Supporting Information

for

Enzyme-mediated bioorthogonal labeling method for genome-wide mapping of 5-hydroxymethyluracil

Cheng-Jie Ma,[1, †] Lin Li,[2, †] Wen-Xuan Shao,[1] Jiang-Hui Ding,[1] Xiao-Li Cai,[3] Zhao-Rong Lun,[3] Bi-Feng Yuan,*[,1,4] Yu-Qi Feng[1,4]

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

[2] School of Pharmacy, Shanghai Jiao Tong University, Shanghai 200240, P.R. China.

[3] Center for Parasitic Organisms, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, P.R. China.

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

[†] These authors contributed equally to this work.

* Corresponding author: Bi-Feng Yuan. Tel: +86-27-68755595; Fax: +86-27-68755595; E-mail: bfyuan@whu.edu.cn

The Supporting Information includes following items:

Table S1. Sequences of oligonucleotides.

Name	Sequence (5' to 3')
34nt T-DNA	CGATCTCACCATGGCCTTAACTCAGAAGCCCTAG
34nt U-DNA	CGATCTCACCAUGGCCTTAACTCAGAAGCCCTAG
34nt 5hmU-DNA	CGATCTCACCA5hmUGGCCTTAACTCAGAAGCCCTAG
34nt complementary-A-DNA	FAM-CTAGGGCTTCTGAGTTAAGGCCATGGTGAGATCG
34nt complementary-G-DNA	FAM-CTAGGGCTTCTGAGTTAAGGCCGTGGTGAGATCG
60nt 5hmU-DNA	CCTCACCATCTCAACCAATATTATACCA5hmUGGTATGCGT TATAATGAGGGAGAAGTGGTGA
60nt complementary-A-DNA	TCACCACTTCTCCCTCATTATAACGCATACCATGGTATAAT ATTGGTTGAGATGGTGAGG
60nt complementary-G-DNA	TCACCACTTCTCCCTCATTATAACGCATACCGTGGTATAAT ATTGGTTGAGATGGTGAGG
129nt control-DNA	CGGACACTATCAGTCTACGTTAATCACTATAAGCGACGTTC ATTTGGTGTCGGGCTGTGTTCGTAATCGTACCAGAGTATGT GCGGGTGGTGGCCCTAGGGCTTCTGAGTTAAGGCCATGGT GAGATCG

Table S2. Primers for PCR.

Name	Sequence (5’ to 3’)
IP241 forward primer	TCAGCTCAAGCTCTTCTGCA
IP241 reverse primer	CTTGAGCTGAAGGCTGCCGA
IP1081 forward primer	GCACATCACGCTTATATGCG
IP1081 reverse primer	CCAGTCACTTCCGCTCAACC
IP630 forward primer	CTTCATCATGGCTTCGCTGA
IP630 reverse primer	AAATGTCGTGAAGTTATGGA
IP263 forward primer	CTTCCTCAGCCTCCACGC
IP263 reverse primer	AGAAGATGCTGAGCGCATGAA
IP674 forward primer	AGATGCCGTGTGTGCGTGTT
IP674 reverse primer	GCTGTTACCTCCTCGCGCAC
IP1228 forward primer	CGGAGGACTTCATAGCGGAGG
IP1228 reverse primer	TCCATGATGGCTTCGGTTTGCTG
IP1309 forward primer	CCGATAGCCGCAGCTTCCTTG
IP1309 reverse primer	GACGCTCCTCTTCCATGGTGC
Up-IP241 forward primer	CGGCAACTACGCCAAAACTG
Up-IP241 reverse primer	GACAAATGGGGCAGAAGCAC
Up-IP1081 forward primer	CGTGGCTATCACTGGACACAA
Up-IP1081 reverse primer	CTCCTGTGCTAACTGTGCTGA
Up-IP630 forward primer	GCCCACTGCTCTGAATGTGTAT
Up-IP630 reverse primer	CCGCATATTTGTAGCTTCCGT
Up-IP263 forward primer	GCACTACCAGTCGGTTCTCG
Up-IP263 reverse primer	AGTAGAGTTAGCGGTTTGCCC
Up-IP674 forward primer	TCACATGACTACGTTGGTCG
Up-IP674 reverse primer	ATGGCCAGTGCCAGTTCTTT
Up-IP1228 forward primer	CAGAAGGCACCAGCGACAA
Up-IP1228 reverse primer	GTAGGTTCTGCGTAAGCGATGT
Up-IP1309 forward primer	AGTCGTTTGGCTGAGGTTGA
Up-IP1309 reverse primer	AAGGAAGCAACTCAAGGGGG
60nt forward primer	CCTCACCATCTCAACCAATA
60nt reverse primer	TCACCACTTCTCCCTCATTA
129nt forward primer	CGGACACTATCAGTCTACGT
129nt reverse primer	CGATCTCACCATGGCCTTAA
P7-index primer (input 1)	CAAGCAGAAGACGGCATACGAGATAACGTGAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC T
P7-index primer (input 2)	CAAGCAGAAGACGGCATACGAGATAAACATCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC T
P7-index primer (pull-down 1)	CAAGCAGAAGACGGCATACGAGATAGTGGTCA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC T
P7-index primer (pull-down 2)	CAAGCAGAAGACGGCATACGAGATACCACTGT

	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC T
P5-universal primer	AATGATACGGCGACCACCGAGATCTACACTCTT TCCCTACACGACGCTCTTCCGATCT

Table S3. High-throughput sequencing data.

Library name	Raw reads	Clean reads	Mapped reads
Input 1	14302993	9423177	7303202 (77.50%)
Input 2	15365837	8710311	6437940 (73.91%)
Pull-down 1	22144857	15042102	10838743 (72.06%)
Pull-down 2	21858322	11347664	7245368 (63.85%)

Figure S1. Quantitative evaluation of the phosphorylation of 5hmU by 5hmUDK using the mixture of duplex T-DNA and duplex 5hmU-DNA. (A) Analysis of the phosphorylation of 5hmU by 5hmUDK with NcoI cleavage assay. Duplex T-DNA and 5hmU-DNA were mixed at different ratios (0, 20, 40, 60, 80, and 100% of 5hmU-DNA) and then processed with 5hmUDK treatment at 37°C for 30 min. The resulting DNA was incubated with NcoI followed by polyacrylamide gel electrophoresis analysis. (B) The cleaved fractions of 5hmU-DNA were plotted against the percentage of 5hmU-DNA. The slope of -0.96 indicated that the levels of 5hmU-DNA in DNA mixture could be quantitatively obtained using the 5hmUDK assay.

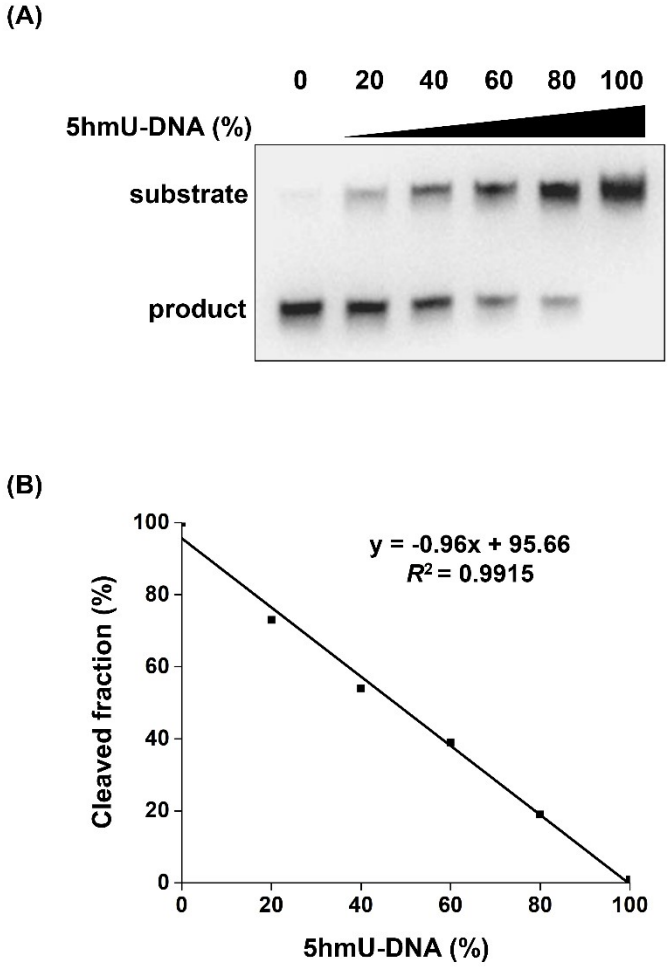
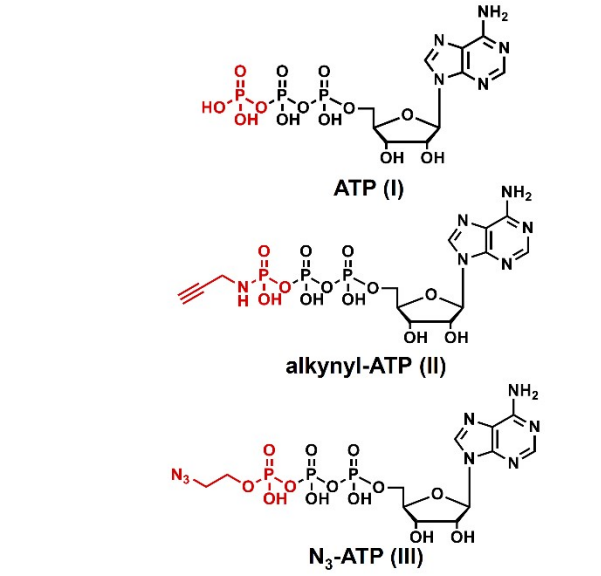


Figure S2. Phosphorylation of 5hmU in single-stranded DNA (ssDNA) by 5hmUDK at different times (0.5 h and 12 h) using ATP, alkynyl-ATP, or N_3-ATP as the cofactor. NcoI cleavage assay was used to evaluate the phosphorylation of 5hmU. I, ATP; II, alkynyl-ATP; III, N_3-ATP.



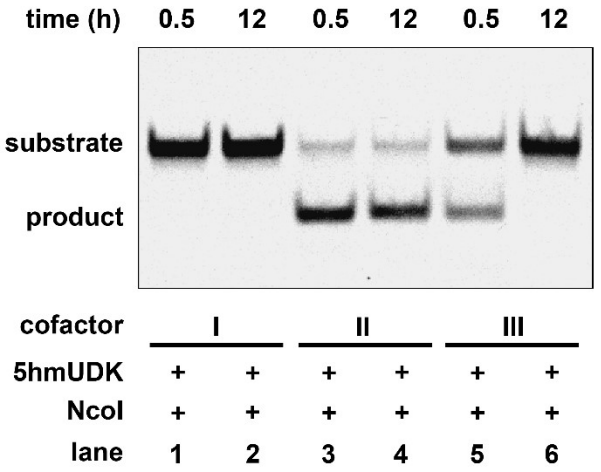


Figure S3. Reversal of phosphorylation of 5hmU in 5hmU-DNA by alkaline phosphatase. (A) Schematic diagram of the phosphorylation of 5hmU by 5hmUDK and reversal of phosphorylation of 5hmU by alkaline phosphatase. (B) Analysis of the reversal of phosphorylation of 5hmU by different alkaline phosphatases. Duplex 5hmU-DNA was firstly incubated with 5hmUDK using different cofactor of ATP, alkynyl-ATP, or N_3-ATP. Then the phosphorylated DNA was treated with different alkaline phosphatases, including CIAP, and SAP, and EAP. The resulting DNA was cleaved by NcoI followed by polyacrylamide gel electrophoresis analysis. I, ATP; II, alkynyl-ATP; III, N_3-ATP.

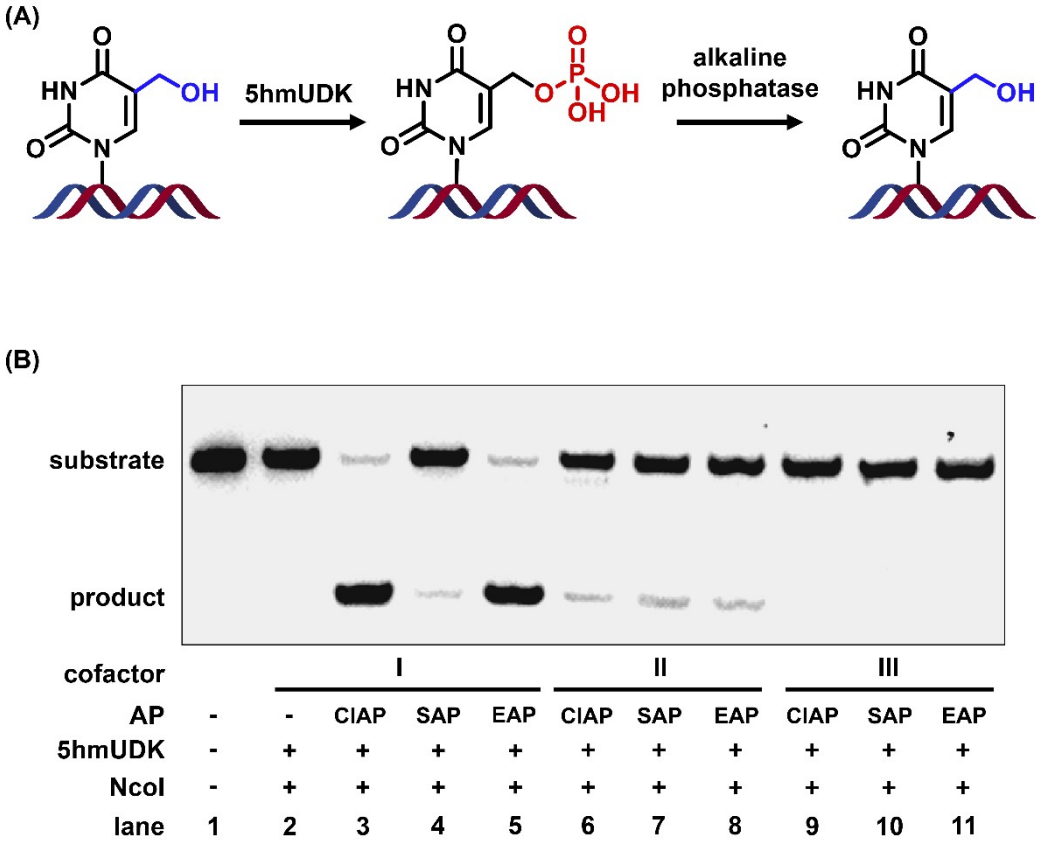


Figure S4. Bioorthogonal reaction between N_3-5hmU and DBCO-Cy3. (A) The schematic diagram of the reaction between N_3-5hmU and DBCO-Cy3. (B) Evaluation of the bioorthogonal reaction between N_3-5hmU in different DNA forms and DBCO-Cy3. Single-stranded 5hmU-DNA, duplex 5hmU-DNA (5hmU:A base pair) and duplex 5hmU-DNA (5hmU:G mispair) were incubated with or without 5hmUDK and N_3-ATP. Then the obtained DNA was reacted with DBCO-Cy3 followed by polyacrylamide gel electrophoresis analysis. Green light excitation was employed to detect Cy3 fluorescence. The reaction between N_3-5hmU DNA with DBCO-Cy3 led to the slow shift compared to the unlabeled DNA (comparing lanes 2 and 3, lanes 4 and 5), indicating the successful bioorthogonal reaction between N_3-5hmU DNA and DBCO-Cy3.

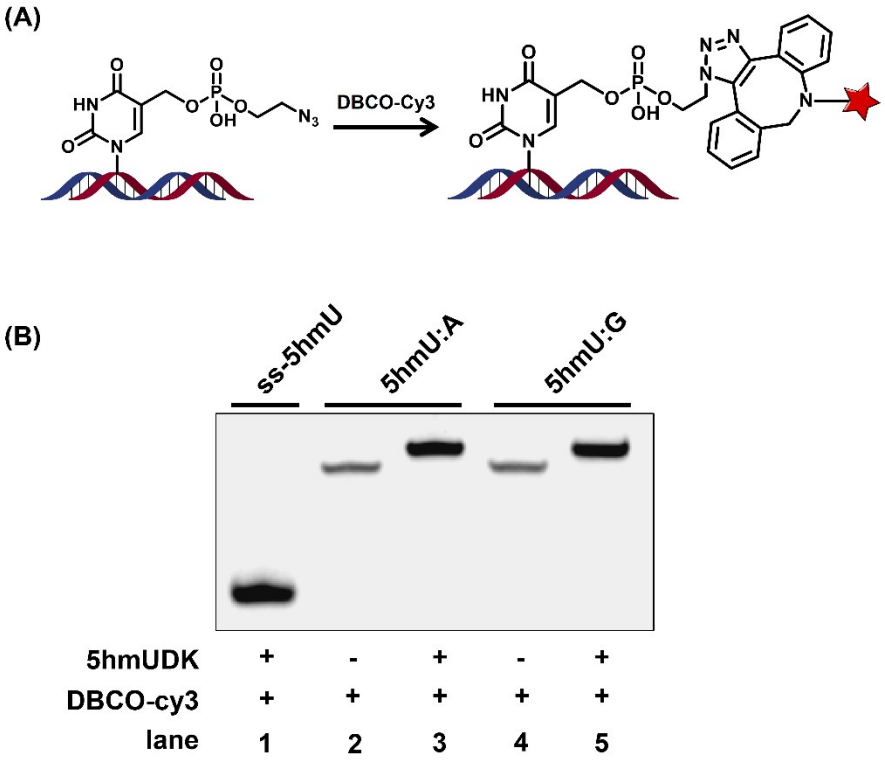


Figure S5. LC-UV trace of 5hmU-DNA before (A) and after reaction with N_3-ATP and DBCO-SS-biotin (B). The LC separation was performed on a Shimadzu VP-ODS column (250 mm × 2.1 mm i.d, 5 μm) with a flow rate of 0.2 mL/min at 40 ℃. 100 mM TEAA in water (solvent A) and 100 mM TEAA in acetonitrile (6/4, v/v, solvent B) were employed as mobile phases. A gradient of 0−0.5 min 5% B, 0.5−5 min 5−80% B, 5−7 min 80% B, and 7−16 min 5% B was used.

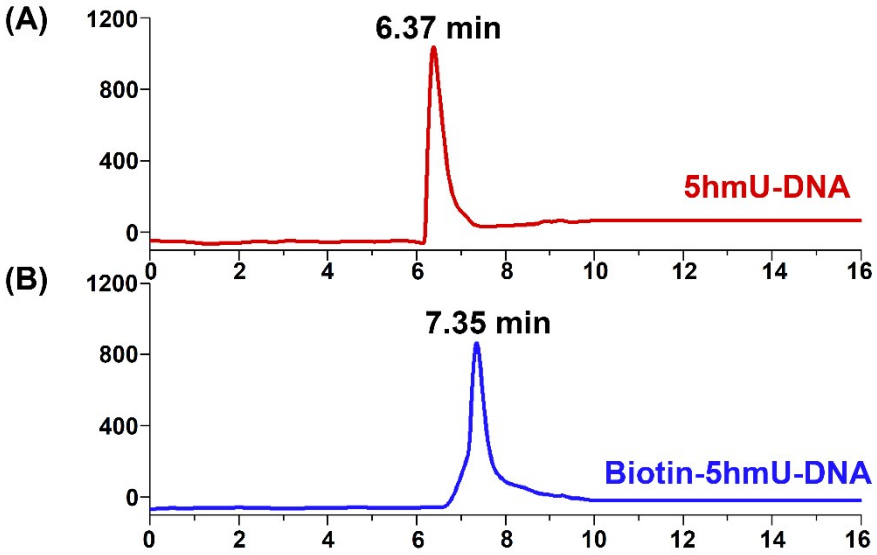


Figure S6. Evaluation of the preference of the phosphorylation of 5hmU with ATP or N_3-ATP as the cofactor by 5hmUDK. (A) ATP and N_3-ATP were mixed at different ratios (0, 20, 40, 60, 80, and 100%) and incubated with 5hmUDK at 37°C for 30 min. The resulting DNA was incubated with DBCO-SS-biotin followed by polyacrylamide gel electrophoresis analysis. (B) The fractions of biotinylated 5hmU-DNA were plotted against the percentage of N_3-ATP in the mixture. The slope of 1.05 indicated that 5hmUDK had not preference between ATP and N_3-ATP.

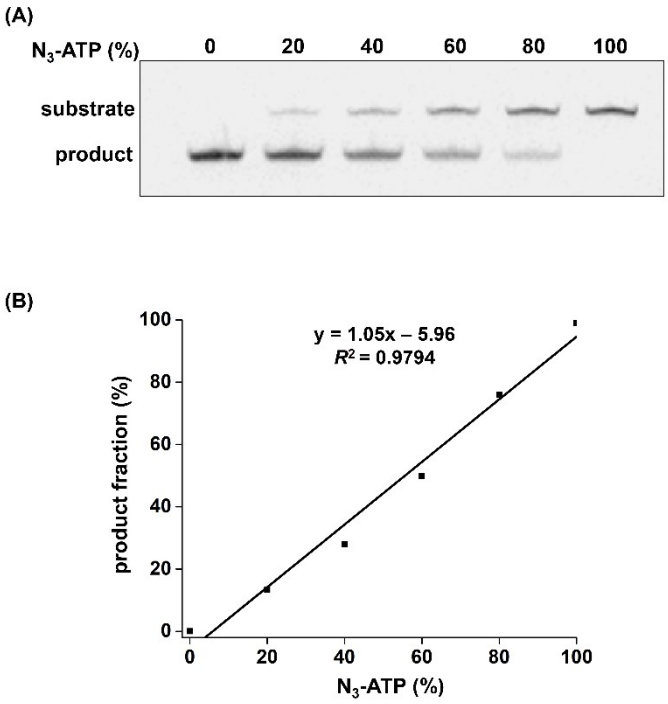


Figure S7. Evaluation of the real-time qPCR amplification efficiencies. (A) Standard curve of 60-bp 5hmU-DNA. (B) Standard curve of 129-bp control DNA. The standard curves were constructed with using a series of amounts of DNA templates (0.01, 0.1 1, 10, and 100 pg) for real-time qPCR. The obtained amplification efficiencies of 60-bp 5hmU-DNA and 129-bp control DNA were 103.1% and 90.2%, respectively.

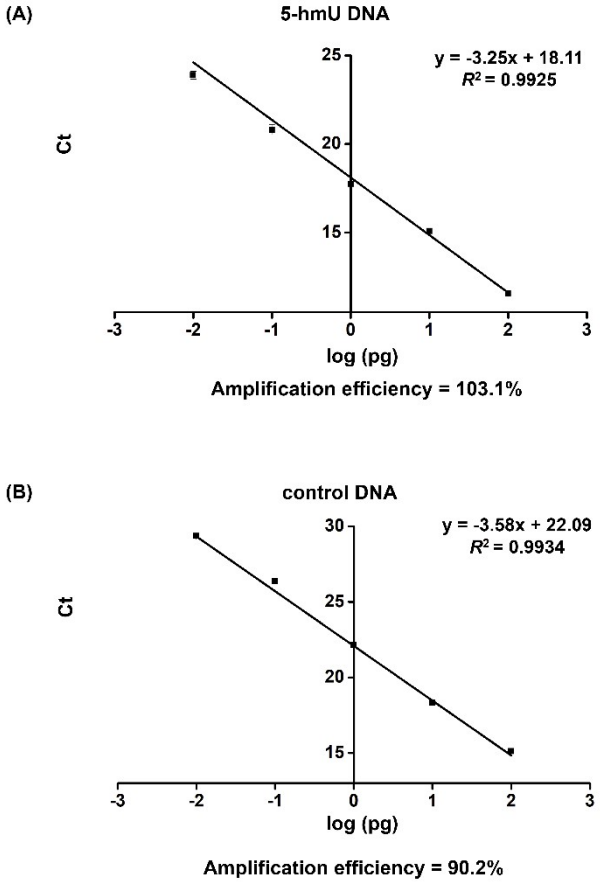


Figure S8. The schematic diagram of library preparation for high-throughput sequencing.

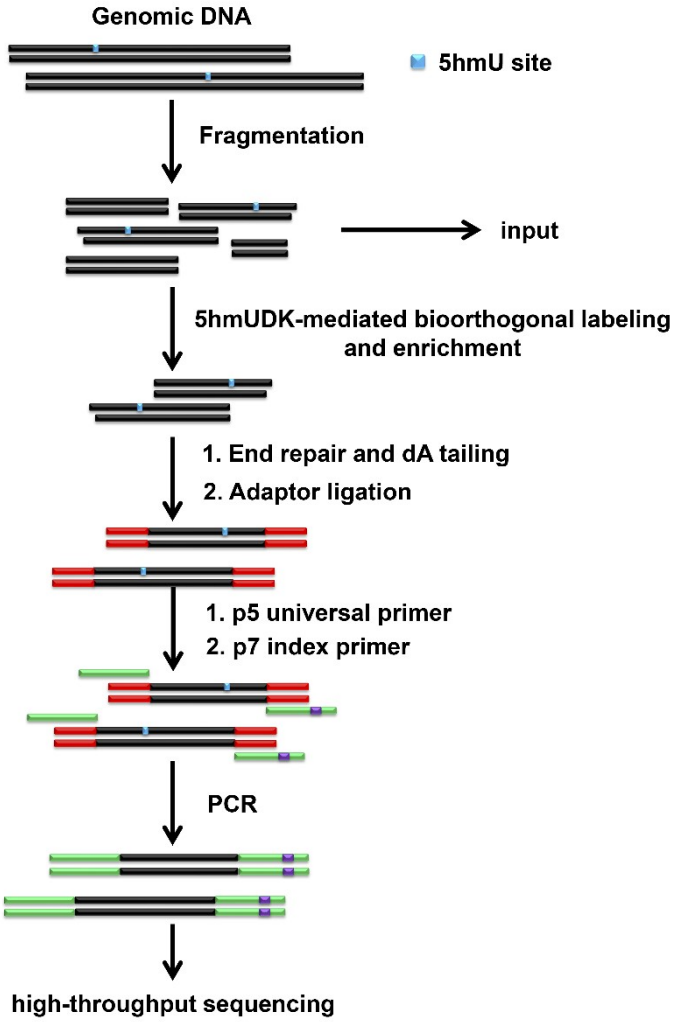


Figure S9. The adaptors and primers used for high-throughput sequencing.

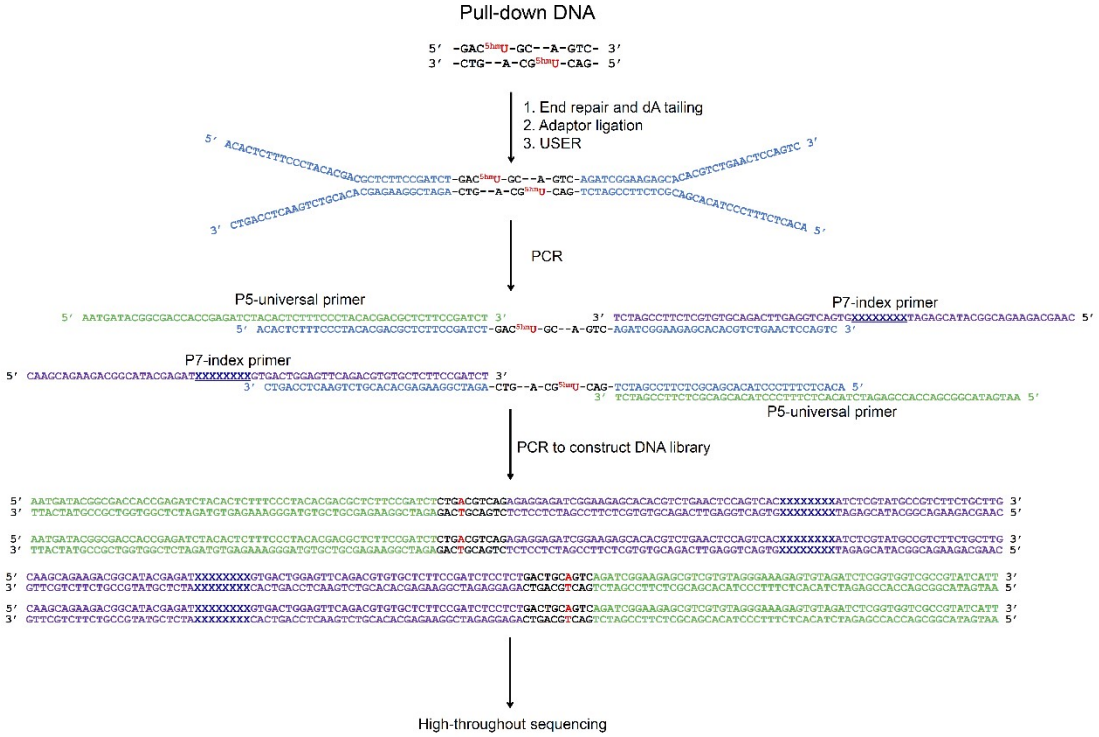


Figure S10. The spearman correlation analysis between four samples (input 1, input 2, pull-down 1 and pull-down 2).

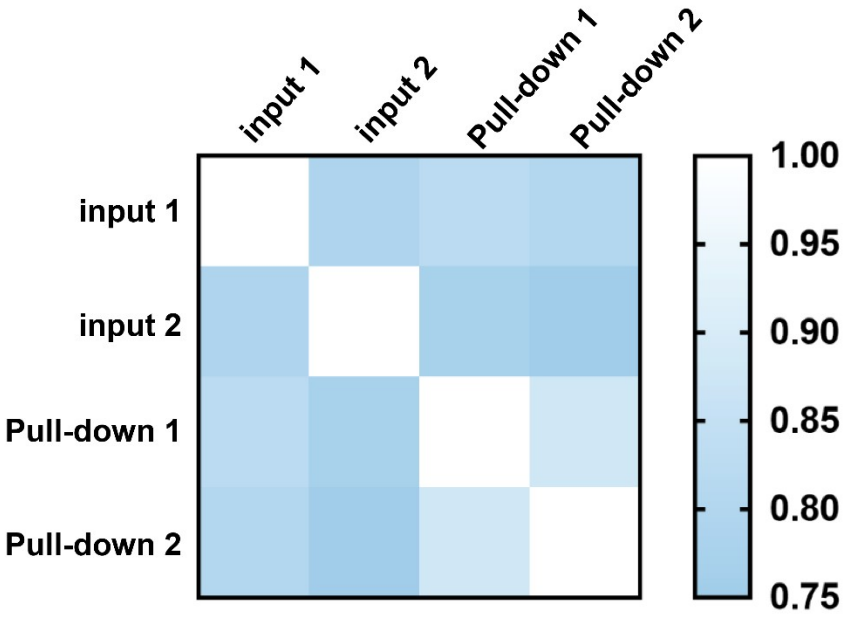


Figure S11. Verification of 5hmU peaks in *T. brucei* genome by real-time qPCR. A mixture of 0.1 ng of synthesized 129-bp control DNA and 10 μg of fragmented genomic DNA was used for the verification. The mixture was processed sequentially with 5hmUDK treatment (using N_3-ATP), bioorthogonal labeling (with DBCO-SS-biotin), and streptavidin pull-down. Then 7 different 5hmU-rich regions observed from sequencing results as well as the control DNA were amplified by real-time qPCR to calculate the enrichment folds. (A) The regions amplified by real-time qPCR are from the sequencing peaks that contain the 5hmU. (B) The regions amplified by real-time qPCR are vicinal to the regions of 5hmU peaks (~ 1000 bp upstream of the regions of 5hmU peaks).

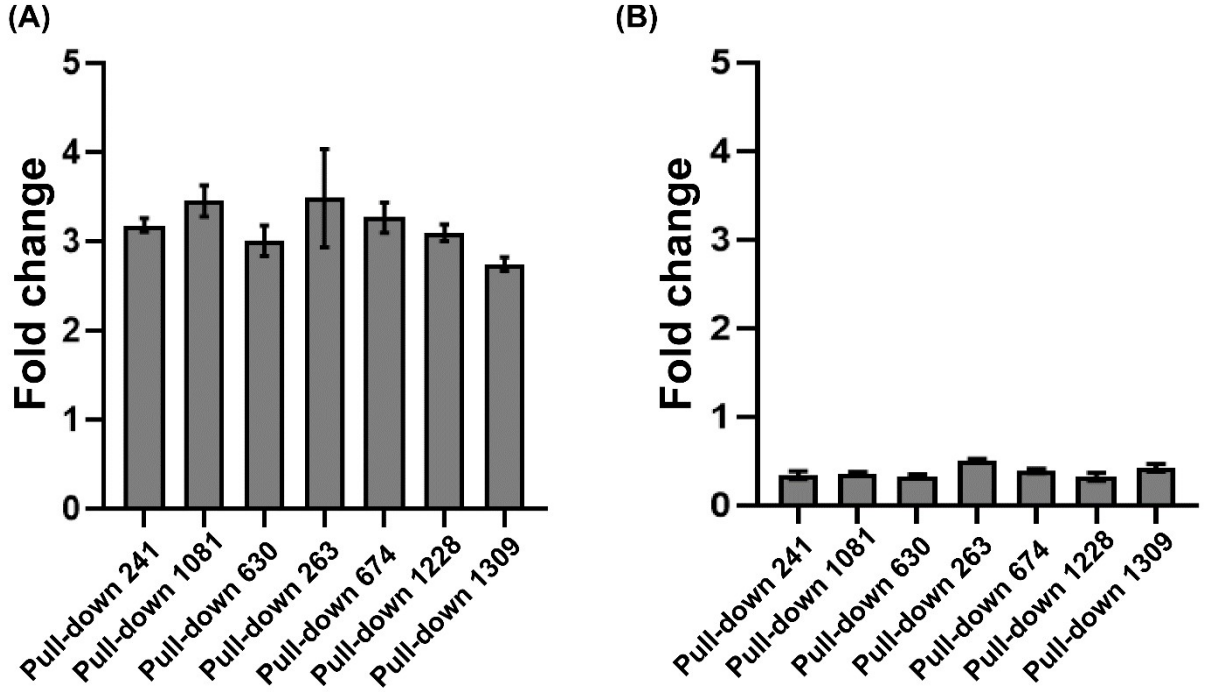


Figure S12. The IGVtools view of 5hmU peaks in pull-down samples compared to input sample.

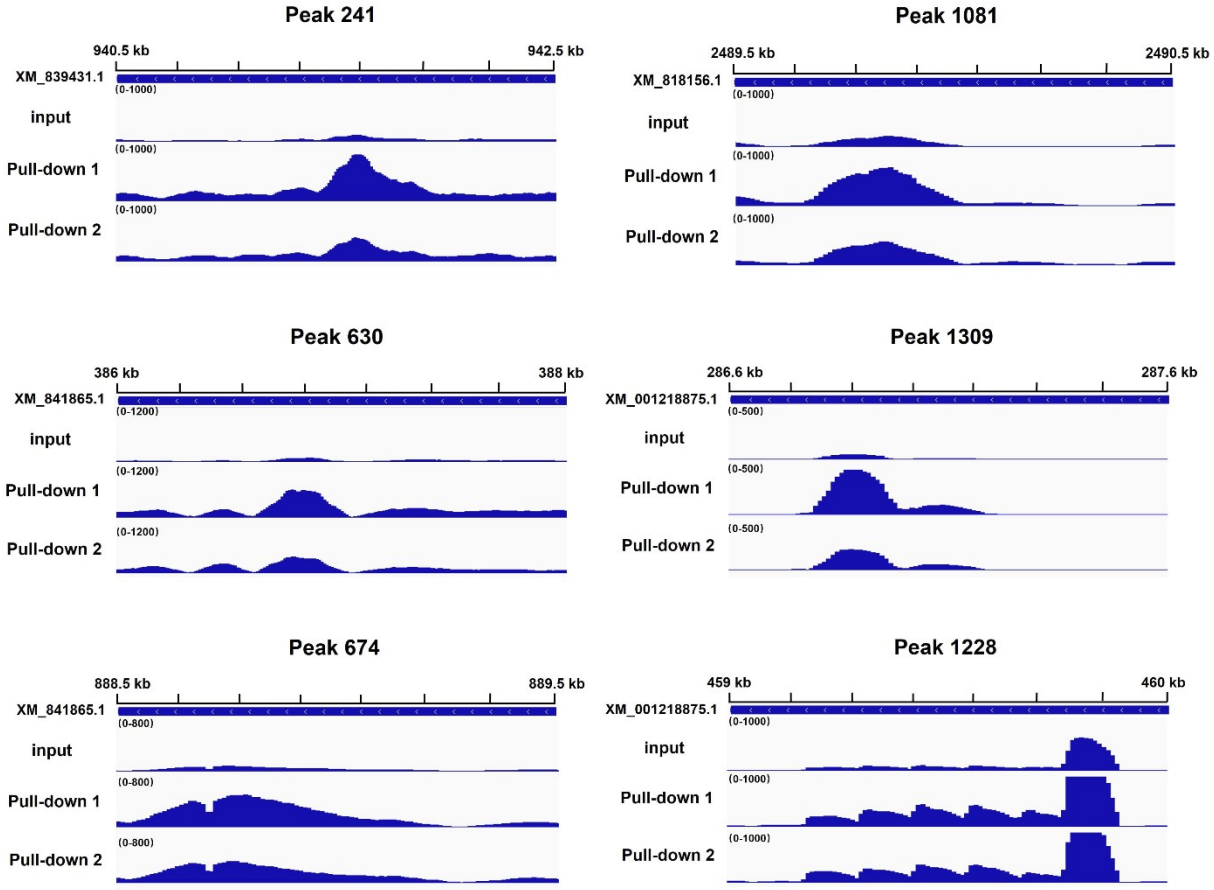


Figure S13. Enrichment of 5hmU-containing DNA fragments with using ATP instead of N_3-ATP. A mixture of 0.1 ng of synthesized 129-bp control DNA and 10 μg of fragmented genomic DNA was used. The mixture was processed sequentially with 5hmUDK treatment (using ATP instead of N_3-ATP), bioorthogonal labeling (with DBCO-SS-biotin), and streptavidin pull-down. Then 7 different 5hmU-rich regions observed from sequencing results as well as the control DNA were amplified by real-time qPCR to calculate the enrichment folds. Theoretically, ATP doesn't carry the N_3 group, which therefore couldn't be labeled by DBCO-SS-biotin. Thus, 5hmU-containing DNA fragments would not get enriched.