ISCI, Volume *9*

Supplemental Information

Highly Selective 5-Formyluracil Labeling and Genome-wide Mapping Using (2-Benzimidazolyl)Acetonitrile Probe

Yafen Wang, Chaoxing Liu, Fan Wu, Xiong Zhang, Sheng Liu, Zonggui Chen, Weiwu Zeng, Wei Yang, Xiaolian Zhang, Yu Zhou, Xiaocheng Weng, Zhiguo Wu, and Xiang Zhou

Supplemental figures and legends

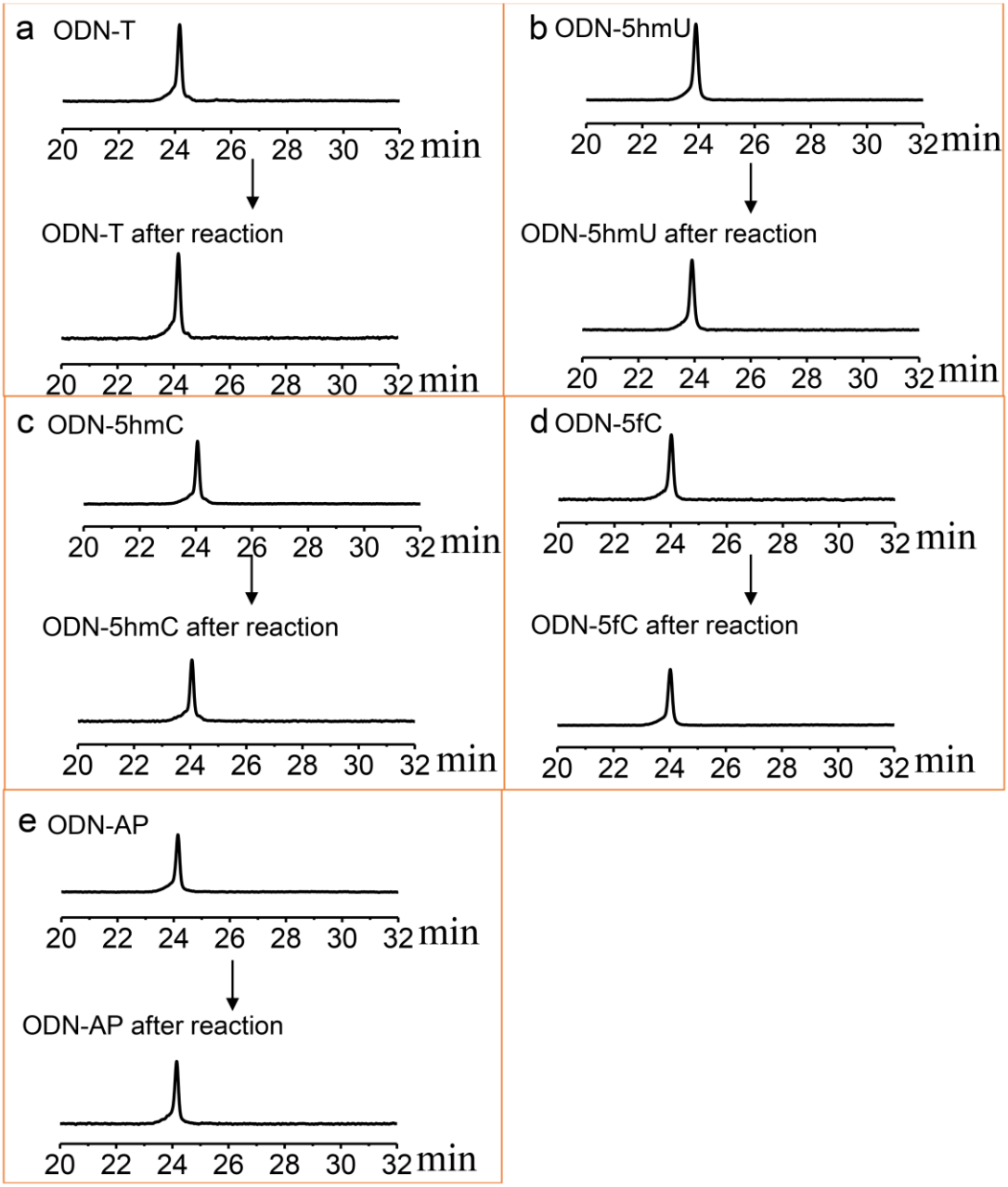


Figure S1. HPLC detection of various ODNs before and after treatment with azi-BIAN (Related to Figure 2a-d).

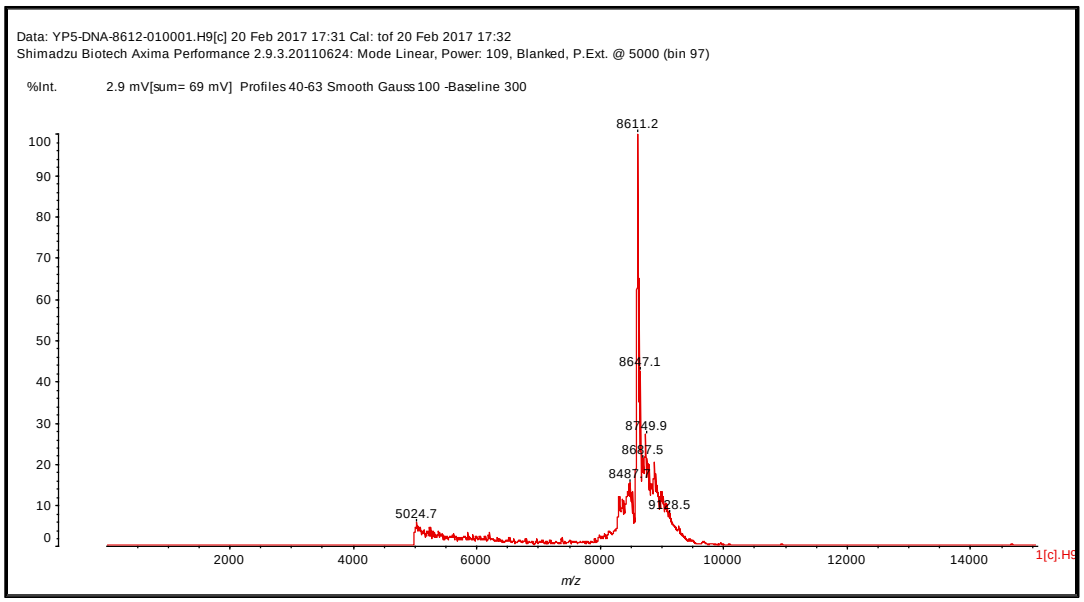
(a) RP-HPLC trace at λ = 260 nm of ODN-T before and after treatment with azi-BIAN. (b) RP-HPLC trace at λ = 260 nm of ODN-5hmU before and after treatment with azi-BIAN. (c) RP-HPLC trace at λ = 260 nm of ODN-5hmC before and after treatment with azi-BIAN (d) RP-HPLC trace at λ = 260 nm of ODN-5fC before and after treatment with azi-BIAN. (e) RP-HPLC trace at λ = 260 nm of ODN-AP before and after treatment with azi-BIAN.

a

5’-CATAG**fU**GCTCAAGAGAAATCTCGATGG-3’ ⟶

5’-CATAG**azi-biaU**GCTCAAGAGAAATCTCGATGG-3’

calculated 8612.6, found 8611.2.



b

5’-CATAG**fU**GCTCAAGAGAAATCTCGATGG-3’ ⟶

5’-CATAG**biotin-U**GCTCAAGAGAAATCTCGATGG-3’

calculated 9480.9, found 9484.8.

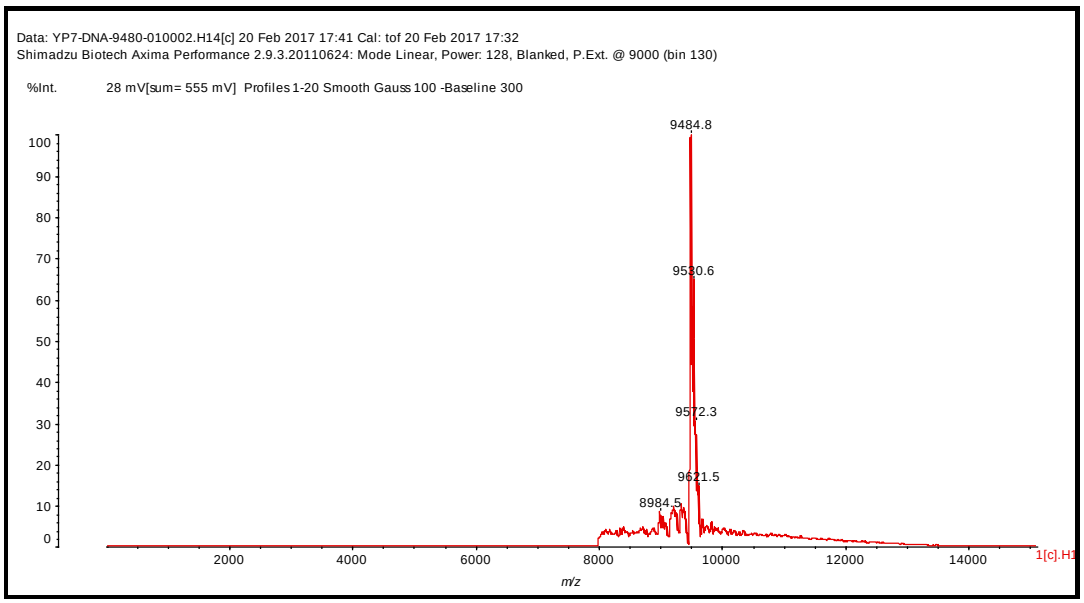
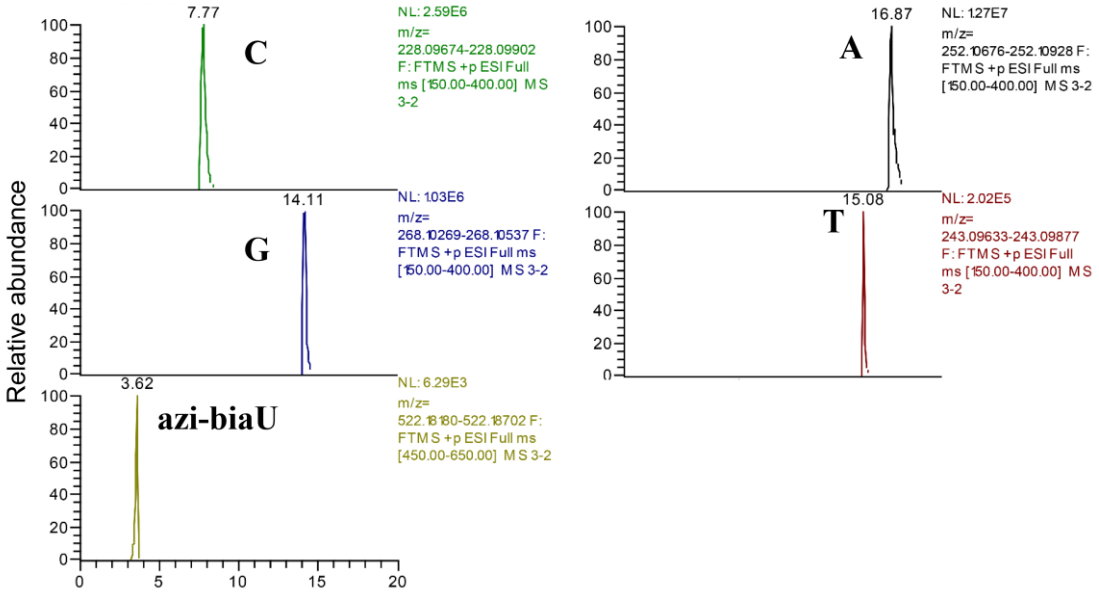


Figure S2. DNA MALDI-TOF Mass Spectra (Related to Figure 2a-d). (a) MALDI-TOF-spectrum of ODN2-5fU after incubation with azi-BIAN; (b) MALDI-TOF-spectrum of ODN2-5fU after incubation with azi-BIAN, then reacted with DBCO-biotin.

a



b

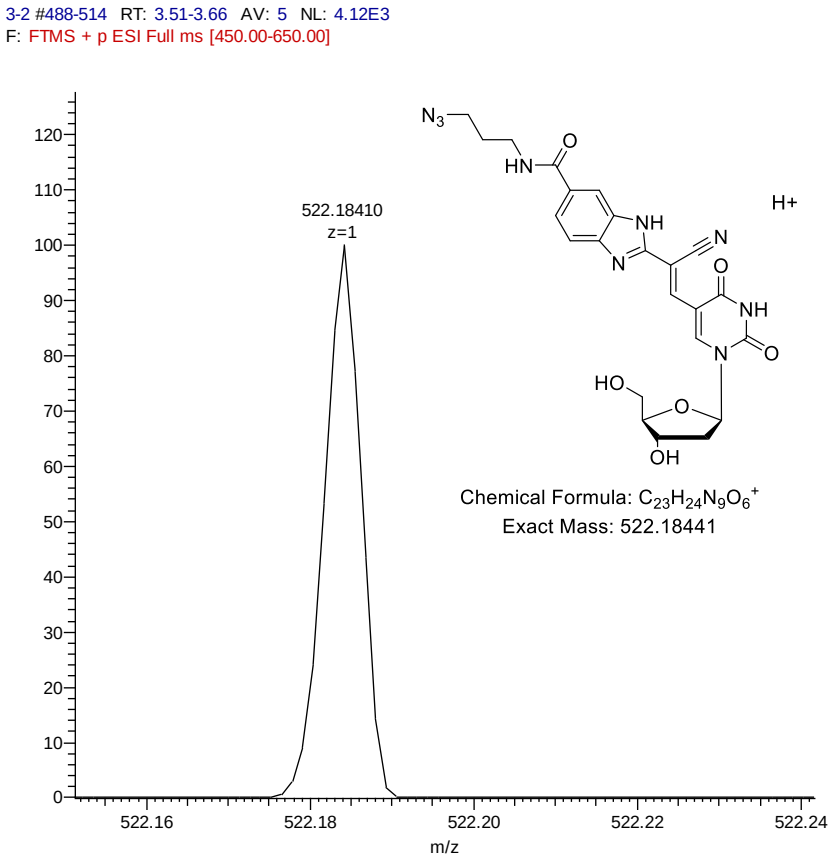
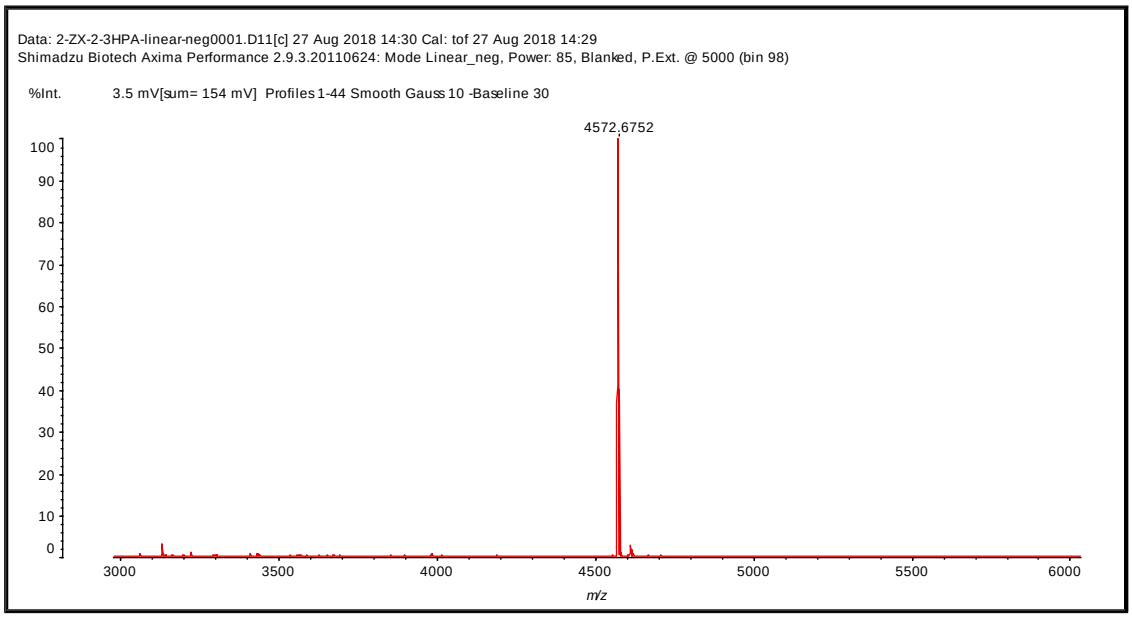


Figure S3. HPLC-MS detection of digestion of ODN-azi-biaU (Related to Figure 2a-d). (a) HPLC-MS extracted $[M+H]^+$ ion count for A, T, C, G, azi-biaU after digestion of DNA from the ODN-azi-biaU. (b) HRMS (ESI+) of azi-biaU in HPLC-MS after digestion, HRMS $C_{23}H_{24}N_9O_6^+$ $[M+H]^+$ calculated 522.18441, found 522.18410.

a

5’-GACTCAA5fCAGCCGTA-3’ caculated 4571.8, found 4572.7.



b

5’-GACTCAA5fCAGCCGTA-3’ —azi-BIAN→ 5’-GACTCAAazi-biaUAGCCGTA-3’

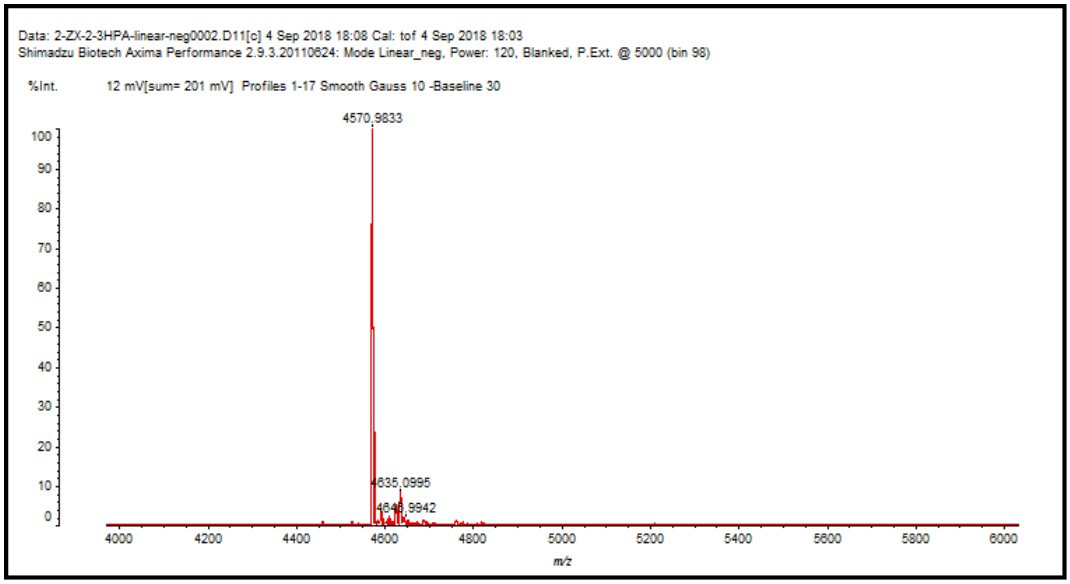


Figure S4. DNA MALDI-TOF Mass Spectra (Related to Figure 2a-d). (a) MALDI-TOF-spectrum of ODN-5fC; (b) MALDI-TOF-spectrum of ODN-5fC after incubation with azi-BIAN. No mass of ODN-azi-biaU appeared.

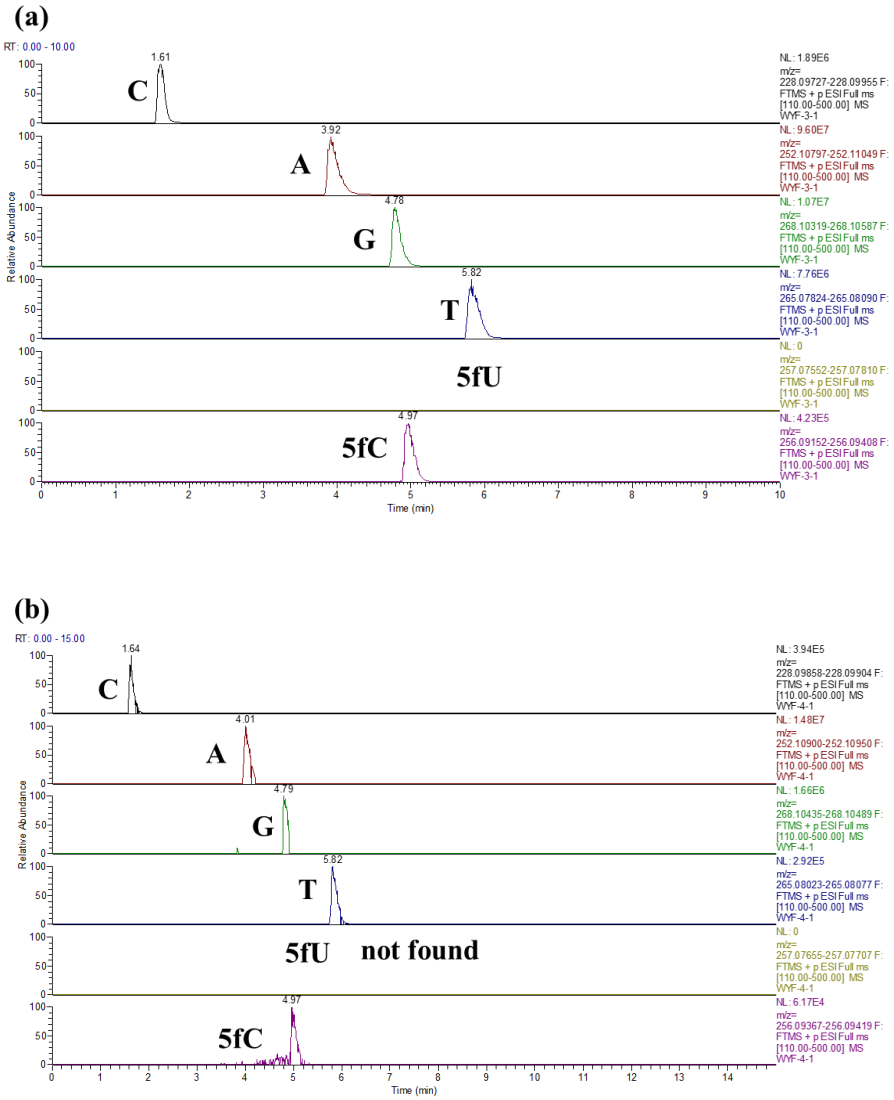


Figure S5. HPLC-MS detection of digestion of extracted model DNA (Related to Figure 2a-d). (a) HPLC-MS extracted $[M+H]^+$ ion count for A, T, C, G, 5fU, 5fC after digestion of the 80bp ds ODN-fC. (b) HPLC-MS extracted $[M+H]^+$ ion count for A, T, C, G, 5fU, 5fC after digestion of the extracted 80bp ds ODN-fC by DNeasy® Blood & Tissue Kit.

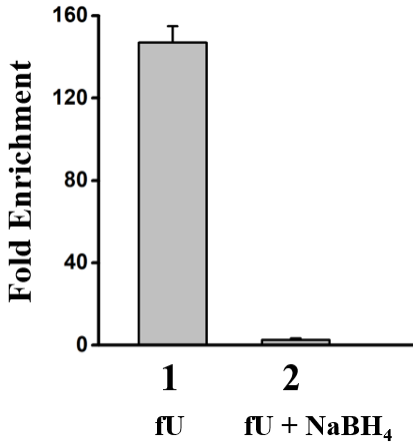


Figure S6. Enrichment tests of reduced 5fU (Related to Figure 2e). 1 represent ODN-5fU without treatment by $NaBH_4$, 2 represent ODN 5fU was treated by $NaBH_4$. Values shown are fold-enrichment over canonical nucleobases. Error bars represent the standard deviations of three parallel measurements.

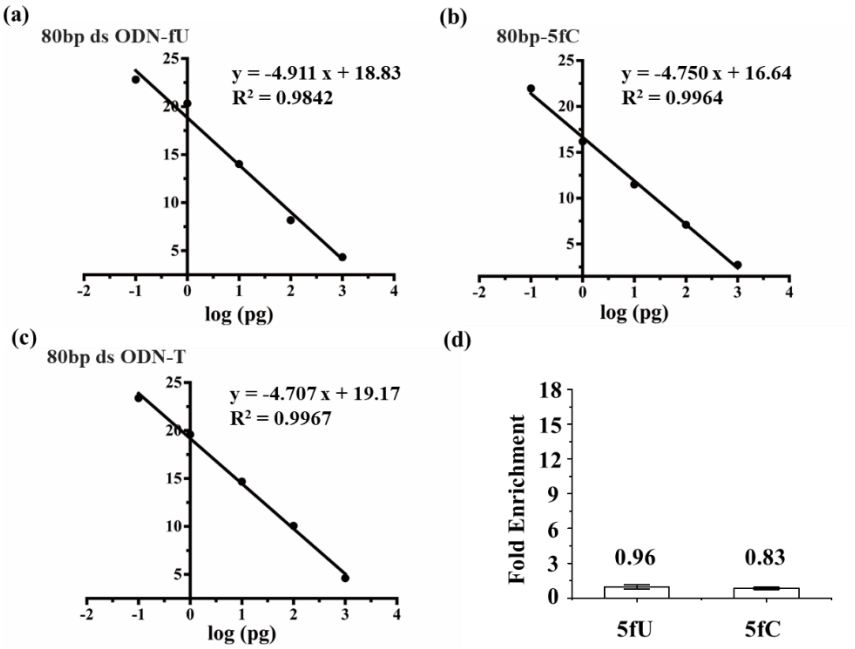


Figure S7. Enrichment tests of DNA containing hydroxylamine ($EtONH_2$) protection of the mixed system (80bp ds ODN-fU, 80bp-5fC, and 80bp ds ODN-T) (Related to Figure 2e). (a)-(c) Example calibration line of 80bp ds ODN-fU, 80bp-5fC, and 80bp ds ODN-T for enrichment analysis. (d) Values shown are fold-enrichment over 80bp ds ODN-T. Error bars represent the standard deviations of three parallel measurements.

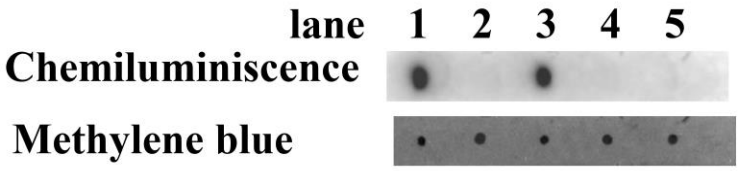
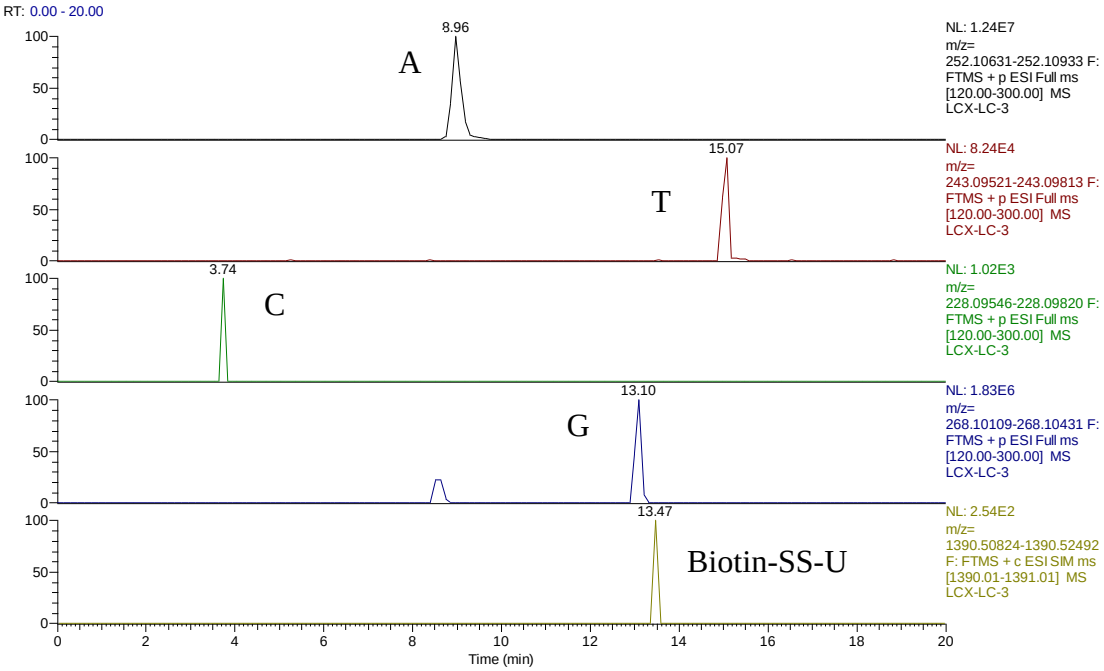


Figure S8. Dot-blot assay of streptavidin-HRP detection of samples containing 5fU (Related to Figure 2). Lane 1: genomic DNA from mouse hippocampus was treated with azi-BIAN, then incubation with DBCO-S-S-PEG3-biotin (Click Chemistry Tools, A112-10) to introduce biotin group. Lane 2: genomic DNA from mouse hippocampus was treated with (2-benzimidazolyl)acetonitrile, then incubation with DBCO-S-S-PEG3-biotin. However, the reagent (2-benzimidazolyl)acetonitrile without azido group can't successfully react with DBCO-S-S-PEG3-biotin through click chemistry. Lane 3: synthesized DNA with 5fU sites was treated with azi-BIAN, then incubation with DBCO-S-S-PEG_3-biotin to introduce biotin group. Lane 4: synthesized DNA with 5fC sites was treated with azi-BIAN, then incubation with DBCO-S-S-PEG3-biotin in the same conditions. Lane 5: synthesized DNA (only containing canonical nucleosides) was treated with azi-BIAN, then incubation with DBCO-S-S-PEG3-biotin in the same conditions. Only the biotin labeled DNAs can be get a dot. And after methylene blue incubation, we can verify the existence of DNA of every dot.

a



b

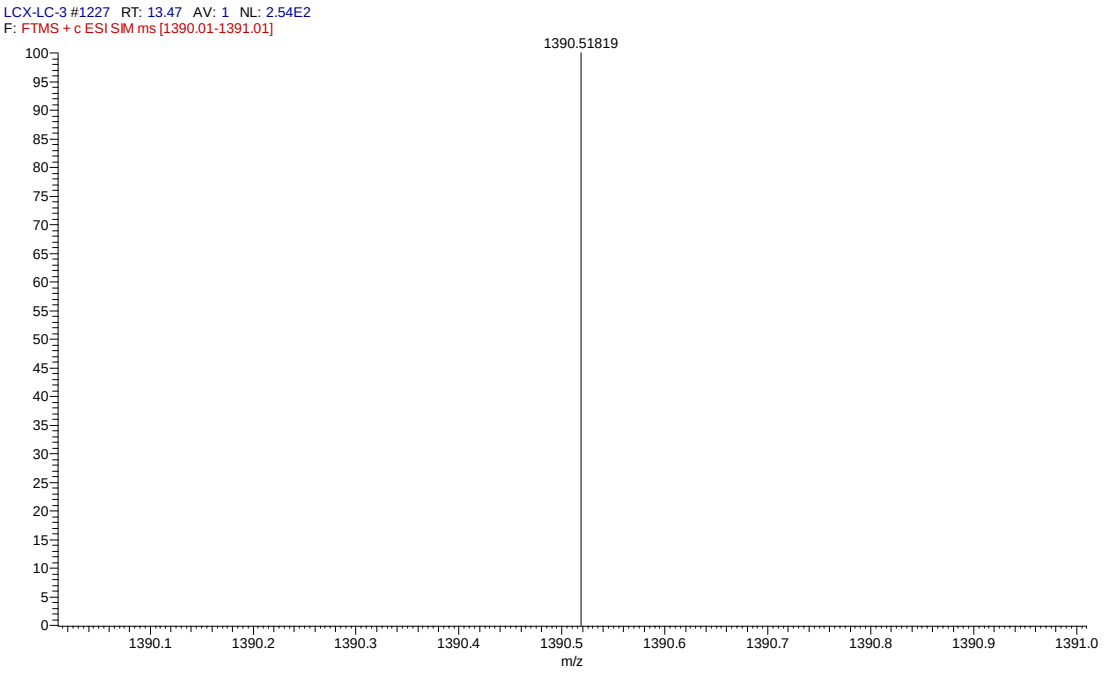


Figure S9. HPLC-MS detection of digestion of biotin labeled genomic DNA (Related to Figure 2). (a) HPLC-MS extracted $[M+H]^+$ ion count for A, T, C, G, Biotin-SS-U after digestion of the biotin labeled genomic DNA. (b) HRMS (ESI+) of Biotin-SS-U in HPLC-MS after digestion, HRMS $C_{65}H_{80}N_{15}O_{14}S_3^+$ $[M+H]^+$ calculated 1390.51658, found 1390.51819.

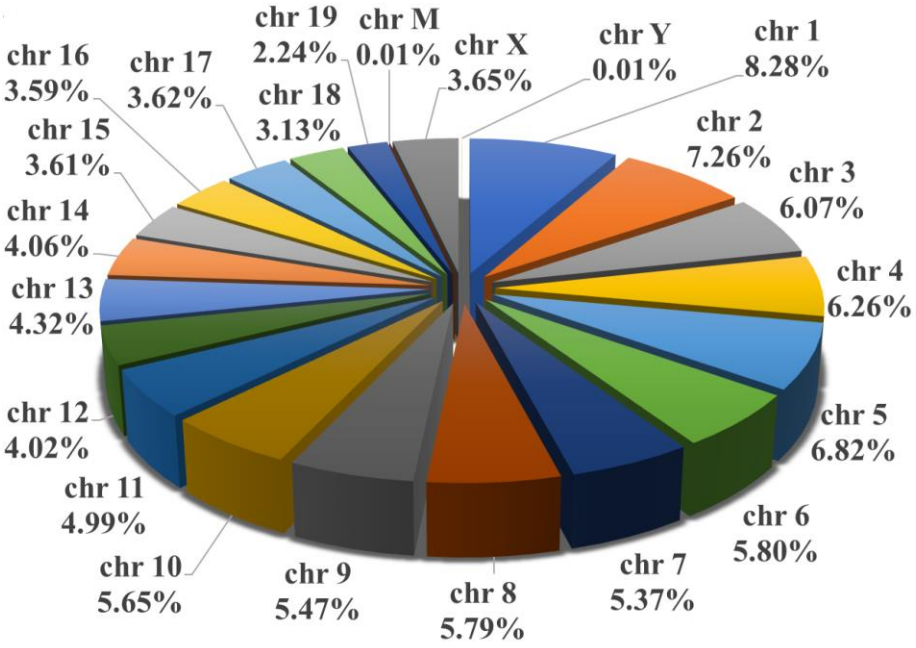


Figure S10. The pie chart shows the percentage of peak number of every chromosome in whole genome (Related to Figure 3).

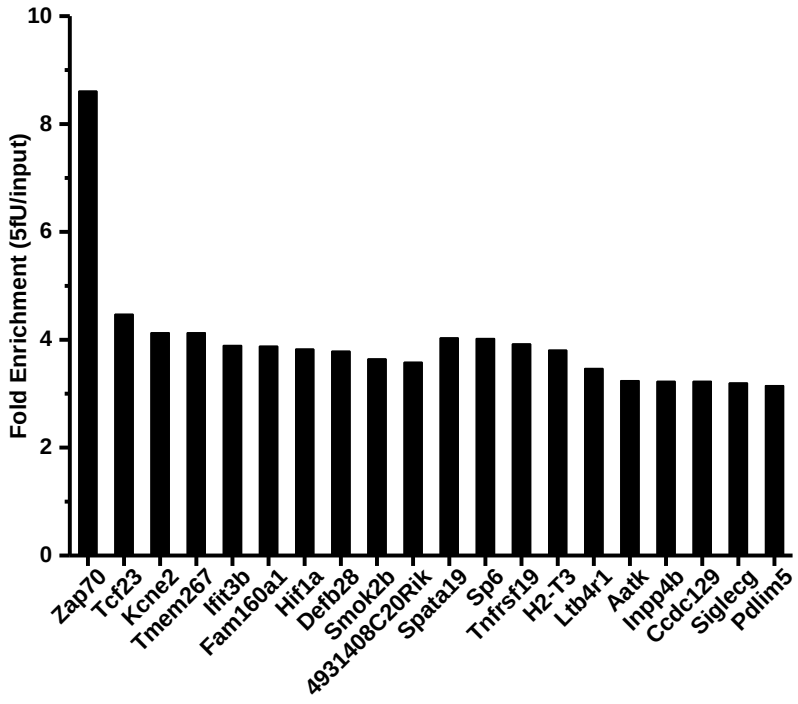


Figure S11. Verification 5fU-enriched regions in mouse hippocampus by qPCR (Related to Figure 4a&b). X-axis is labeled with the gene names within which the identified peak was identified. Fold enrichment is calculated as 2^-ddCt, where dCt_1 = Ct (5fU enriched) – Ct (Input), $dCt_2 = Ct_{ref}$ (5fU enriched) – Ct_{ref} (Input), ddCt = $dCt_1 - dCt_2$ (Ct_{ref} represented the referenced gene). The value shown for each biological replicate was the average of three pulldown-qPCR technical replicates.

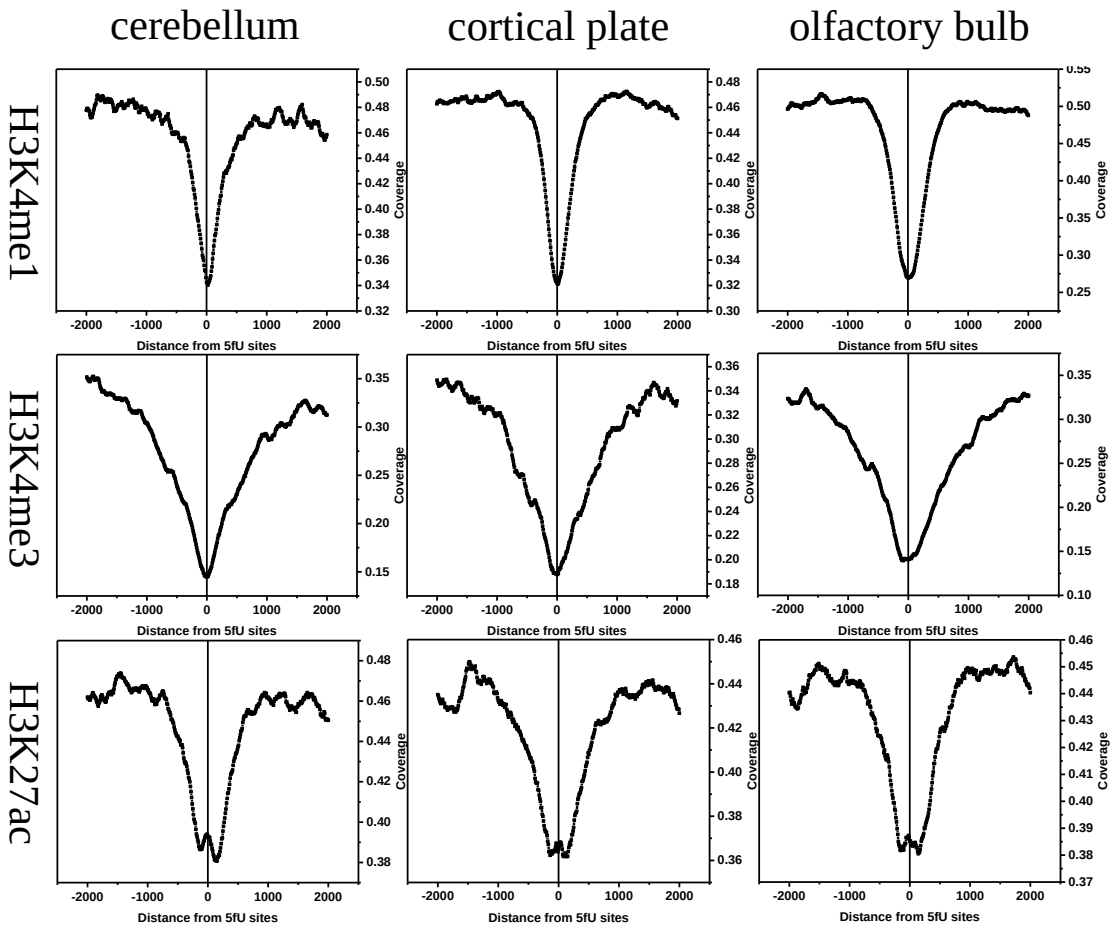


Figure S12. Distribution patterns of 5fU at different histone modification sites of varied brain tissues (Related to Figure 4c). Distribution patterns of 5fU with respect to H3K4me1, H3K27me3 and H3K27ac modification sites in the cerebellum, cortical plate and olfactory bulb, respectively. From these, negative peak showed negative correlation between 5fU and the histone modification.

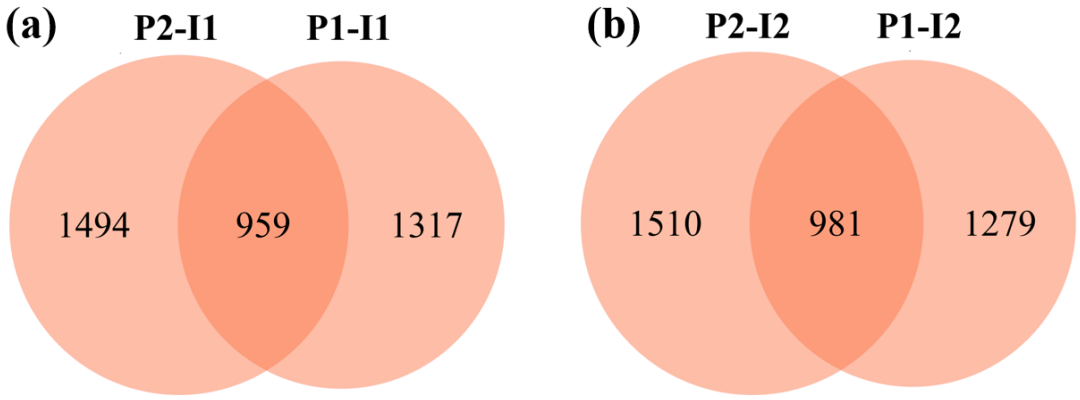


Figure S13. Venn diagrams showing the peak-merging results of two biological replicates, pull-down data 1 (P1) and pull-down data 2 (P2) against a) Input data 1 (I1) and b) Input data 2 (I2) (Related to Figure 5).

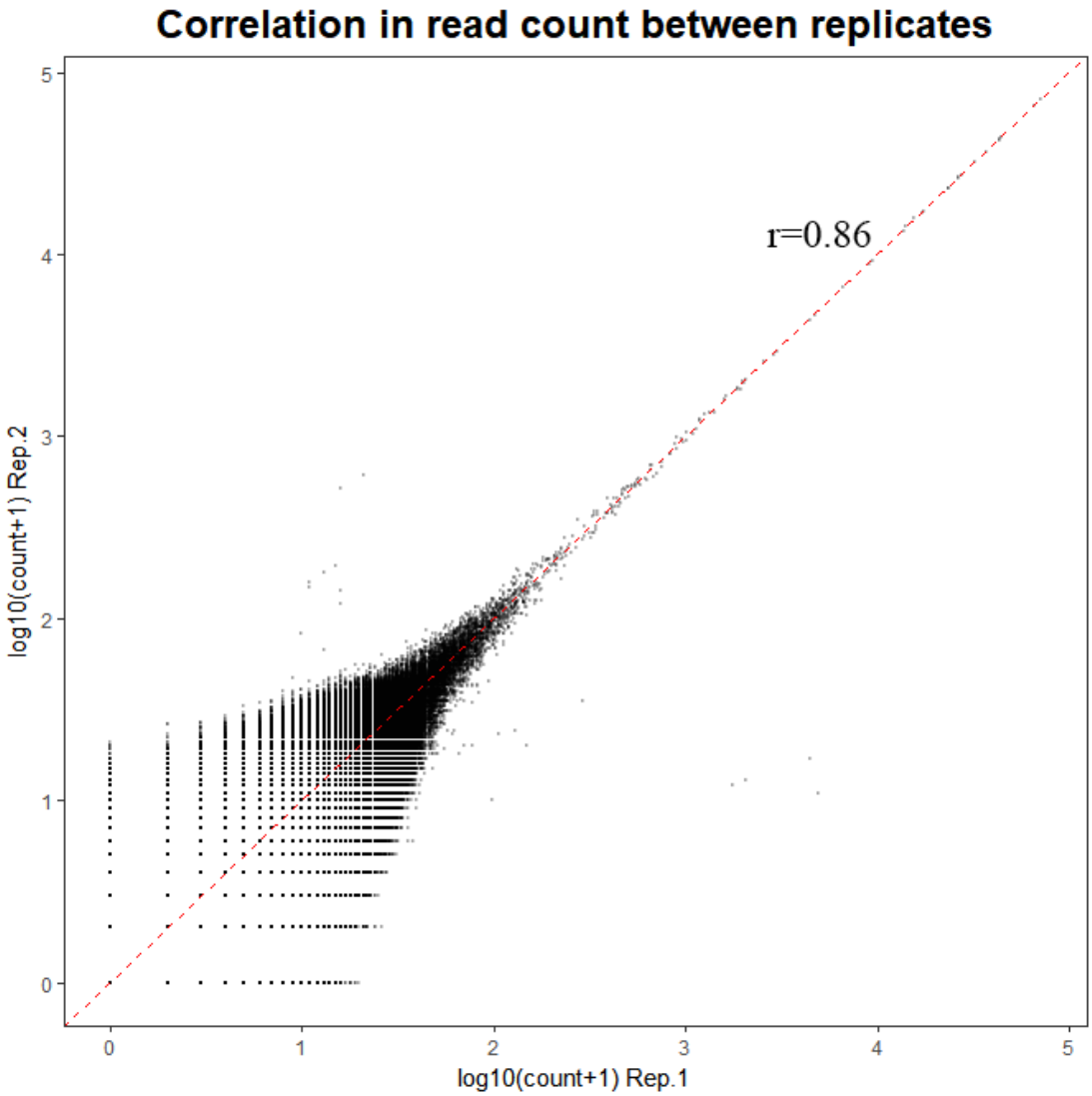


Figure S14. Scatter plots showing the correlation in read counts of peaks between biological replicates (Related to Figure 5).

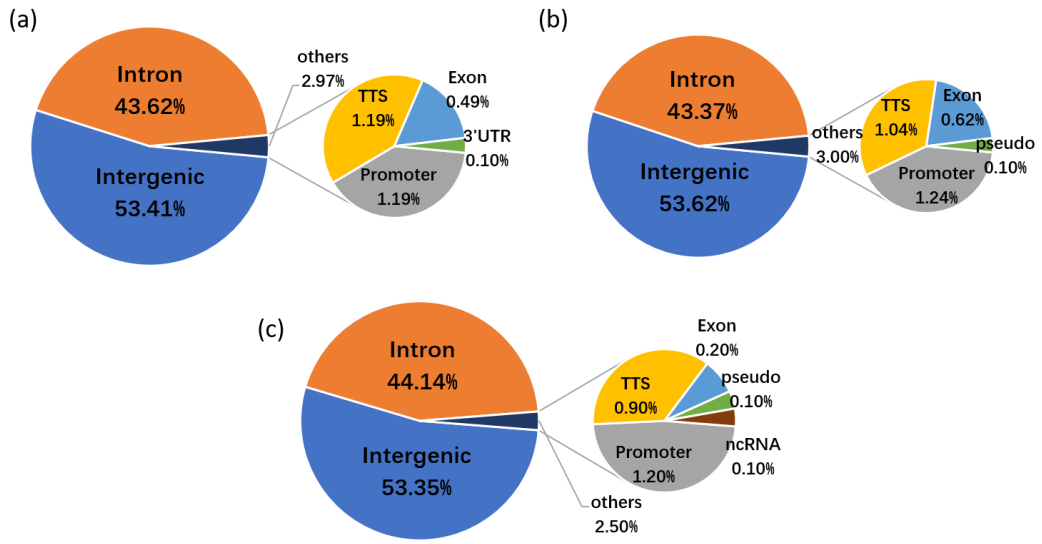


Figure S15. Overall distribution of four sets of filtered 5fU peaks according to merging results in human genome, combined two by two within two pull-down data (P1, P2) and two input data (I1, I2): (a) P1-I1, (b) P2-I1, (c) P2-I2 (Related to Figure 5c).

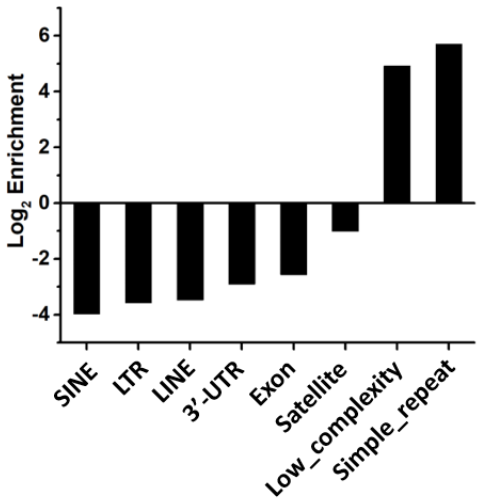


Figure S16. Relative enrichment of 5fU peaks in different genomic elements (Related to Figure 5c).

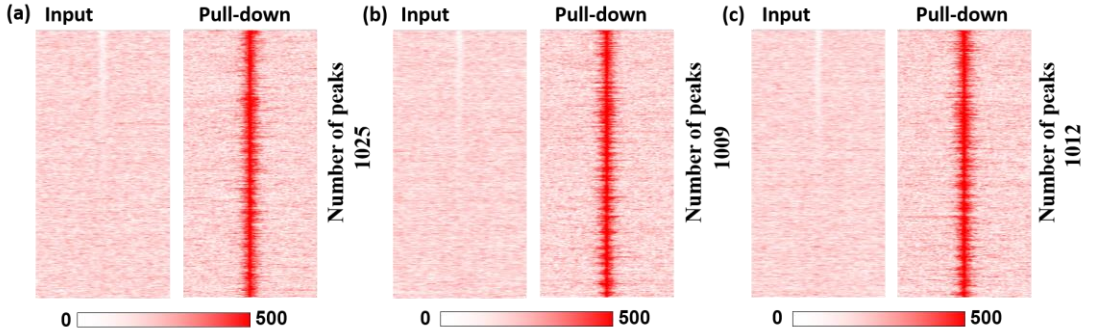


Figure S17. Heatmap shows 5fU-normalized read densities (reads/million/base) across human genome. signals ranked by RPKM in default chromosome sort order, heatmap scales correspond to normalized read densities. (a) P1-I1, (b) P1-I2, (c) P2-I2 (Related to Figure 6b).

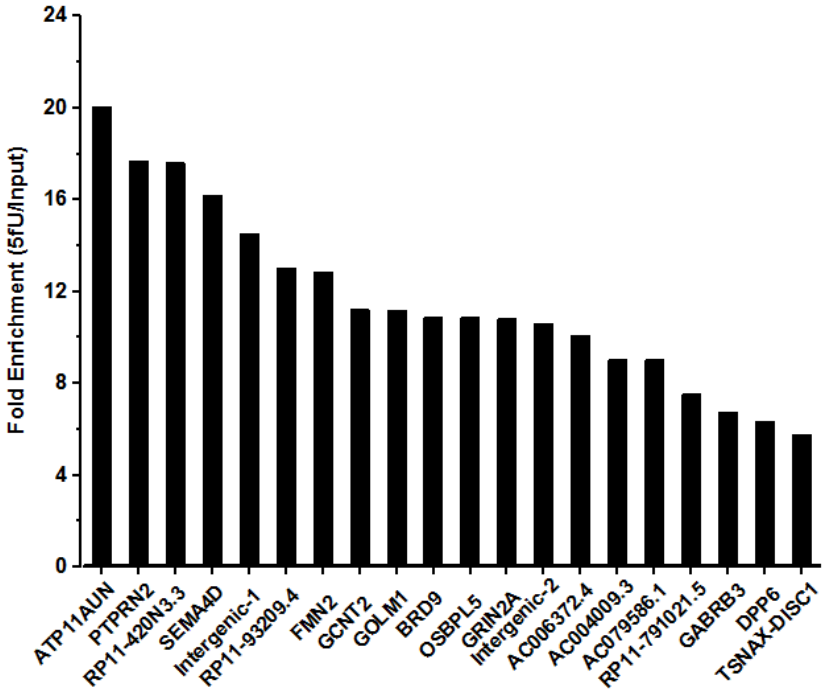


Figure S18. Verification 5fU-enriched regions in human thyroid carcinoma tissues by qPCR (Related to Figure 6a&b).

Table S1. Models of oligonucleotides sequences (Related to Figure 2).

Note: FP is short for forward primer, RP is short for reverse primer.

Oligomer	**Sequence (from 5'to 3')**
ODN-T	GACTCAA**T**AGCCGTA
ODN-AP	GACTCAA**AP**AGCCGTA
ODN-5fU	GACTCAA**5fU**AGCCGTA
ODN-5hmU	GACTCAA**5hmU**AGCCGTA
ODN-5hmC	GACTCAA**5hmC**AGCCGTA
ODN-5fC	GACTCAA**5fC**AGCCGTA
ODN2-5fU	CATAG**5fU**GCTCAAGAGAAATCTCGATGG
80bp ds ODN-fU（Template 1, FP 1, RP 1）	a) TCCTCGGCGGTGTTGCTCTCTGTTGTGCCTCCGCCCG**5fU**CAGG CAG**5fU**GGGCAGGACAAGGACGCAGAGCCACAGCCAAGAA b) TTCTTGGCTGTGGCTCTGCGTCCTTGTCCTGCCCAC**5fU**GCC**5fU** GACGGGCGGAGGCACAACAGAGAGCAACACCGCCGAGGA
80bp ds ODN-fC（Template 2, FP 2, RP 2）	a) CCTATCATCTTATATCTACTACTACTACCTTTAA**5fC**TAAGA TT**5fC**ATGTATAGAATAGATTTAGAGGATTTAGTAGATTTAG b) CTAAATCTACTAAATCCTCTAAATCTATTCTATA**5fC**ATGAAT**5fC** TTAGTTAAAGGTAGTAGTAGTAGATATAAGATGATAGG
80bp ds ODN-T（Template 3, FP 3, RP 3）	a) GCTCGCTTTGTTGGTTTCCTTGTTCTCTGTGCCCACTGCCTG ACGGGCGGAAAGCAGCGCGAGCAAGCGAGACAGGACAC b) GTGTCCTGTCTCGCTTGCTCGCGCTGCTTTCCGCCCGTCAG GCAGTGGGCACAGAGAACAAGG AAACCAACAAAGCGAGC
80bp-5fC（Template 4, FP 4, RP 4）	a) TCCTCCTACATCATTCCTCTCTAACCCCTTATATGTA5fCTTAGA AT5fCAATTGAGTGATTGAAGGTAGTTAGTGGTGGTAGA b) TCTACCACCACTAACTACCTTCAATCACTCAATTGATT5fCTA AGTA5fC ATATAAGGGGTTAGAGAGGAATGATGTAGGAGGA
Template 1	TTCTTGGCTGTGGCTCTGCGTCCTTGTCCTGCCCACTGCCTGAC GGGCGGAGGCACAACAGAGAGCAACACCGCCGAGGA
Template 2	CTAAATCTACTAAATCCTCTAAATCTATTCTATACATGAATCTTAGT TAAAGGTAGTAGTAGTAGATATAAGATGATAGG
Template 3	GCTCGCTTTGTTGGTTTCCTTGTTCTCTGTGCCCACTGCCTGACG GGCGGAAAGCAGCGCGAGCAAGCGAGACAGGACAC
Template 4	TCCTCCTACATCATTCCTCTCTAACCCCTTATATGTACTTAGAATCA ATTGAGTGATTGAAGGTAGTTAGTGGTGGTAGA
Forward primer 1	TTCTTGGCTGTGGCTCTGCGTCCTTGTCCT
Reverse primer 1	TCCTCGGCGGTGTTGCTCTCTGTTGTGCCT
Forward primer 2	CTAAATCTACTAAATCCTCTAAATCTATTC
Reverse primer 2	CCTATCATCTTATATCTACTACTACTACCT
Forward primer 3	GCTCGCTTTGTTGGTTTCCTTGTTCTCTGT
Reverse primer 3	GTGTCCTGTCTCGCTTGCTCGCGCTGCTTT
Forward primer 4	TCCTCCTACATCATTCCTCTCTAACCCCT
Reverse primer 4	TCTACCACCACTAACTACCTTCAATCACTC

Table S2. Primers for pull down-qPCR of mouse hippocampus tissues (Related to Figure 4a&b).

Chr	Peak ID	Gene Name	PCR-Primer(5'–3')	
Chr1	Chr1-41	*4931408C20Rik*	F	AACACCAGACCCCAATAGCAAC
			R	GGATAGGATACTTCAAGCAGCAGA
Chr1	Chr1-42	*Zap70*	F	TGGGGGAGTAGGGATGGG
			R	TGTGTGCGTCTGTCTGTG
Chr2	Chr2-39	*Defb28*	F	TTAAGTTCTCAGGACATTCAGGTCA
			R	TTTAGGCAGAGGAGTTCAGAAAGG
Chr3	Chr3-46	*Fam160a1*	F	AGAGCCAAACACTTCCCATAAAT
			R	ACGTGTTAAAGTGGAGCATTGTG
Chr3	Chr3-142	*Pdlim5*	F	CAGGCCTATGTAGCTTTTGTTCA
			R	TAGTGACCATGACTGGTTTGGA
Chr5	Chr5-106	*Tcf23*	F	GCTCTCAGCAAGAGACTGTCTAAAA
			R	AAGTGTCTTTCTTCAGGGTGCTG
Chr6	Chr6-42	*Ccdc129*	F	GGGCTAAGAATCCGCAGGTG
			R	CCTGTCTATCGCTTCTCTGACC
Chr7	Chr7-17	*Siglecg*	F	AAAAAGAAGCACCACAGTGAGTCAA
			R	TGTGTGTCTGTCTTCAAGTTTCCC
Chr8	Chr8-23	*Inpp4b*	F	CAATCCACCCACCCACTCAA
			R	GGTGTGAGGGTGTCTTTGGT
Chr9	Chr9-55	*Spata19*	F	CAAGCACAAAGATACATACCACAC
			R	GTGCCTCTGTATGTGCCTC
Chr11	Chr11-43	*Sp6*	F	GGTGAGAACAGACTCCTGGG
			R	CCAGTCTTTTCCCAAAAGCTCC
Chr11	Chr11-203	*Aatk*	F	ACATACCAGCCTGAATAACCCTG
			R	GCAGACACGGATGTGAAAAACC
Chr12	Chr12-91	*Hif1a*	F	CATGTGTACGTGTTGCAGGAAATA
			R	CTCGCAGGAGACTACCGCAT
Chr13	Chr13-20	*Tmem267*	F	AAGCCGGATGAAGGCAGTT
			R	GCACTCAGGGAACAATCACGA
Chr14	Chr14-17	*Tnfrsf19*	F	TGAAGTTCAGTGTAAAGATGTGTGT
			R	ACAAACACACTCAAAGACATGC
Chr14	Chr14-71	*Ltb4r1*	F	AAAAAGAAGCACCACAGTGAGTCAA
			R	TGTGTGTCTGTCTTCAAGTTTCCC
Chr16	Chr16-17	*Kcne2*	F	AGCACAGGCACAGACACA
			R	GTGAATGTACGTCAGTTGGTGT
Chr17	Chr17-4	*Smok2b*	F	AAGTAGCAAAGTACAAGTCGTCCTC
			R	TGTAGTGTCTGTGTGTGAGATGCAG
Chr17	Chr17-60	*H2-T3*	F	TACAAGCACAGGCACAGAC
			R	TGAATGTTCTTGAGTGAGTTGTGA
Chr19	Chr19-26	*Ifit3b*	F	AAACTGCTCACTGGATGCCC
			R	GTGACTCCTCTAGTGCTAGCTT

Table S3 Primers of reference gene of mouse hippocampus tissues (Related to Figure 4a&b).

Chr13	Start position: 106872886	End position: 106873053	F	CAAGTGAAAGTTAGTTTGAAGGGTA
			R	TTACTCCACACAGAACTCCAGG

Table S4. Primers for pull down-qPCR of human thyroid carcinoma tissues (Related to Figure 6a&b).

Chr	Peak ID	Gene Name	PCR-Primer(5'–3')	
Chr13	Chr13-12	*ATP11AUN*	F	ACTCTGACAGCTGAGTAAAGCAAGG
			R	CCTGCATTCACTCACACGCTTC
Chr7	Chr7-3	*PTPRN2*	F	CCTCTTACTCTGCCTGTGGTGG
			R	TCCTGAAGCTCTCATCTGTCCCT
Chr16	Chr16-38	*RP11-420N3.3*	F	GGAGTACGGTCGGGTGG
			R	ACAATCCTAACCAACTACGCTT
Chr9	Chr9-10	*SEMA4D*	F	TGTGCGGTATACGTCATGCT
			R	CCAACACACACCATCCCAGT
Chr15	Chr15-3	*RP11-93209.4*	F	TGTGTGTGGCGTGGACTG
			R	TGCTCCCTTCCCCCTCC
Chr1	Chr1-26	*FMN2*	F	TGGTACATCAATTTCTCTCCGCT
			R	GGAGAAAGTAAAGGGAACATGAAAA
Chr6	Chr6-9	*GCNT2*	F	GATCACTTAGAGCTGGAGATGAGG
			R	CAAGCATATTGGGACACCACAC
Chr9	Chr9-6	*GOLM1*	F	GAGGGCTAATCAGCAACCTCA
			R	CATGGTGTAGTGTCGTGTAGA
Chr5	Chr5-21	*BRD9*	F	GAGAATGTGGCAATTCCAGGGC
			R	GCTTGCACTGTAGTTCTTTTGGA
Chr11	Chr11-41	*OSBPL5*	F	TCTTAGTCTGTGAGTGCCACCAT
			R	GGAGCGGGAGACTGGTGTA
Chr16	Chr16-50	*GRIN2A*	F	CCACGTCATGCACATCCAAA
			R	TGAGTCCAGGGTTGTGAGTG
Chr7	Chr7-42	*AC006372.4*	F	GCGTCGTGTGTCGTGTCA
			R	ACTTCCTACACATAACACACCAGAT
Chr7	Chr7-24	*AC004009.3*	F	ACCCACCTCACACACTCCTAT
			R	TTCTGCCCACAACAACTTGC
Chr2	Chr2-69	*AC079586.1*	F	TCATCCACCAAGGATCACTCAC
			R	TGGATGACTGGGCGTCT
Chr9	Chr9-7	*RP11-791021.5*	F	CCCACAACCACCCAACAACC
			R	TGTGTGGCATGTGTTGTGTCA
Chr15	Chr15-22	*GABRB3*	F	GTAGCGTGTGGCGGGACTA
			R	ACATAGGGACCTGAGTAGGGG
Chr7	Chr7-25	*DPP6*	F	ACTTCCAACACTACAACCTTCA
			R	TGGTGGGGTGCTGGTG
Chr1	Chr1-28	*TSNAX-DISC1*	F	GTAATGGTAGAAGCAGTGGGTAGA
			R	AACAACACTCCTACACTCTTCCTG
Chr9	Chr9-12	*Intergenic-1*	F	GGGTTGTTCAGGCATCACAC
			R	GTCTTCCGGTCCAGACAACA
Chr3	Chr3-8	*Intergenic-2*	F	TGGGGGTGTGGTATGTAAAGA
			R	ATCACACATGTGCTCTCACCA

Table S5. Primers of reference gene of human thyroid carcinoma tissues (Related to Figure 6a&b).

Chr2	Start position: 32946833	End position: 32947116	F	ATTAGGATCTGCCGCCTGAC
			R	CCGGCGGCGAAGTCTTTTA

Transparent methods

1. Materials

All chemicals were purchased from Adamas-beta® (Shanghai, China) and Shanghai Shaoyuan Co. Ltd. (Shanghai, China) unless stated otherwise. The nucleic acid stains (Super GelRed, NO.: S-2001) were bought from US Everbright Inc. (Suzhou, China). 2× Hieff™ PCR SYBR® Green Master Mix were ordered from YEASEN (Shanghai, China). ^{1}H NMR, ^{13}C NMR spectra were acquired with Varian Mercury 400 spectrometers. HRMS was recorded on Thermo Scientific™ Dionex Ultimate 3000 hybrid LTQ Orbitrap Elite Velos Pro (Thermo Scientific, USA). DNA MALDI-TOF Mass Spectra were collected on MALDI-TOF-MS (Shimadzu, Japan). Degradase Plus and enzyme reaction buffer were purchased from Zymo Research (Zymo Research, USA). EasyPure® PCR Purification Kit was purchased from TransGen Biotech (Beijing, China). Gel Imaging was collected in Pharos FX Molecular imager (Bio-Rad, USA). TLC plates were monitored with portable UV-LAMP (GL-9406, Jiangsu, China). LC-MS data were collected with the Agilent™ 1220 Infinity LC combined with the 6120 Single Quadrupole mass spectrometer (Agilent Technologies). pH was measured with Mettler Toledo, FE20-Five Easy™ pH (Mettler Toledo, Switzerland). The mouse hippocampus tissues were approved by the Institutional Animal Care and Use Committee of Wuhan University. The human thyroid carcinoma tissues were approved by the ethics committee of Hubei Cancer Hospital (Wuhan, China).

2. Experimental section

5fU modification oligonucleotides synthesis and model DNA preparation. Shorter oligonucleotides containing 5fU were synthesized using Dr. Oligo 192 DNA/RNA synthesizer (provided by GeneCreate Co., Ltd. Wuhan, China). The modified nucleotide was incorporated at the designed sites with the synthesized phosphoramidites using our previous report. Purified oligonucleotides were characterized by mass spectra. 80-bp dsDNAs (containing 5fC or 5fU sites) were obtained by incorporation dfUTP or dfCTP during the process of PCR amplification. Shorter oligonucleotides containing 5fC were purchased from Takara Biotechnology (Dalian, China). General oligonucleotides and primers were synthesized and purified from GeneCreate Co., Ltd. Wuhan, China. dfCTP and dfUTP were purchased from Trilink Biotechnologies.

Genomic DNA extraction. The adult mice were bought from Hubei Research Center of Laboratory Animals. The tissues were picked out under the image of Nikon SMZ1500 Microscope followed by washing with 1× PBS three times. The human thyroid carcinoma tissues were collected from Hubei Cancer Hospital. Genomic DNAs were extracted and purified by DNeasy® Blood & Tissue Kit (Qiagen, Germany) according to the manufacturer's instructions.

5fU labeling and click chemistry. Generally, 5fU labeling reaction can be divided into two steps. Firstly, ODNs containing 5fU were performed in 100 mM NaOAc buffer (pH 5.0) with 12.5 mM azi-BIAN (self-synthesized) at 37℃ for 6 hr in a 1.5 mL tube in a thermo-shaker (Ningbo Biocotek Scientific Instrument Co., Ltd., China, 1500 r.p.m.). After purification with the mini quick spin oligo column (Roche), DBCO-S-S-PEG3-biotin (Click Chemistry Tools, A112-10) was added into the system for click reaction to a final concentration of 20 mM and incubated at 37℃ for 2 hr in a thermo-shaker (1500 r.p.m.). The excess compounds were removed by the mini quick spin oligo

column (Roche). The purification DNA was characterized with RP-HPLC chromatography (Shimadzu LC-6AD) at 260 nm. Column: Inertsil ODS-SP column (5 μm, 250 mm ×4.6 mm) (GL Science Inc., Japan); Eluent: mobile phase A (100 mM TEAA buffer, pH 7.0) and B (acetonitrile); Concentration of B: 5%–5%–35%/0–5–35 min; Flow rate: 1.0 mL·min^{-1}.

Enzymatic of labeled 5fU. To verify the successful reaction between azi-BIAN and ODN-5fU, the labeled DNA was digested to use LC-MS for testing the product. Typically, DNAs and degradase plus (1 μL, 5U/μL) (Zymo Research) were mixed in 1× degradase plus reaction buffer in a final volume of 25 μL at 37℃ for 2 hr. Then filtered by an ultrafiltration tube (3 kDa cutoff, Amicon, Millipore) to remove the enzymes followed by LC-MS assay.

Biotin labeling of 5fU in genomic DNA samples. Genomic DNA was fragmented by sonication with Covaris sonicator under the condition of 175 W for 7 min (Thermo Fisher) to obtain 250 to 450 bp fragments. Typically, 30 μg fragmented genomic DNAs were added into the mixture of 5 μL NaOAc-HOAc buffer (1 M, pH 5.0) and 5 μL azi-BIAN (100 mM in DMSO) and then added H_2O to get a final volume of 50 μL. After concussion and centrifugation, the mixture was incubated at 37℃ for 10 hr in a thermo-shaker (Ningbo Biocotek Scientific Instrument Co., Ltd., China, 1500 r.p.m.). Then the excess chemicals were removed by the mini quick spin oligo column (Roche). DBCO-S-S-PEG3-biotin (Click Chemistry Tools, A112-10) was added into the purified product at 37℃ for 2 hr for click chemistry to form the biotin labeling of 5fU in genomic DNA. After that the mixture was purified by the mini quick spin oligo column (Roche) for further enrichment.

Enrichment of labeling 5fU-containing DNA or genomic DNA. Dynabeads™ M-280 Streptavidin (Invitrogen) were used to pull down the biotin-labelled DNA as the protocol suggested with minor modifications to the 1 × binding and washing (B&W) buffer (pH 7.5). Briefly, the B&W was added with 0.05% Tween-20. For the releasing biotinylated nucleic acids, 50 mM freshly prepared DTT was added into the beads and incubated at 37℃ for 2 hr. Then the beads were segregated with a magnet to obtain the 5fU containing DNA in suspension. The released DNA solution was then applied to DNA Clean & Concentrator™-5 kit (zymo research, Orange County, California, USA) to remove DTT. From 30 μg genomic DNA, 35 ng pull-down DNA was obtained for library construction.

Selective labeling of 5fU test by quantitative PCR. 80 bp 5fU-DNA, 5fC-DNA and T-DNA were labeled with biotin as described previous. After purification, 1 ng of the labeled DNA was mixed with 10 μg ctDNA for enrichment as the protocol described above. The enriched DNA was dissolved in 25 μL ddH_2O. 3 μL of enriched DNA was added into a mixture of Hieff qPCR SYBR Green Master Mix (5 μL) (YEASEN), forward primer (1 μM), reverse primer (1 μM) to give a final volume of 10 μL. Each sample test was repeated three times independent. The mixture was subject to qPCR according to the protocol by the manufacturer's instructions. DNA concentration was quantified by comparison with calibration lines of known concentration of input ODNs.

Dot-blot assay. Synthesized model DNA or genomic DNA was treated using biotin-labeling protocol as described above. For the dot-blot assay, different DNAs were spotted on Amersham Hybond-N+ membrane (GE Healthcare). After dried, the membrane was UV-crosslinked with 254 nm at RT for 5 min twice and then washed with 1× TBST twice. Then the membrane was blocked with 5% BSA at 37℃ for 1 hr and washed with 1× TBST five times. After incubation with streptavidin-HRP (1:1500)

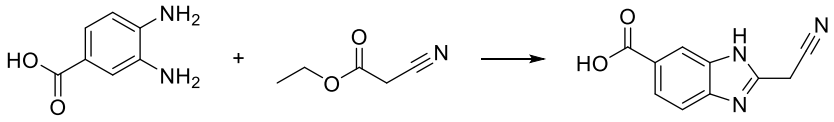
(Thermo Scientific) at 37℃ for 1 hr and washed with 1× TBST four times, the results were visualized by enhanced chemiluminescence (SuperSignal™ West Pico Chemiluminescent Substrate, Cat: 34077, Thermo Scientific) using Molecular Imager® ChemiDocTM XRS+ Imaging System (Bio-Rad). Finally, the membrane was soaked in methylene blue (in NaOAc buffer) to verify the existence of DNA of every dot.

Library preparation and next-generation sequencing of labeled 5fU-enriched DNA samples. The fU-Seq enriched genomic DNAs were quantified using a Qubit Fluorometer (Thermo). Then the DNAs were used directly for library preparation with a Thruplex DNA-Seq kit (Rubicon Genomics) according to the manufacturer's instructions. AMPure XP beads (Beckman) were used for library purification. The purified libraries were subjected into NGS using Hiseq PE150. A pair-end sequencing mode was suggested for maximal data collection. Each biological sample was prepared replicates in parallel, two non-labeled input DNAs (input: pre-fU-Seq), two enriched by pull-down output samples (output: fU-Seq enriched genomic DNA) were sequenced according to the same procedure.

Sequencing data processing and analysis. Raw data were first analyzed with FastQC (Version 0.11.5, Babraham Bioinformatics) to check out the overall sequencing quality, followed with trimming for residual adapter sequence at 3' end and bases whose sequencing quality score were lower than 28 using cutadapt. Processed reads were then mapped to the mouse genome (Mus musculus, GRCm38.p5.genome.fa, downloaded from GENCODE database) by Bowtie2 (version 1.2.1.1). Mapping results can be obtained through the options -N 1 -L 20 in both paired-end and single-end modes. The 5fU-enriched regions in each output file were detected using HOMER (v4.9) findPeaks algorithm, and a total 42954 peaks were found. After screening all peaks with standards of fold change of pull-down vs control > 4 and p-value < 10^{-5}, there remains 39829 peaks. Peak annotation analysis was done using HOMER annotatePeaks algorithm while reads visualization was done with Integrative Genomics Viewer (IGV) under the help of SAMtools (Li et al., 2009) and bedtools (Quinlan and Hall, 2010). To plot the distribution of fU-Seq signals around 5fU sites more accurately, we ignored the sites located within anshul.blacklist.mm10.bed and psublacklist.mm10.bed (downloaded from ENCODE database). To study the relationship between 5fU sites and histone modification, annotatePeaks algorithm of HOMER was used with options -size 4000 -hist 10.

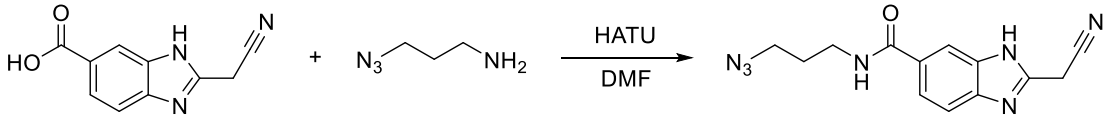
3. Synthesis

Scheme S1: Synthesis of azide modified (2-benzimidazolyl)acetonitrile derivative (azi-BIAN)



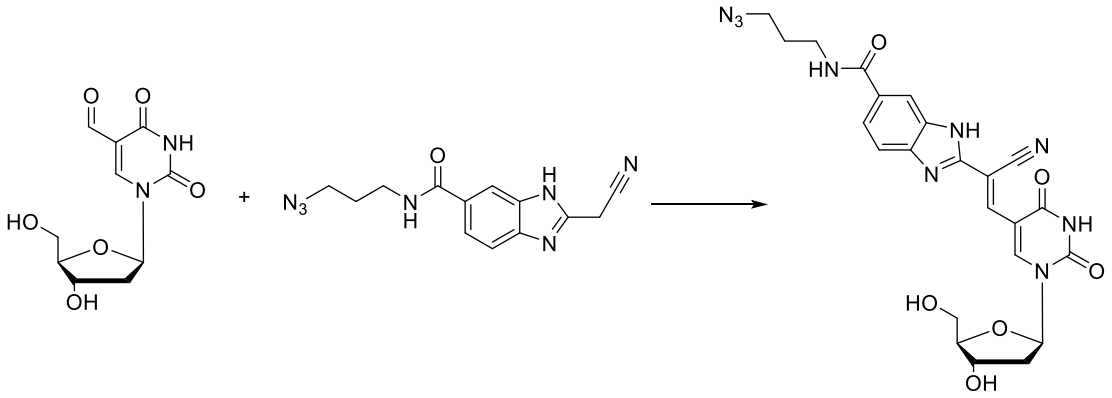
3,4-diaminobenzoic acid (3 g, 19.7 mmol) and ethyl cyanoacetate (7 mL, 65.8 mmol) were dispersed in 100 mL round bottom flask. Then, the reaction mixture was kept stirring and stayed at 180°C. After 1 hr, the reaction mixture was poured into 100 mL diethyl ether, the precipitate was filtered off and purified by silica gel chromatography, eluting with 1% methanol and 0.1% acetic acid in

dichloromethane to give BIAN (yield 30%) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 12.92 (s, 1H), 8.15 (s, 1H), 7.83 (d, J = 8.4 Hz, 1H), 7.61 (d, J = 7.7 Hz, 1H), 4.46 (s, 2H). HRMS(ESI+) $C_{10}H_8N_3O_2^+$ $[M+H]^+$ calculated 202.06110, found 202.06082. This result is in reasonable agreement with the precious report of Refaat, H.M (Refaat, 2010).



BIAN (290 mg, 1.4 mmol), 3-azidopropan-1-amine (720 mg, 7.2 mmol)(Schatz et al., 2009) and HATU (1.1 g, 2.9 mmol) were dissolved into 10 mL DMF which containing 5 drops of TEA in 25 mL round bottom flask. Then, the reaction mixture was kept stirring and stayed at 25°C. After 4 hr, the reaction mixture was evaporated under vacuum and purified by silica gel chromatography, eluting with dichloromethane: methanol from 100:1 to 40:1 to give azi-BIAN (yield 80%) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 12.86 (s, 1H), 8.53 (t, J = 5.5 Hz, 1H), 8.08 (s, 1H), 7.74 (dd, J = 8.4, 1.3 Hz, 1H), 7.59 (d, J = 8.2 Hz, 1H), 4.44 (s, 2H), 3.43 (t, J = 6.8 Hz, 2H), 3.34 (m, 2H), 1.80 (p, J = 6.8 Hz, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 167.12, 147.33, 129.10, 122.12, 116.96, 49.05, 37.19, 28.96, 18.95. HRMS(ESI+) $C_{13}H_{14}N_7O^+$ $[M+H]^+$ calculated 284.12543, found 284.12497.

Scheme S2: Synthesis of 5-formyl-2'-deoxyuridine and azi-BIAN adduct (azi-biaU)



5-formyl-2'-deoxyuridine (30 mg, 0.12 mmol) and azi-BIAN (33 mg, 0.12 mmol) were dissolved into 10 mL methanol which containing 5 drops of acetic acid in 25 mL round bottom flask. Then, the reaction mixture was kept stirring and stayed at 50°C. After 15 hr, the reaction mixture was evaporated under vacuum and purified by silica gel chromatography, eluting with dichloromethane: methanol from 50:1 to 15:1 to give the product azi-biaU (yield 80%) as a yellow solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 13.39 (s, 1H), 11.98 (s, 1H), 8.95 (s, 1H), 8.57 (s, 1H), 8.36 – 7.90 (m, 2H), 7.67 (dd, J = 78.5, 12.8 Hz, 2H), 6.20 (t, J = 6.6 Hz, 1H), 5.35 (s, 1H), 4.99 (s, 1H), 4.35 – 4.25 (m, 1H), 3.90 (dd, J = 6.8, 4.1 Hz, 1H), 3.63 (d, J = 4.0 Hz, 2H), 3.52 – 3.40 (m, 4H), 2.37 – 2.11 (m, 2H), 1.90 – 1.72 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 167.04, 162.12, 149.76, 143.38, 142.24, 138.00, 129.33, 129.22, 123.61, 122.02, 118.71, 116.68, 111.55, 107.86, 100.39, 88.68, 86.20, 71.21, 61.86, 49.06, 40.78, 37.22, 28.94.

HRMS(ESI+) $C_{23}H_{24}N_9O_6^+$ $[M+H]^+$ calculated 522.18441, found 522.18419.

Supplemental References

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Refaat, H.M. (2010). Synthesis and anticancer activity of some novel 2-substituted benzimidazole derivatives. Eur J Med Chem *45*, 2949-2956.

Schatz, C., Louguet, S., Le Meins, J.F., and Lecommandoux, S. (2009). Polysaccharide‐block‐polypeptide Copolymer Vesicles: Towards Synthetic Viral Capsids. Angew Chem Int Ed *48*, 2572-2575.