACATCCTTTTAAATT
215 bp DNA substrates for the method development		
DNA-C	GAGTGAGTGAGGGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCT CTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCC CGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTC ACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGT GTGGAATTGTGAGTGG	
DNA-4mC	GAGTGAGTGAGGGAGGAAG^{4m}CGGAAGAG^{4m}CG^{4m}C^{4m}C^{4m}CAATA^{4m}CG^{4m}CA AA^{4m}C^{4m}CG^{4m}C^{4m}CT^{4m}CT^{4m}C^{4m}C^{4m}C^{4m}CG^{4m}CG^{4m}CGTTGG^{4m}C^{4m}CGATT^{4m}CATTA ATG^{4m}CAG^{4m}CTGG^{4m}CA^{4m}CGA^{4m}CAGGTTT^{4m}C^{4m}C^{4m}CGA^{4m}CTGGAAAG^{4m}CG GG^{4m}CAGTGAG^{4m}CG^{4m}CAA^{4m}CG^{4m}CAATTAATGTGAGTTAG^{4m}CT^{4m}CA^{4m}CT ^{4m}CATTAGG^{4m}CA^{4m}C^{4m}C^{4m}C^{4m}CAGG^{4m}CTTTA^{4m}CA^{4m}CTTTATG^{4m}CTT^{4m}C^{4m}CGG ^{4m}CT^{4m}CGTATGTTGTGTGGAATTGTGAGTGG	
DNA-5mC	GAGTGAGTGAGGGAGGAAG^{5m}CGGAAGAG^{5m}CG^{5m}C^{5m}C^{5m}CAATA^{5m}CG^{5m}CA AA^{5m}C^{5m}CG^{5m}C^{5m}CT^{5m}CT^{5m}C^{5m}C^{5m}C^{5m}CG^{5m}CG^{5m}CGTTGG^{5m}C^{5m}CGATT^{5m}CATTA ATG^{5m}CAG^{5m}CTGG^{5m}CA^{5m}CGA^{5m}CAGGTTT^{5m}C^{5m}C^{5m}CGA^{5m}CTGGAAAG^{5m}CG GG^{5m}CAGTGAG^{5m}CG^{5m}CAA^{5m}CG^{5m}CAATTAATGTGAGTTAG^{5m}CT^{5m}CA^{5m}CT ^{5m}CATTAGG^{5m}CA^{5m}C^{5m}C^{5m}C^{5m}CAGG^{5m}CTTTA^{5m}CA^{5m}CTTTATG^{5m}CTT^{5m}C^{5m}CGG ^{5m}CT^{5m}CGTATGTTGTGTGGAATTGTGAGTGG	
Spike-in DNA for Illumina sequencing		
C spike-in	TGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAG CACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACG GGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAG GTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATAT ACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAG ATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTC	
4mC spike-in	GAGTGAGTGAGGGAGGAAG^{4m}CGGAAGAG^{4m}CG^{4m}C^{4m}C^{4m}CAATA^{4m}CG^{4m}CA AA^{4m}C^{4m}CG^{4m}C^{4m}CT^{4m}CT^{4m}C^{4m}C^{4m}C^{4m}CG^{4m}CG^{4m}CGTTGG^{4m}C^{4m}CGATT^{4m}CATTA ATG^{4m}CAG^{4m}CTGG^{4m}CA^{4m}CGA^{4m}CAGGTTT^{4m}C^{4m}C^{4m}CGA^{4m}CTGGAAAG^{4m}CG GG^{4m}CAGTGAG^{4m}CG^{4m}CAA^{4m}CG^{4m}CAATTAATGTGAGTTAG^{4m}CT^{4m}CA^{4m}CT ^{4m}CATTAGG^{4m}CA^{4m}C^{4m}C^{4m}C^{4m}CAGG^{4m}CTTTA^{4m}CA^{4m}CTTTATG^{4m}CTT^{4m}C^{4m}CGG ^{4m}CT^{4m}CGTATGTTGTGTGGAATTGTGAGTGG	
5mC spike-in	TTCTGACAACGATCGGAGGA^{5m}C^{5m}CGAAGGAG^{5m}CTAA^{5m}C^{5m}CG^{5m}CTTTTTT G^{5m}CA^{5m}CAA^{5m}CATGGGGGAT^{5m}CATGTAA^{5m}CT^{5m}CG^{5m}C^{5m}CTTGAT^{5m}CGTTGG GAA^{5m}C^{5m}CGGAG^{5m}CTGAATGAAG^{5m}C^{5m}CATA^{5m}C^{5m}CAAA^{5m}CGA^{5m}CGAG^{5m}C GTGA^{5m}CA^{5m}C^{5m}CA^{5m}CGATG^{5m}C^{5m}CTGTAG^{5m}CAATGG^{5m}CAA^{5m}CAA^{5m}CGTTG ^{5m}CG^{5m}CAAA^{5m}CTATTAA^{5m}CTGG^{5m}CGAA^{5m}CTA^{5m}CTTA^{5m}CT^{5m}CTAG^{5m}CTT^{5m}C ^{5m}C^{5m}CGG^{5m}CAA^{5m}CAATTAATAGA^{5m}CTGGATGGAGGCGGATAAAGTTG	

Table S2. The mass spectrometry parameters for the analysis of nucleosides.

Analytes	Precursor ion	Product ion	Q1 Pre/V	CE / V	Q3 Pre / V
dG	268.2	152.1	22	10	29
dA	252.2	136.1	20	15	20
dC	228.2	112.1	11	10	20
T	243.2	127.0	12	10	22
3mdC/4mdC/5mdC	242.2	126.1	20	11	24
4mdC	242.2	95.1	19	39	17
D_3-4mdC	245.2	129.1	20	11	24
5hmdC	258.1	142.1	11	9	23
5fdC	256.2	140.2	10	9	27
5cadC	272.1	156.1	20	11	30

Table S3. Conditions used to for steady-state kinetics study. Enzyme concentrations and incubation time were chosen with < 20% substrates being converted to products.

Substrate	[A3A] (nM)	Time (min)
DNA-C	2.4	5
DNA-5mC	2.4	5
DNA-4mC	240	30

Table S4. Oligonucleotides for library construction for the 4mC-AMD-seq.

Oligonucleotides	Sequence (5′ to 3′)
Adaptor-OH	TGAGAGAGAGGAGAATATAAATATGACATCGAT
Adaptor-P	*TCGATGTCATATTTATATTAGGAGGAGAGGA
pre-P5 primer	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAG AGAGAGGAGAATATAAATATGATATT
pre-P5 primer	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCC TCTCCTCCTAATATAAATATAACATCA
P5-universal primer	AATGATACGGCGACCACCGAGATCTACACTCTTTCC CTACACGACGCTCTTCCGATC**T
P7-index primer	CAAGCAGAAGACGGCATACGAGATXXXXXXXXGT GACTGGAGTTCAGACGTGTGCTCTTCCGATC**T

Note: * 5′-phosphorylation. ** phosphorothioate. In the sequence of P7-index primer, "XXXXXXXX" refers to index sequence which is consistent with NEBNext Index Primers for Illumina (Catlog E6611A, NEB).

Table S5. Statistical data from 4mC-AMD-seq for *D. radiodurans.*

Sample	Raw reads	Mapping rate (%)	Coverage (%)	Sequencing depth	Identified number of 4mC sites
Rep 1	28,540,290	63.1	99.2	604	2067
Rep 2	53,587,212	79.4	99.2	498	2148
Rep 3	49,126,578	74.6	99.4	747	2345

Table S6. Overall 4mC level in *D. radiodurans* DNA by 4mC-AMD-seq. FP, false positive; FN, false negative.

Sample	Spike-ins		*D. radiodurans*		
	False positive (FP)	False negative (FN)	Total N4-unmethylated C reads number	Total N4-methylated C reads number	Genomic methylation reads level [a]
Rep 1	0.19%	3.6%	921,946,965	5,894,438	0.33%
Rep 2	0.09%	3.1%	630,321,544	3,542,119	0.41%
Rep 3	0.19%	2.7%	1,684,156,094	5,657,228	0.18%

[a] The genomic methylation reads level was calculated as follows:

$$\text{Level (\%)} = \frac{\text{Number of total methylated reads} - (\text{Number of total unmethylated reads} \times \text{FP})}{\text{Number of total unmethylated reads}} \times 100\%$$

Table S7. Primers for locus-specific amplification of regions that contain C(4mC)GCGG motif detected by 4mC-AMD-seq in *D. radiodurans*.

Position of C4mCGCGG	Assay	Sequence (5′ to 3′)
Chr1: 291 642 (-)	4mC-AMD-seq	Fw TTATGATGTATTTGGTGAGGATGGG Rv AACCCTCAAACCACTTTACACTCAC
	Original sequence	Fw ACCCTGTTCCAGCCCCACC Rv CGAGCCGGACACGCTCAAT
Chr1: 1 912 062 (-)	4mC-AMD-seq	Fw GGGGTGTTGGGTGATTTTG Rv CTCATCATTCACCAAAACCACA
	Original sequence	Fw GGTGAGGGGTCTTTTTTACGGC Rv CAGTTTCCCATCGAGAACGAGG
Chr2: 336 034 (+)	4mC-AMD-seq	Fw GGGTGAGTAGTTATGGTATGGTG Rv CTAAATCTACCCCACCTCCATA
	Original sequence	Fw CTGCAGTTCGCCAACTTCAGG Rv CAGCGACATCATGTTGCTCAGC
Chr2: 336 037 (-)	4mC-AMD-seq	Fw TGAGTGTTGAAAGTTGGGTGT Rv CACTCTACTCAACTCCAAAACCT
	Original sequence	Fw CTGCAGTTCGCCAACTTCAGG Rv CAGCGACATCATGTTGCTCAGC
Plasmid MP1: 95 556 (-) Plasmid MP1: 95 634 (-)	4mC-AMD-seq	Fw AGTTTGGAAATGTAGGGTTTGAGTG Rv CACATCAAACATCATCTTATCACCC
	Original sequence	Fw CATCTGGTCGGCAACTATGTCA Rv CGTCCTGTGTCCACTCGTCTTT
Plasmid SP1: 8995 (+)	4mC-AMD-seq	Fw TTTGGTTGAGGTGGATTGATG Rv CACCCTCAATCATCACAACAACT
	Original sequence	Fw GGCAGTTCTGGAACACCCAG Rv CGACGAGCAGGACTTCATGC

Figure S1. SDS-PAGE analysis of the purified A3A protein.

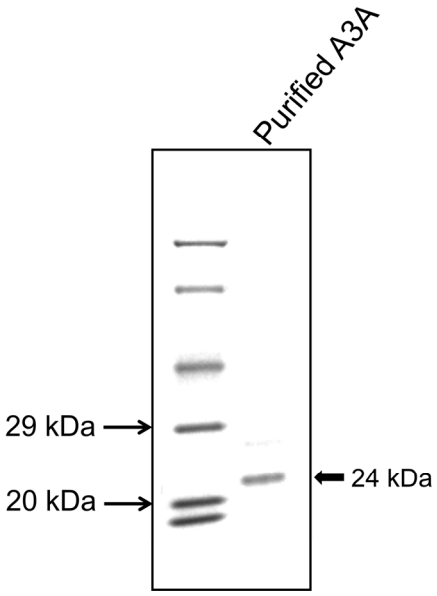


Figure S2. Preparation of 215-bp dsDNA substrates (DNA-C, DNA-4mC and DNA-5mC). (A) Sequence of the 215 bp DNA. (B) 0.5 ng of pUC19 vector DNA (Takara) was used as template for PCR amplification with using dATP/dGTP/dTTP and dCTP (or 5mdCTP, 4mdCTP). (C) Agarose gel electrophoresis analysis of PCR products.

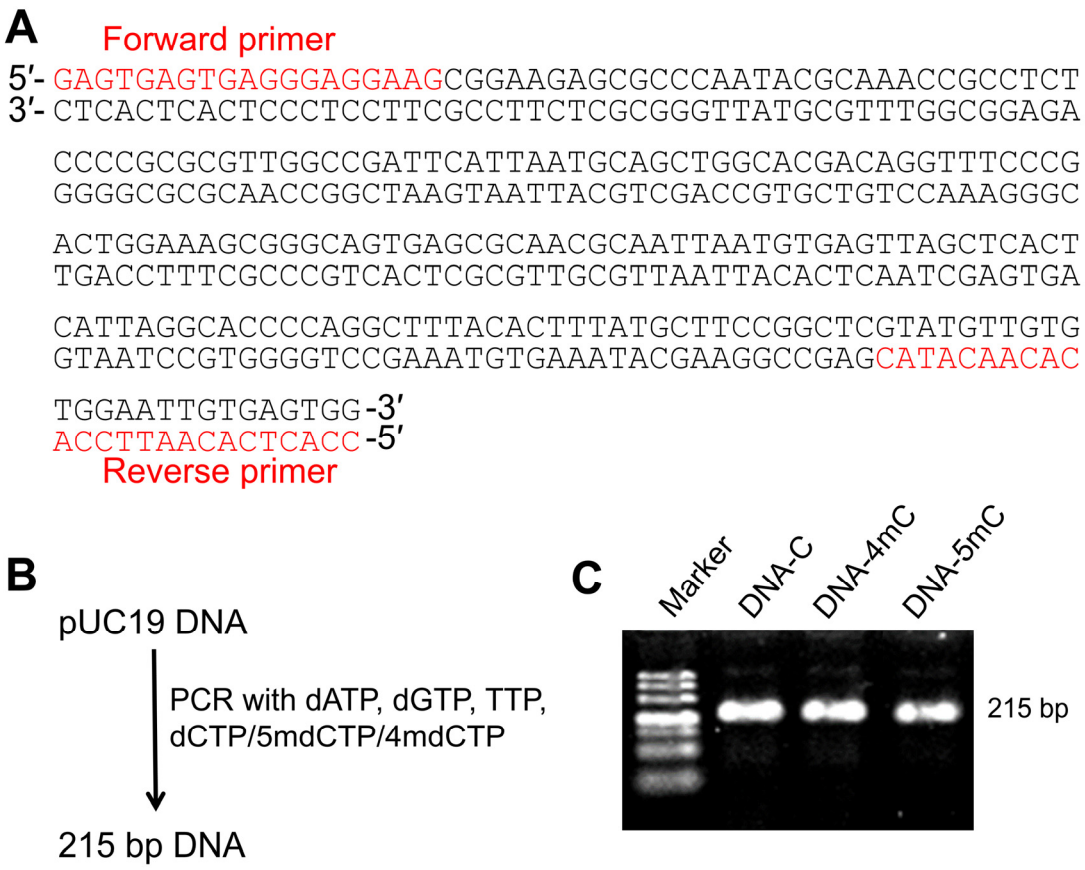


Figure S3. Colony sequencing using DNA-C without A3A treatment. Forty clones were picked for sequencing. All the 2120 cytosine sites (53 cytosines/strand × 40 clones = 2120 cytosines) in DNA-C were read as cytosine. Red, read as C.

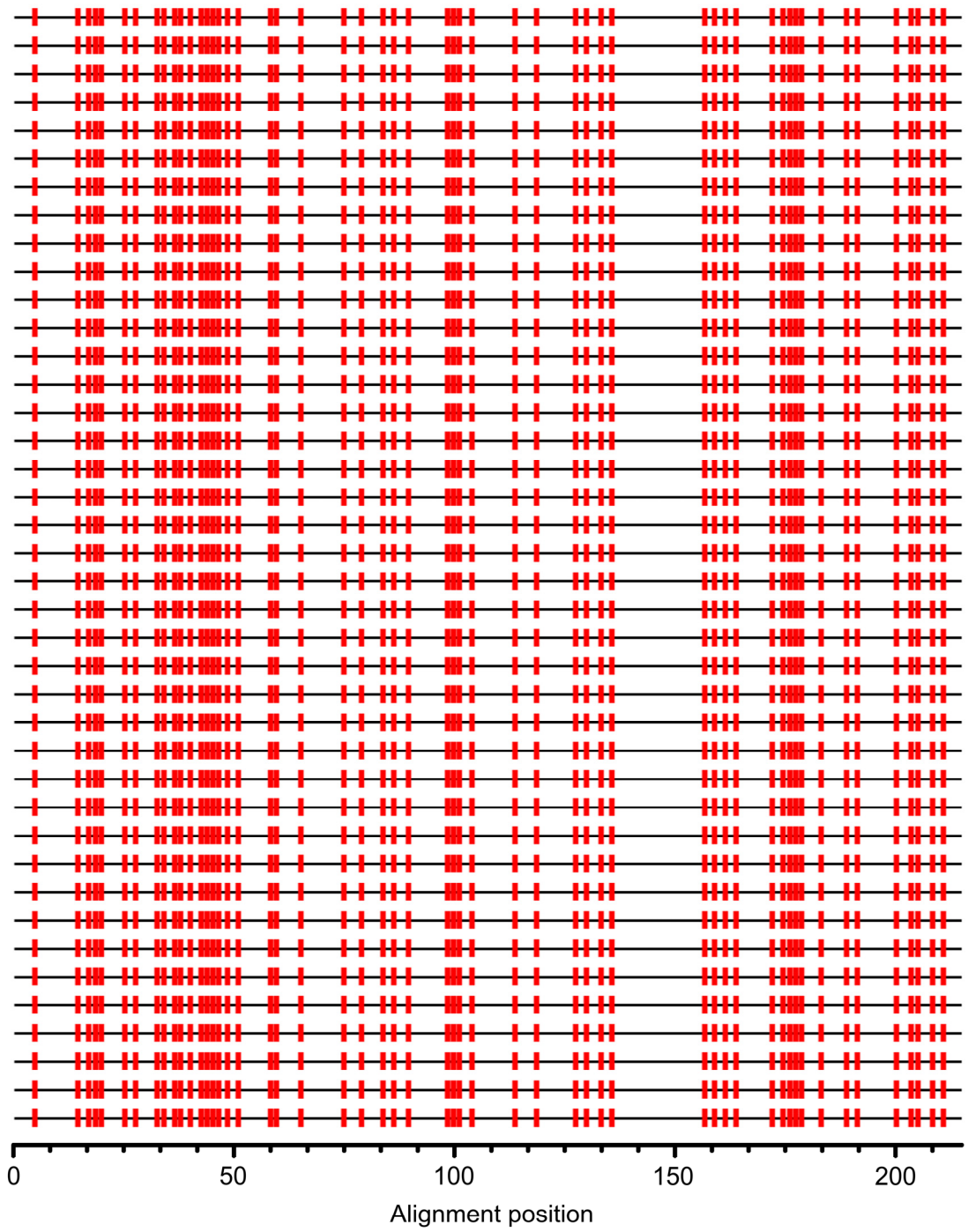


Figure S4. Colony sequencing using DNA-C with A3A treatment. Forty clones were picked for sequencing. All the 2120 cytosine sites (53 cytosines/strand × 40 clones = 2120 cytosines) in DNA-C were read as thymine, indicating that 100% cytosine sites were deaminated by A3A protein. Blue, read as T.

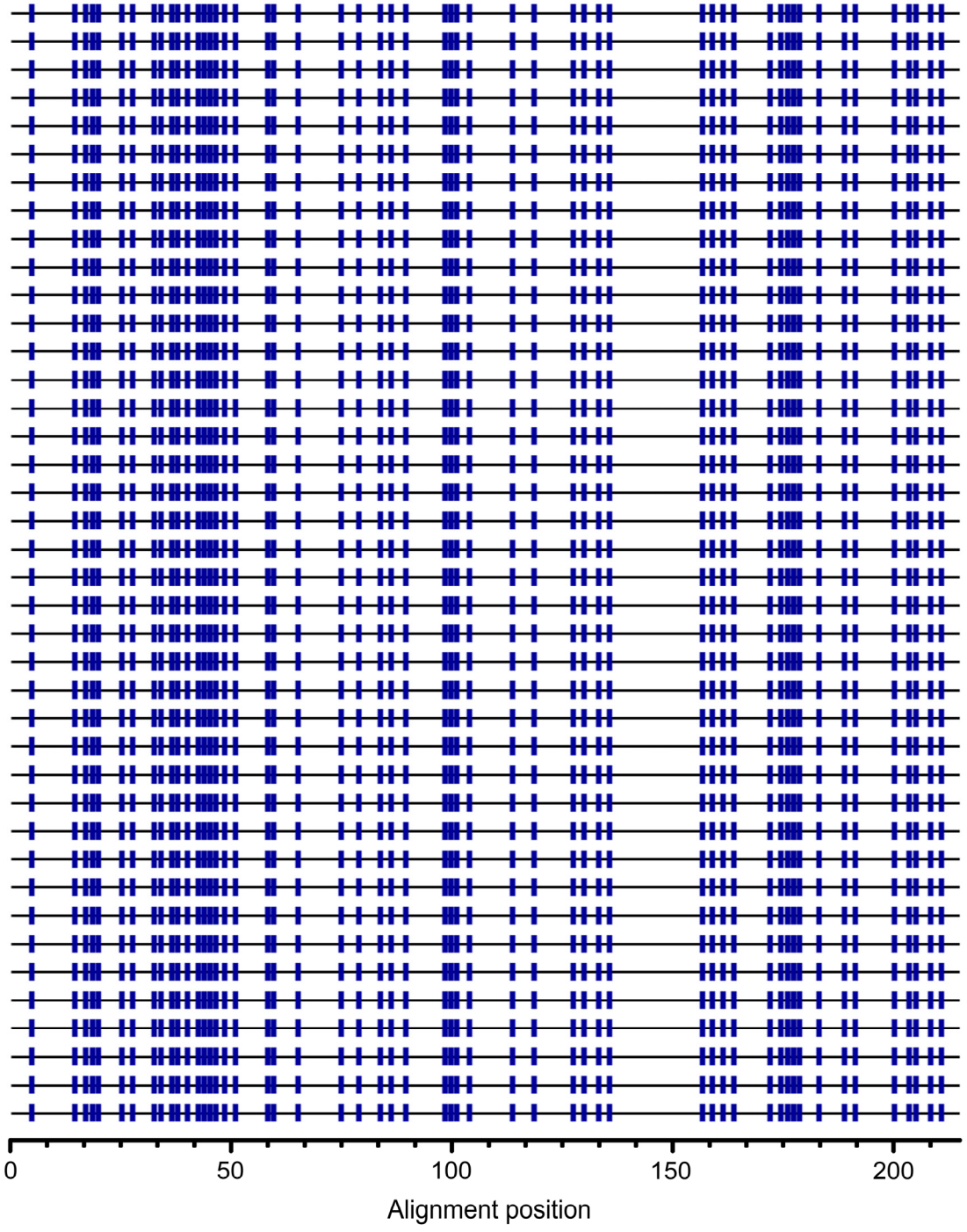


Figure S5. Colony sequencing using DNA-5mC without A3A treatment. Forty clones were picked for sequencing. 2119 5mC sites in DNA-5mC were read as cytosine. Red, read as C; blue, read as T.

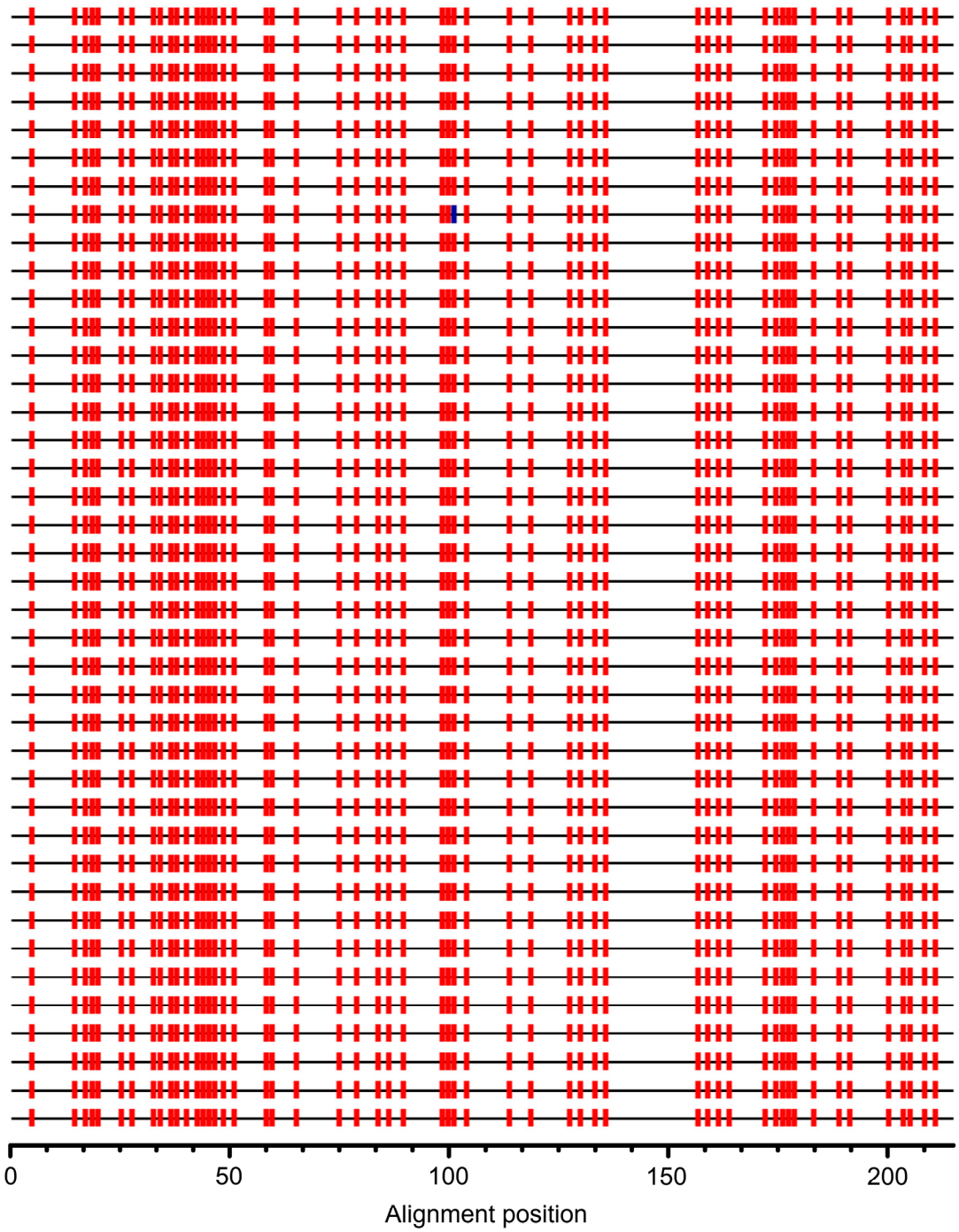


Figure S6. Colony sequencing using DNA-5mC with A3A treatment. Forty clones were picked for sequencing. Only 26 5mC sites in DNA-5mC (53 cytosines/strand × 40 clones = 2120 cytosines) were read as cytosine after A3A treatment, indicating that only 1.2% (26/2120 = 1.2%) residual 5mC sites were left after A3A treatment. Red, read as C; blue, read as T.

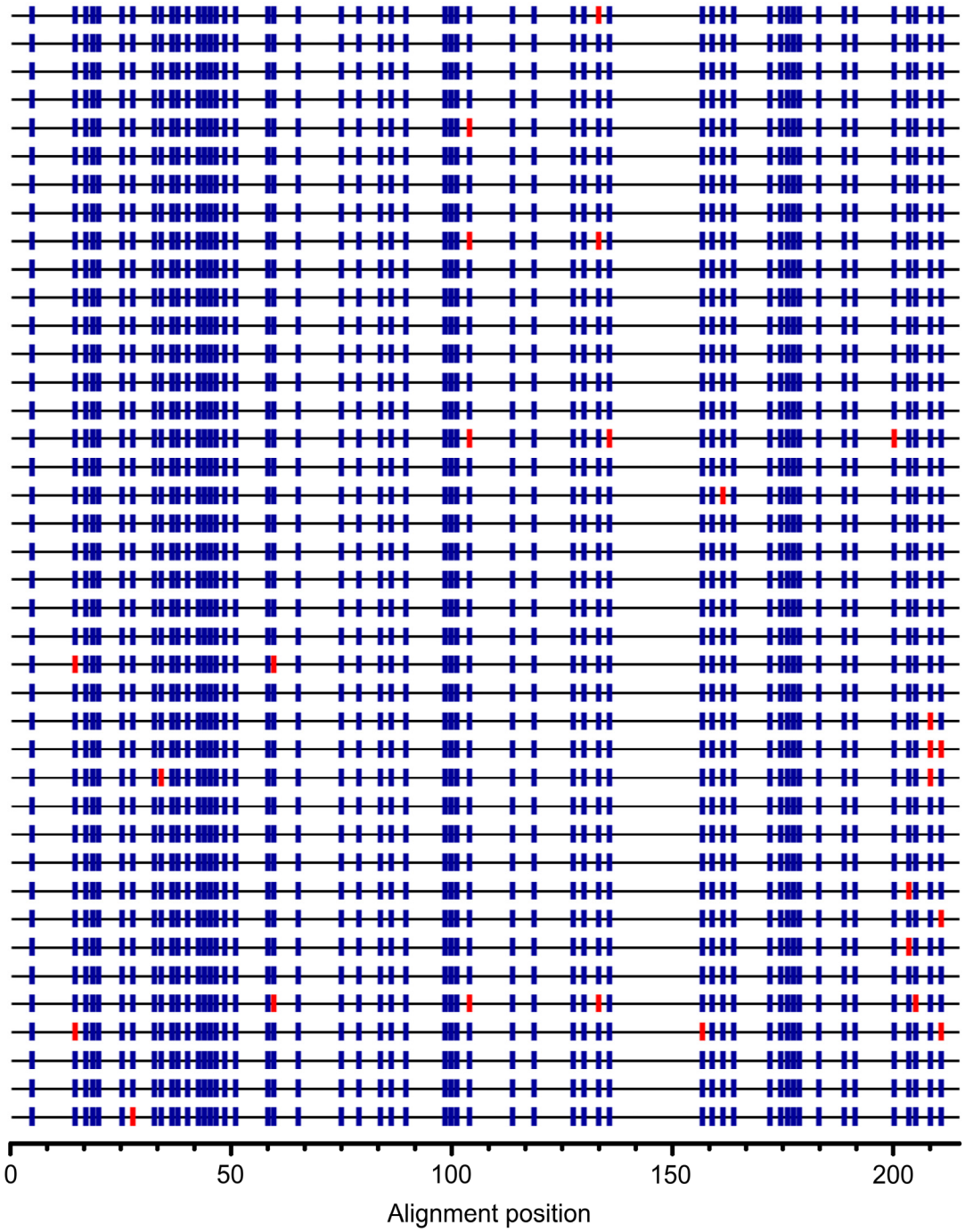


Figure S7. Colony sequencing using DNA-4mC without A3A treatment. Forty clones were picked for sequencing. All the 2120 4mC sites (53 cytosines/strand × 40 clones = 2120 cytosines) in DNA-4mC were read as cytosine. Red, read as C.

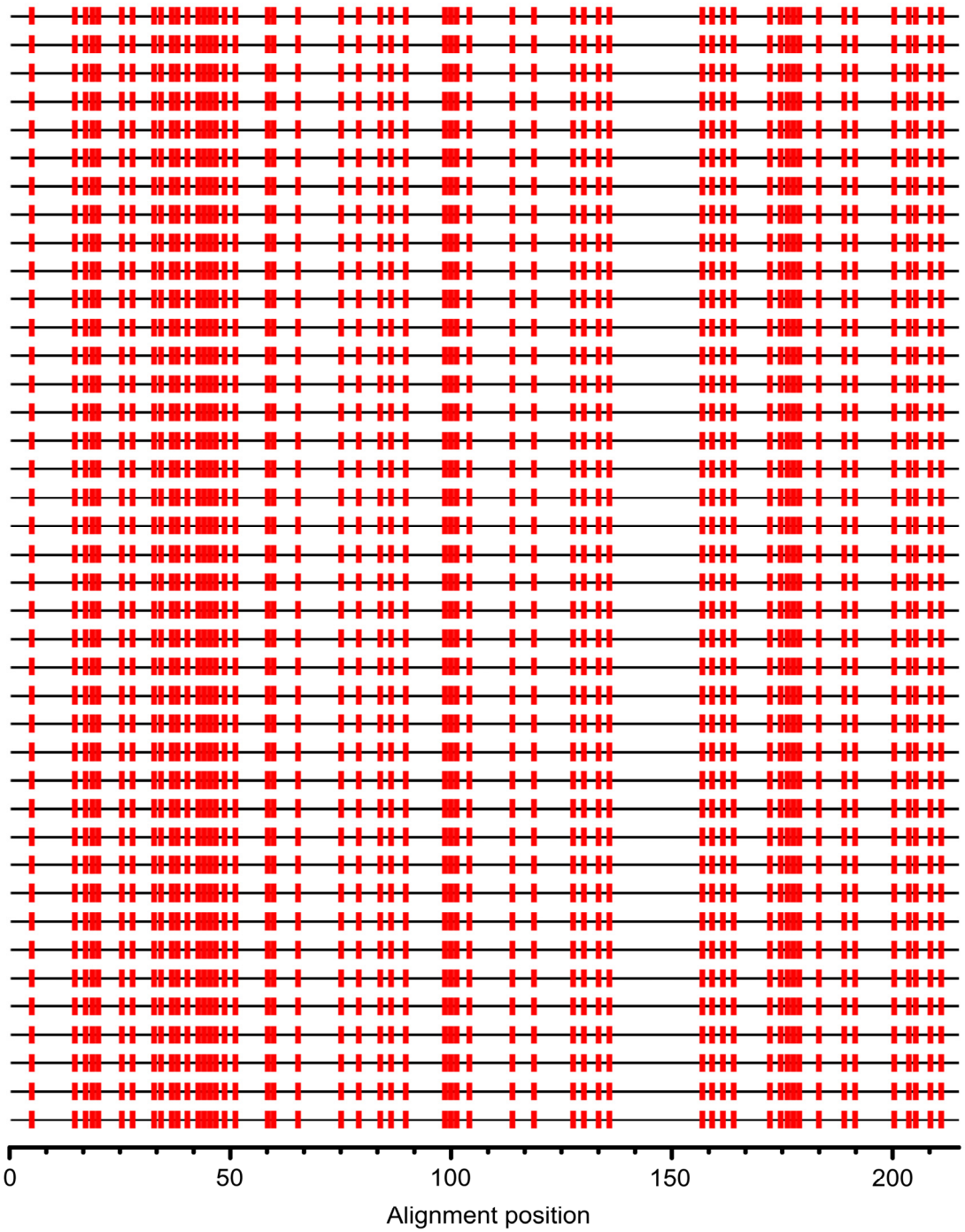


Figure S8. Colony sequencing using DNA-4mC with A3A treatment. Forty clones were picked for sequencing. 2056 4mC sites (53 cytosines/strand × 40 clones = 2120 cytosines) in DNA-4mC were read as cytosine, indicating that 97.0% (2056/2120 = 97.0%) 4mC sites were kept after by A3A treatment. Red, read as C; blue, read as T.

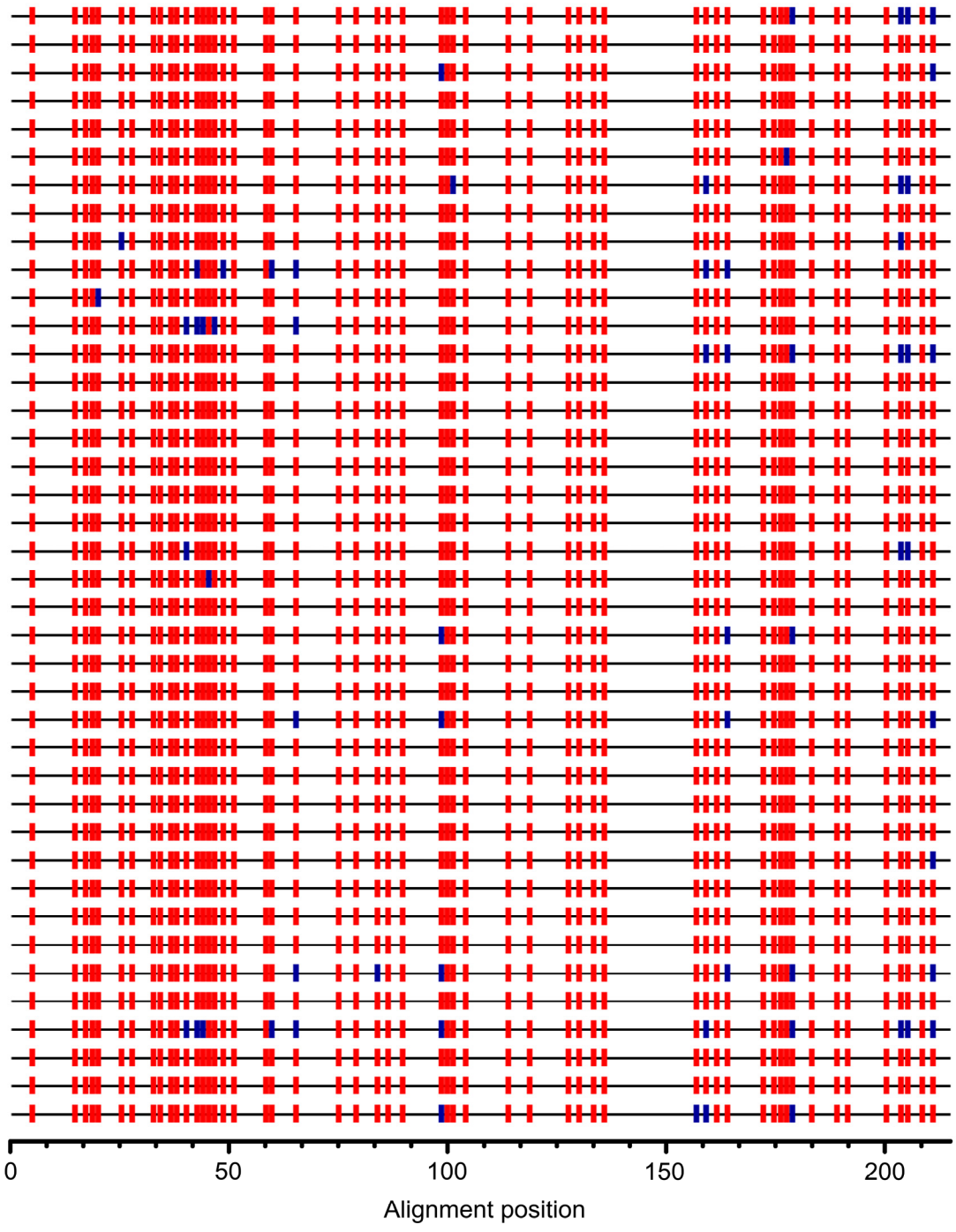
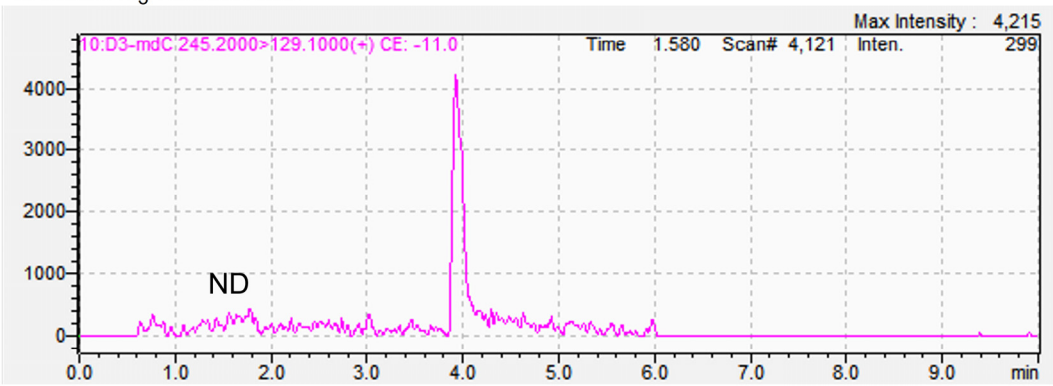
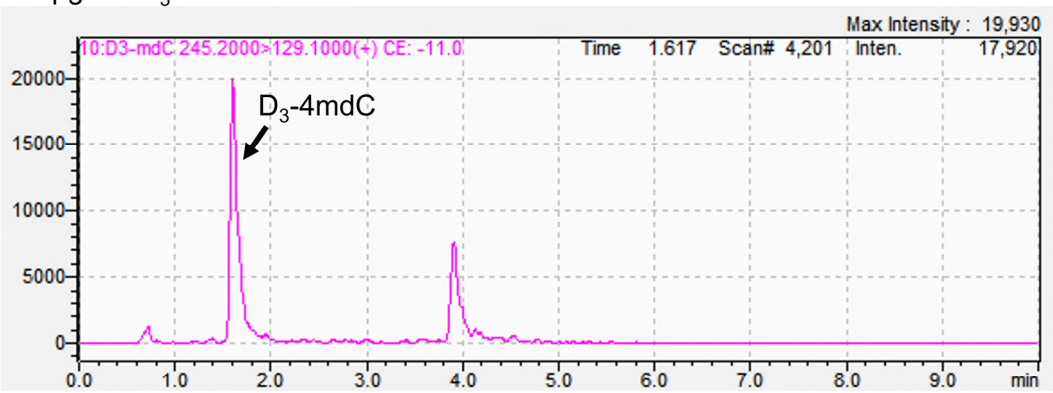


Figure S9. Raw data for detection of D_3-4mdC from *D. radiodurans* DNA with LC-MS/MS analysis. (A-C) Different concentrations of D_3-methionine was added in the cell culture medium. The MRM detection mode (mass transition of *m/z* 245.2→129.1) was used for the detection of D_3-4mdC.

A Without D_3-methionine



B 50 μg/mL D_3-methionine



C 500 μg/mL D_3-methionine

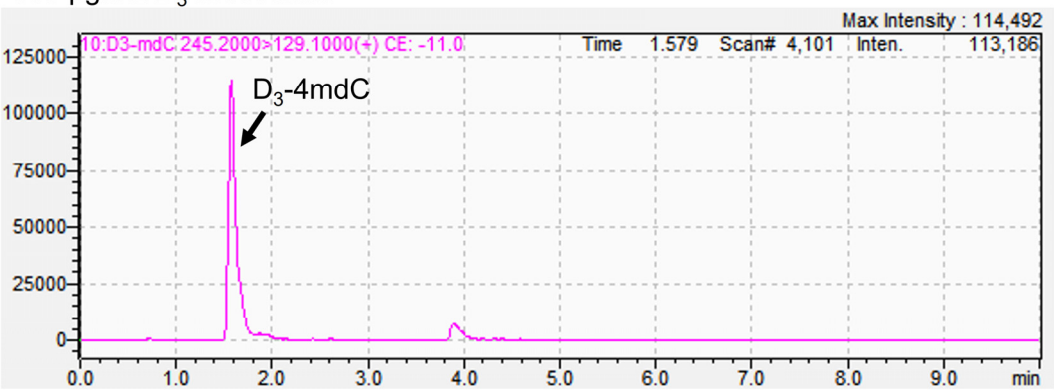


Figure S10. Determination of 4mC in *D. radiodurans* DNA by LC-MS/MS. (A, B) Product-ion spectra of 4mdC standard under collision energy of 10 V and 40 V. (C, D) Extracted-ion chromatograms of 4mdC standard with using mass transition of *m/z* 242.2 → 126.1 or *m/z* 242.2 → 95.1. (E, F) Product-ion spectra of 5mdC standard under collision energy of 10 V and 40 V. (G, H) Extracted-ion chromatograms of 5mdC standard with using mass transition of *m/z* 242.2 → 126.1 or *m/z* 242.2 → 95.1. (I, J) Product-ion spectra of 4mdC/5mdC from *D. radiodurans* DNA under collision energy of 10 V and 40 V. (K, L) Extracted-ion chromatograms of 4mdC/5mdC from *D. radiodurans* DNA with using mass transition of *m/z* 242.2 → 126.1 or *m/z* 242.2 → 95.1.

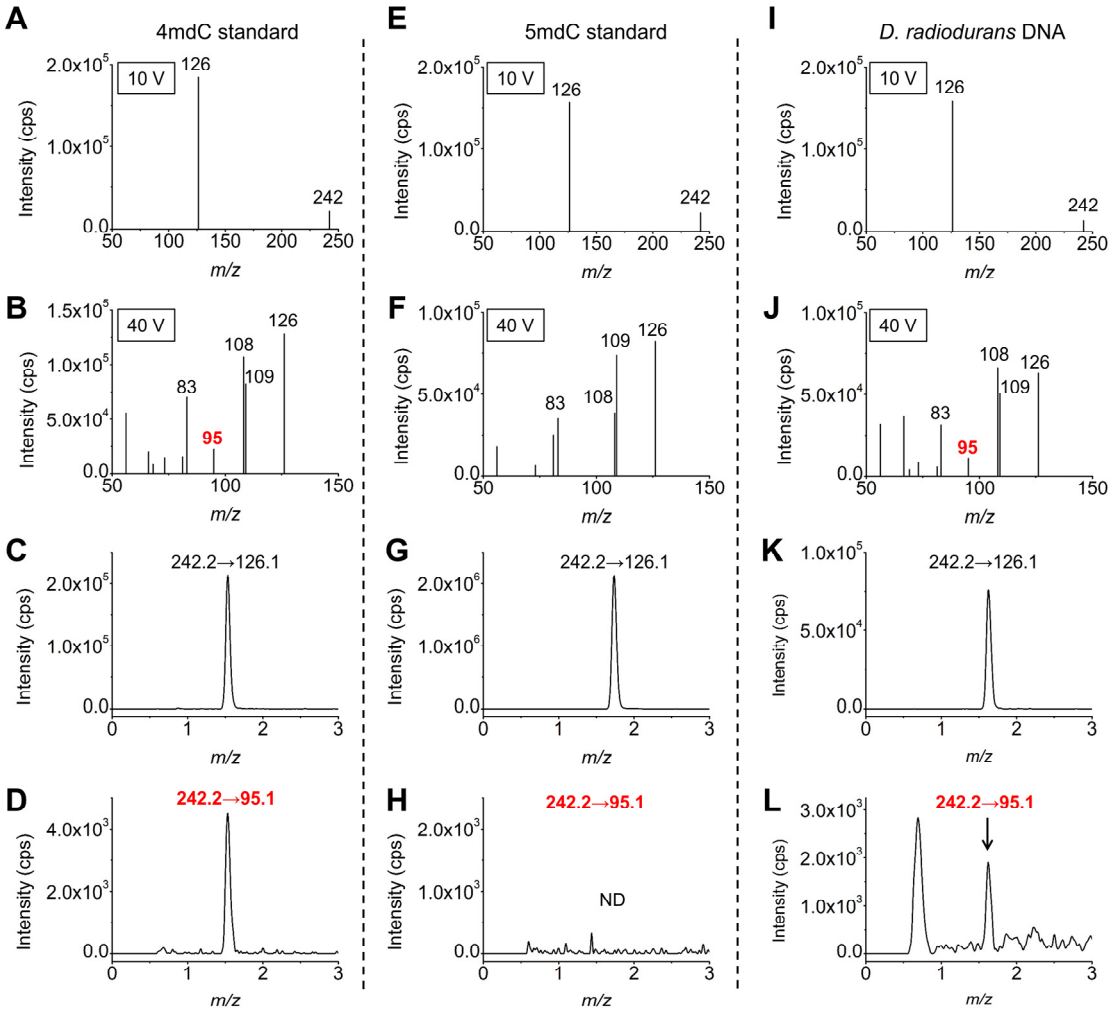


Figure S11. Confirmation of 4mC modification in *D. radiodurans* DNA by high-resolution mass spectrometry. (A) Extracted-ion chromatogram (*m/z* 242.1135) of 4mdC standard. (B–D) MS^1, MS^2 and MS^3 spectra of 4mdC standard. (E) Extracted-ion chromatogram (*m/z* 242.1135) of 5mdC standard. (F–H) MS^1, MS^2 and MS^3 spectra of 5mdC standard. (I) Extracted-ion chromatogram (*m/z* 242.1135) of digested nucleosides from *D. radiodurans* DNA. (J–L) MS^1, MS^2 and MS^3 spectra of digested nucleosides from *D. radiodurans* DNA. The fragment ion of *m/z* 95.0238 is unique in 4mdC standard and in digested nucleosides from *D. radiodurans* DNA, indicating the presence of 4mdC in *D. radiodurans* DNA.

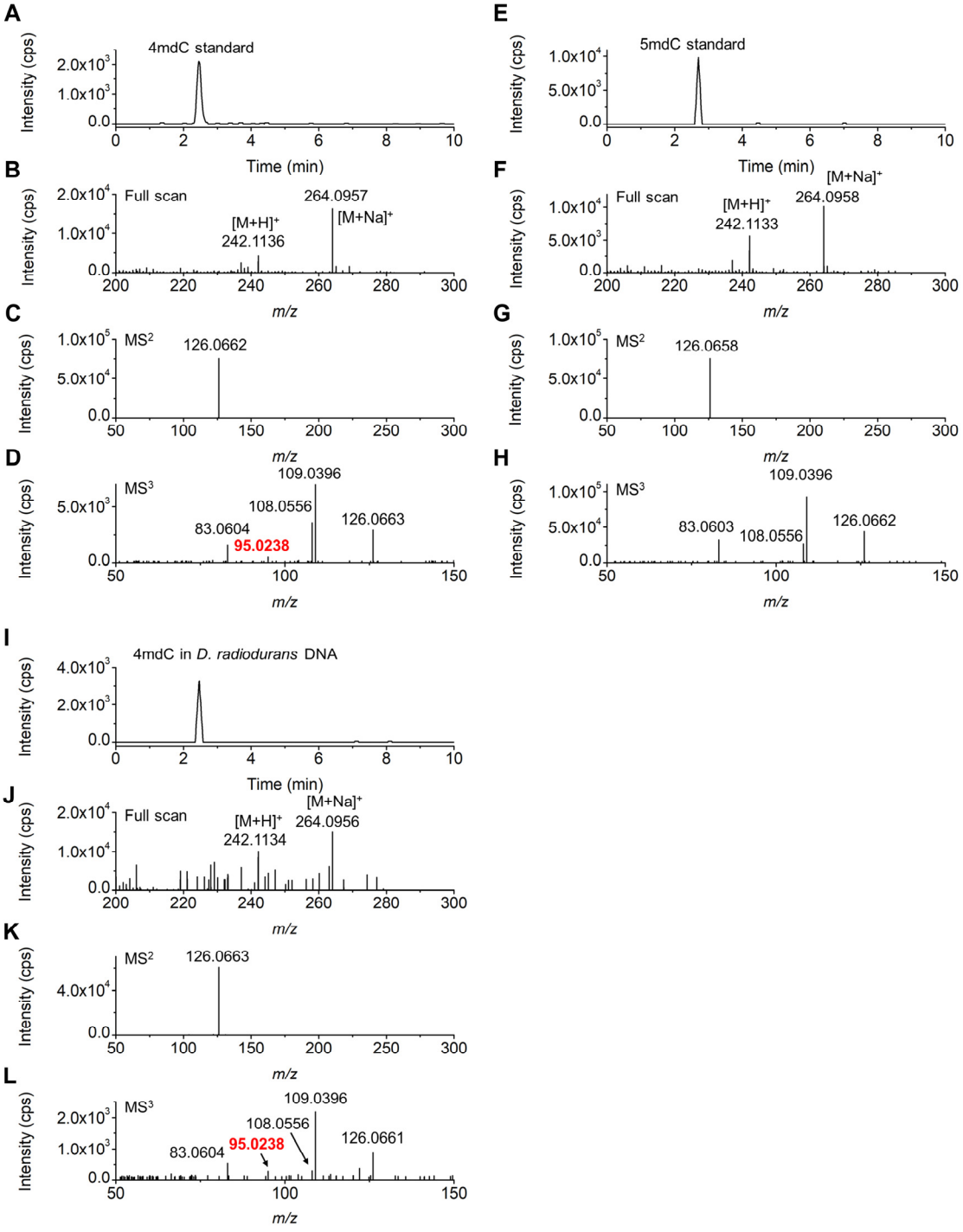


Figure S12. The proposed fragmentation pathways of (A) 4mdC and (B) 5mdC.

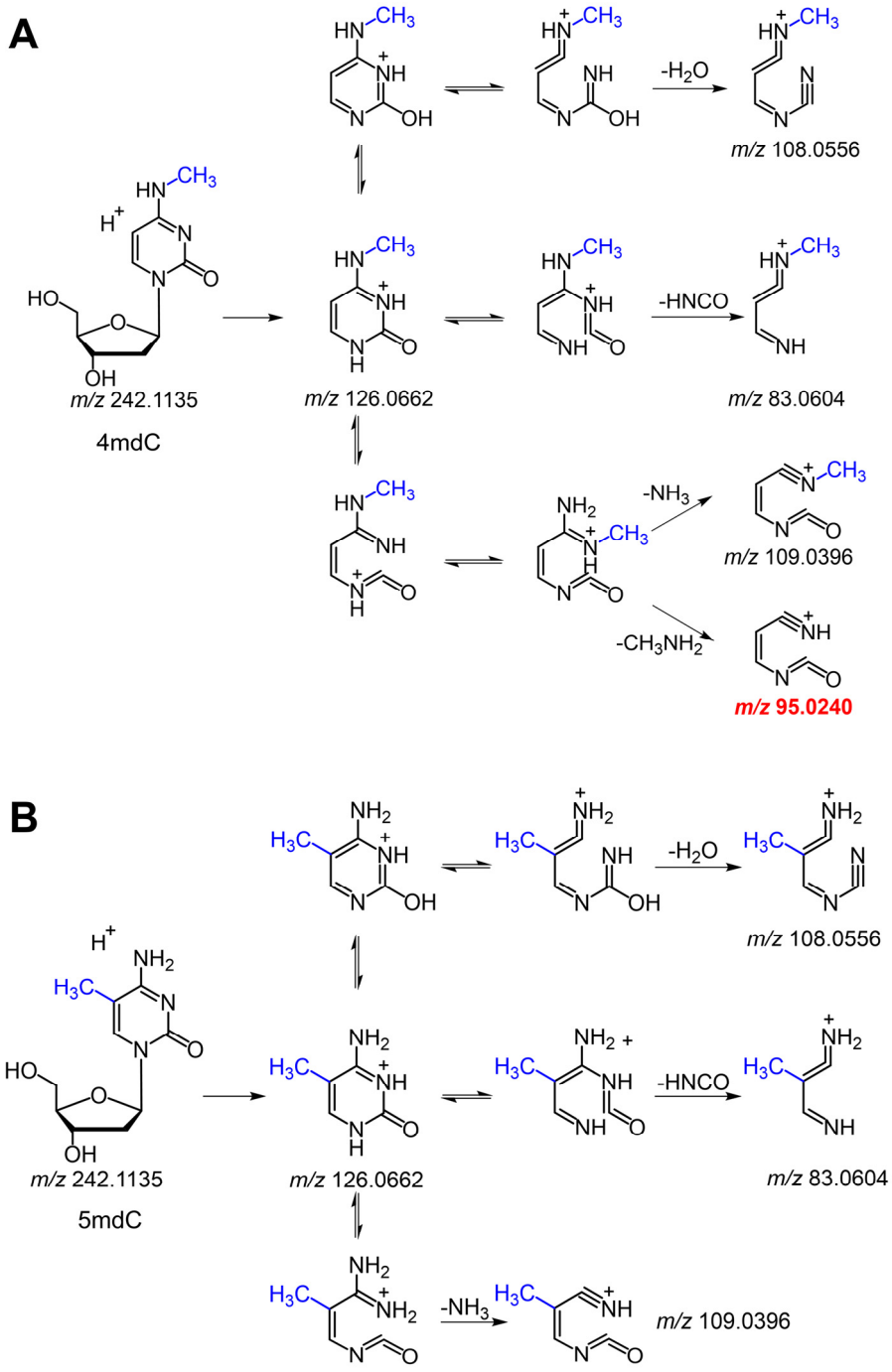


Figure S13. Determination of DNA cytosine methylation in *D. radiodurans* DNA by LC-MS/MS. The extracted-ion chromatograms of 3mC, 4mC and 5mC standards and DNA from *E. coli* and *D. radiodurans*. The LC separation was performed using mobile phases of water (solvent A) and methanol (solvent B). A gradient of 0–1.5 min 5% B, 1.5–3 min 5–40% B, 3–6.5 min 40% B, 6.5–7 min 40–5% B, and 7–12 min 5% B was used. Under the optimized conditions, 3mC, 4mC and 5mC were well separated. The results showed that 5mC existed in *E. coli* DNA, but was absent in *D. radiodurans* DNA. However, 4mC existed in *D. radiodurans* DNA, but was absent in *E. coli* DNA. The mobile phases and gradient used in this LC separation were different from those used in the main text (Figures 2 and 5), therefore leading to the different elution and separation resolution of these modified cytosines.

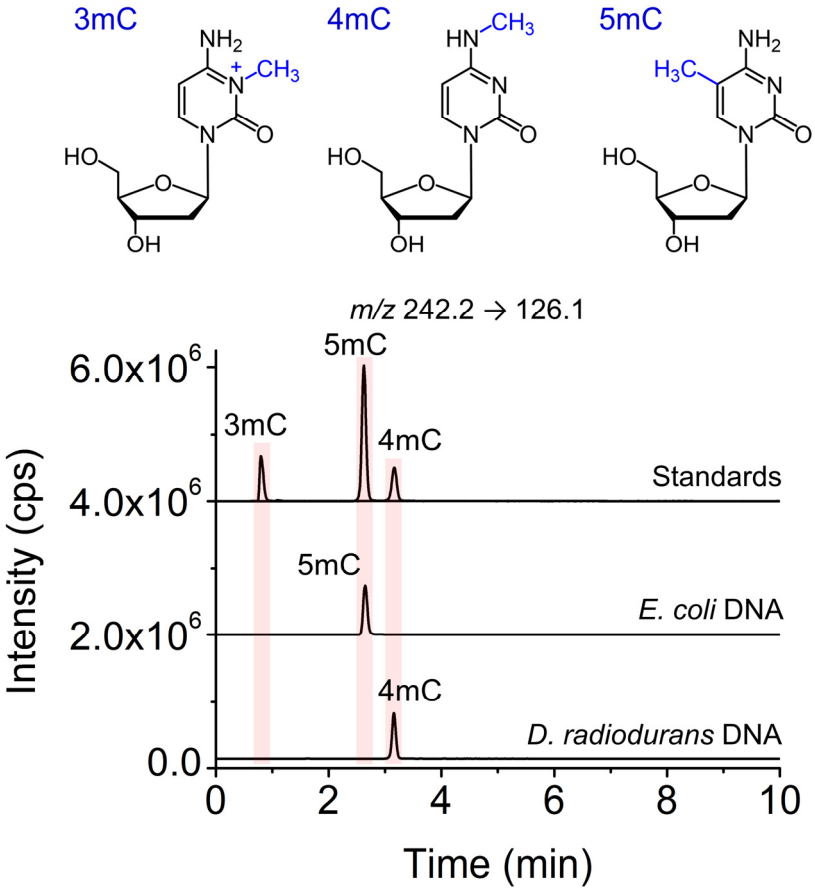


Figure S14. Determination of 5hmC (A), 5fC (B), and 5caC (C) in *D. radiodurans* DNA by LC-MS/MS. The mobile phases and gradient used in this LC separation were same as those used in the main text (Figure 2 and 4)

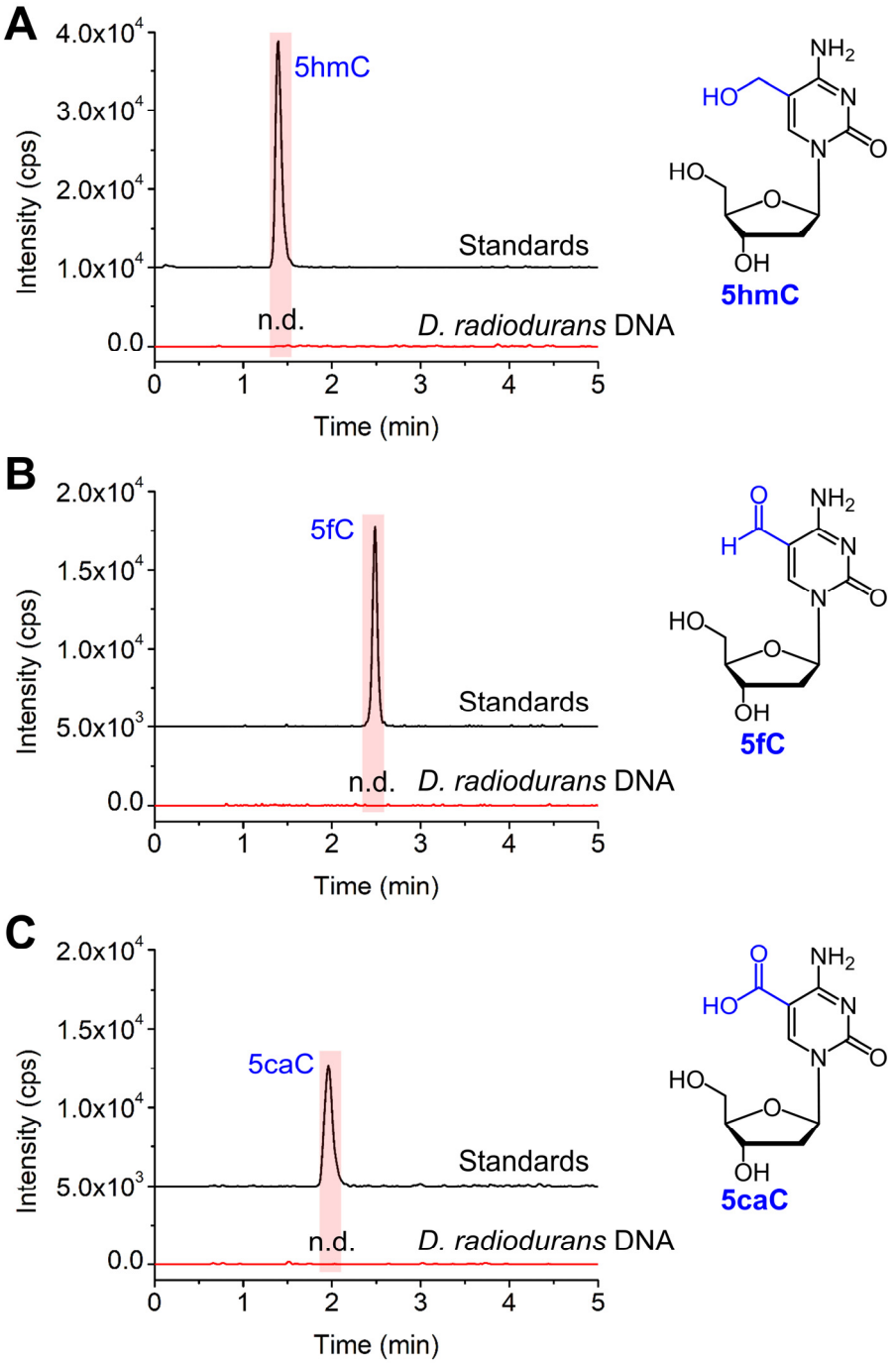


Figure S15. Growth curve of *D. radiodurans*.

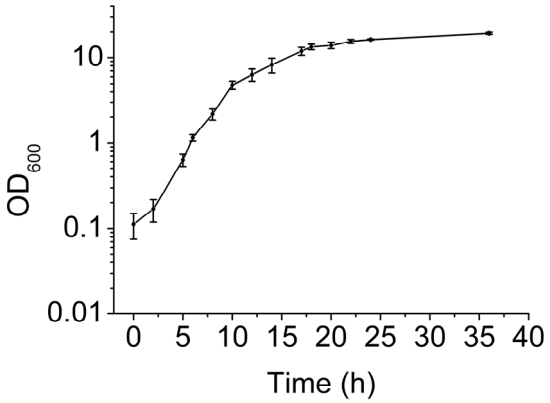


Figure S16. Schematic illustration of the library construction for 4mC-AMD-seq.

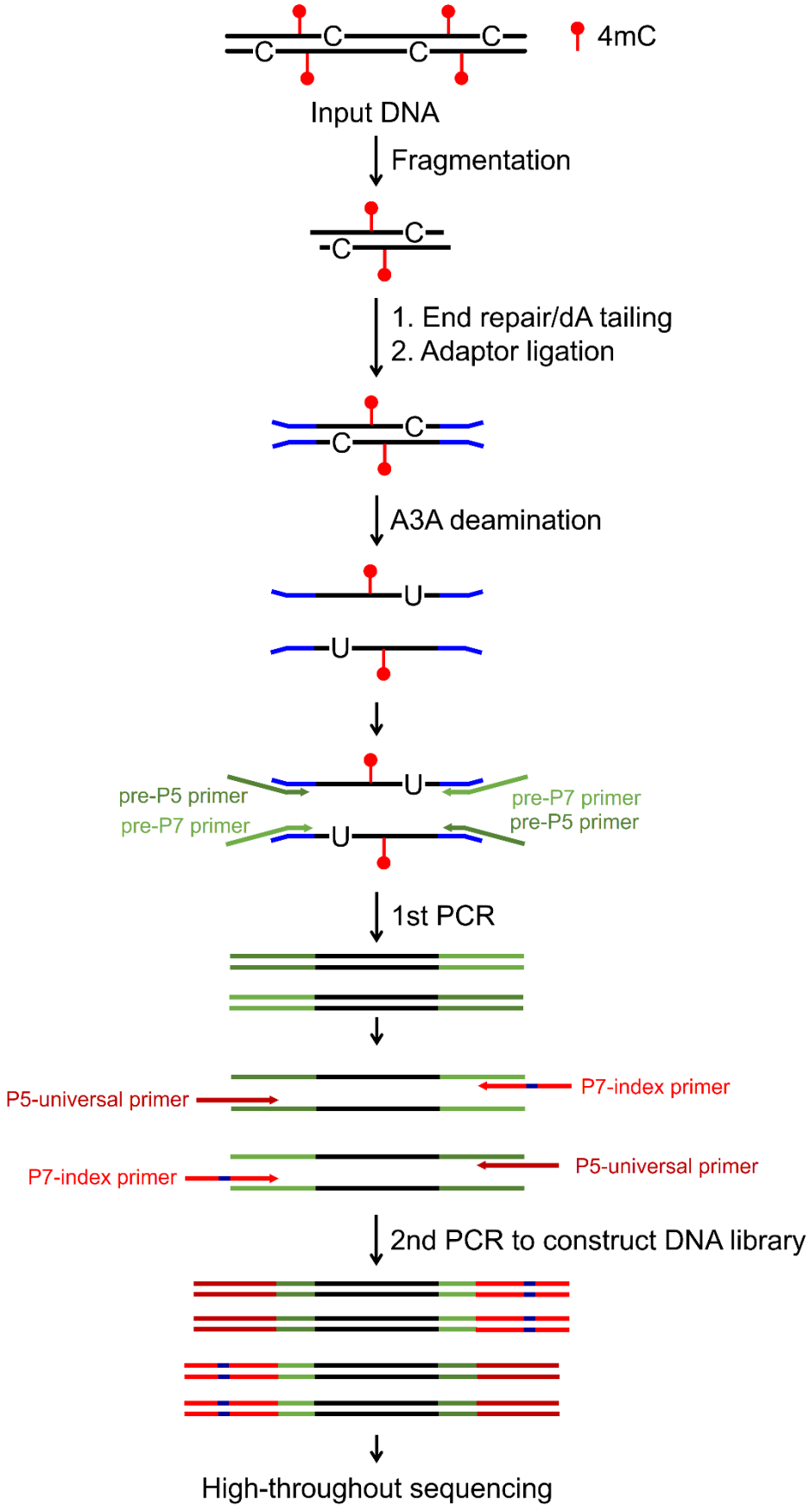


Figure S17. Correlation of methylation level of the identified 4mC sites in different replicates (Rep 1, Rep 2, Rep 3). Red, 4mC in C(4mC)GCGG motif; blue, 4mC in non-CCGCGG context. Pearson correlation coefficient (r) was calculated by SPSS version25.

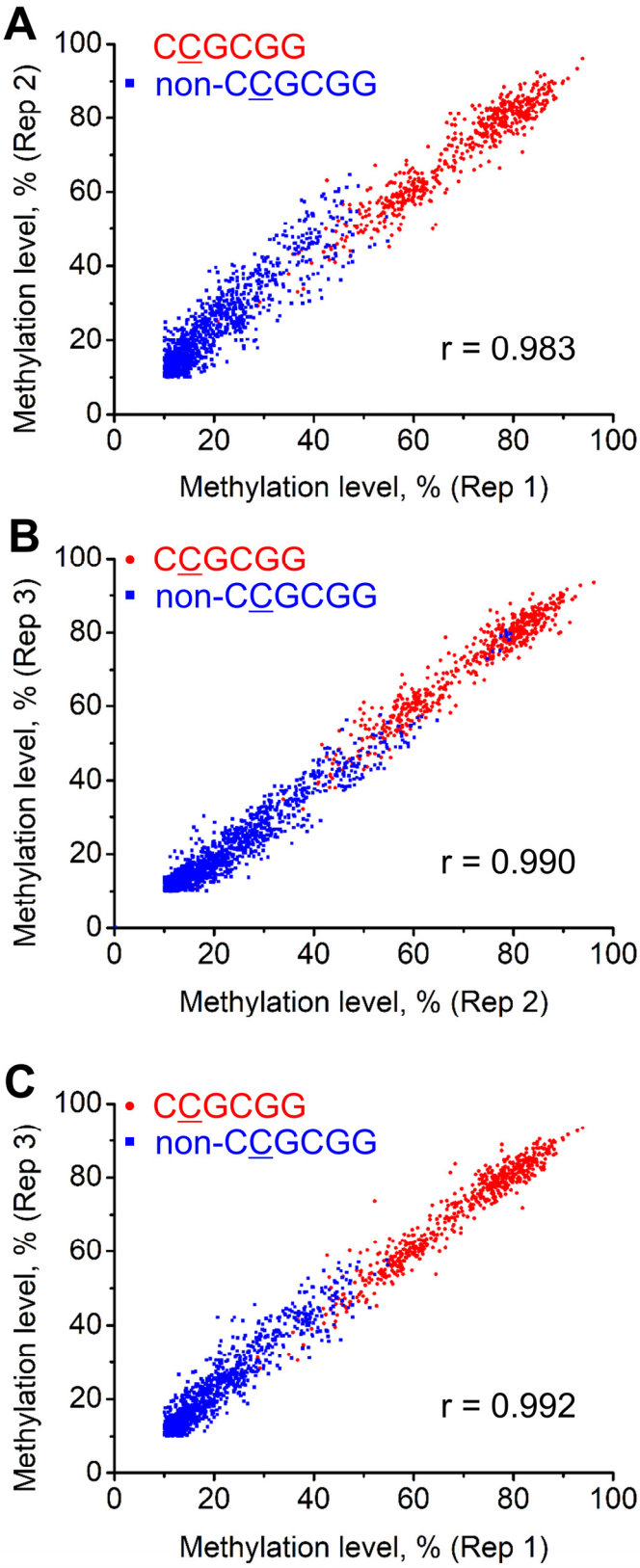


Figure S18. Distribution of 4mC sites in the different regions of genes.

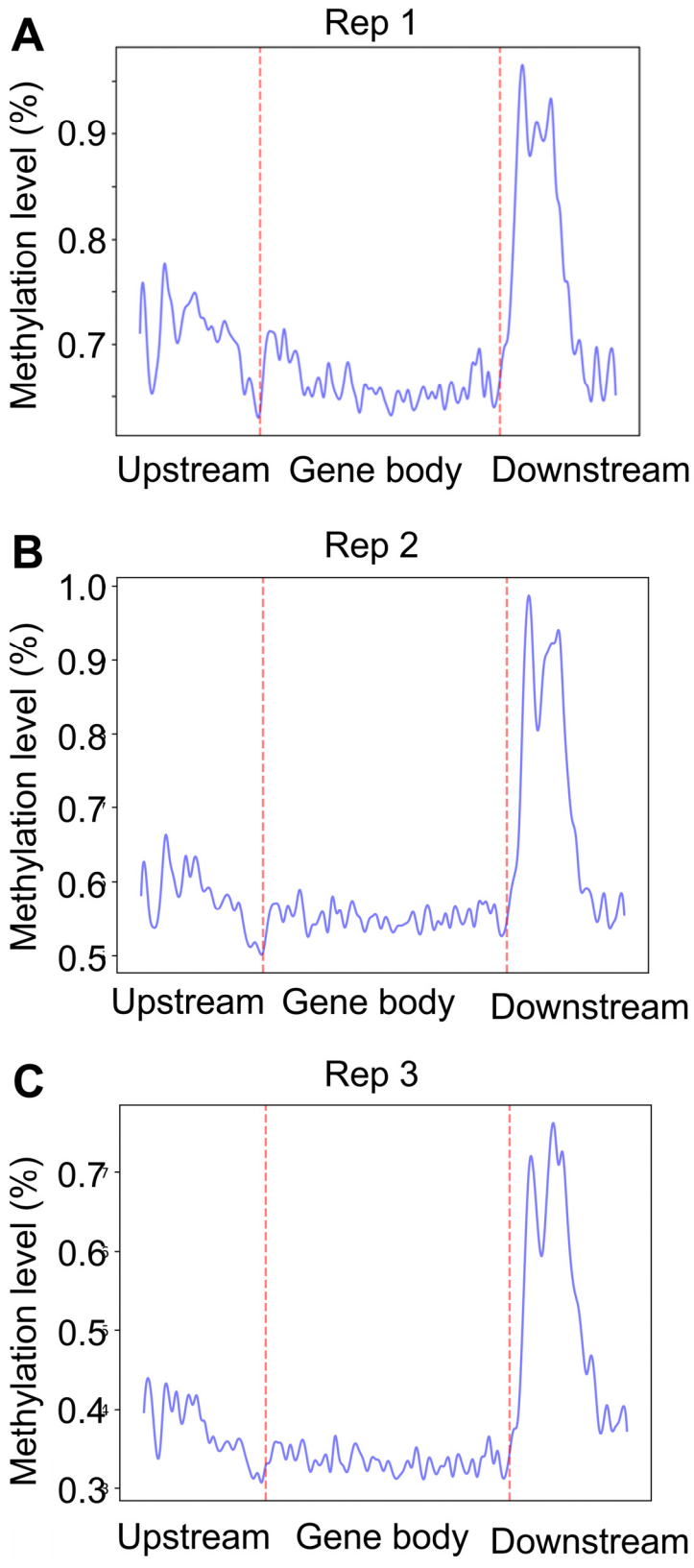


Figure S19. Distribution of the methylation level of 4mC site in C(4mC)GCGG motif in different replicates (Rep1, Rep 2, and Rep 3).

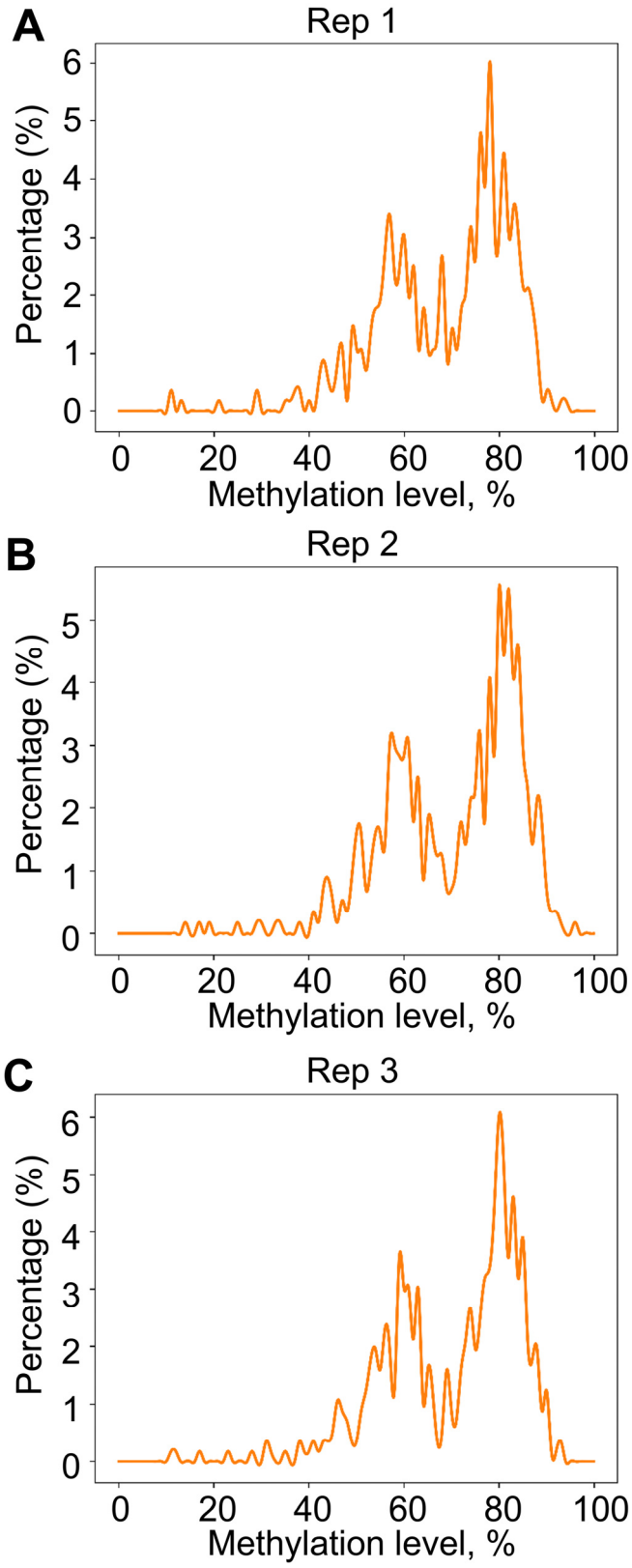
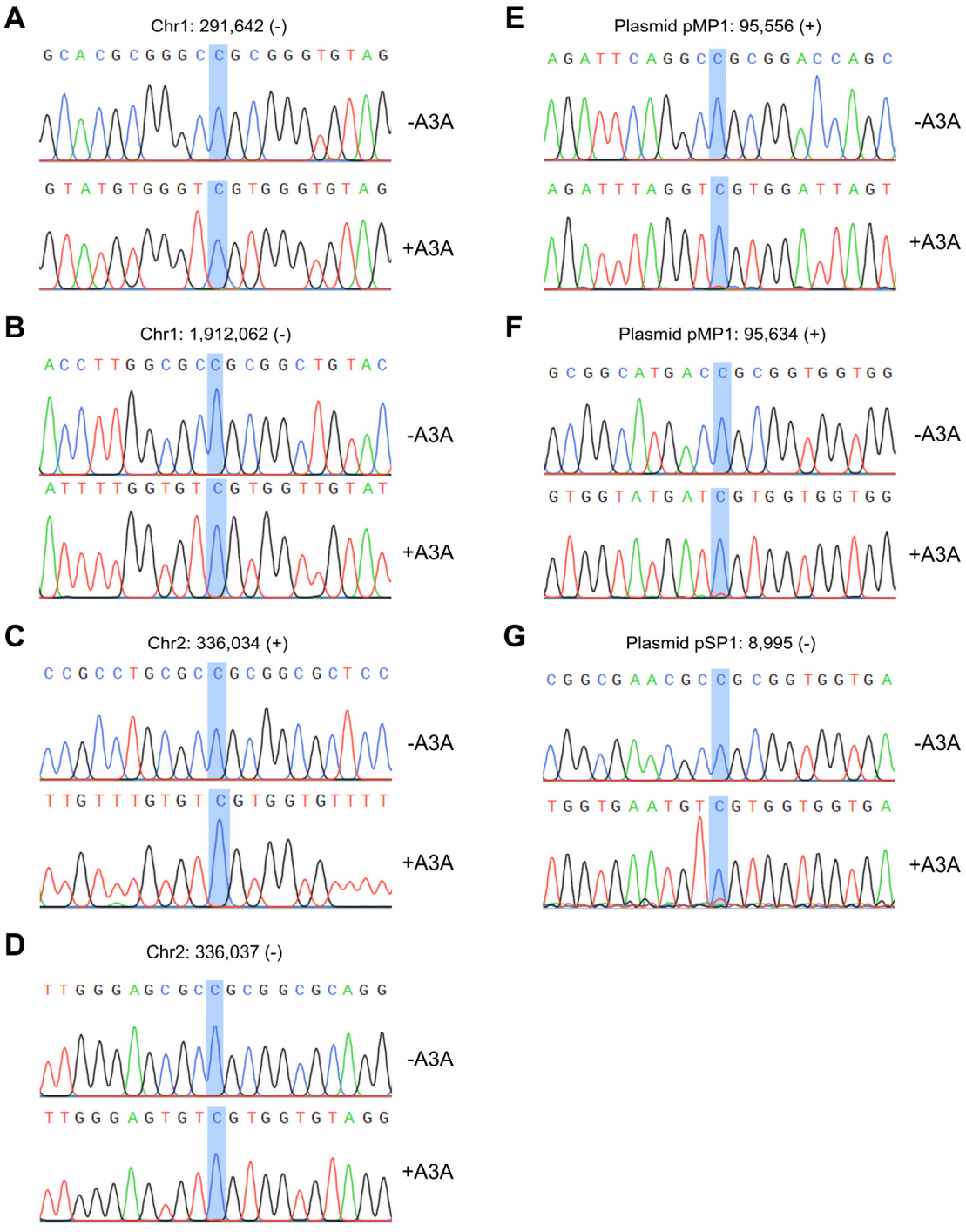


Figure S20. Locus-specific analysis of 4mC in seven C(4mC)GCGG motif by Sanger sequencing.



References

1. M. Y. Cheng, X. J. You, J. H. Ding, Y. Dai, M. Y. Chen, B. F. Yuan and Y. Q. Feng, *Chem Sci*, 2021, **12**, 8149-8156.
2. E. K. Schutsky, C. S. Nabel, A. K. F. Davis, J. E. DeNizio and R. M. Kohli, *Nucleic Acids Res*, 2017, **45**, 7655-7665.