Supporting Information

Bisulfite-Free, Nanoscale Analysis of 5-Hydroxymethylcytosine at Single Base Resolution

Hu Zeng,[†,#] Bo He,[‡,#] Bo Xia,[†] Dongsheng Bai,[†] Xingyu Lu,[§] Jiabin Cai,[//] Lei Chen,[⊥] Ankun Zhou,[†] Chenxu Zhu,[†] Haowei Meng,[†] Yun Gao,[○] Hongshan Guo,[○] Chuan He,[*,▽,¶] Qing Dai,[*,▽] Chengqi Yi[*,†,‡,¶]

[†]State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China

[‡] Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

[§] Shanghai Epican Genetech, Co. Ltd., Zhangjiang Hi-Tech Park, Shanghai 201203, China

[//]Department of Liver Surgery, Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai 200032, China

[⊥]International Co-operation Laboratory on Signal Transduction, Eastern Hepatobiliary Surgery Institute, Second Military Medical University, Shanghai 200438, China

[○] Biodynamic Optical Imaging Center and Beijing Advanced Innovation Center for Genomics, School of Life Sciences, Peking University, Beijing 100871, China

[▽]Department of Chemistry, Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, Howard Hughes Medical Institute, The University of Chicago, Chicago, Illinois 60637, United States

[¶] Department of Chemical Biology and Synthetic and Functional Biomolecules Center, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

Corresponding Author: *chuanhe@uchicago.edu, *daiqing@uchicago.edu, *chengqi.yi@pku.edu.cn.

[#] These authors contributed equally

Table of Contents

Experimental section S3
Oligonucleotide synthesis and model DNA preparation S3
hESC culture and genomic DNA extraction S3
Hydroxylamine-mediated blocking of endogenous 5fC S3
Preparation of potassium ruthenate (K_2RuO_4) S3
K_2RuO_4-mediated oxidation of genomic DNA S4
Quantification of 5mC, 5hmC and 5fC level by LC-MS/MS S4
Synthesis of 5-(2-azidoethyl)-1,3-indandione (AI) S4
AI-mediated labeling of the newly generated 5fC S4
Library preparation and click chemistry S4
Sanger sequencing on model DNA S5
Pull-down for enrichment of 5hmC-containing DNA S5
Quantitative PCR (qPCR) to test selectivity of 5hmC enrichment S5
The calculation of relative gene body 5hmC level by hmC-CATCH and TAB-seq S5
Preparation of cfDNA samples S5
Data processing and analysis S5
External data S7
Supplementary Figures and Table S8
Figure S1 ..... S8
Figure S2 ..... S9
Figure S3 ..... S10
Figure S4 ..... S11
Figure S5 ..... S12
Figure S6 ..... S13
Figure S7 ..... S14
Figure S8 ..... S15
Figure S9 ..... S16
Figure S10 ..... S17
Figure S11 ..... S18
Figure S12 ..... S19
Figure S13 ..... S20
Table S3 ..... S21
References S22

Experimental section

Oligonucleotide synthesis and model DNA preparation

Oligonucleotides containing 5mC, 5hmC, 5fC or 5caC were synthesized using the ABI Expedite 8909 Nucleic Acid Synthesizer. The modified nucleotides were site-specifically incorporated at the desired positions (**Table S1**) using commercially available phosphoramidites (Glen Research). Subsequent deprotection and purification were carried out with Glen-Pak Cartridges (Glen Research) according to the manufacturer's instructions. Purified oligonucleotides were characterized by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF, ABI 7500). Regular oligonucleotides (and PCR primers) were purchased from Sangon Biotech (Shanghai).

Long duplex DNAs (**Table S1**) were prepared through ligation of short duplexes fragments (20-40 bp) with sticky overhangs[1]. In brief, the ligation-site oligonucleotides were phosphorylated with T4 polynucleotide kinase (NEB) and then annealed with the corresponding complementary strands. Annealed duplexes with sticky overhangs were mixed and ligated with T7 DNA ligase (NEB) for 4 h at 16 °C, followed by purification with DNA Clean & Concentrator 5 (Vistech) and native PAGE (10%).

hESC culture and genomic DNA extraction

Human embryonic stem cells (hESCs) H1 were derived from WiCell research institute (Madison, WI). H1 hESCs were maintained on Mitomycin C-treated MEFs in hESC culture medium consisting of 80% DMEM/F12 (Invitrogen), 20% Knockout serum replacement (KSR) (Invitrogen), 1% non-essential amino acids,1 mM L-glutamine, 0.1 mM beta-mercaptoethanol (all from Invitrogen) and 4 ng/ml basic fibroblast growth factor bFGF (P&A Biotech). H1 hESCs were passaged by dispase II (Invitrogen). The medium was changed every day. The genomic DNA was prepared by SDS–proteinase K digestion followed by phenol-chloroform extraction and ethanol precipitation.

Hydroxylamine-mediated blocking of endogenous 5fC

For O-ethylhydroxylamine-mediated selective blocking of 5fC, 9-mer and 76-mer chemically synthesized model sequences containing 5fC and 5hmC respectively were used (Table S1). 4 μg oligonucleotide or model DNA was incubated with 10 mM O-ethylhydroxylamine (Aldrich, 274992) in 50 μl 100 mM MES buffer (pH 5.0) at 37°C (850rpm) for 4 h in a thermomixer (Eppendorf). After the reaction, we used ethanol precipitation to purify the short DNAs with the help of glycogen (Invitrogen), whereas genomic DNA samples were purified by AMPure XP (Beckman Coulter). Reaction were monitored by MALDI-TOF (ABI 7500) and sanger sequencing.

Preparation of potassium ruthenate (K_2RuO_4)

0.15mmol Potassium perruthenate (Aldrich, 334537) was added to 0.5 M NaOH solution (1 mL) and vortexed to make sure all solid was dissolved. The color of the solution should be yellow/brown (Figure S1d). Incubate the solution at 25 °C for 2 days to produce potassium ruthenate (K_2RuO_4). There are bubbles produced and vortex to get rid

of the bubble twice in a day. After 2 days, the color should change into very deep red color (Figure S1d). The color of the oxidant solution is indicative of a successful preparation of K_2RuO_4. Divide the K_2RuO_4 solution into 100 parts and store at -20 °C refrigerator as 10x oxidant. The solution remains stable for at least two months under the storage condition. If color change or precipitation is found, oxidant then has deteriorated and should be freshly prepared again (Figure S1d).

K_2RuO_4-mediated oxidation of genomic DNA

Before 5hmC oxidation, genomic DNA was purified with AMPure XP beads (Beckman Coulter) and Micro Bio-Spin P-6 SSC column (Bio-Rad). Then DNA was denatured in 0.05 M NaOH (total volume 48.5 μL) for 30 min at 37°C. The reaction was snap cooled on ice-water bath (0 °C) for 5 min. Add 1.5ul 1x oxidant to the denatured DNA sample and briefly mix the reaction. Leave the oxidation reaction on ice for 1 h. The solution should remain orange during the 1h incubation period. Any other color (green, brown or black) of the solution implies a faulty oxidation and the prep then should not be used further for experiments (Figure S1d). The oxidized sample was purified by a Bio-Rad Micro Bio-Spin P-6 SSC column.

Quantification of 5mC, 5hmC and 5fC level by LC-MS/MS.

The pre-oxidation and oxidized genomic DNA samples were digested into nucleosides by 0.5 U nuclease P1 (Sigma, N8630) in 20 μL buffer containing 10 mM ammonium acetate, pH 5.3 at 42 °C for 6 h, followed by the addition of 2.5 μL 0.5 M MES buffer, pH 6.5 and 0.5 U alkaline phosphatase (Sigma, P4252). The mixture was incubated at 37 °C for another 6 h and diluted to 50 μL. 5 μL of the solution was injected into LC-MS/MS. The nucleosides were separated by ultra-performance liquid chromatography on a C18 column, and then detected by triple quadrupole mass spectrometer (AB SCIEX QTRAP 5500) in the positive ion multiple reaction monitoring (MRM) mode. The mass transitions of m/z 227.9 to 112.0 (C), m/z 242.0 to 126.0 (mC), m/z 257.9 to 142.1 (hmC), m/z 255.8 to 140.1 (fC) were monitored and recorded. Concentrations of nucleosides in DNA samples were deduced by fitting the signal intensities into the stand curves.

Synthesis of 5-(2-azidoethyl)-1,3-indandione (AI)

The synthesis AI was performed as previously described[2].

AI-mediated labeling of the newly generated 5fC

The reaction was performed in a suspension of AI (self-synthesized) in 20 mM Tris-HCl buffer (pH 8.0). We used 1 μg oligonucleotide or model DNA per 100-μL reaction and incubated the reaction mixture at 37 °C for 24 h in an Eppendorf tube in a thermomixer (Eppendorf, 850 r.p.m.). After reaction, we used ethanol precipitation to purify the short DNAs with the help of glycogen (Invitrogen), whereas genomic DNA samples were purified with MinElute PCR Purification Kit (QIAGEN) with the help of carrier RNA (Thermo Fisher Scientific).

Library preparation and click chemistry

The AI-labeled ssDNA was used directly for library preparation with the TELP protocol except PCR amplification[3]. Click chemistry was performed by adding the DBCO-S-S-PEG3-Biotin (Click Chemistry Tools, Cat. No. A112-10) to a final concentration of 400 mM, and incubated in the thermomixer (Eppendorf) for 1 h at 37°C (800 rpm). Purification steps were performed with DNA Clean & Concentrator 5 (Vistech) with the help of carrier RNA (Thermo Fisher Scientific).

Sanger sequencing on model DNA

Chemically labelled model DNAs that contain either 5mC, 5hmC, or 5fC were prepared as described above. PCR amplification was performed under common reaction conditions. PCR products were purified with DNA Clean & Concentrator 5 (Vistech) and Sanger-sequenced with unified Sequencing Primer. For oligonucleotides and primers, see **Table S1**.

Pull-down for enrichment of 5hmC-containing DNA

After Hydroxylamine-mediated blocking of endogenous 5fC, K_2RuO_4-mediated oxidation, AI-mediated labeling of the newly generated 5fC and click chemistry, the original 5hmC sites were labeled with a biotin group. The Dynabeads MyOne Streptavidin C1 (Invitrogen) was used to pull-down the biotin-labelled DNA with minor modifications on the suggested immobilizing procedure for nucleic acids. Specifically, the 1× binding and washing buffer (B&W buffer, pH 7.5) was added with 0.1% Tween-20. The canonical washing step was repeated for 4 times and twice with 1ml 1× TE buffer (pH 7.5) followed by 50 μL 1× SSC buffer (pH 7.0) washing. Then the beads were then sequentially washed once with 50 μL of 1× B&W buffer and once with 50 μL 1× TE buffer (pH 7.5). Beads were then resuspended and incubated in freshly prepared 50 mM DTT to release the 5hmC-containing strand for 1 h at 37 °C. Then the supernatant containing the desired DNA was purified with Micro Bio-Spin P-6 Gel Columns (Bio-Rad) to remove DTT, and we can obtain ~1% enriched DNA after pull down for hESC genomic DNA samples.

Quantitative PCR (qPCR) to test selectivity of 5hmC enrichment

The spike-in DNAs and primers for qPCR were prepared as described above. 0.1 pg of 5hmC and 5fC spike-in DNAs and 1 pg of Reference spike-in (with only regular C's) was added per 1ng of fragmented genomic DNA backgrounds. qPCR test was run in duplicates with SYBR Premix Ex TaqTM II (TAKARA) according to the manufacturer's instructions. Reactions were run on the Roche LightCycler® 96 Instrument. Three biological replicates were repeated to validate the results. For spike-in sequences and primers, see **Table S1**.

The calculation of relative gene body 5hmC level by hmC-CATCH and TAB-seq

We first calculated the mean 5hmC reads number of gene bodies using the hmC-CATCH data. We then calculated the relative 5hmC level for each gene by dividing the gene body reads number by the mean reads number. For the TAB-seq data, we first calculated the mean 5hmC density of gene bodies, and then also calculated the relative 5hmC level for each gene by dividing the gene body density by the mean density.

Preparation of cfDNA samples

Cell-free DNA (cfDNA) was isolated from peripheral blood of healthy individuals and HCC patients. Briefly, peripheral blood (about 4 ml per sample) was collected into EDTA anticoagulant tubes, and plasma was prepared within 6 h by centrifuging twice at 1,350×g for 12 min, and then centrifuging the supernatant at 13,500 × g for 12 min at 4 °C. cfDNA was isolated using QIAamp Circulating Nucleic Acid Kit (Qiagen) according to the company manual. In general, we could obtain about 10 ng cfDNA as the starting materials for hydroxymethylation analysis.

Data processing and analysis.

Firstly, low-quality bases, adaptor (trim_galore v.0.4.1,using options: -q 30, --stringency 3) and PCR duplicates (FASTX Toolkit v 0.0.13) were removed. Unique reads were trimmed for the poly-C sequence at the 3' end of reads (cutadapt v.1.16, using options: -a CCCCCCCCCCCCCCCCCCCCCCCCCC -g GATCTGGGGGGGGG -m 40). Reads with minimum length of 40 bp were mapped to the human (hg19) reference genome by bismark (v.0.15.0)[4] using options –bowtie2 -N 1 -L 30. Read visualization was done with IGV (v.2.3.90)[5]. The identification of 5hmC-enriched regions (peaks) in each sample was performed using MACS (v.1.4.2, using options: -g hs -p 0.0005)[6], with the corresponding input sample. HOMER (v.4.5)[7] calculated tag counts in each corresponding peaks to show the correlation between CATCH-Seq replicates, or between CATCH-Seq and hmC-Seal in hESC. The tag counts were transformed to $\log_2$ values to generate scatterplots and calculate the Pearson correlation. For each site, the numbers of converted "T" bases (NT) and unconverted "C" bases (NC) were further extracted. 5hmC sites with known SNPs (database, v. dbSNP 142) were discarded. For samples with severe deamination such as Formalin-fixed, Paraffin-embedded (FFPE) tissue genomic DNA, whole-genome sequencing before chemical labeling is required to discard deamination sites. All sites with less than ten total bases (NT+NC) or three NT were discarded for 5hmC calling. Then, we used the binomial distribution having parameters N as the sequencing depth (NC+NT) and p as the normal cytosine conversion rate (p_{100ng_1}= 0.61%, p_{100ng_2}= 0.58%, p_{50ng_1}= 0.70%, p_{50ng_2}= 0.71%, p_{10ng_1}= 0.72%, p_{10ng_2}= 0.72%, $p_{Healthy\_1}$= 0.82%, $p_{Healthy\_2}$= 1.02%, $p_{Healthy\_3}$= 0.95%, p_{Cancer_1}= 0.84%, p_{Cancer_2}=0.87%, p_{Cancer_3}=0.92%) to assess the probability of observing NT by chance. We considered 5hmC with Holm-Bonferroni method–adjusted $P < 0.01$ in both replicate samples that were located within 5hmC-enriched regions as genuine 5hmC sites. Genome annotation was done with HOMER. We also used RepeatMasker to remove sites in repeat regions to plot the distribution of ChIP-Seq signals around 5hmC sites.

We used featureCounts of subread (v.1.6.0)[8] to count cfDNA mapped reads in gene body regions (annotated by GENCODE V24[9]). Differentially modified genes in autosomes between healthy samples and cancer patients were analyzed using DESeq2 package (v. 1.20.0)[10] (|FC| >= 1.41 and p.adj =< 0.05). Furthermore, unbiased gene analysis was performed by PCA using rlog function to normalize reads counts in DESeq2 package[10]. Heatmaps of differential genes were analyzed in R (v.3.4.1). The deeptools software (v.3.0.1)[11] was used to generate cfDNA profiles and heatmaps. The DREME algorithm in

MEME suite (v.4.11.4)[12] was used to discover enriched motifs at ±15bp around 5hmC sites.

External data

H3K4me1(GSM466739), H3K27ac(GSM466732), CTCF(GSM822297), Pol II (GSM822300), chromatin accessibility (GSM878616), MYC (GSM935509) ChIP-Seq and RNA-seq (GSM2400152) data produced by the ENCODE Project Consortium[13]. TAB-Seq data sets were derived from GSE36173[14]. hmC-Seal data sets of H1 hESC genomic DNA were derived from GSE30173[15].

Supplementary Figures

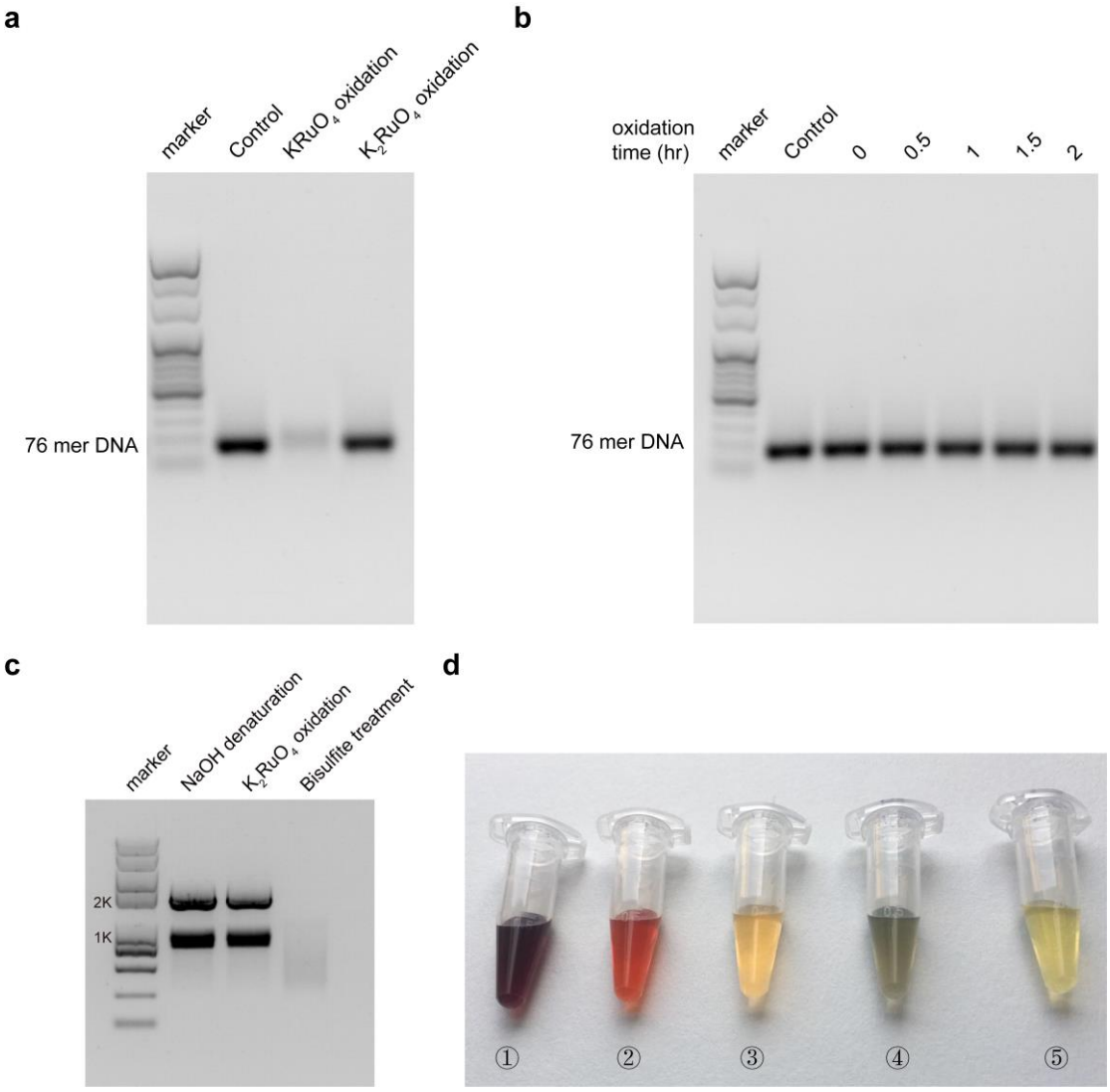

Figure S1. Optimized condition of 5hmC oxidation does not cause DNA degradation. (a) Agarose gel electrophoresis shows that K_2RuO_4-mediated oxidation causes no noticeable DNA degradation, while $KRuO_4$-mediated oxidation causes significant degradation. (b) Agarose gel electrophoresis shows that K_2RuO_4-mediated oxidation causes no noticeable degradation even at extended time points. (c) Agarose gel electrophoresis shows that K2RuO4-mediated oxidation causes no noticeable DNA degradation for 2K lambda DNA fragment, while bisulfite treatment causes significant degradation. Note that incomplete denaturation for long DNA may influence the oxidation efficiency; hence we recommend to fragment to less than 500 bp for maximal oxidation efficiency. (d) The accepted colors of 10X stock oxidant solution (①), 1X stock oxidant solution (②), oxidation reaction (③) and colors of bad oxidation reaction with ethanol (④) or Tris (⑤)contamination.

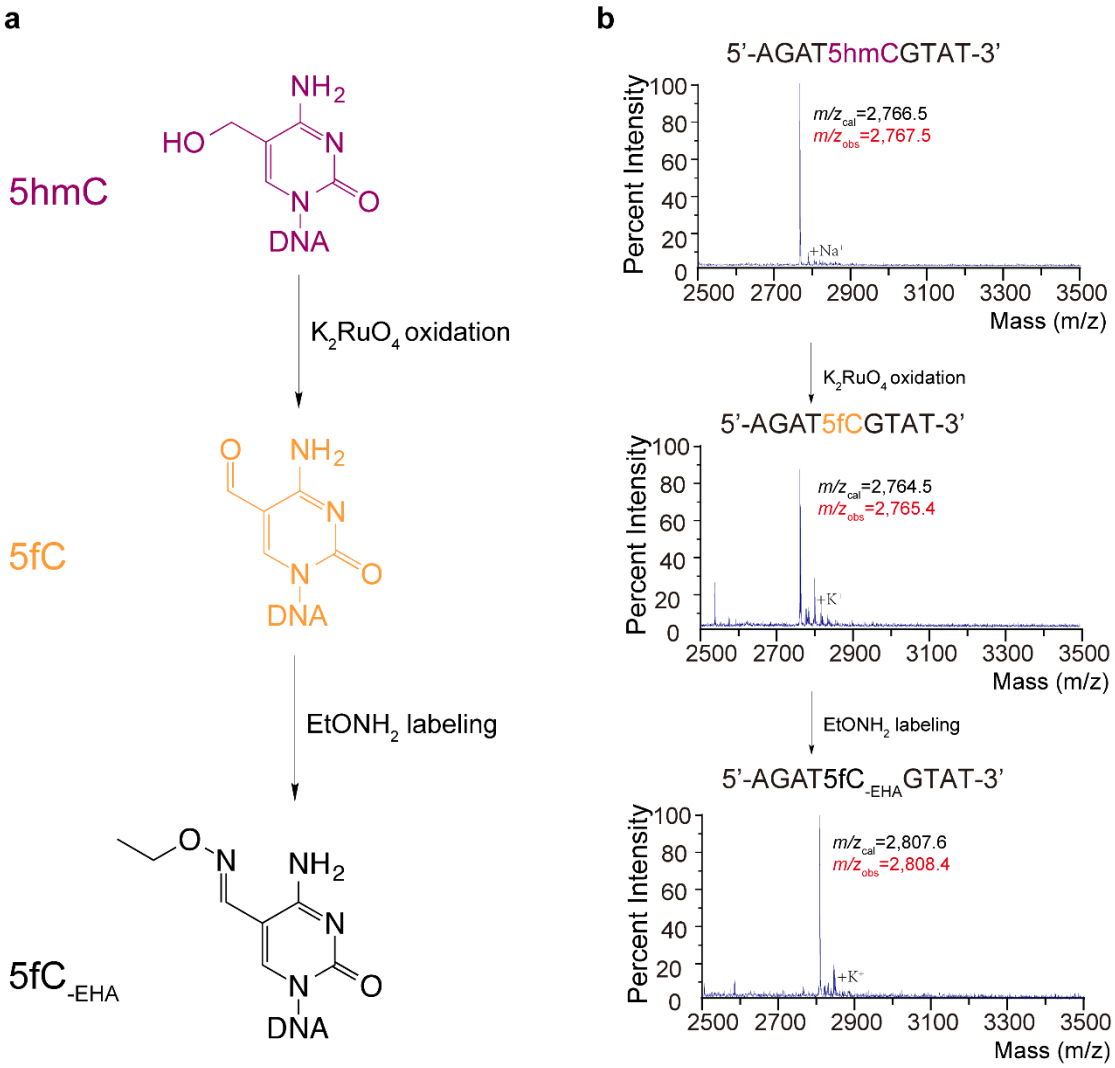

Figure S2. Optimized condition of K_2RuO_4 oxidation converts 5hmC to 5fC in model DNA. (a) A 5hmC-containing model sequence was oxidized to 5fC by K_2RuO_4. To fully separate 5fC from 5hmC in mass spectrometry, the K_2RuO_4-treated oligo was then labeled with $EtONH_2$ to enlarge the mass differences. (b) MALDI-TOF characterizations of K_2RuO_4-mediated oxidation of 5hmC in a 9mer model DNA. Calculated and observed molecular weights are shown.

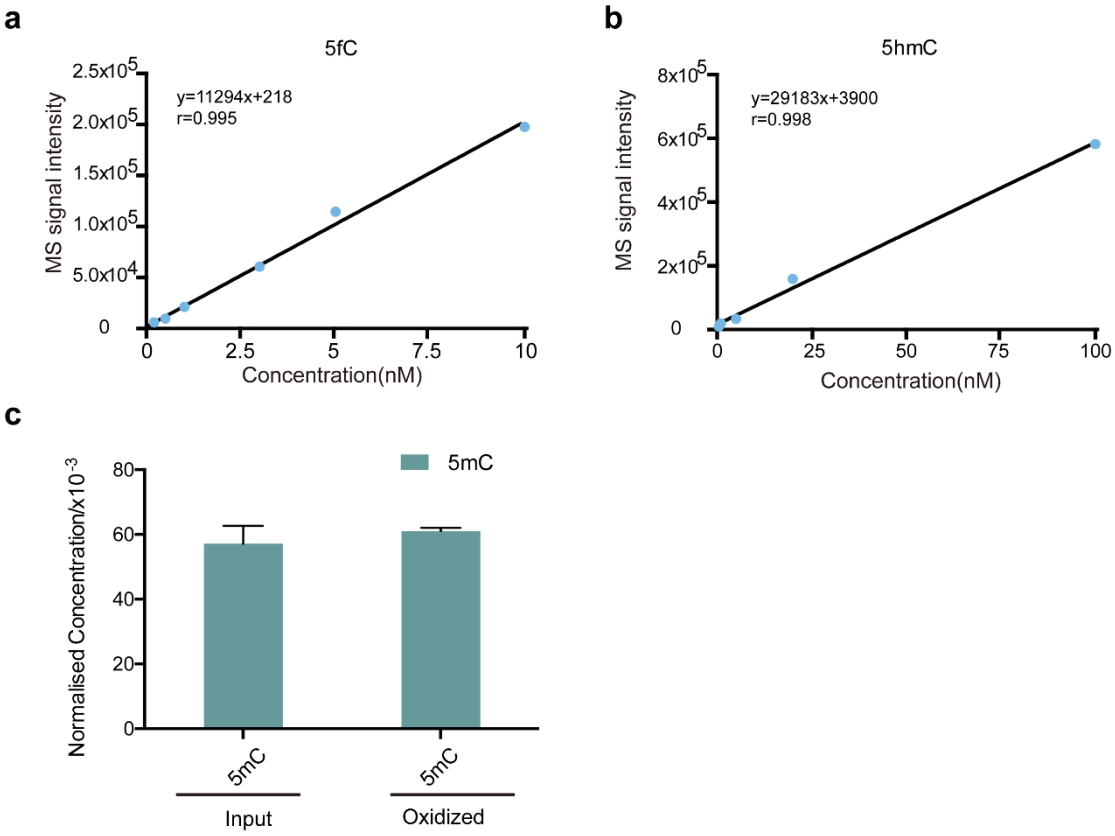

Figure S3. The stand curve of LC-MS/MS and the measurement of 5mC level. (a,b) The 5fC (a) and 5hmC (b) standard calibration curves were obtained using standard samples with different concentrations of pure 5fC or 5hmC, respectively. (c) 5mC level (normalized to C) of hESC genomic DNA before and after oxidation were quantified by LC-MS/MS. Data are represented as mean ± SEM, n=2.

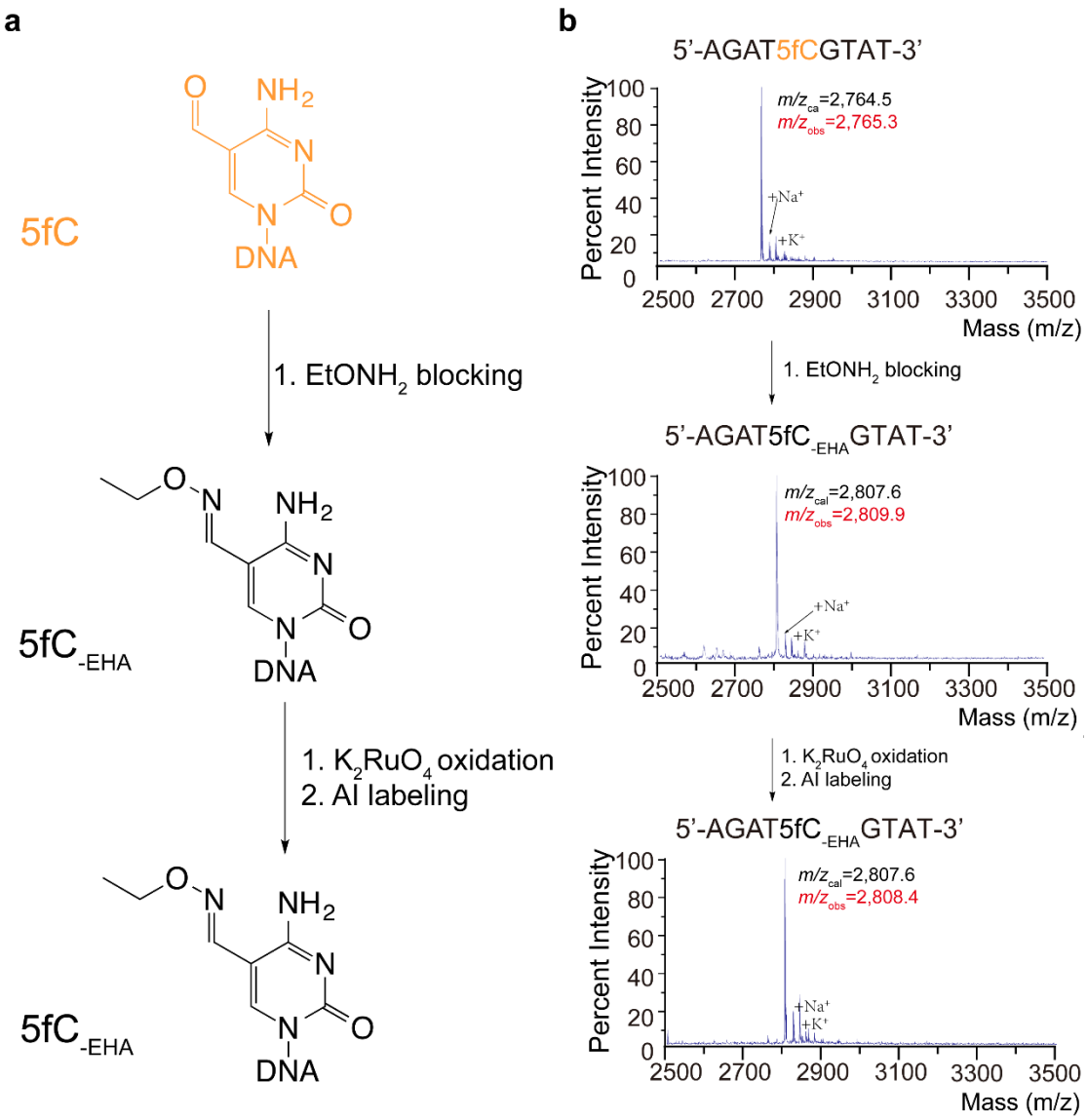

Figure S4. Blocking the endogenous 5fC. (a) $EtONH_2$-mediated reaction blocked the endogenous 5fC from being labeled by Al. (b) MALDI-TOF characterizations of a 9mer 5fC-containing model DNA to demonstrate the specificity of hmC-CATCH chemistry.

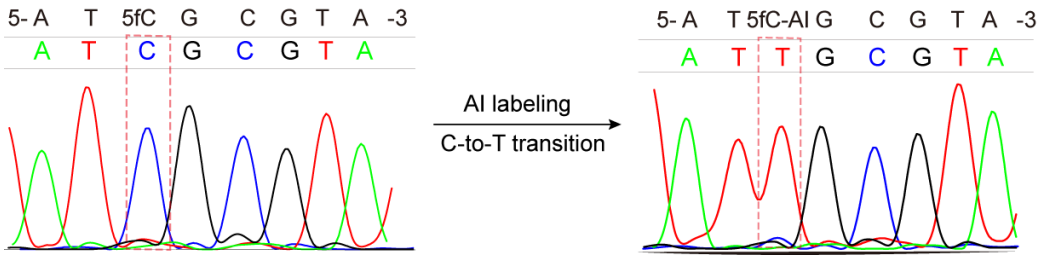

Figure S5. A representative Sanger sequencing result showing the C-to-T conversion after AI labeling. AI, an azido derivative of 1,3-indandione.

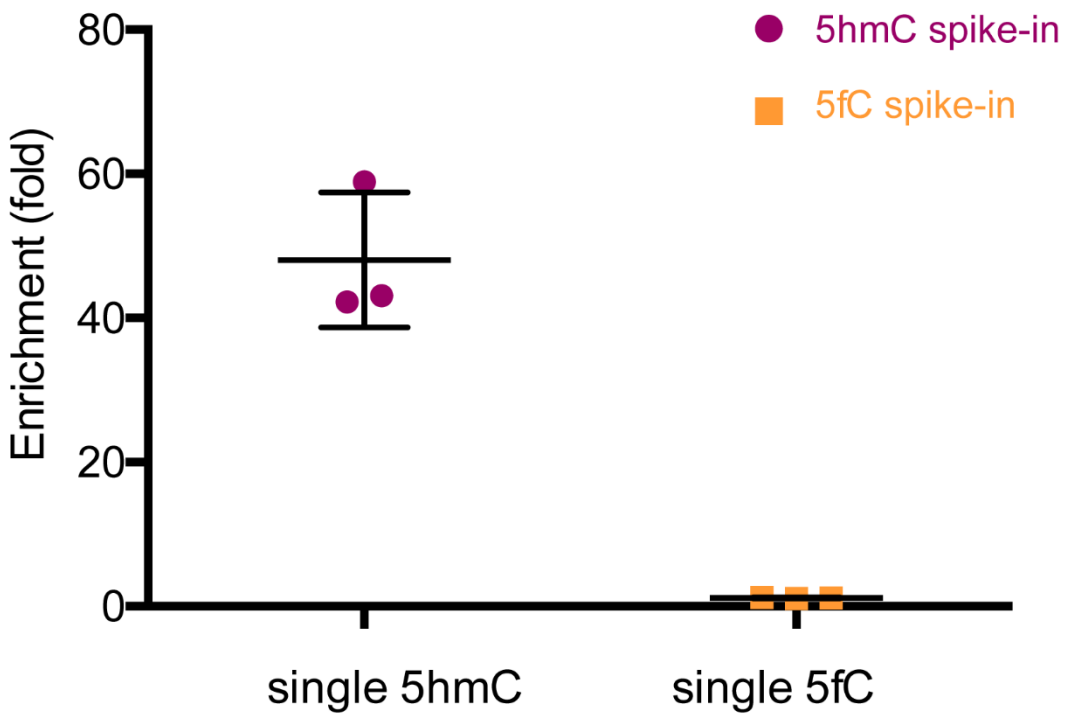

Figure S6. Enrichment of spike-in probes by qPCR. Values represent fold enrichment over the input (*n* = 3), normalized to the C-Ref probe (reference with only regular C's). "Single 5hmC": a synthetic DNA probe with a single 5hmC site; "single 5fC": a synthetic DNA oligo with a single 5fC site. Data are represented as mean ± SEM, n=3.

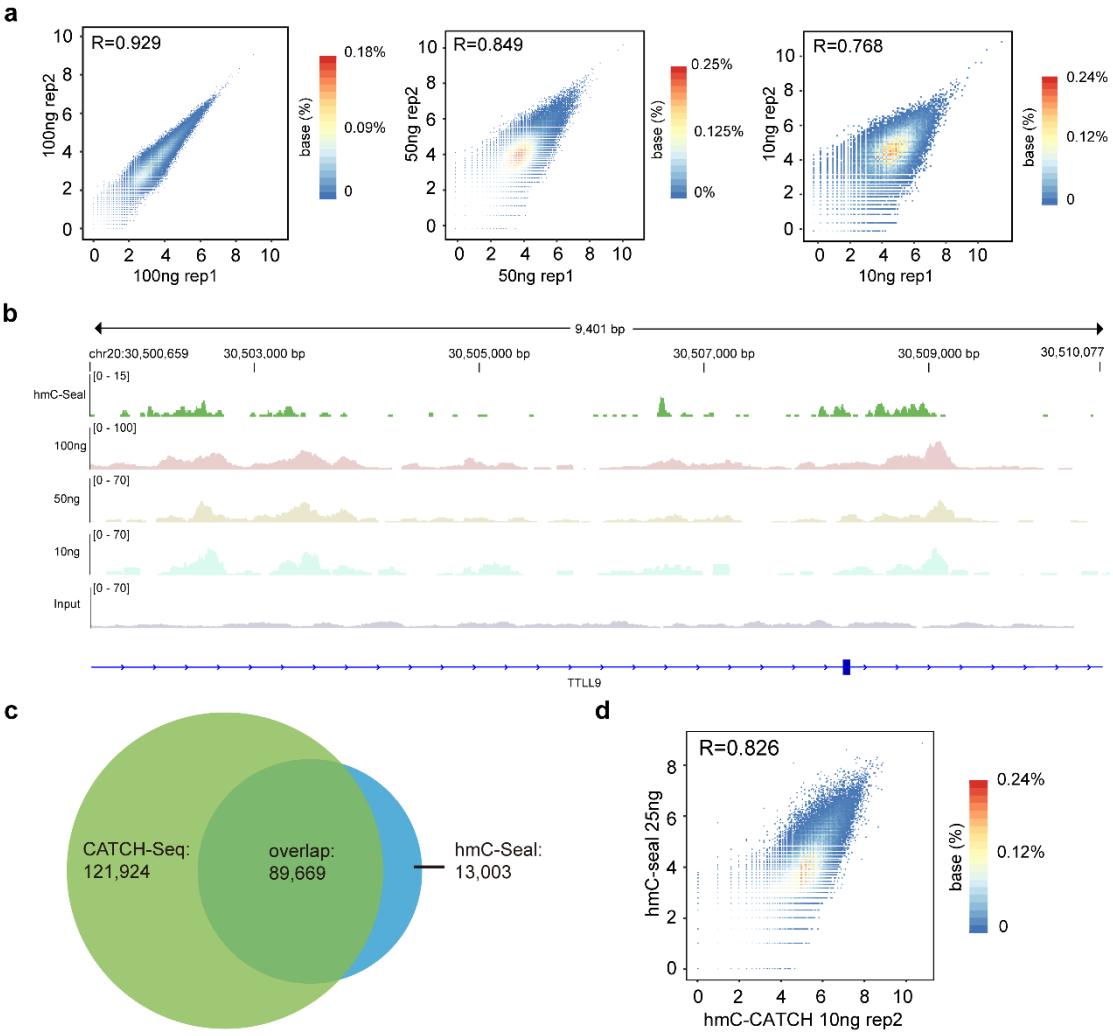

Figure S7. hmC-CATCH generates genome-wide 5hmC-enriched peaks from nano-scale hESC genomic DNA. (a) Scatterplots showing correlation between hmC-CATCH replicates, with Pearson correlation (r) displayed. Each dot represents a 5hmC-enriched peak and the number of points plotted is 211,593 for 100ng replicates, 105,322 for 50ng replicates and 60,338 for 10ng replicates, respectively. The read counts were transformed to $\log_2$ base. (b) Genome browser views of 5hmC signals detected in a 9,401 bp region from libraries generated with 10 ng, 50ng, 100 ng of starting genomic DNA from hESCs. The green view at the top is the 5hmC profile obtained using hmC-Seal with 25 ng genomic DNA, showing the similar 5hmC pattern. (c) A Venn diagram showing that hmC-Seal detected 5hmC-enriched peaks (102,672) largely overlap with hmC-CATCH detected 5hmC-enriched peaks (211,593). (d) Scatterplots showing high correlation between hmC-CATCH (using 10ng genomic DNA, replicate 2) and hmC-Seal (using 25ng genomic DNA) with Pearson correlation (r) displayed. Each dot represents a 5hmC-enriched peak and the number of points plotted is 90,218. The read counts were transformed to $\log_2$ base.

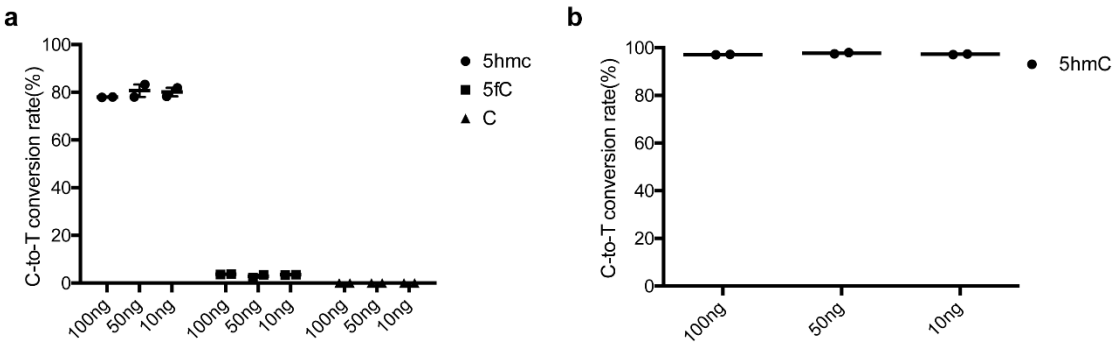

Figure S8. The C-to-T conversion rate of spike-in probes by high throughput sequencing. Rates in the input materials(a) and enriched samples(b) are shown. Data are represented as mean ± SEM, n=2.

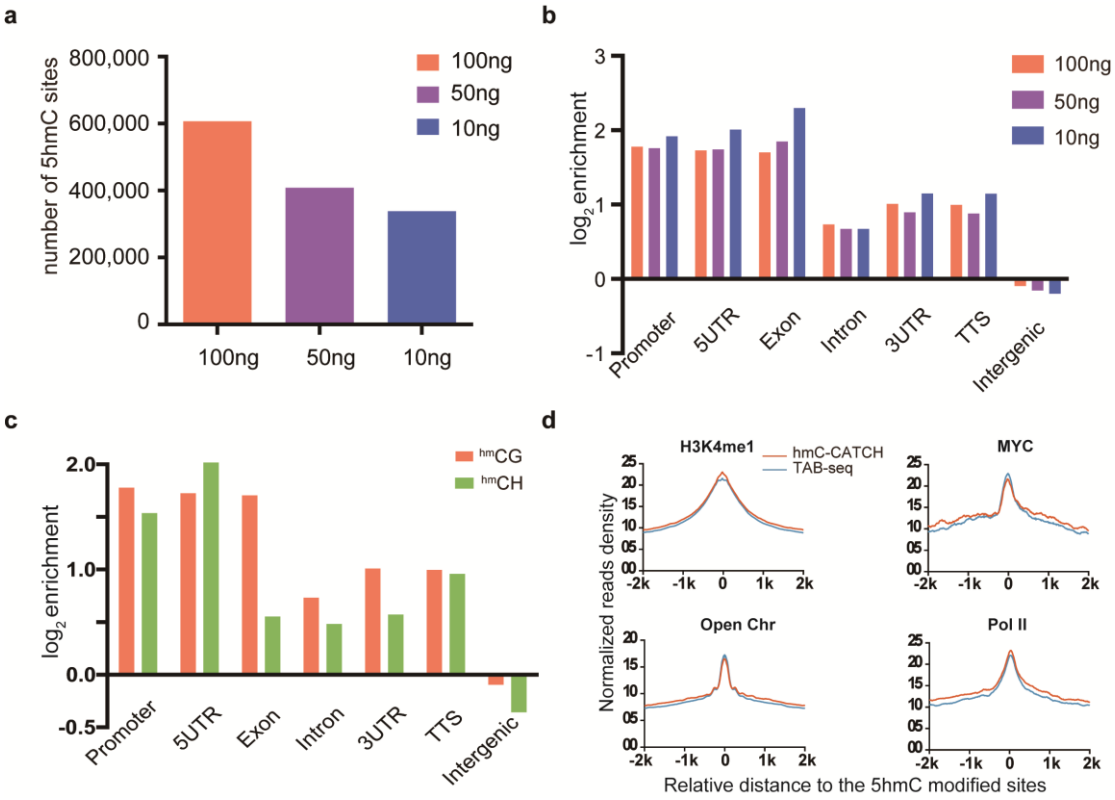

Figure S9. Distribution of 5hmC in the genome of hESC. (a) The number of detected 5hmC sites with 100ng, 50ng and 10ng of starting genomic DNA. (b) The relative enrichment of 5hmC sites detected with 100ng, 50ng and 10ng of starting genomic DNA in different genomic elements. (c) The relative enrichment of 5hmCG sites and 5hmCH sites in different genomic elements. TTS, transcription termination site. (d) Normalized read densities of 5hmC for H3K4me1, H3K27ac, open chromatin, Pol II, CTCF and MYC enriched regions in hESCs. H3K4me1(GSM466739), PolII (GSM822300), chromatin accessibility (GSM878616) and MYC (GSM935509) ChIP-Seq data produced by the ENCODE Project Consortium.

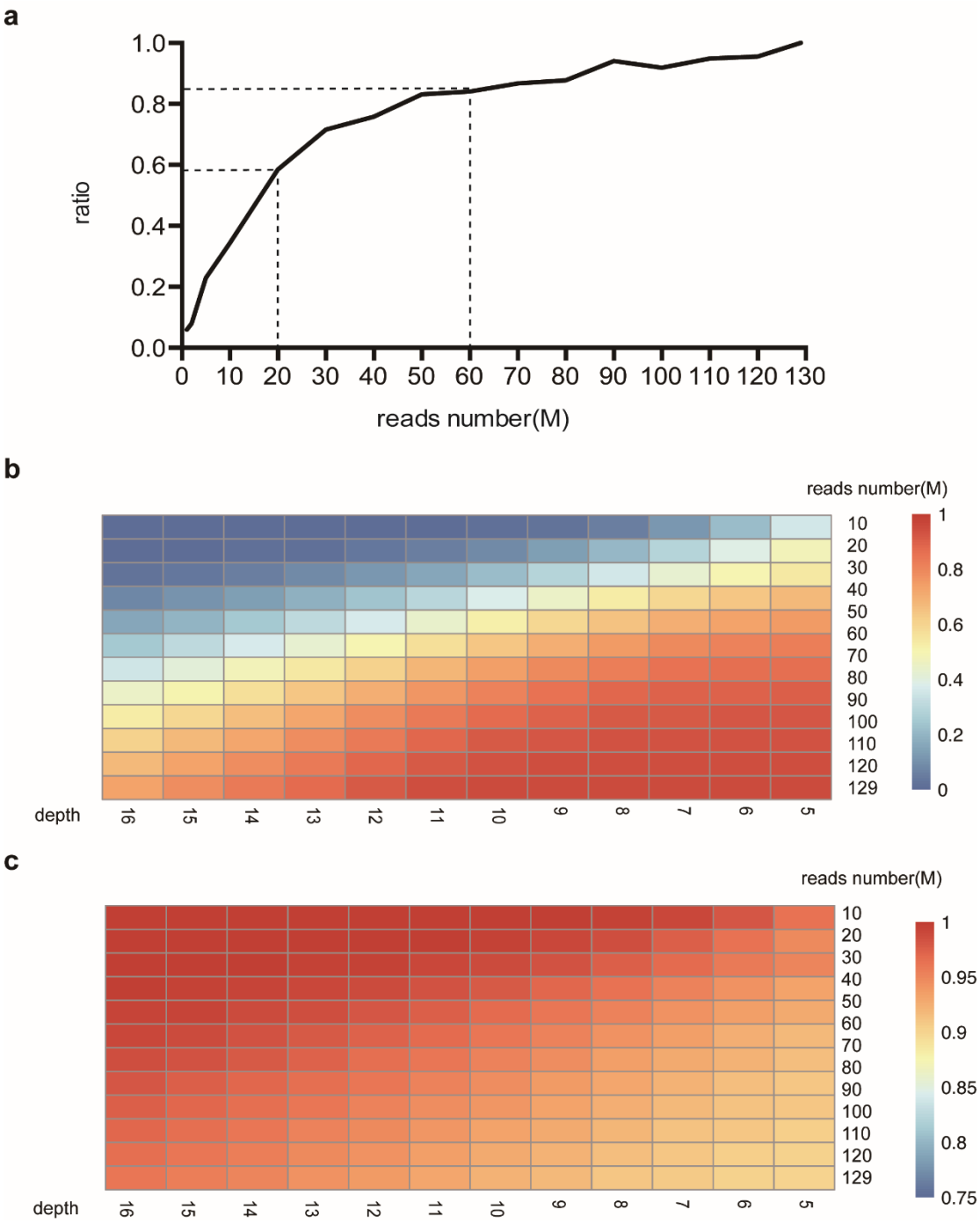

Figure S10. Downsampling of hmC-CATCH data to demonstrate its cost-effective feature. (a) Ratio of recovered 5hmC peaks during the downsampling process. With 60M sequencing reads, ~84% peaks can be detected; with even 20M reads, ~58% 5hmC peaks can still be found. (b,c) The sensitivity(b) and specificity(c) of base-resolution information of hmC-CATCH data in the downsampling process.

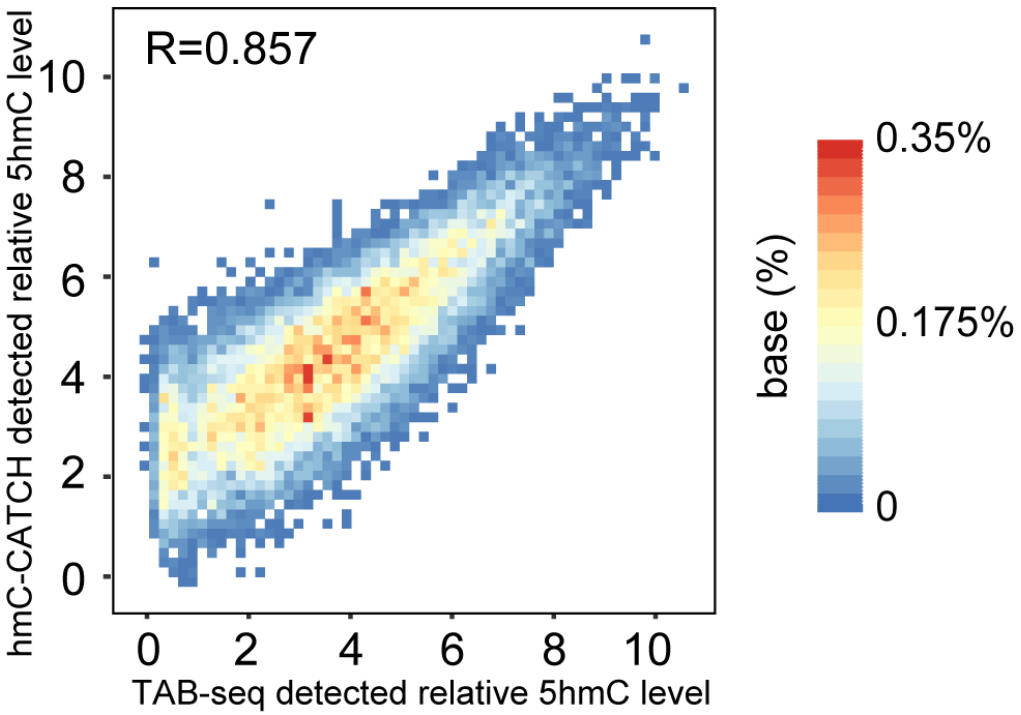

Figure S11. 5hmC signals detected by hmC-CATCH agreed well with that of TAB-seq. Scatterplots showing correlation between the relative 5hmC level of gene bodies detected by hmC-CATCH and TAB-seq, with Pearson correlation (r) displayed. Each dot represents the relative 5hmC level of a gene body and the number of points plotted is 15,956. The read counts were transformed to $\log_2$ base.

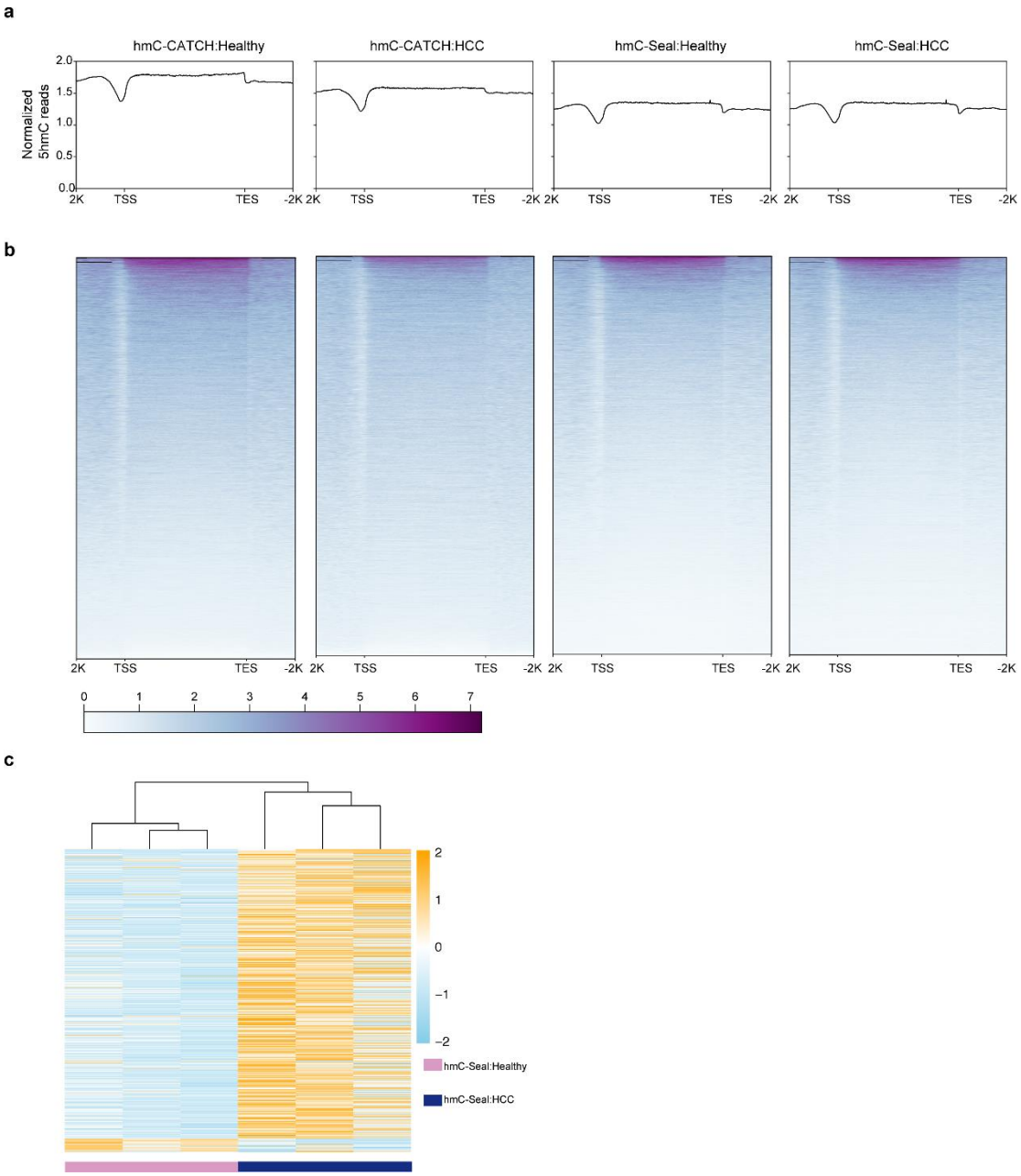

Figure S12. Comparison of hmC-CATCH and hmC-Seal sequencing data of cfDNA samples. (a,b) Average profiles (a) and heatmap (b) across gene regions ±2,000 bp for hmC-CATCH and hmC-Seal libraries with cfDNA. Each row represents a gene, ordered by the mean value signals. (c) The 675 differential genes detected by hmC-CATCH in healthy and HCC samples also show differential hydroxymethylation in healthy and HCC cfDNA datasets by hmC-Seal.

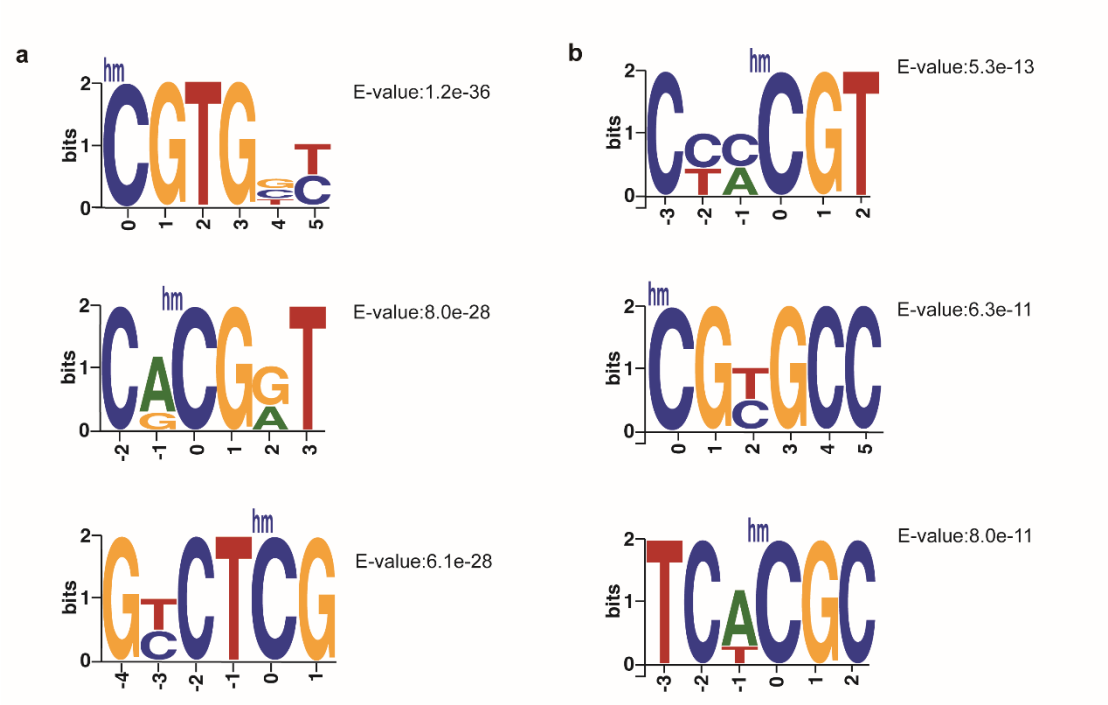

Figure S13. Promoter regions of 5hmC motif in cfDNA of healthy individuals (a) and HCC patients (b).

Table S3. Comparison of three sequencing techniques for genome-scale mapping of 5hmC.

	nano-hmC-Seal	hmC-CATCH	TAB-seq
Resolution	tens to few hundreds	single-base	single-base
Enrichment	yes	yes	no
Quantitative	no	no	yes
Mean-depth	NA	4.5	26.5
Costs	cost-effective	cost-effective	high
Bisulfite treatment	No	No	Yes

NA: Not available.

References

1. Wang, D.; Hara, R.; Singh, G.; Sancar, A.; Lippard, S. J., Nucleotide excision repair from site-specifically platinum-modified nucleosomes. *Biochemistry* **2003,** *42* (22), 6747-6753.

2. Xia, B.; Han, D.; Lu, X.; Sun, Z.; Zhou, A.; Yin, Q.; Zeng, H.; Liu, M.; Jiang, X.; Xie, W.; He, C.; Yi, C., Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nat Methods* **2015,** *12* (11), 1047-50.

3. Peng, X.; Wu, J. Y.; Brunmeir, R.; Kim, S. Y.; Zhang, Q. Y.; Ding, C. M.; Han, W. P.; Xie, W.; Xu, F., TELP, a sensitive and versatile library construction method for next-generation sequencing. *Nucleic acids research* **2015,** *43* (6).

4. Krueger, F.; Andrews, S. R., Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **2011,** *27* (11), 1571-1572.

5. Robinson, J. T.; Thorvaldsdottir, H.; Winckler, W.; Guttman, M.; Lander, E. S.; Getz, G.; Mesirov, J. P., Integrative genomics viewer. *Nature Biotechnology* **2011,** *29* (1), 24-26.

6. Zhang, Y.; Liu, T.; Meyer, C. A.; Eeckhoute, J.; Johnson, D. S.; Bernstein, B. E.; Nussbaum, C.; Myers, R. M.; Brown, M.; Li, W.; Liu, X. S., Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **2008,** *9* (9).

7. Heinz, S.; Benner, C.; Spann, N.; Bertolino, E.; Lin, Y. C.; Laslo, P.; Cheng, J. X.; Murre, C.; Singh, H.; Glass, C. K., Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* **2010,** *38* (4), 576-589.

8. Liao, Y.; Smyth, G. K.; Shi, W., featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014,** *30* (7), 923-930.

9. Harrow, J.; Frankish, A.; Gonzalez, J. M.; Tapanari, E.; Diekhans, M.; Kokocinski, F.; Aken, B. L.; Barrell, D.; Zadissa, A.; Searle, S.; Barnes, I.; Bignell, A.; Boychenko, V.; Hunt, T.; Kay, M.; Mukherjee, G.; Rajan, J.; Despacio-Reyes, G.; Saunders, G.; Steward, C.; Harte, R.; Lin, M.; Howald, C.; Tanzer, A.; Derrien, T.; Chrast, J.; Walters, N.; Balasubramanian, S.; Pei, B. K.; Tress, M.; Rodriguez, J. M.; Ezkurdia, I.; van Baren, J.; Brent, M.; Haussler, D.; Kellis, M.; Valencia, A.; Reymond, A.; Gerstein, M.; Guigo, R.; Hubbard, T. J., GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Research* **2012,** *22* (9), 1760-1774.

10. Love, M. I.; Huber, W.; Anders, S., Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **2014,** *15* (12).

11. Ramirez, F.; Dundar, F.; Diehl, S.; Gruning, B. A.; Manke, T., deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* **2014,** *42* (W1), W187-W191.

12. Bailey, T. L.; Boden, M.; Buske, F. A.; Frith, M.; Grant, C. E.; Clementi, L.; Ren, J. Y.; Li, W. W.; Noble, W. S., MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* **2009,** *37*, W202-W208.

13. Myers, R. M.; Stamatoyannopoulos, J.; Snyder, M.; Dunham, I.; Hardison, R. C.; Bernstein, B. E.; Gingeras, T. R.; Kent, W. J.; Birney, E.; Wold, B.; Crawford, G. E.; Bernstein, B. E.; Epstein, C. B.; Shoresh, N.; Ernst, J.; Mikkelsen, T. S.; Kheradpour, P.; Zhang, X. L.; Wang, L.; Issner, R.; Coyne, M. J.; Durham, T.; Ku, M. C.; Truong, T.; Ward, L. D.; Altshuler, R. C.; Lin, M. F.; Kellis, M.; Gingeras, T. R.; Davis, C. A.; Kapranov, P.; Dobin, A.; Zaleski, C.; Schlesinger, F.; Batut, P.;

Chakrabortty, S.; Jha, S.; Lin, W.; Drenkow, J.; Wang, H. E.; Bell, K.; Gao, H.; Bell, I.; Dumais, E.; Dumais, J.; Antonarakis, S. E.; Ucla, C.; Borel, C.; Guigo, R.; Djebali, S.; Lagarde, J.; Kingswood, C.; Ribeca, P.; Sammeth, M.; Alioto, T.; Merkel, A.; Tilgner, H.; Carninci, P.; Hayashizaki, Y.; Lassmann, T.; Takahashi, H.; Abdelhamid, R. F.; Hannon, G.; Fejes-Toth, K.; Preall, J.; Gordon, A.; Sotirova, V.; Reymond, A.; Howald, C.; Graison, E. A. Y.; Chrast, J.; Ruan, Y. J.; Ruan, X. A.; Shahab, A.; Poh, W. T.; Wei, C. L.; Crawford, G. E.; Furey, T. S.; Boyle, A. P.; Sheffield, N. C.; Song, L. Y.; Shibata, Y.; Vales, T.; Winter, D.; Zhang, Z. C.; London, D.; Wang, T. Y.; Birney, E.; Keefe, D.; Iyer, V. R.; Lee, B. K.; McDaniell, R. M.; Liu, Z.; Battenhouse, A.; Bhinge, A. A.; Lieb, J. D.; Grasfeder, L. L.; Showers, K. A.; Giresi, P. G.; Kim, S. K. C.; Shestak, C.; Myers, R. M.; Pauli, F.; Reddy, T. E.; Gertz, J.; Partridge, E. C.; Jain, P.; Sprouse, R. O.; Bansal, A.; Pusey, B.; Muratet, M. A.; Varley, K. E.; Bowling, K. M.; Newberry, K. M.; Nesmith, A. S.; Dilocker, J. A.; Parker, S. L.; Waite, L. L.; Thibeault, K.; Roberts, K.; Absher, D. M.; Wold, B.; Mortazavi, A.; Williams, B.; Marinov, G.; Trout, D.; Pepke, S.; King, B.; McCue, K.; Kirilusha, A.; DeSalvo, G.; Fisher-Aylor, K.; Amrhein, H.; Vielmetter, J.; Sherlock, G.; Sidow, A.; Batzoglou, S.; Rauch, R.; Kundaje, A.; Libbrecht, M.; Margulies, E. H.; Parker, S. C. J.; Elnitski, L.; Green, E. D.; Hubbard, T.; Harrow, J.; Searle, S.; Kokocinski, F.; Aken, B.; Frankish, A.; Hunt, T.; Despacio-Reyes, G.; Kay, M.; Mukherjee, G.; Bignell, A.; Saunders, G.; Boychenko, V.; Brent, M.; Van Baren, M. J.; Brown, R. H.; Gerstein, M.; Khurana, E.; Balasubramanian, S.; Zhang, Z. D.; Lam, H.; Cayting, P.; Robilotto, R.; Lu, Z.; Guigo, R.; Derrien, T.; Tanzer, A.; Knowles, D. G.; Mariotti, M.; Kent, W. J.; Haussler, D.; Harte, R.; Diekhans, M.; Kellis, M.; Lin, M.; Kheradpour, P.; Ernst, J.; Reymond, A.; Howald, C.; Graison, E. A. Y.; Chrast, J.; Valencia, A.; Tress, M.; Rodriguez, J. M.; Snyder, M.; Landt, S. G.; Raha, D.; Shi, M. Y.; Euskirchen, G.; Grubert, F.; Kasowski, M.; Lian, J.; Cayting, P.; Lacroute, P.; Xu, Y. H.; Monahan, H.; Patacsil, D.; Slifer, T.; Yang, X. Q.; Charos, A.; Reed, B.; Wu, L. F.; Auerbach, R. K.; Habegger, L.; Hariharan, M.; Rozowsky, J.; Abyzov, A.; Weissman, S. M.; Gerstein, M.; Struhl, K.; Lamarre-Vincent, N.; Lindahl-Allen, M.; Miotto, B.; Moqtaderi, Z.; Fleming, J. D.; Newburger, P.; Farnham, P. J.; Frietze, S.; O'Geen, H.; Xu, X. Q.; Blahnik, K. R.; Cao, A. R.; Iyengar, S.; Stamatoyannopoulos, J. A.; Kaul, R.; Thurman, R. E.; Wang, H.; Navas, P. A.; Sandstrom, R.; Sabo, P. J.; Weaver, M.; Canfield, T.; Lee, K.; Neph, S.; Roach, V.; Reynolds, A.; Johnson, A.; Rynes, E.; Giste, E.; Vong, S.; Neri, J.; Frum, T.; Johnson, E. M.; Nguyen, E. D.; Ebersol, A. K.; Sanchez, M. E.; Sheffer, H. H.; Lotakis, D.; Haugen, E.; Humbert, R.; Kutyavin, T.; Shafer, T.; Dekker, J.; Lajoie, B. R.; Sanyal, A.; Kent, W. J.; Rosenbloom, K. R.; Dreszer, T. R.; Raney, B. J.; Barber, G. P.; Meyer, L. R.; Sloan, C. A.; Malladi, V. S.; Cline, M. S.; Learned, K.; Swing, V. K.; Zweig, A. S.; Rhead, B.; Fujita, P. A.; Roskin, K.; Karolchik, D.; Kuhn, R. M.; Haussler, D.; Birney, E.; Dunham, I.; Wilder, S. P.; Keefe, D.; Sobral, D.; Herrero, J.; Beal, K.; Lukk, M.; Brazma, A.; Vaquerizas, J. M.; Luscombe, N. M.; Bickel, P. J.; Boley, N.; Brown, J. B.; Li, Q. H.; Huang, H. Y.; Gerstein, M.; Habegger, L.; Sboner, A.; Rozowsky, J.; Auerbach, R. K.; Yip, K. Y.; Cheng, C.; Yan, K. K.; Bhardwaj, N.; Wang, J.; Lochovsky, L.; Jee, J.; Gibson, T.; Leng, J.; Du, J.; Hardison, R. C.; Harris, R. S.; Song, G.; Miller, W.; Haussler, D.; Roskin, K.; Suh, B.; Wang, T.; Paten, B.; Noble, W. S.; Hoffman, M. M.; Buske, O. J.; Weng, Z. P.; Dong, X. J.; Wang, J.; Xi, H. L.; Tenenbaum, S. A.; Doyle, F.; Penalva, L. O.; Chittur, S.; Tullius, T. D.; Parker, S. C. J.; White, K. P.; Karmakar, S.; Victorsen, A.; Jameel, N.; Bild, N.; Grossman, R. L.; Snyder, M.; Landt, S. G.; Yang, X. Q.; Patacsil, D.; Slifer, T.; Dekker, J.; Lajoie, B. R.; Sanyal, A.; Weng, Z. P.; Whitfield, T. W.; Wang, J.; Collins, P. J.; Trinklein, N. D.; Partridge, E. C.; Myers, R. M.; Giddings, M. C.; Chen, X.; Khatun, J.; Maier, C.; Yu, Y. B.; Gunawardena, H.; Risk, B.; Feingold,

E. A.; Lowdon, R. F.; Dillon, L. A. L.; Good, P. J.; Consortium, E. P., A User's Guide to the Encyclopedia of DNA Elements (ENCODE). *Plos Biol* **2011,** *9* (4).

14. Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C. X.; Zhang, L.; Kim, A.; Li, X.; Dai, Q.; Shen, Y.; Park, B.; Min, J. H.; Jin, P.; Ren, B.; He, C., Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **2012,** *149* (6), 1368-80.

15. Szulwach, K. E.; Li, X. K.; Li, Y. J.; Song, C. X.; Han, J. W.; Kim, S.; Namburi, S.; Hermetz, K.; Kim, J. J.; Rudd, M. K.; Yoon, Y. S.; Ren, B.; He, C.; Jin, P., Integrating 5-Hydroxymethylcytosine into the Epigenomic Landscape of Human Embryonic Stem Cells. *Plos Genetics* **2011,** *7* (6).