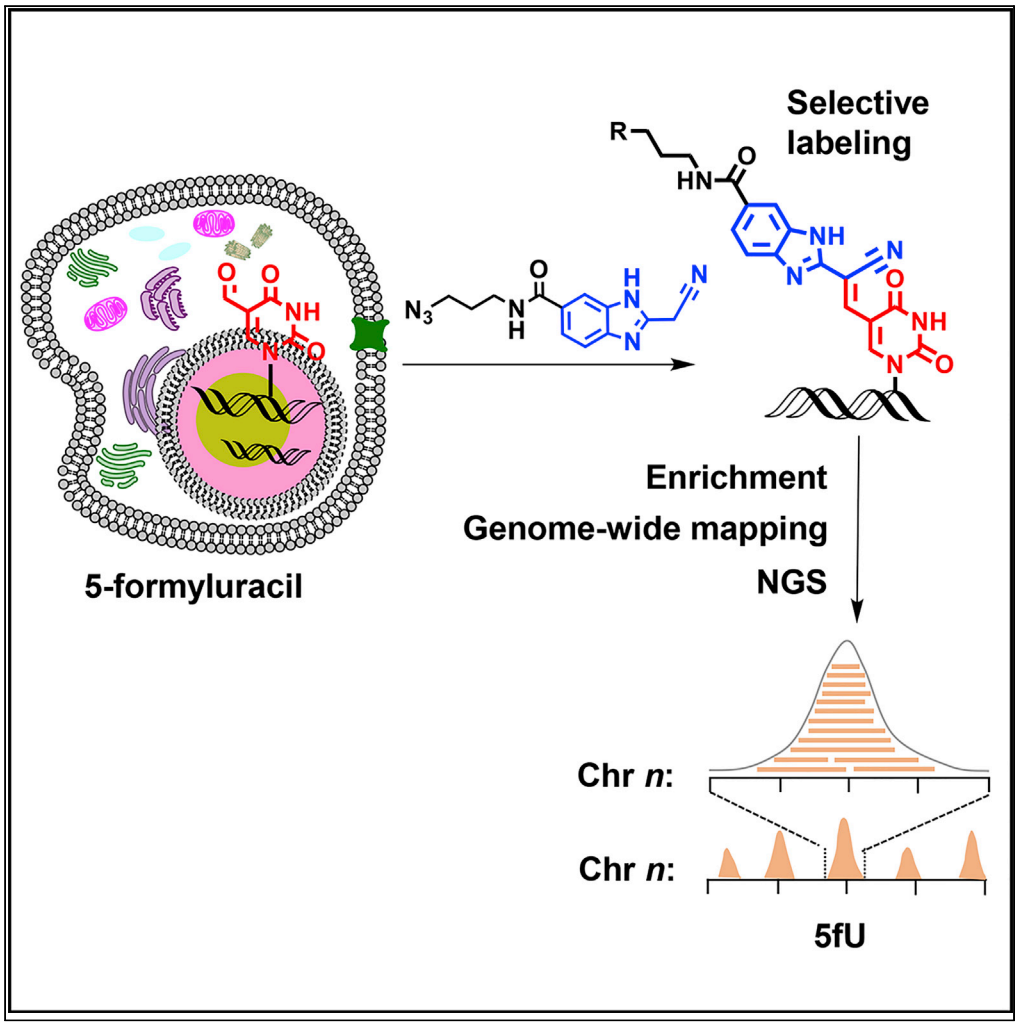
iScience

CellPress

Article

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



Yafen Wang, Chaoxing Liu, Fan Wu, ..., Xiaocheng Weng, Zhiguo Wu, Xiang Zhou

xzhou@whu.edu.cn

HIGHLIGHTS

The derivative of (2-benzimidazolyl) acetonitrile (azi-BIAN) can selectivity label 5fU

Azi-BIAN can selectively label and pull down 5fU in the genome for NGS

The first genome-wide map of 5fU in mammalian genomic DNA

5fU is highly enriched at intergenic regions and introns

DATA AND SOFTWARE AVAILABILITY

GSE115918

Wang et al., iScience 9, 423–432
November 30, 2018
https://doi.org/10.1016/j.isci.2018.10.024

iScience

CellPress

Article

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

Yafen Wang,[1,4] Chaoxing Liu,[1,4] Fan Wu,[1,4] Xiong Zhang,[1] Sheng Liu,[2] Zonggui Chen,[3] Weiwu Zeng,[1] Wei Yang,[1] Xiaolian Zhang,[2] Yu Zhou,[3] Xiaocheng Weng,[1] Zhiguo Wu,[3] and Xiang Zhou[1,5,*]

SUMMARY

Chemical modifications to nucleobases have a great influence on various cellular processes, by making gene regulation more complex, thus indicating their profound impact on aspects of heredity, growth, and disease. Here, we provide the first genome-wide map of 5-formyluracil (5fU) in living tissues and evaluate the potential roles for 5fU in genomics. We show that an azido derivative of (2-benzimidazolyl)acetonitrile has high selectivity for enriching 5fU-containing genomic DNA. The results have demonstrated the feasibility of using this method to determine the genome-wide distribution of 5fU. Intriguingly, most 5fU sites were found in intergenic regions and introns. Also, distribution of 5fU in human thyroid carcinoma tissues is positively correlated with binding sites of POLR2A protein, which indicates that 5fU may distributed around POLR2A-binding sites.

INTRODUCTION

Chemical modifications to nucleobases play important roles in mediating fundamental biological processes (Booth et al., 2015; Shu et al., 2018; Suzuki and Bird, 2008; Wu and Zhang, 2017) and are regarded as hallmarks of many diseases (Chen et al., 2017; Jackson and Bartek, 2009; Johnson et al., 2017). Therefore, a detailed analysis of natural nucleobase modifications is essential for a complete understanding of genetic and epigenetic regulation (Hong et al., 2018; Iwan et al., 2017; Liu et al., 2016; Shen et al., 2014). However, a global mapping of the modified nucleobases in the genome is often missing because of the low abundance of these modifications and lack of sensitive, selective, and genome-wide detection methods (Wyrick and Roberts, 2015).

5-Formyluracil (5fU), which is present in many cells and tissues (Hong and Wang, 2007; Pfaffeneder et al., 2014), can be generated by exposure to UV light (Decarroz et al., 1986), ionizing radiation (Hong and Wang, 2007; Kasai et al., 1990), Fenton-type reagents, reactive oxygen species attack (Hong et al., 2006), or enzyme oxidation (Pais et al., 2015). It has been reported that 5fU modification can introduce gene mispairing (Yoshida et al., 1997), alter DNA structures (Kawasaki et al., 2017b), modulate protein-DNA interactions (Kittaka et al., 2001; Rogstad et al., 2004), and induce perturbations of DNA function (Rogstad et al., 2004). Recently, 5-formylcytosine (5fC), the modified cytosine counterpart of 5fU, has been identified as a vital epigenetic modification involved in gene regulation (Kitsera et al., 2017; Song et al., 2013; Wang et al., 2018b), cell differentiation, and development (Wagner et al., 2015; Zhu et al., 2017). 5fU may be an oxidation product of 5-hydroxymethyluracil (5hmU) *in vivo*, and 5hmU has been identified as not only an oxidized nucleobase but also an essential epigenetic mediator that influences transcription factors, changes the physical properties of local DNA duplex in the genome, and helps in binding of chromatin remodeling proteins (Kawasaki et al., 2017a, 2018; Modrzejewska et al., 2016; Pfaffeneder et al., 2014). Whether 5fU also acts as an epigenetic mediator like 5fC and 5hmU remains an open question. Although many reagents such as aminothiophenol (Hirose et al., 2010), phenylenediamine (Hardisty et al., 2015; Liu et al., 2017b; Wang et al., 2018a), hydrazine (Kawasaki et al., 2017a; Liu et al., 2017a), and indole (Samanta et al., 2015) derivatives have been utilized to selectively label 5fU, genome-wide profiling of 5fU remains a challenge owing to its low abundance in the genome. The development of an efficient, rapid, sensitive, environment-friendly, and catalyst-free method to analyze 5fU is highly desired.

Herein, we present a novel method termed fU-Seq for determining the genome-wide distribution of 5fU in mouse hippocampus and human thyroid carcinoma tissue via an azido-modified reagent that

[1]College of Chemistry and Molecular Sciences, Key Laboratory of Biomedical Polymers of Ministry of Education, The Institute for Advanced Studies, Hubei Province Key Laboratory of Allergy and Immunology, Wuhan University, Wuhan, Hubei 430072, P. R. China

[2]State Key Laboratory of Virology and Hubei Province Key Laboratory of Allergy and Immunology and Department of Immunology, School of Medicine, Wuhan University, Wuhan, Hubei 430072, P. R. China

[3]College of Life Science, Wuhan University, Wuhan, Hubei 430072, P. R. China

[4]These authors contributed equally

[5]Lead Contact

*Correspondence: xzhou@whu.edu.cn

https://doi.org/10.1016/j.isci.2018.10.024

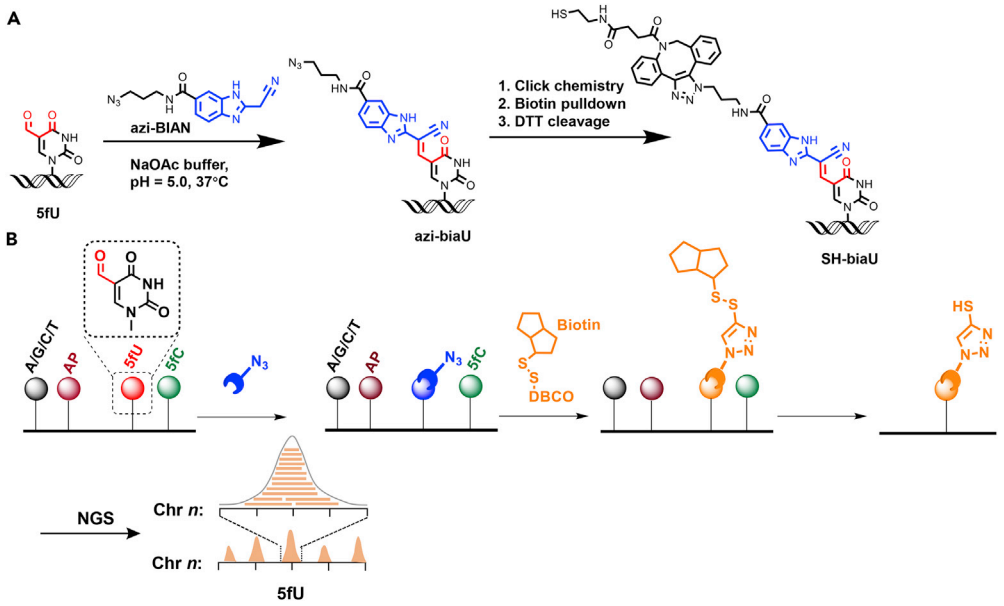


Figure 1. fU-Seq, a Method that Utilizes Selective Tagging of 5fU with an Azido-Modified Reagent

(A) The chemical structures of 5fU, the azido-modified reagent, DBCO derivatives, and their products.

(B) Schematic illustration of fU-Seq. Genomic DNA is fragmented, sequentially labeled with azi-BIAN, conjugated to biotin for pull-down enrichment, and released following enrichment using DTT to define 5fU sites throughout the genome. NGS, next-generation sequencing.

selectively labels 5fU (Figures 1A and 1B). The challenge of selective capturing and profiling 5fU is therefore addressed by employing copper-free click chemistry between azido-modified 5fU-containing genomic DNA and commercial dibenzocyclooctyne (DBCO)-modified biotin (Click Chemistry Tools). After enrichment using streptavidin-coated magnetic beads, dithiothreitol can be used to cleave the biotin linker (Song et al., 2011) to release the pulled down 5fU-containing genomic DNA for further next-generation sequencing (NGS).

RESULTS

Evaluating the Reactivity of azi-BIAN with ODN-fU

First, we screened for chemicals that have the ability to efficiently tag 5fU with high yield and selectivity under warm conditions, when compared with other aldehyde modifications present in DNA. We successfully identified several chemicals that reacted efficiently with 5fU (Liu et al., 2017a, 2017b, 2018). The inherent chemical properties of (2-benzimidazolyl)acetonitrile (azi-BIAN) make it high selectivity for 5fU in both nucleosides and oligonucleotides. More importantly, azi-BIAN could not react with abasic site and 5fC. We thus designed an azido derivative of azi-BIAN for enriching 5fU-containing genomic DNA (Schemes S1 and S2).

Next, we incubated a 15-mer oligodeoxyribonucleotide containing one 5fU site (ODN-5fU) with azi-BIAN in NaOAc buffer (pH 5.0) at 37°C for 6 hr. Complete conversion to the new product ODN-azi-biaU was recorded by reversed-phase (RP)-high-performance liquid chromatography (HPLC) (monitored at 260 nm) (Figure 2C). The integrity of labeled DNAs was confirmed by MALDI-TOF mass spectrometry (MS) (Figures 2D, S2A, and S2B). In addition, enzymatically digested mononucleosides were analyzed through HPLC-MS to ensure that the reaction of 5fU yielded 5-formyl-2′-deoxyuridine and azi-BIAN adduct (azi-biaU) (Figure S3). In control experiments, ODN-T, ODN-5hmU, ODN-5hmC, ODN-5fC, and ODN-AP (in which the 5fU site was replaced with a T, 5hmU, 5hmC, 5fC, and abasic site, respectively) were also selected to react with azi-BIAN under the same conditions. The high selectivity of azi-BIAN for 5fU was verified by RP-HPLC (monitored at 260 nm) (Figure S1) and denaturing polyacrylamide gel electrophoresis analysis (Figures 2A and 2B), indicating that the substitution of 5fU with 5fC or an abasic site containing aldehydes did not disturb selective labeling of 5fU by azi-BIAN. To manifest whether the oxidative damage occurs during

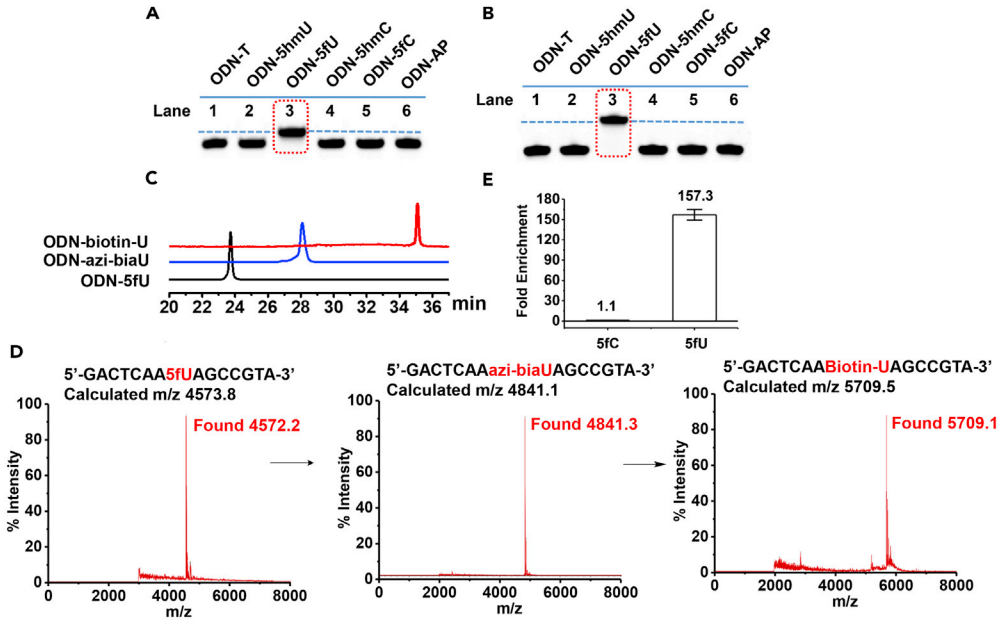


Figure 2. azi-BIAN Selectively Labels and Enriches 5fU

(A) PAGE analysis of ODN-5fU after incubation with azi-BIAN (lane 3) (dashed line) after being stained with nucleic acid stains (fluorescence mode, λ_{ex}: 532 nm) when compared with other control DNAs such as ODN-T (lane 1), ODN-5hmU (lane 2), ODN-5hmC (lane 4), ODN-5fC (lane 5), and ODN-AP (lane 6) under the same conditions.

(B) PAGE analysis of ODN-5fU after incubation with azi-BIAN and reaction with DBCO-biotin (lane 3) (dash line) after being stained with nucleic acid stains (fluorescence mode, λ_{ex}: 532 nm) when compared with other control DNAs such as ODN-T (lane 1), ODN-5hmU (lane 2), ODN-5hmC (lane 4), ODN-5fC (lane 5), and ODN-AP (lane 6) under the same conditions.

(C) RP-HPLC trace (260 nm) of ODN-5fU (black line); ODN-azi-biaU (blue line), which was generated by reaction with azi-BIAN; and ODN-azi-biaU, which was further labeled with DBCO-S-S-PEG3-bitoin (ODN-biotin-U, red line).

(D) MALDI-TOF-spectra of ODN-5fU, ODN-5fU after incubation with azi-BIAN, and ODN-5fU after incubation with azi-BIAN and reaction with DBCO-biotin.

(E) Enrichment tests of a single pool of spike-in amplicons containing 5fU, 5fC, or only canonical nucleobases, using fU-Seq. Values shown are fold enrichment over canonical nucleobases. Data are represented as mean ± SD of biological triplicate.

sample workup, ODN-5fC was incubated with azi-BIAN and then the mixture was subjected to DNA MALDI-TOF MS analysis. No mass spectra of ODN-azi-biaU appeared (Figure S4B). Meanwhile, the model DNA (80 bp double-stranded [ds] ODN-fC) was subjected to extraction by DNeasy Blood & Tissue Kit for simulating the process of genomic DNA extraction. After that, the extracted DNA was digested for liquid chromatography (LC)-MS analysis. No 5fU peaks was found (Figure S5). These results indicated that the oxidative damage did not occur during sample workup.

Enriching 5fU-Containing DNA Fragments

Besides the advantage of 5fU selective labeling, the labeled 5fU containing biotin can be used to enrich DNA fragments bearing 5fU. Because most biological samples bearing 5fU are in ds form, it was also vital to determine whether azi-BIAN selectively labels 5fU in dsDNAs. Therefore, we used a series of 80 bp dsDNAs (containing two 5fC or 5fU sites per strand or only canonical nucleosides) as a test of specificity under conditions described previously (Hardisty et al., 2015) to evaluate the enrichment efficiency. These ODNs were made to react with azi-BIAN followed by biotinylation. fU-DNA was enriched over C-DNA by ~157-fold with azi-BIAN, whereas fC-DNA was enriched over C-DNA by 1.1-fold, based on qPCR quantitation. These results confirmed that our pull-down method fU-Seq has specificity for enriching 5fU-containing DNA (Figure 2E). Because of $NaBH_4$ can reduce 5fU to 5hmU and hydroxylamine ($EtONH_2$) can react with formyl group of 5fU, we applied the 80bp-dsDNA which contain 5fU was reduced by $NaBH_4$ or blocked by $EtONH_2$ as the control experiments to validate the effectively enrichment on model DNA containing 5fU (Figures S6 and S7). Taken together, these experiments demonstrate that covalent chemical labeling coupled with biotin-based affinity purification ensures accurate and comprehensive capture of 5fU-containing DNA fragments.

Genome-wide Mapping of 5fU in Mouse Hippocampal Tissues

Recently, the LC-tandem MS (MS/MS) quantification results indicated that 5fU levels are slightly higher in mouse hippocampus (2×10^{-6} per nucleoside) than in other tissues (Pfaffeneder et al., 2014). Thus, we performed selective labeling of 5fU in genomic DNA isolated from mouse hippocampus. Genomic DNA from mouse hippocampus was sonicated into small fragments (~250–450 base pairs), treated with azi-BIAN to yield azido-modified 5fU-containing genomic DNA, and labeled with DBCO-biotin to install biotin (Figure 1B). Because each step is bio-orthogonal and efficient, this protocol ensures selective labeling of most 5fU sites present in genomic DNA. The presence of the introduced biotin group was confirmed by avidin-horseradish peroxidase tagging and enhanced chemiluminescence visualization to obtain a dot in the dot blot assays (Figure S8), and the HPLC-MS analyses of enzymatically digested mononucleosides from labeled genome DNA proved that the reaction of 5fU yielded the target product (Figure S9).

Pull-down genomic DNA and an input control obtained from the same genomic DNA sample were subjected to high-throughput sequencing. We first removed adapter sequences in sequencing reads with cutadapt and only kept the reads with acceptable sequencing quality using FastQC to obtain clean data (version 0.11.5, Babraham Bioinformatics) (Martin, 2011). Following these steps, Bowtie2 (version 1.2.1.1) (Langmead et al., 2009) was used to map the remaining reads to the reference genome of *Mus musculus* (GRCm38.p5.genome, downloaded from GENCODE) in single-end alignment mode.

To determine the pull-down efficiency, we identified the peaks with read enrichment in pull-down sample relative to the input control using HOMER (v4.9) software (Heinz et al., 2010). Using the findPeaks command with default parameters, 42,954 peaks were found across the genome, of which 39,829 peaks remained after filtering with the following criteria: fold change of pull down versus control > 4 and p value of pull-down versus control < 10^{-5}.

A chromosome-level analysis of 5fU-enriched peaks indicated that the 5fU sites occur in a near-uniform distribution, although their presence was relatively higher in chromosomes 1 (8.28%), 2 (7.26%), and 5 (6.82%) (Figure S10). In consideration of the difference in chromosome size, a rather higher distribution in chrM (peak number/genome size = 1.23×10^{-4}) and an extremely lower distribution in chrY (peak number/genome size = 3.27×10^{-8}) were found (Figure 3A). The fold change versus control for most peaks (75%) was found to fall between 6.46 and 24.47 (Figure 3B). We further examined the distribution of 5fU sites within different genomic element groups and found that 62.12% of the sites occurred in intergenic regions, 36.02% in introns, and 1.87% in other regions, including promoters, transcriptional termination sites (TTSs), and exons (Figure 3C). Enriched peaks were inspected in the Integrative Genomics Viewer (IGV) (Robinson et al., 2011; Thorvaldsdóttir et al., 2013) using the input control and pull-down data as shown in Figure 4A. We also obtained heatmaps of both the input and pull-down data (Figure 4B) with the script annotatePeaks.pl, from which the pull-down efficiency could be calculated. This result was also further confirmed by qPCR (Figure S11). These results indicated that the selective enrichment of 5fU in genome by the fU-Seq strategy is effective.

To speculate the potential genetic significance of 5fU on histone modifications, with the annotatePeaks.pl (-size: 4,000; -hist: 10), several major existing histone modification peak data of brain tissues in adult *M. musculus*, including H3K27ac, H3K27me3, H3K4me1, and H3K4me3 (downloaded from ENCODE database), were compared with 5fU sites. Interestingly, the appearance of 5fU sites in the genome negatively correlated with H3K27ac modification peak but positively correlated with H3K27me3 modification peak (Figures 4C and S12). H3K27me3 is known for preventing transcription. These two histone modifications have been reported to be physiologically antagonistic. When H3K27 is trimethylated, it is tightly associated with inactive gene promoters, whereas acetylation of H3K27 is associated with active transcription (Barski et al., 2007; Ferrari et al., 2014). Thus, it is reasonable to speculate that 5fU sites might also play an inhibitory role in gene transcription. Also, more efforts need to be made for testifying this speculation.

Genome-wide Mapping of 5fU in Human Thyroid Carcinoma Tissues

With the advances of new techniques for whole-genome sequencing, carcinomas are discovered to be associated with modified nucleobases. Recently, researchers found that 5fU levels are about ten

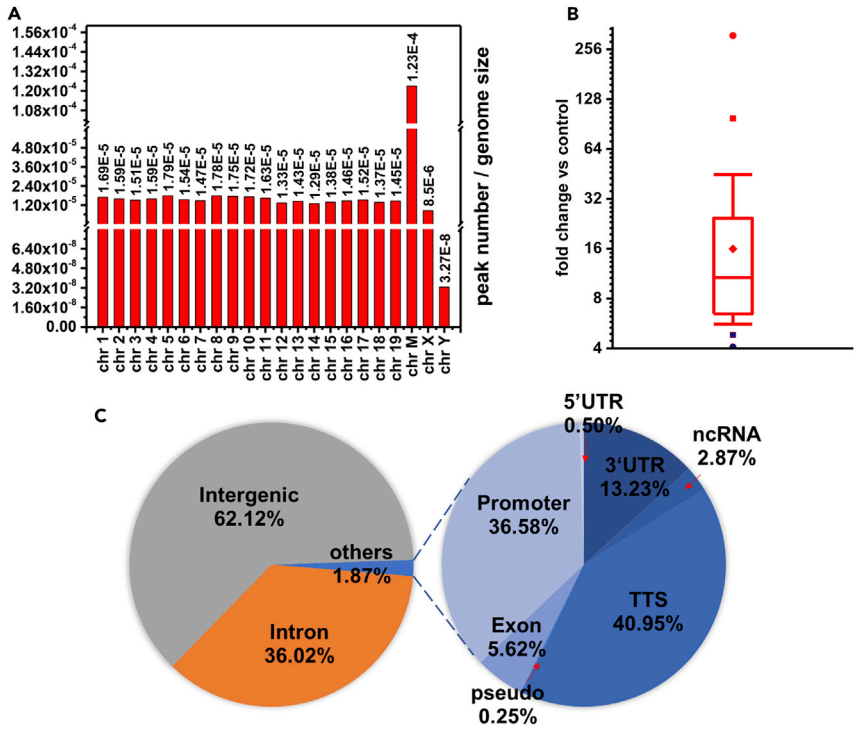


Figure 3. The 5fU Peak Distribution in Mouse Hippocampal Tissues

(A) Normalized to chromosome size at the whole-genome level.

(B) Box chart of 5fU distribution based on fold change versus control. ● is the maximum value, equal to 309.98; ● is minimum value, equal to 4.08; ■ is the 99% value, equal to 97.87; ■ is the 1% value, equal to 4.82; and ◆ is the mean, equal to 15.96; the box stands for interquality (12.5%–87.5%) area, ranging from 6.46 to 24.47; the red line within is the median, equal to 10.71; upper split line is the 95% point, equal to 44.87; and lower split line is the 5% point, equal to 5.61.

(C) Percentage of 5fU peak lengths overlapping with genomic features.

5-formyldeoxyuridine per 10^6 nucleotides in human thyroid carcinoma tissues (Jiang et al., 2017). So exploring the distribution of 5fU in cancer tissues might be vital for understanding the relationship between diseases and 5fU. Encouraged by the results of genome-wide mapping of 5fU in mouse hippocampal tissues, we further analyzed the distribution of 5fU in human thyroid carcinoma tissues. Similarly, the genomic DNA was fragmented to 250–450 bp, labeled with azi-BIAN, and biotinylated for enrichment. The pull-down samples are applied for library construction and NGS.

To verify if the methods of fU-Seq are reproducible, two biological replicates were prepared to validate the results. Input 1 (I1) and input 2 (I2) represented the input groups, and pull down 1 (P1) and pull down 2 (P2) represented the pull-down groups. The NGS reads were aligned to the reference human genome (GRCh38.p7, downloaded from GENCODE). After overlapping the two biological replicates, about 950 peaks were identified (Figures S13 and S14).

We next analyzed the distribution pattern of 5fU peaks along each chromosome (Figure 5A). The results show that the distribution of 5fU in human thyroid carcinoma tissues is uniform in most chromosomes. The fold change of pull-down versus control for most peaks (50%) was found to fall between 6.32 and 15.30 (Figure 5B). To locate 5fU sites within the genome, we examined the distribution of 5fU sites within different gene fragment groups and found that 54.17% of the sites occurred in intergenic segments and 43.02% in introns (Figures 5C and S15). We then examined whether 5fU is enriched at specific types of different genomic elements. We found that 5fU is highly enriched at low_complexity and simple_repeat but is depleted at transposable elements including long terminal repeats (LTRs), short interspersed nuclear elements, and long interspersed nuclear elements (Figure S16).

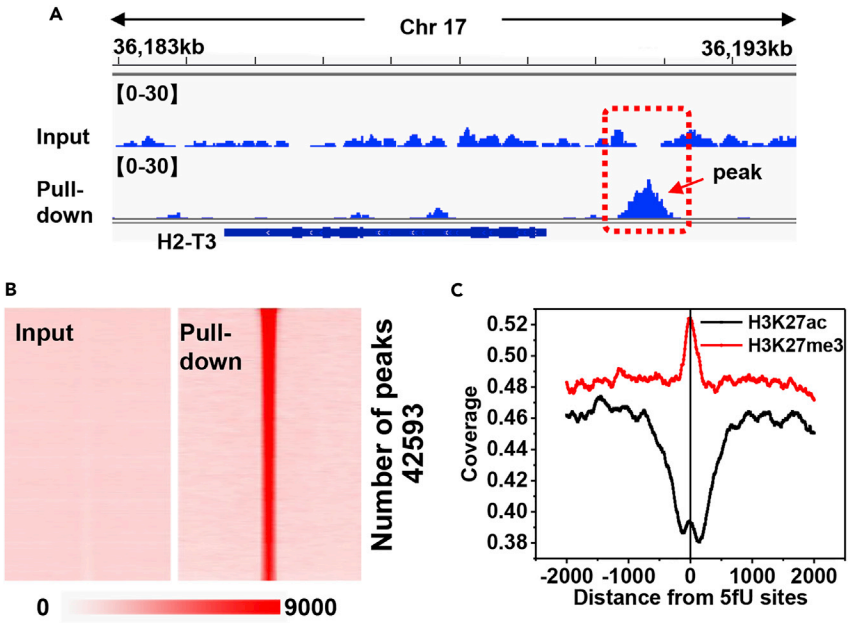


Figure 4. fU-Seq Reveals 5fU Maps in the Whole Genome of Mouse Hippocampal Tissues

(A) Visual representation of the enrichment peak coverage of fU-Seq (below) and the input control (above) are shown.

(B) Heatmap representations of 5fU-normalized read densities (reads/million/base) across the genome. 5fU-containing read signals across the genome ranked by Reads Per Kilobase per Million mapped reads (RPKM) in default chromosome sort order. Heatmap scales correspond to normalized read densities.

(C) Distribution patterns of 5fU with respect to H3K27me3 and H3K27ac modification sites in the cerebellum.

Enriched peaks were visualized in the IGV using the input control and pull-down data as shown in Figure 6A. To determine the pull-down efficiency, the heatmap results were obtained (Figures 6B and S17). We also validated the 5fU-specific enrichments observed in the peaks with qPCR (Figure S18). All these results manifested that the selective enrichment of 5fU in genome by the fU-Seq strategy is effective. We further investigated whether 5fU in human thyroid carcinoma tissues is associated with histone modification or protein-binding sites, which has genetic significance. With the assistance of annotatePeaks.pl (-size: 4,000; -hist: 2), several major existing histone modification peak and protein-binding site data of thyroid gland tissues in adults, including H3K27ac, H3K27me3, H3K4me1, H3K4me3, POLR2A (all downloaded from ENCODE database), were compared with 5fU sites. Surprisingly, the appearance of 5fU sites in the genome negatively correlated with H3K4me1 modification peak, whereas it positively correlated with binding sites of POLR2A protein (Figure 6C). H3K4me1 is enriched at active enhancers and acts as a marker for many cell-type-specific enhancer sites (Creyghton et al., 2010; Rada-Iglesias, 2018; Shen et al., 2016). Hence, it is tempting to speculate that 5fU sites in human thyroid carcinoma tissues are indicative of suppressed enhancer overactivation. Besides, a lot of researchers had reported that POLR2A encodes the largest and catalytic subunit of the RNA polymerase II complex (Bradner, 2015; Liu et al., 2015). Because 5fU sites act as a potential repressor, they may be distributed around POLR2A-binding sites. It is reasonable to guess that the existence of 5fU in human thyroid carcinoma may impede protein binding with DNA and then inhibit the transcription of some specific genes. Further endeavor is needed for studying the causal relationship between 5fU and cancers.

DISCUSSION

The current study developed the first selective and efficient approach (fU-Seq) to label and capture 5fU from mouse and human genomic DNA and investigated the relationship between 5fU and histone modifications. We have demonstrated the feasibility of using this method to determine the genome-wide distribution of 5fU. Intriguingly, most 5fU sites were found in intergenic regions and introns. In addition, the analysis of histone modifications and 5fU sites suggested that 5fU might play an inhibitory role in gene transcription. Also, the distribution of 5fU in human thyroid carcinoma tissues is positively correlated with binding sites of POLR2A protein, which indicates that 5fU may distributed around POLR2A-binding

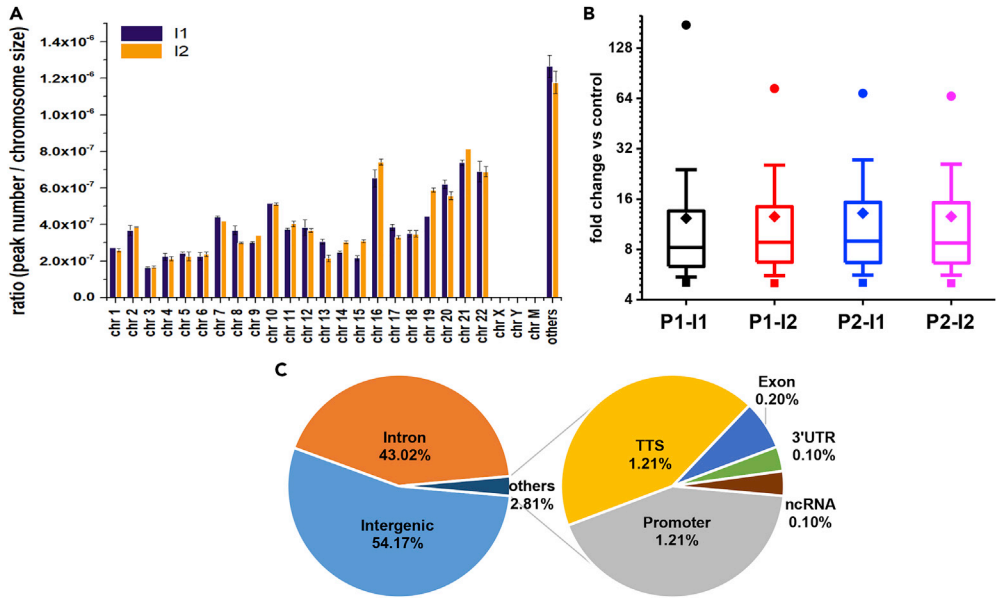


Figure 5. The 5fU Peaks Distribution in Human Thyroid Carcinoma Tissues

(A) The average distribution of merged 5fU peak (the peaks in pull-down 1 was merged with the peaks in pull-down 2) calculated from two biological replicates normalized to chromosome size in whole-genome level, using two sets of input data (I1 and I2) as control, respectively.

(B) Box chart showing the distribution of four sets of 5fU peaks calculated from two pull-down data (P1, P2) and two input data (I1, I2) based on fold change versus control. ● is the maximum value, ■ is minimum value, ◆ is mean, the box stands for interquality (25–75%) area, the line within is median, upper split line is 90% point, and lower split line is 10% point. I1, input 1; I2, input 2; P1, pull-down 1; P2, pull-down 2.

(C) Percentage of 5fU peak lengths overlapping with genomic features.

sites. What's more, recently, Zhou et al. and Tretyakova et al. reported that aldehydes present in 5fC can conjugate histone to yield DNA-protein cross-linking (Ji et al., 2017; Li et al., 2017). Aldehydes present in 5fU are more active than 5fC (Habgood et al., 2011; Hardisty et al., 2015), which might also conjugate histone to yield DNA-protein cross-linking and influence transcriptional regulation and chromatin remodeling. Although the method we proposed could not realize single-base-resolution analysis of 5fU, we believe that this goal can be realized in the future, according to the recent report of single-base-resolution detection method of 5hmU proposed by Balasubramanian's group (Kawasaki et al., 2018). Further studying and applying this method will advance our understanding of the role played by 5fU modification in genomics.

Limitations of the Study

Although this method can profile the distribution of 5fU in genomic DNA, it does not have single-base resolution. In addition, it is unavoidable for nonspecific enrichment of DNA during pull-down process. We are currently in the process of realizing the single-base-resolution analysis of 5fU, as per the recent report by Balasubramanian's group (Kawasaki et al., 2018). Combining with the single-base-resolution analysis, the nonspecific enrichment of DNA can be ruled out.

External Data

H3K27ac (ENCSR000CCE, ENCSR000CDC, ENCSR000CDD), H3K27me3 (ENCSR000CFM), H3K4me1 (ENCSR000CCF, ENCSR000CAL, ENCSR000CAI, ENCFF710BOL), H3K4me3 (ENCSR000CAK, ENCSR000CAJ, ENCSR000CDS), and POLR2A (ENCFF079QGD) chromatin immunoprecipitation sequencing datasets were obtained from ENCODE Project database.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

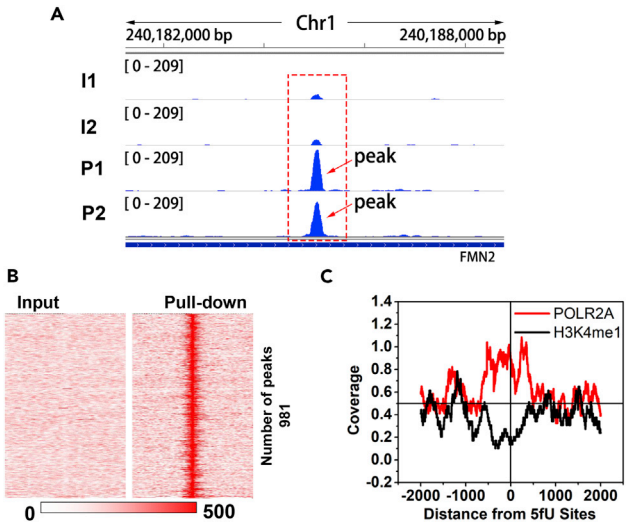


Figure 6. fU-Seq Reveals 5fU Maps in the Whole Genome of Human Thyroid Carcinoma Tissues

(A) Visual representation of the enrichment peak coverage of fU-Seq and the input control are shown. I1, input 1; I2, input 2; P1, pull-down 1; P2, pull-down 2.

(B) Heatmap representations of 5fU-normalized read densities (reads/million/base) across the genome. 5fU-containing read signals across the genome ranked by RPKM in default chromosome sort order. Heatmap scales correspond to normalized read densities.

(C) The correlation between 5fU sites with H3K4me1 modification and POLR2A in human thyroid carcinoma tissues.

DATA AND SOFTWARE AVAILABILITY

Sequencing data have been deposited into the Gene Expression Omnibus (GEO). The accession number is GSE115918.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, 18 figures, 2 schemes, and 5 tables and can be found with this article online at https://doi.org/10.1016/j.isci.2018.10.024.

ACKNOWLEDGMENTS

We thank the National Natural Science Foundation of China (21432008, 91753201, and 21721005 to X Zhou) and the China Postdoctoral Innovative Talent Support Program (No. BX20180228 to Y.W.). The numerical calculations in this article have been done on the supercomputing system in the Supercomputing Center of Wuhan University. We also thank Dr. Haifang Li (Analysis Center, Tsinghua University) who provided DNA MALDI-TOF test instructions.

AUTHOR CONTRIBUTIONS

X. Zhou, Y.W., and C.L. conceived the original idea and designed the experiments with the help of Xiong Zhang and Z.W.; Y.W., C.L., F.W., and Xiong Zhang performed the experiments. F.W., Z.C., W.Z., Y.Z., X.W., and Z.W. performed bioinformatics analysis; C.L., Xiong Zhang, and W.Y. synthesized the chemicals; S.L., and Xiaolian Zhang helped with the tissues. X. Zhou, Y.W., C.L., and F.W. wrote the paper.

DECLARATION OF INTERESTS

I and three of my authors are applying for a Chinese patent. The authors declare no competing interests.

Received: July 31, 2018
Revised: September 4, 2018
Accepted: October 23, 2018
Published: November 30, 2018

REFERENCES

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Booth, M.J., Raiber, E.-A., and Balasubramanian, S. (2015). Chemical methods for decoding cytosine modifications in DNA. Chem. Rev. *115*, 2240–2254.

Bradner, J.E. (2015). An essential passenger with p53. Nature *520*, 626.

Chen, Y., Hong, T., Wang, S., Mo, J., Tian, T., and Zhou, X. (2017). Epigenetic modification of nucleic acids: from basic studies to medical applications. Chem. Soc. Rev. *46*, 2844–2872.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U S A *107*, 21931.

Decarroz, C., Wagner, J.R., Van Lier, J.E., Krishna, C.M., Riesz, P., and Cadet, J. (1986). Sensitized photo-oxidation of thymidine by 2-methyl-1,4-naphthoquinone. Characterization of the stable photoproducts. Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med. *50*, 491–505.

Ferrari, Karin J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Mol. Cell *53*, 49–62.

Habgood, M., Price, S.L., Portalone, G., and Irrera, S. (2011). Testing a variety of electronic-structure-based methods for the relative energies of 5-formyluracil crystals. J. Chem. Theory Comput. *7*, 2685–2688.

Hardisty, R.E., Kawasaki, F., Sahakyan, A.B., and Balasubramanian, S. (2015). Selective chemical labeling of natural T modifications in DNA. J. Am. Chem. Soc. *137*, 9270–9272.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hirose, W., Sato, K., and Matsuda, A. (2010). Selective detection of 5-formyl-2′-deoxyuridine, an oxidative lesion of thymidine, in DNA by a fluorogenic reagent. Angew. Chem. Int. Ed. *122*, 8570–8572.

Hong, H., Cao, H., Wang, Y., and Wang, Y. (2006). Identification and quantification of a guanine−thymine intrastrand cross-link lesion induced by Cu(II)/H2O2/ascorbate. Chem. Res. Toxicol. *19*, 614–621.

Hong, H., and Wang, Y. (2007). Derivatization with girard reagent T combined with LC−MS/MS for the sensitive detection of 5-Formyl-2′-deoxyuridine in cellular DNA. Anal. Chem. *79*, 322–326.

Hong, T., Yuan, Y., Chen, Z., Xi, K., Wang, T., Xie, Y., He, Z., Su, H., Zhou, Y., Tan, Z.-J., et al. (2018). Precise antibody-independent m6A identification via 4SedTTP-involved and FTO-assisted strategy at single-nucleotide resolution. J. Am. Chem. Soc. *140*, 5886–5889.

Iwan, K., Rahimoff, R., Kirchner, A., Spada, F., Schröder, A.S., Kosmatchev, O., Ferizaj, S., Steinbacher, J., Parsa, E., Müller, M., et al. (2017). 5-Formylcytosine to cytosine conversion by C–C bond cleavage in vivo. Nat. Chem. Biol. *14*, 72–78.

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071–1078.

Ji, S., Shao, H., Han, Q., Seiler Christopher, L., and Tretyakova Natalia, Y. (2017). Reversible DNA–protein cross-linking at epigenetic DNA marks. Angew. Chem. Int. Ed. *56*, 14130–14134.

Jiang, H.-P., Liu, T., Guo, N., Yu, L., Yuan, B.-F., and Feng, Y.-Q. (2017). Determination of formylated DNA and RNA by chemical labeling combined with mass spectrometry analysis. Anal. Chim. Acta *981*, 1–10.

Johnson, R.P., Fleming, A.M., Perera, R.T., Burrows, C.J., and White, H.S. (2017). Dynamics of a DNA mismatch site held in confinement discriminate epigenetic modifications of cytosine. J. Am. Chem. Soc. *139*, 2750–2756.

Kasai, H., Iida, A., Yamaizumi, Z., Nishimura, S., and Tanooka, H. (1990). 5-Formyldeoxyuridine: a new type of DNA damage induced by ionizing radiation and its mutagenicity to Salmonella strain TA102. Mutat. Res. *243*, 249–253.

Kawasaki, F., Beraldi, D., Hardisty, R.E., McInroy, G.R., van Delft, P., and Balasubramanian, S. (2017a). Genome-wide mapping of 5-hydroxymethyluracil in the eukaryote parasite *Leishmania*. Genome Biol. *18*, 23.

Kawasaki, F., Martínez Cuesta, S., Beraldi, D., Mahtey, A., Hardisty, R.E., Carrington, M., and Balasubramanian, S. (2018). Sequencing 5-hydroxymethyluracil at single-base resolution. Angew. Chem. Int. Ed. *130*, 9842–9844.

Kawasaki, F., Murat, P., Li, Z., Santner, T., and Balasubramanian, S. (2017b). Synthesis and biophysical analysis of modified thymine-containing DNA oligonucleotides. Chem. Commun. (Camb) *53*, 1389–1392.

Kitsera, N., Allgayer, J., Parsa, E., Geier, N., Rossa, M., Carell, T., and Khobta, A. (2017). Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter. Nucleic Acids Res. *45*, 11033–11042.

Kittaka, A., Takayama, H., Kurihara, M., Horii, C., Tanaka, H., Miyasaka, T., and Inoue, J.-I. (2001). DNA sequence recognition by NFκB p50 homodimer: strict and obscure recognition sites in the binding sequence. Nucleosides Nucleotides Nucleic Acids *20*, 669–672.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Li, F., Zhang, Y., Bai, J., Greenberg, M.M., Xi, Z., and Zhou, C. (2017). 5-formylcytosine yields DNA–protein cross-links in nucleosome core particles. J. Am. Chem. Soc. *139*, 10617–10620.

Liu, C., Chen, Y., Wang, Y., Wu, F., Zhang, X., Yang, W., Wang, J., Chen, Y., He, Z., Zou, G., et al. (2017a). A highly efficient fluorescence-based switch-on detection method of 5-formyluracil in DNA. Nano Res. *10*, 2449–2458.

Liu, C., Wang, Y., Zhang, X., Wu, F., Yang, W., Zou, G., Yao, Q., Wang, J., Chen, Y., Wang, S., et al. (2017b). Enrichment and fluorogenic labelling of 5-formyluracil in DNA. Chem. Sci. *8*, 4505–4510.

Liu, C., Zou, G., Peng, S., Wang, Y., Yang, W., Wu, F., Jiang, Z., Zhang, X., and Zhou, X. (2018). 5-formyluracil as a multifunctional building block in biosensor designs. Angew. Chem. Int. Ed. *130*, 9837–9841.

Liu, M.Y., DeNizio, J.E., Schutsky, E.K., and Kohli, R.M. (2016). The expanding scope and impact of epigenetic cytosine modifications. Curr. Opin. Chem. Biol. *33*, 67–73.

Liu, Y., Zhang, X., Han, C., Wan, G., Huang, X., Ivan, C., Jiang, D., Rodriguez-Aguayo, C., Lopez-Berestein, G., Rao, P.H., et al. (2015). TP53 loss creates therapeutic vulnerability in colorectal cancer. Nature *520*, 697–701.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 10–12, Next Generation Sequencing Data Analysis.

Modrzejewska, M., Gawronski, M., Skonieczna, M., Zarakowska, E., Starczak, M., Foksinski, M., Rzeszowska-Wolny, J., Gackowski, D., and Olinski, R. (2016). Vitamin C enhances substantially formation of 5-hydroxymethyluracil in cellular DNA. Free Radic. Biol. Med. *101*, 378–383.

Pais, J.E., Dai, N., Tamanaha, E., Vaisvila, R., Fomenkov, A.I., Bitinaite, J., Sun, Z., Guan, S., Corrêa, I.R., Noren, C.J., et al. (2015). Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. Proc. Natl. Acad. Sci. U S A *112*, 4316–4321.

Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S.K., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotljarova, O., et al. (2014). Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. Nat. Chem. Biol. *10*, 574–581.

Rada-Iglesias, A. (2018). Is H3K4me1 at enhancers correlative or causative? Nat. Genet. *50*, 4–5.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Rogstad, D.K., Heo, J., Vaidehi, N., Goddard, W.A., Burdzy, A., and Sowers, L.C. (2004). 5-Formyluracil-induced perturbations of DNA function. Biochemistry *43*, 5688–5697.

Samanta, B., Seikowski, J., and Höbartner, C. (2015). Fluorogenic labeling of 5-formylpyrimidine

nucleotides in DNA and RNA. Angew. Chem. Int. Ed. *55*, 1912–1916.

Shen, H., Xu, W., Guo, R., Rong, B., Gu, L., Wang, Z., He, C., Zheng, L., Hu, X., Hu, Z., et al. (2016). Suppression of enhancer overactivation by a RACK7-histone demethylase complex. Cell *165*, 331–342.

Shen, L., Song, C.-X., He, C., and Zhang, Y. (2014). Mechanism and function of oxidative reversal of DNA and RNA methylation. Annu. Rev. Biochem. *83*, 585–614.

Shu, X., Liu, M., Lu, Z., Zhu, C., Meng, H., Huang, S., Zhang, X., and Yi, C. (2018). Genome-wide mapping reveals that deoxyuridine is enriched in the human centromeric DNA. Nat. Chem. Biol. *14*, 680–687.

Song, C.-X., Clark, T.A., Lu, X.-Y., Kislyuk, A., Dai, Q., Turner, S.W., He, C., and Korlach, J. (2011). Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. Nat. Methods *9*, 75–77.

Song, C.-X., Szulwach, Keith E., Dai, Q., Fu, Y., Mao, S.-Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., et al. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell *153*, 678–691.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat. Rev. Genet. *9*, 465–476.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

Wagner, M., Steinbacher, J., Kraus Theo, F.J., Michalakis, S., Hackner, B., Pfaffeneder, T., Perera, A., Müller, M., Giese, A., Kretzschmar Hans, A., et al. (2015). Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. Angew. Chem. Int. Ed. *54*, 12511–12514.

Wang, Y., Liu, C., Yang, W., Zou, G., Zhang, X., Wu, F., Yu, S., Luo, X., and Zhou, X. (2018a). Naphthalimide derivatives as multifunctional molecules for detecting 5-formylpyrimidine by both PAGE analysis and dot-blot assays. Chem. Commun. (Camb) *54*, 1497–1500.

Wang, Y., Liu, C., Zhang, X., Yang, W., Wu, F., Zou, G., Weng, X., and Zhou, X. (2018b). Gene specific-loci quantitative and single-base resolution analysis of 5-formylcytosine by compound-mediated polymerase chain reaction. Chem. Sci. *9*, 3723–3728.

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. Nat. Rev. Genet. *18*, 517–534.

Wyrick, J.J., and Roberts, S.A. (2015). Genomic approaches to DNA repair and mutagenesis. DNA Repair (Amst) *36*, 146–155.

Yoshida, M., Makino, K., Morita, H., Terato, H., Ohyama, Y., and Ide, H. (1997). Substrate and mispairing properties of 5-formyl-2′-deoxyuridine 5′-triphosphate assessed by in vitro DNA polymerase reactions. Nucleic Acids Res. *25*, 1570–1577.

Zhu, C., Gao, Y., Guo, H., Xia, B., Song, J., Wu, X., Zeng, H., Kee, K., Tang, F., and Yi, C. (2017). Single-cell 5-formylcytosine landscapes of mammalian early embryos and ESCs at single-base resolution. Cell Stem Cell *20*, 720–731.e5.

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

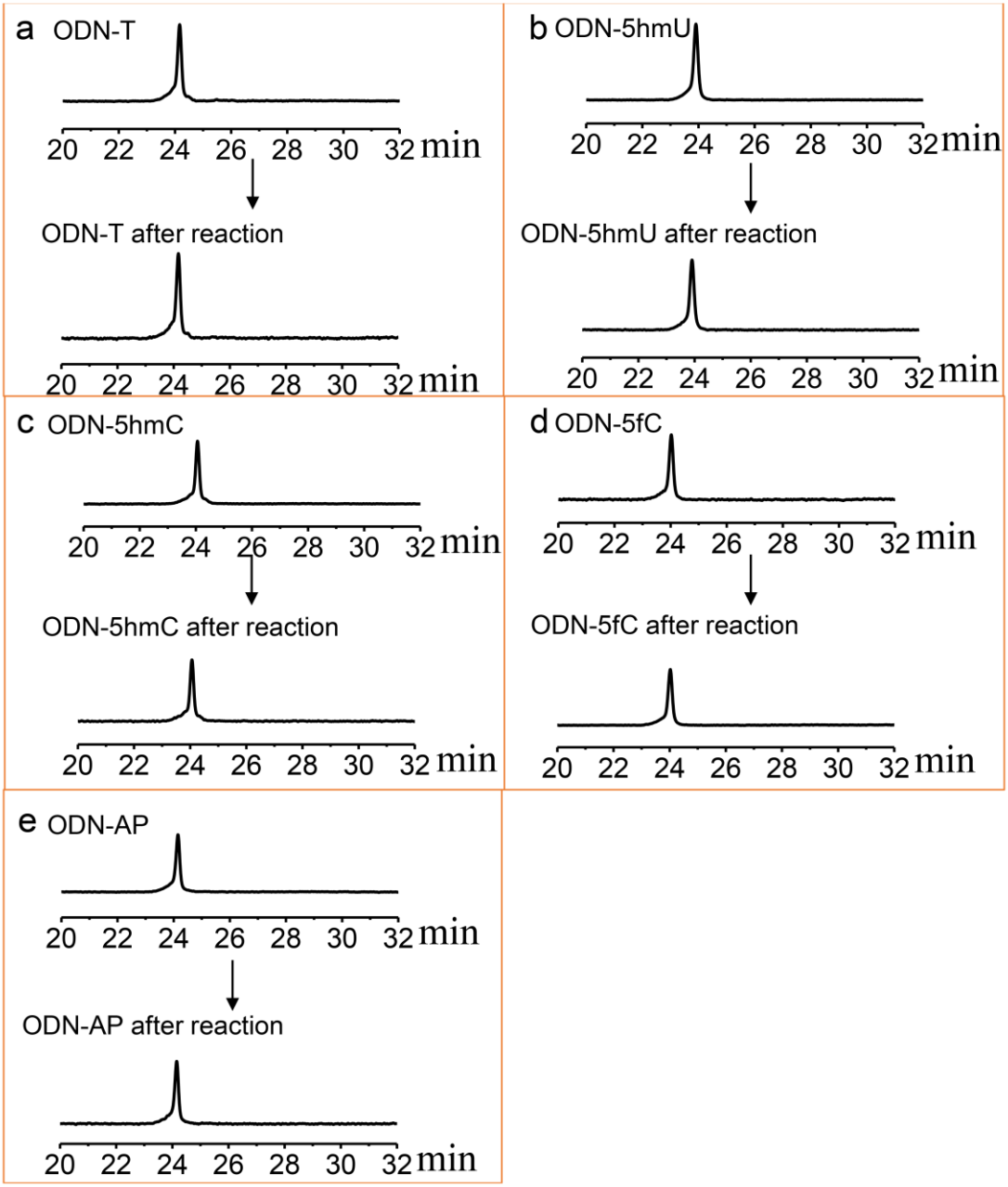


Figure S1. HPLC detection of various ODNs before and after treatment with azi-BIAN (Related to Figure 2a-d).

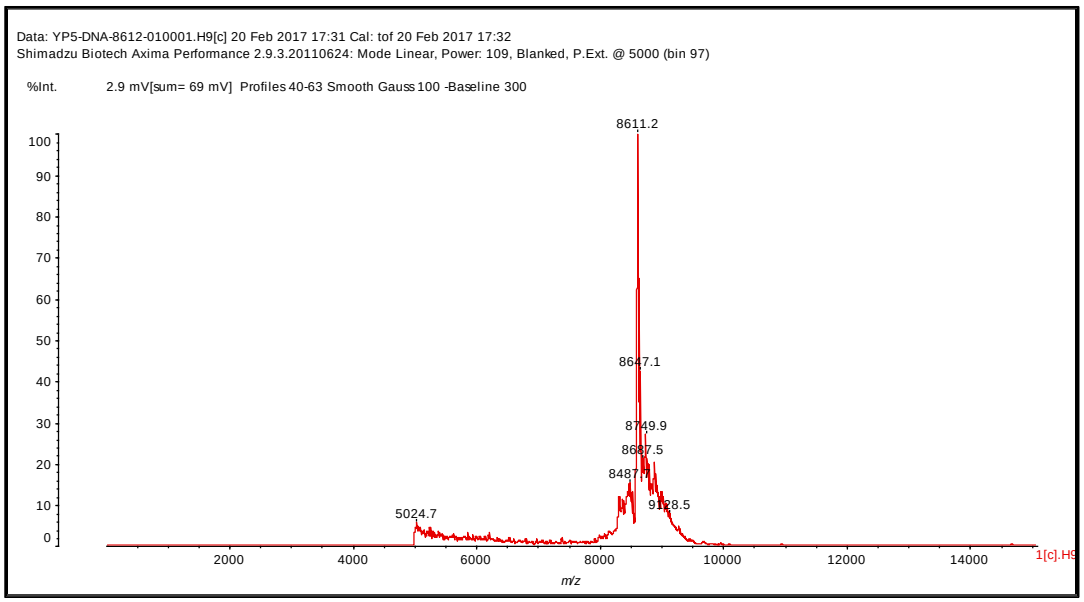
(a) RP-HPLC trace at λ = 260 nm of ODN-T before and after treatment with azi-BIAN. (b) RP-HPLC trace at λ = 260 nm of ODN-5hmU before and after treatment with azi-BIAN. (c) RP-HPLC trace at λ = 260 nm of ODN-5hmC before and after treatment with azi-BIAN (d) RP-HPLC trace at λ = 260 nm of ODN-5fC before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

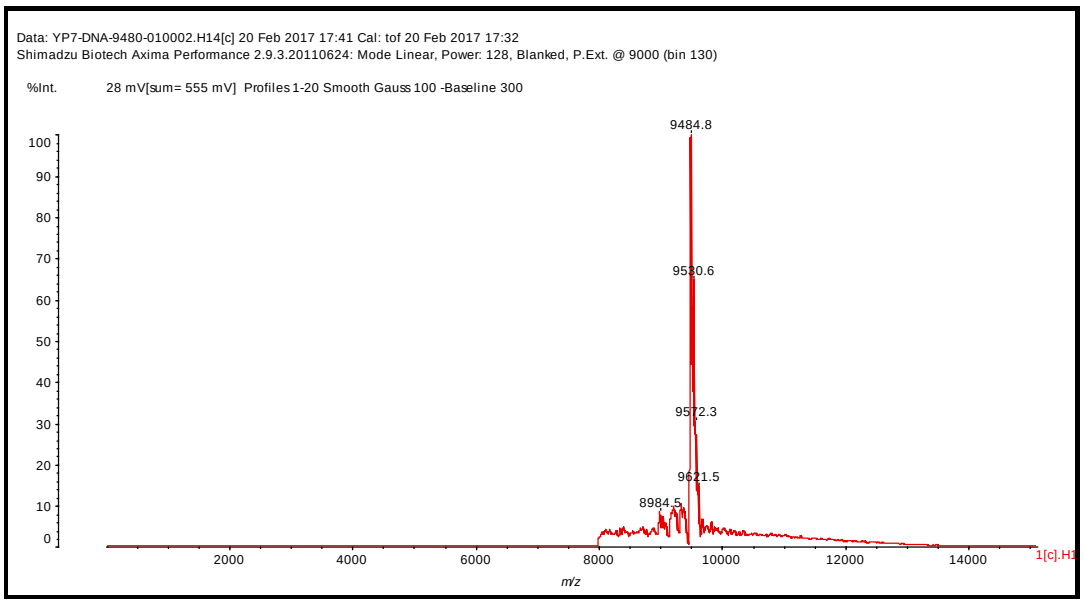
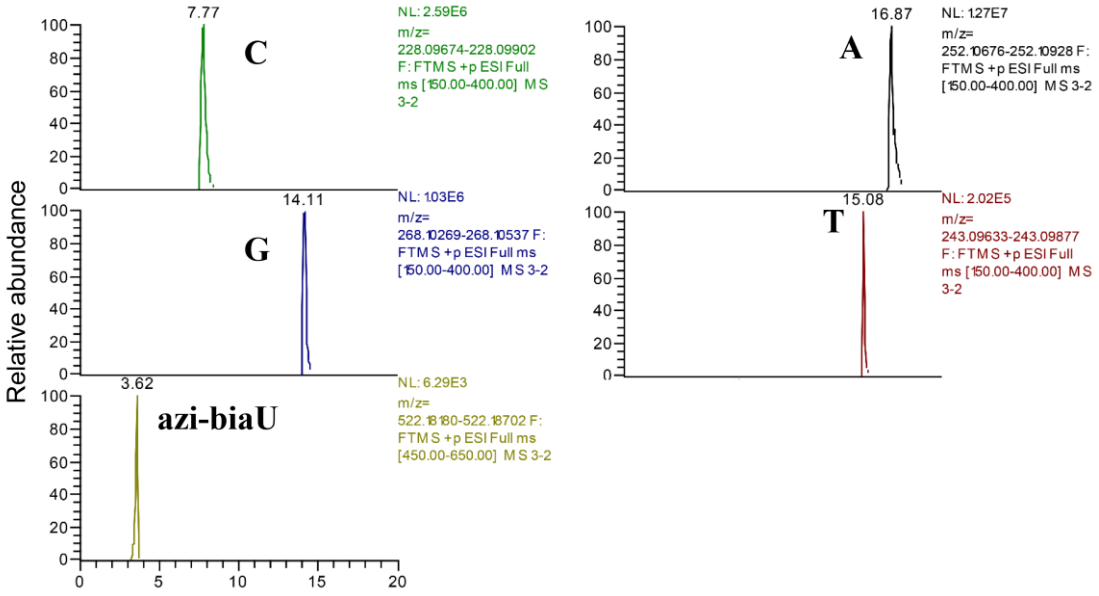


Figure S2. DNA MALDI-TOF Mass Spectra (Related to Figure 2a-d). (a) MALDI-TOF-spectrum of ODN2-5fU after incubation with azi-BIAN; (b) MALDI-TOF-spectrum of ODN2-5fU after incubation with azi-BIAN, then reacted with DBCO-biotin.

a



b

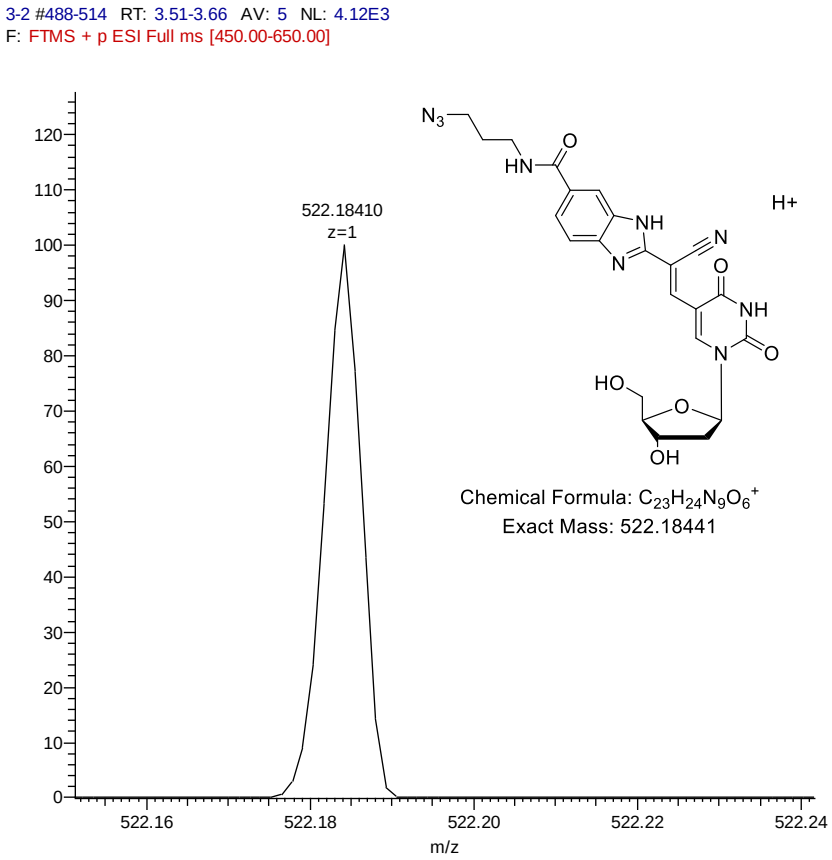
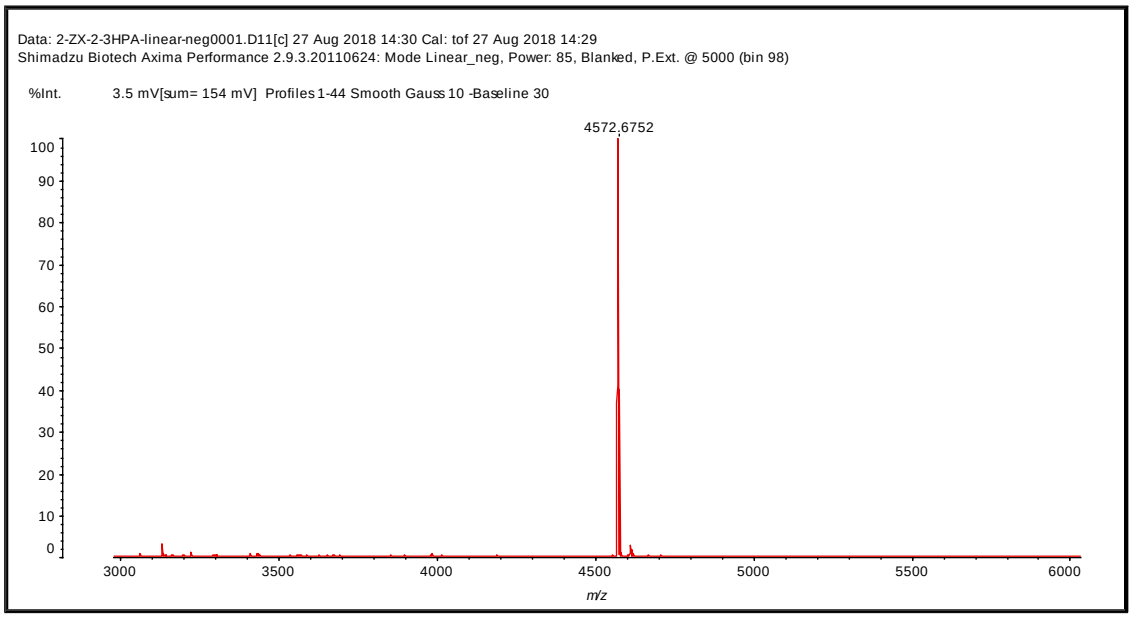


Figure S3. HPLC-MS detection of digestion of ODN-azi-biaU (Related to Figure 2a-d). (a) HPLC-MS extracted $[M+H]^+$ ion count for A, T, C, G, azi-biaU after digestion of DNA from the ODN-azi-biaU. (b) HRMS (ESI+) of azi-biaU in HPLC-MS after digestion, HRMS $C_{23}H_{24}N_9O_6^+$ $[M+H]^+$ calculated 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’

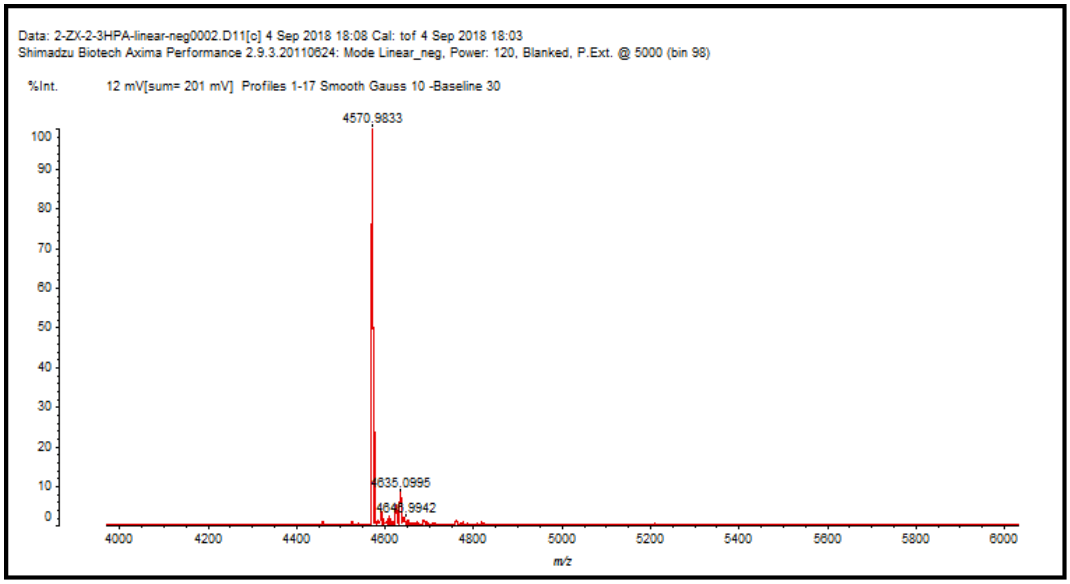


Figure S4. DNA MALDI-TOF Mass Spectra (Related to Figure 2a-d). (a) MALDI-TOF-spectrum of ODN-5fC; (b) MALDI-TOF-spectrum of ODN-5fC after incubation with azi-BIAN. No mass of ODN-azi-biaU appeared.

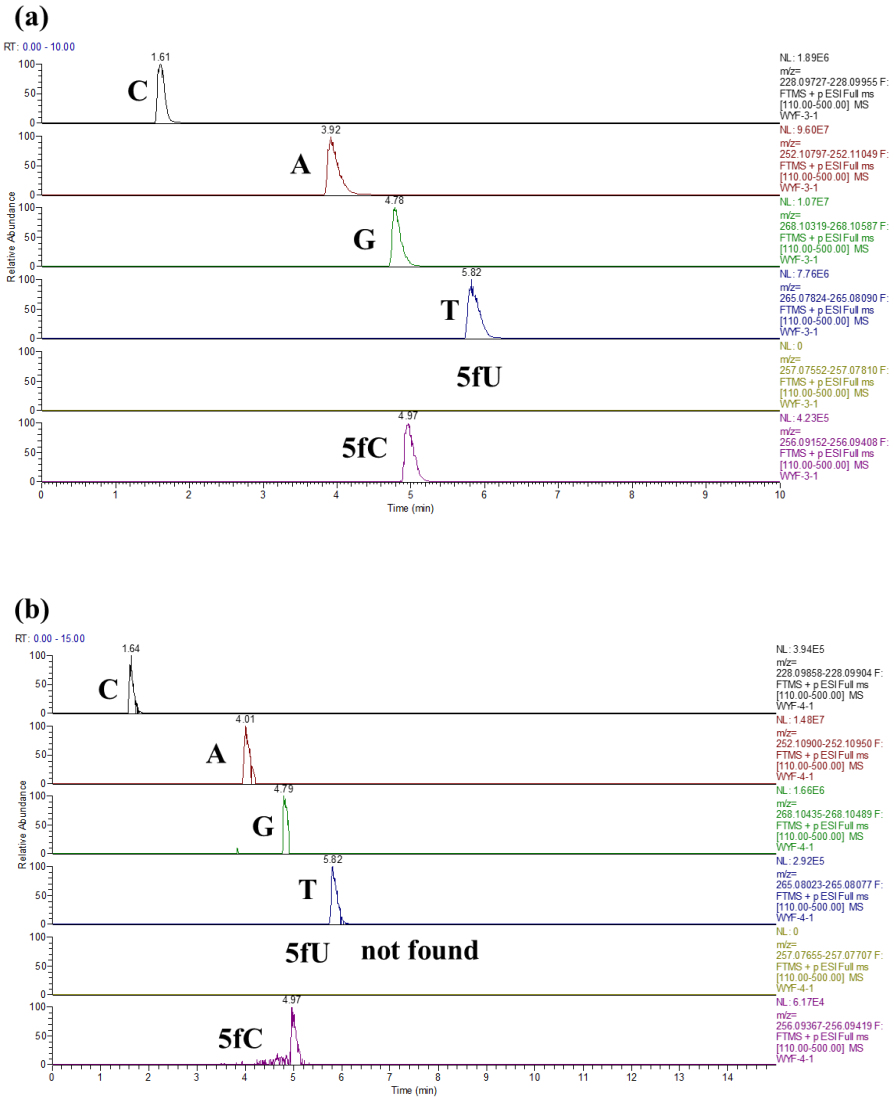


Figure S5. HPLC-MS detection of digestion of extracted model DNA (Related to Figure 2a-d). (a) HPLC-MS extracted $[M+H]^+$ ion count for A, T, C, G, 5fU, 5fC after digestion of the 80bp ds ODN-fC. (b) HPLC-MS extracted $[M+H]^+$ ion count for A, T, C, G, 5fU, 5fC after digestion of the extracted 80bp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

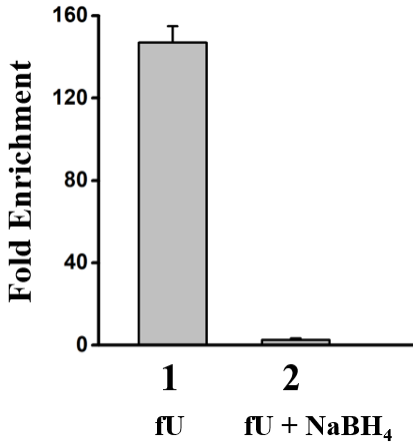


Figure S6. Enrichment tests of reduced 5fU (Related to Figure 2e). 1 represent ODN-5fU without treatment by $NaBH_4$, 2 represent ODN 5fU was treated by $NaBH_4$. Values shown are fold-enrichment over canonical nucleobases. Error bars represent the standard deviations of three parallel measurements.

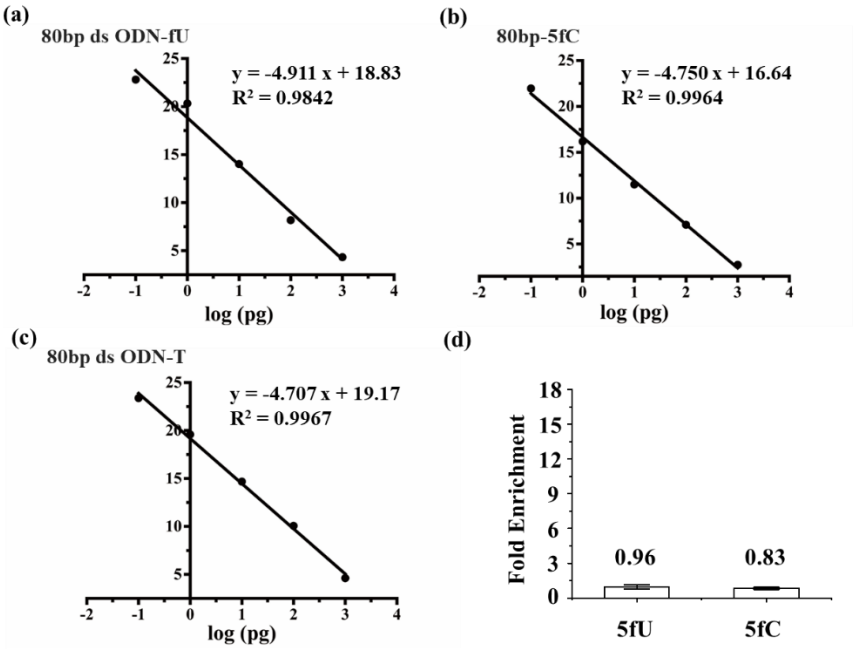


Figure S7. Enrichment tests of DNA containing hydroxylamine ($EtONH_2$) protection of the mixed system (80bp ds ODN-fU, 80bp-5fC, and 80bp ds ODN-T) (Related to Figure 2e). (a)-(c) Example calibration line of 80bp ds ODN-fU, 80bp-5fC, and 80bp ds ODN-T for enrichment analysis. (d) Values shown are fold-enrichment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

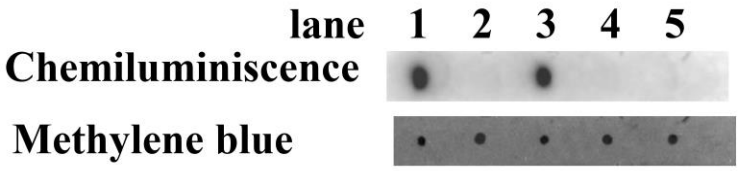
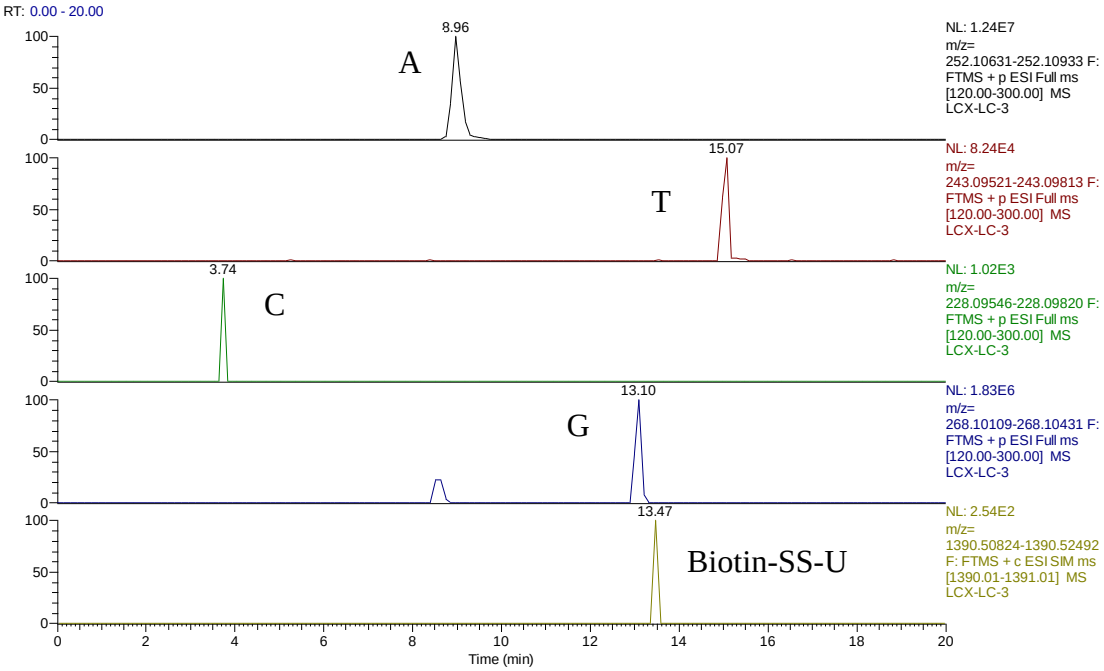


Figure S8. Dot-blot assay of streptavidin-HRP detection of samples containing 5fU (Related to Figure 2). Lane 1: genomic DNA from mouse hippocampus was treated with azi-BIAN, then incubation with DBCO-S-S-PEG3-biotin (Click Chemistry Tools, A112-10) to introduce biotin group. Lane 2: genomic DNA from 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

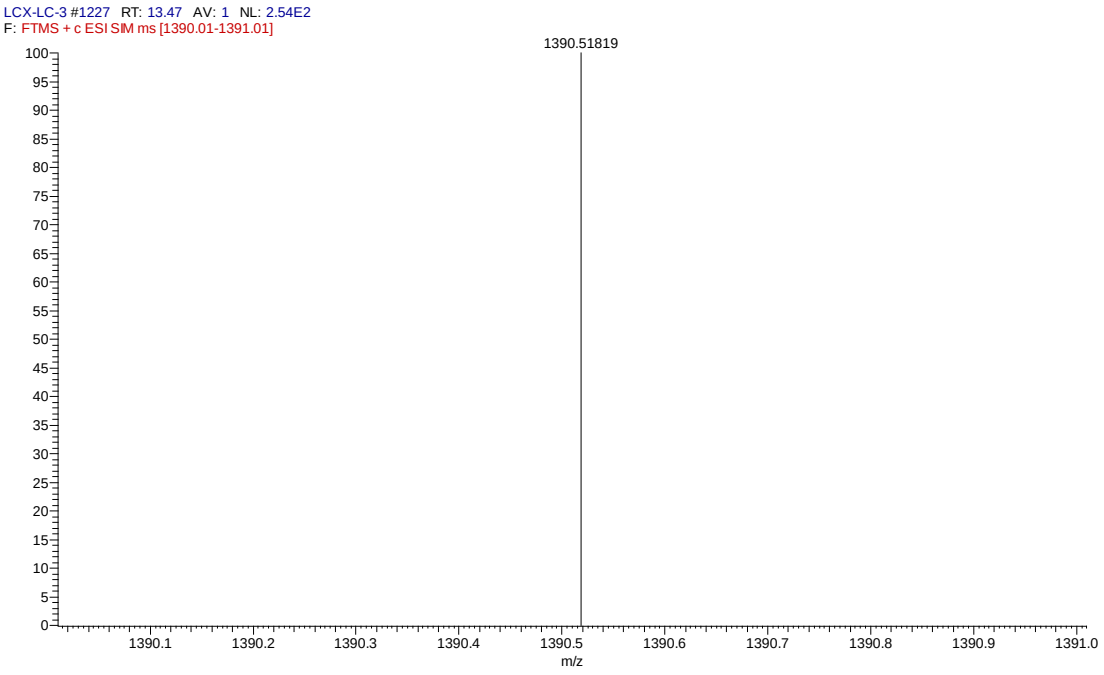


Figure S9. HPLC-MS detection of digestion of biotin labeled genomic DNA (Related to Figure 2). (a) HPLC-MS extracted $[M+H]^+$ ion count for A, T, C, G, Biotin-SS-U after digestion of the biotin labeled genomic DNA. (b) HRMS (ESI+) of Biotin-SS-U in HPLC-MS after digestion, HRMS $C_{65}H_{80}N_{15}O_{14}S_3^+$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

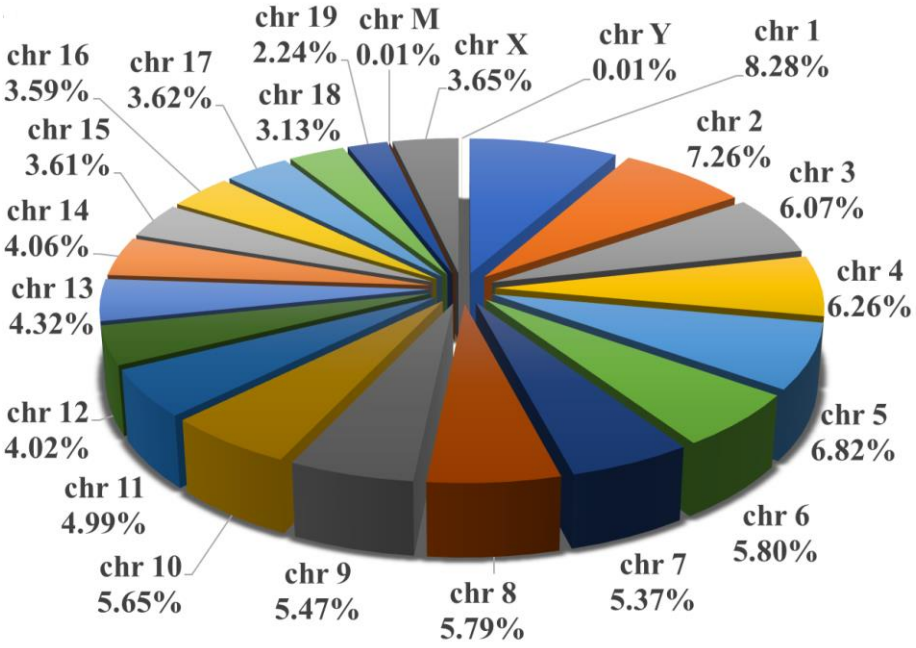


Figure S10. The pie chart shows the percentage of peak number of every chromosome in whole genome (Related to Figure 3).

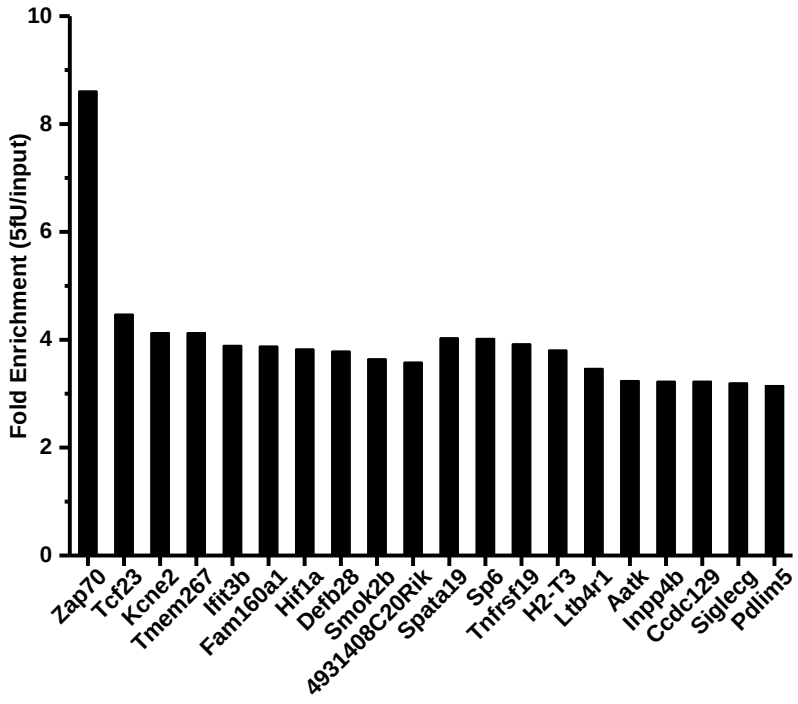


Figure S11. Verification 5fU-enriched regions in mouse hippocampus by qPCR (Related to Figure 4a&b). X-axis is labeled with the gene names within which the identified peak was identified. Fold enrichment is calculated as 2^-ddCt, where dCt_1 = Ct (5fU enriched) – Ct (Input), $dCt_2 = Ct_{ref}$ (5fU enriched) – 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.

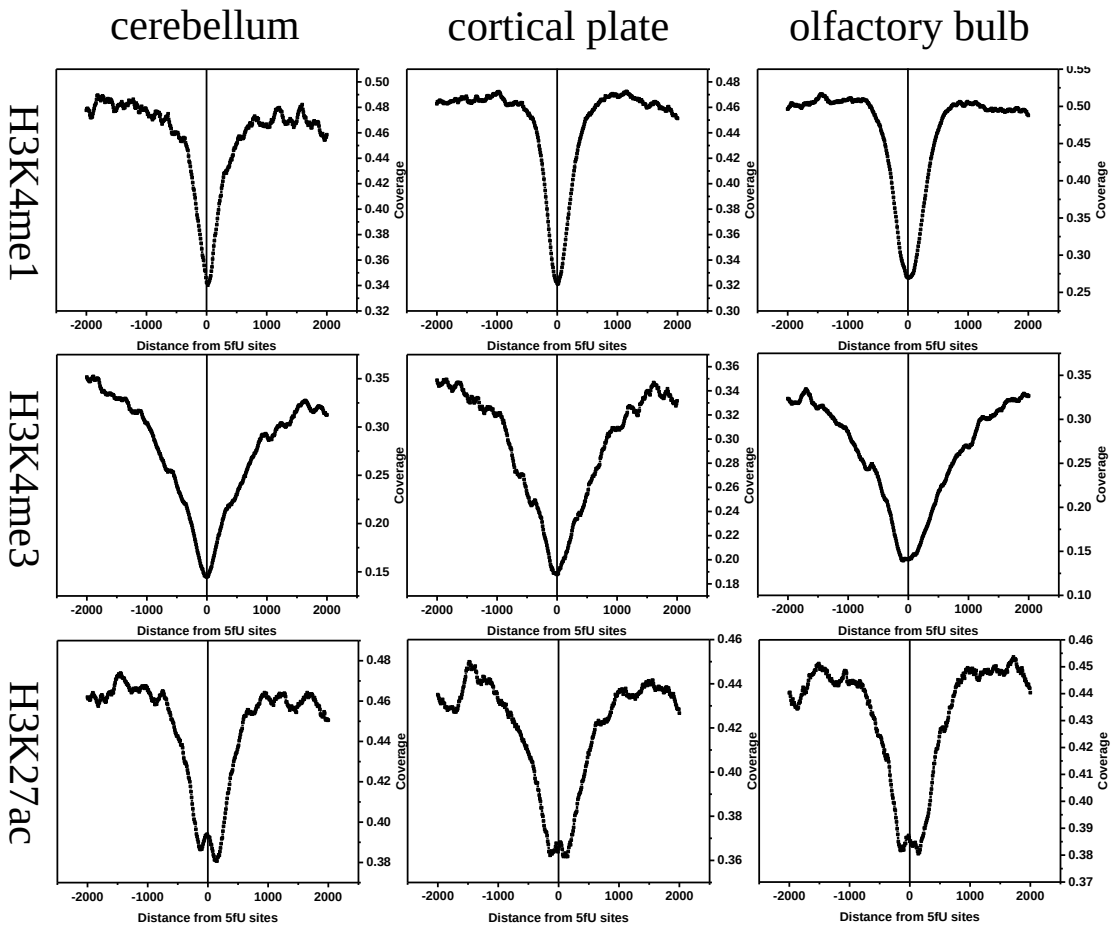


Figure S12. Distribution patterns of 5fU at different histone modification sites of varied brain tissues (Related to Figure 4c). Distribution patterns of 5fU with respect to H3K4me1, H3K27me3 and H3K27ac modification sites in the cerebellum, cortical plate and olfactory bulb, respectively. From these,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

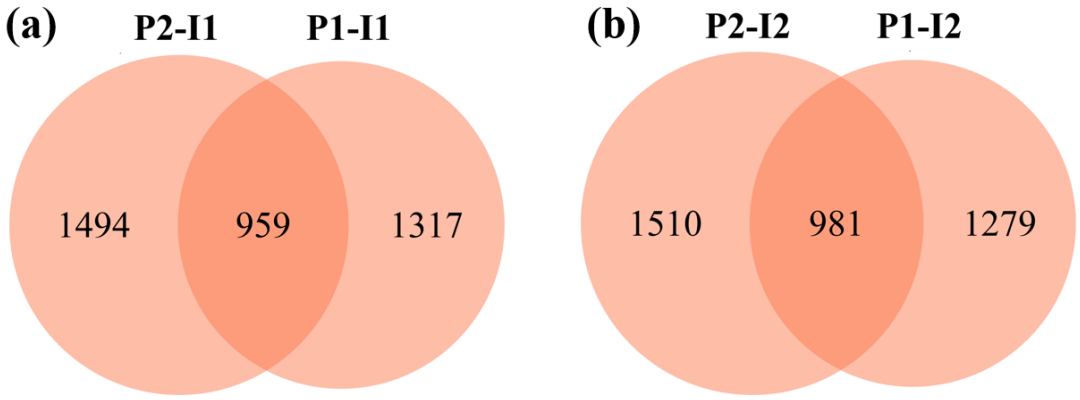


Figure S13. Venn diagrams showing the peak-merging results of two biological replicates, pull-down data 1 (P1) and pull-down data 2 (P2) against a) Input data 1 (I1) and b) Input data 2 (I2) (Related to Figure 5).

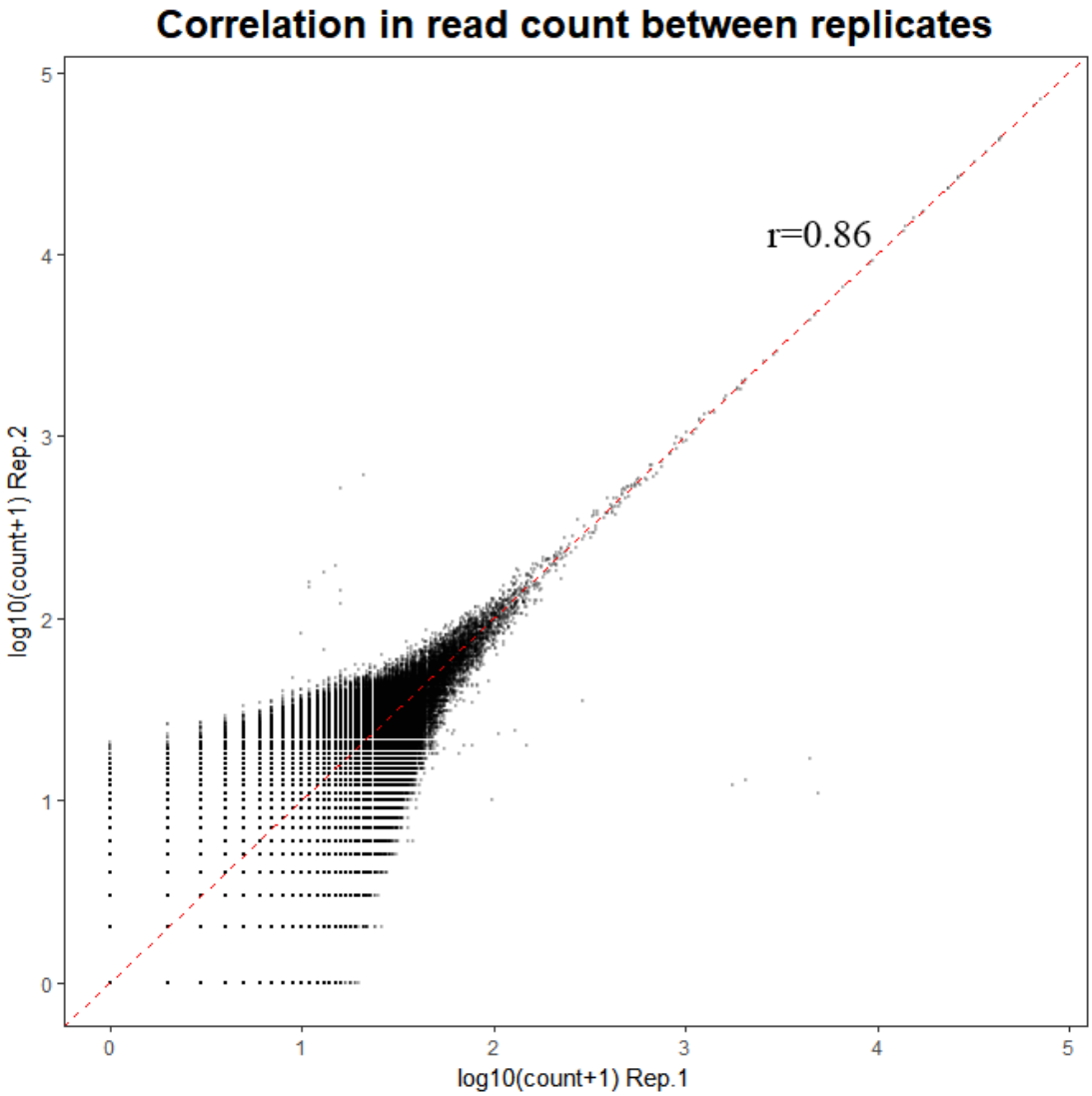


Figure S14. Scatter plots showing the correlation in read counts of peaks between biological replicates (Related to Figure 5).

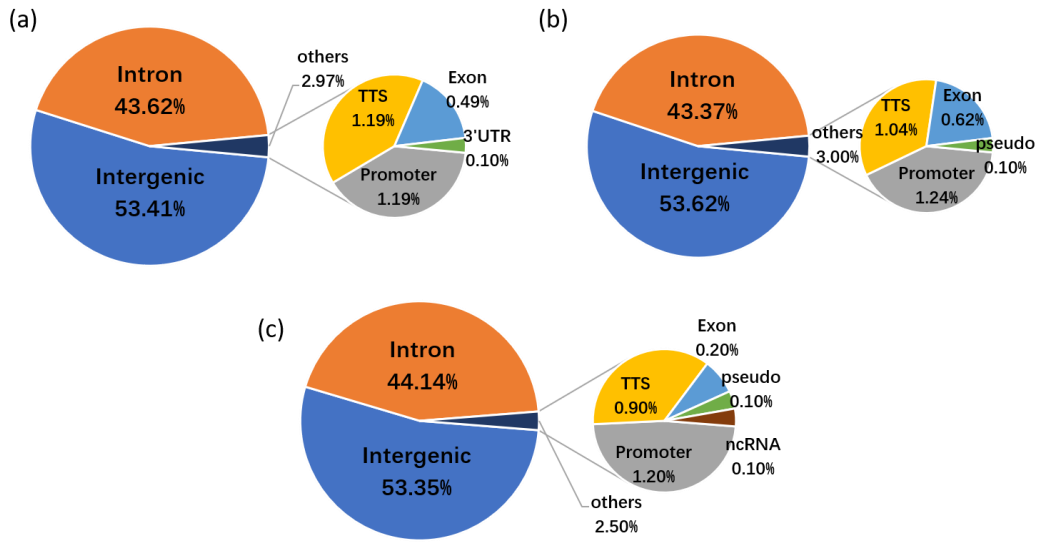


Figure S15. Overall distribution of four sets of filtered 5fU peaks according to merging results in human genome, combined two by two within two pull-down data (P1, P2) and two input data (I1, I2): (a) P1-I1, (b) P2-I1, (c) P2-I2 (Related to Figure 5c).

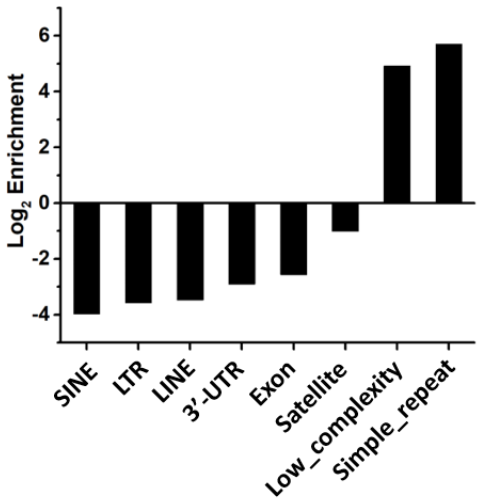


Figure S16. Relative enrichment of 5fU peaks in different genomic elements (Related to Figure 5c).

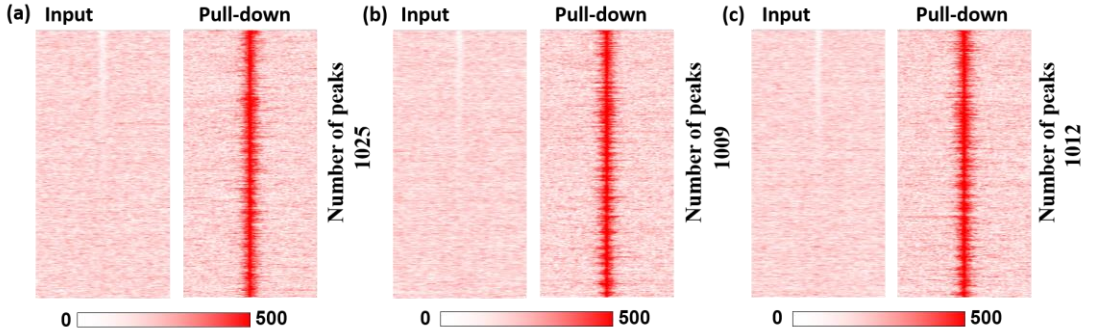


Figure S17. Heatmap shows 5fU-normalized read densities (reads/million/base) across human genome. signals ranked by RPKM in default chromosome sort order, heatmap scales correspond to normalized read densities. (a) P1-I1, (b) P1-I2, (c) P2-I2 (Related to Figure 6b).

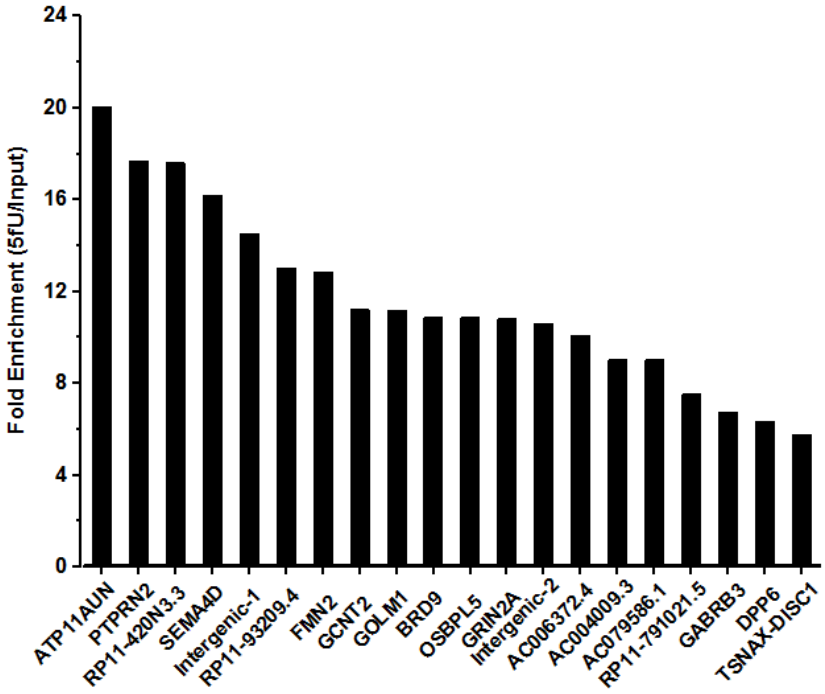


Figure S18. Verification 5fU-enriched regions in human thyroid carcinoma tissues by qPCR (Related to Figure 6a&b).

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)

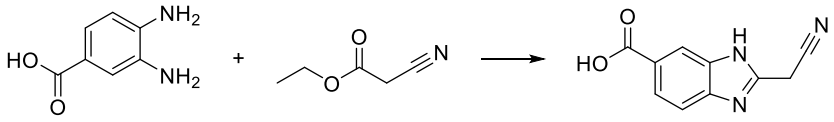
(Thermo Scientific) at 37℃ for 1 hr and washed with 1× TBST four times, the results were visualized by enhanced chemiluminescence (SuperSignal™ West Pico Chemiluminescent Substrate, Cat: 34077, Thermo Scientific) using Molecular Imager® ChemiDocTM XRS+ Imaging System (Bio-Rad). Finally, the membrane 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.

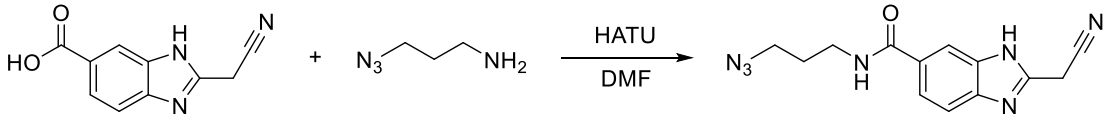
3. Synthesis

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



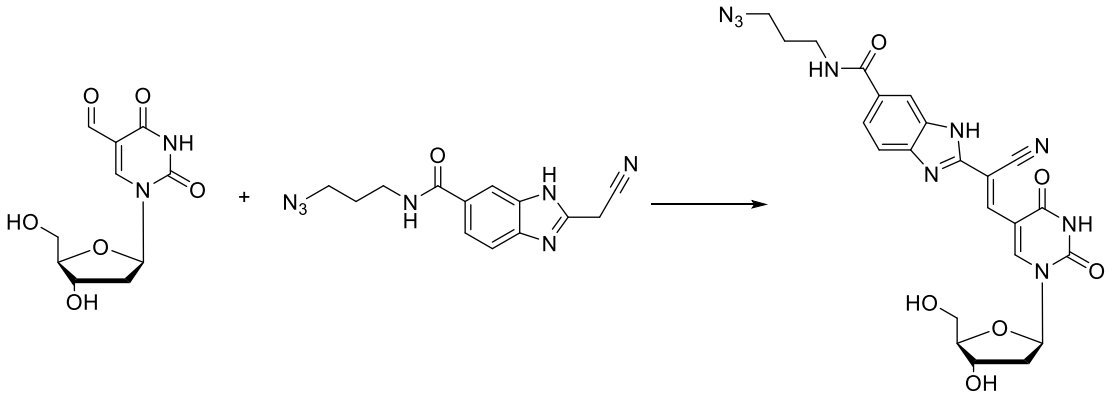
3,4-diaminobenzoic acid (3 g, 19.7 mmol) and ethyl cyanoacetate (7 mL, 65.8 mmol) were dispersed in 100 mL round bottom flask. Then, the reaction mixture was kept stirring and stayed at 180°C. After 1 hr, the reaction mixture was poured into 100 mL diethyl ether, the precipitate was filtered off and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