Supporting Information

Jump-Seq: Genome-wide Capture and Amplification of 5hmC Sites

Lulu Hu[1,2#], Yuwen Liu[6,7#], Shengtong Han[4,7#], Lei Yang[3#], Xiaolong Cui[1,2], Yawei Gao[3], Qing Dai[1,2], Xingyu Lu[1,2], Xiaochen Kou[3], Yanhong Zhao[3], Wenhui Sheng[5], Shaorong Gao[3], Xin He[7*] and Chuan He[1,2*]

1) Department of Chemistry, Department of Biochemistry and Molecular Biology, and Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois, USA.

2) Howard Hughes Medical Institute, The University of Chicago, Chicago, Illinois, USA.

3) Clinical and Translational Research Center of Shanghai First Maternity and Infant Hospital, Shanghai Key Laboratory of Signaling and Disease Research, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China

4) Joseph J. Zilber School of Public Health, University of Wisconsin, Milwaukee, WI, 53205

5) Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, WI, 53233

6) Agricultural Genomes Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, Guangdong, China.

7) Department of Human Genetics, University of Chicago, Chicago, IL

 # These authors contributed equally

Experimental Section

Cell culture

mESCs were cultured in feeder-free gelatin-coated plates in Dulbecco's Modified Eagle Medium (DMEM) (Invitrogen Cat. No. 11995) supplemented with 20% FBS (GIBCO), 2 mM L-glutamine (GIBCO), 0.1 mM 2-mercaptoethanol (Sigma), 1 × nonessential amino acids (GIBCO), 1,000 units/ml LIF (Millipore Cat. No. ESG1107), 1 × pen/strep (GIBCO), 3 mM CHIR99021 (Stemgent), and 1

mM PD0325901 (Stemgent). mESCs were passaged every other day with Trispin (GIBCO) and seeded on gelatin-coated plates. Cells were kept under standard 37 ℃ and 5% CO_2 conditions.

Jump-seq design details

The detailed design of Jump-seq is as follows: 1: We have assembled transposome by incubating transposase with biotin labeled P7 adapter sequence. Genomic DNA is fragmented by transposome and tagged with biotin-P7 via “cut-paste” mechanism. 2: 5hmC in genomic DNA is labeled with a modified azide-glucose using the β-GT mediated selective chemical labeling. 3: a hairpin DNA oligonucleotide with P5 adapter sequence and a unique sequence carrying a DBCO group is covalently connected to the azide-modified 5hmC. The loop part carries three deoxyribose uracils by design; 4: primer extension mediated by DNA polymerases starts from the hairpin DNA attached to 5hmC as indicated and “Jump” happens at the modified 5hmC sites. The “jump” pattern is defined by the behavior of polymerases encountering the "gap” (steric hindrance composed of the azide-glucose and DBCO linker) between the unique DNA sequence attached to 5hmC and genomic DNA. During this “jump” some polymerases land 1∼10 bases 5' ahead of the 5hmC site while others slide back to the genomic strand (1 ~ 10 base towards the 3') and then extend to the 5’ direction on the genomic template until reaching P7 adapter to synthesize new strand. Less frequently, polymerases would land exactly on the modified 5hmC sites containing sugar and DBCO linker, thus forming a "valley" at the exact 5hmC site. The dU linker in the hairpin motif tethered to 5hmC is cleaved by USER® Enzyme (NEB) to release the newly synthesized strand. The extension products with P5 and P7 adapters are subsequently amplified and sequenced; 5: 5hmC single sites are inferred from the juncture connecting the hairpin sequence and any genomic DNA sequence.

Crosslinking and primer extension optimization

We used 76 mer model DNA (Table S1) to optimize the crosslinking and primer extension efficiency as initial exploration (Figure S1). 37 ℃ for 24h was identified as the optimal temperature reaction condition for crosslinking (Figure S1b). Sequenase version 2.0 DNA Polymerase (Thermo Fisher Scientific) could successfully “Jump” over the steric hindrance composed by the glucose and DBCO linker and “land” on the genomic DNA template (Figure S1c). Then we integrated the optimized “Jump” parameters with transposase-based library construction strategy for 5hmC high-throughput sequencing (Figure 1).

Hyperactive Tn5 production and assembly

Hyperactive Tn5 was produced following the published protocol with some details modified[1, 2]. In brief, pTXB1-Tn5 plasmid (Catalog # 60240, Addgene) was transformed into C3013 cells (NEB). 6 L culture was grown at 37 ℃ to OD_{600} ~ 1. The culture was then cooled to 16 ℃ and IPTG was added to 0.1 mM. After continued growth for additional 16 h at 16 ℃, cells were harvested and lysed by French press in lysis buffer (25 mM Tris, pH = 8.0, 0.7 M NaCl, 0.1% Triton X-100). Then the lysate was centrifuged at 13,000 rpm for 30 min, followed by adding 10% PEI (Sigma) to remove E. coli DNA. The supernatant was loaded on a chitin column (NEB), followed by extensively wash by lysis buffer and 36 h cleavage with 100 mM DTT. The elution was purified by Source 15S column

using Buffer A: 20 mM Tris 7.5 and Buffer B: 20 mM Tris 7.5, 200 mM NaCl. Tn5 comes out at conduction ~ 30. Collected fractions were pooled and subjected to gel filtration with gel filtration buffer (20 mM HEPES-KOH pH 7.6, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT). Finally, the 3-step purified Tn5 was concentrated to OD_{280} = 3.0 and stored at -80 ℃ after adding 1.1 vol 100% glycerol and 0.33 vol of 2 × Tn5 dialysis buffer (20 mM HEPES-KOH pH = 7.6, 200 mM NaCl, 0.2 mM EDTA, 2 mM DTT, 20% glycerol). For assembly, oligonucleotides (Jump-P7 adapter and Tn5 bottom adaptor) were resuspended in 1 × TE (pH = 8.0) to a final concentration of 100 μM each. Equimolar amounts of Jump-P7 adapter and Tn5 bottom adaptor were mixed, denatured at 95 ℃ for 10 min and slowly cooled down to anneal. One vol Tn5 was added 0.143 vol annealed adaptor, mixed gently but thoroughly and the Tn5-adaptor complex was incubated for 60 min at RT and stored at -80 ℃.

Genomic DNA preparation and tagmentation

mESCs genomic DNA was purified using Quick-DNA™ Microprep Plus Kit (Zymo Research). The Tn5 reaction protocol has been described previously[1]. In brief, genomic DNA was fragmented and tagmented by mixing 4 μL 5 × TAPS-$MgCl_2$-PEG 8,000 (for DNA < 1 ng) or 5 × TAPS-$MgCl_2$-DMF, 1 μL of the assembled Tn5 and adaptor complex, genomic DNA and suitable volume of H_2O up to 20 μL. Buffer used were: TAPS-$MgCl_2$-PEG 8,000 (50 mM TAPS-NaOH, pH 8.5, 25 mM $MgCl_2$, 40% PEG 8,000) or 5 × TAPS-$MgCl_2$-DMF (50 mM TAPS-NaOH, pH 8.5, 25 mM $MgCl_2$, 50% DMF). For 100 cells, NP-40 was added to final concentration 0.15%. For 48 ng genomic DNA, 1 μL of 2-fold diluted Tn5 is sufficient; for 24 ng genomic DNA, Tn5 needs to be diluted 4 folds and for less genomic DNA, Tn5 should be diluted accordingly. For 600 pg genomic DNA or 100 cells, Tn5 should be diluted 320 folds and added 1 μL into the reaction. The reactions were incubated for 7 min at 55℃, and Tn5 was inactivated for 7 min at 55℃ after adding 1 μL proteinase K (Thermo Fisher Scientific). Typically, all the genomic DNA was converted to fragments of average size 200-300 bp. The fragmented DNA was purified by DNA Clean & Concentrator™-5 kit (Zymo Research) and eluted into H_2O.

β-GT, Tet purification, 5hmC labeling and Jump P5 probe crosslinking

T4 bacteriophage β-GT and Tet was prepared as previously described[3, 4]. The 5hmC labeling reactions were performed in a 20 μL solution containing 50 mM HEPES buffer (pH 8.0), 25 mM $MgCl_2$, tagmented genomic DNA (100-500 bp), 100 μM UDP-6-N3-Glu, and 1 μM wild-type β-GT. The reactions were incubated for 1 h at 37 °C. After the azide-glucose transfer reaction, the DNA substrates were purified by Clean & Concentrator™-5 kit (Zymo Research) and reconstituted in H_2O. The click chemistry was performed with addition of 0.8 nmol dibenzocyclooctyne modified Jump P5 probe into the DNA solution, and the reaction mixture was incubated for 24 h at 37 °C.

Primer extension, Jump and library construction

The click products were incubated with Dynabeads™ MyOne™ Streptavidin C1 (Thermo Fisher Scientific) in 1 × B & W Buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl) at RT for 20min (the click reactions with different barcoded Jump P5 probe could be pooled together for purification), followed by extensively washing steps: 1 × B & W Buffer 1 time, 0.15 M NaOH 1 time, 0.1 M NaOH

twice, 1 × B & W Buffer with 200M Tris-HCl pH 7.5 one time, H_2O three times and finally the beads were resuspended in 6 μL H_2O and 2 μL of 5 × Sequenase Reaction Buffer for annealing (the loop shape Jump probe formed hairpin to initiate primer extension). The reaction was heated at 95 °C and slowly ramped to 4 °C in Thermocycler. 1 μL Sequenase Version 2.0 DNA Polymerase (Thermo Fisher Scientific) and 1 μL dNTP (NEB) were added in the annealed solution, incubated at 37 °C for 2 h. Then 1 μL USER® Enzyme (NEB) was added to cleave at the dU linker region to release the newly synthesized strand. The 11 μL Jump products were subjected to library construction with limited cycle amplification. Next-generation sequencing (NGS) libraries were generated using NEBNext® Ultra™ II Q5® Master Mix (NEB), NEBNext® Multiplex Oligos for Illumina® (NEB) for NEB kit compatible Jump P7 adaptor and Jump P5 probe and Illumina Nextera DNA oligo sets for Nextera DNA kit compatible Jump P7 adaptor and Jump P5 probe.

5hmC Spike-in sequence, preparation and validation

The 200bp spike-in sequences (Table S1) were prepared by ligation using short oligonucleotides synthesized at Integrated DNA Technologies (IDT) (Table S2). As the spike-in sequence is 200 bp, so we separate them into 2 parts (spike-in 1F and spike-in 2F). spike-in 1F is biotinylated, 5hmC modified and spike-in 1F is adenylated at 5' while dideoxy C modified at 3'. Spike-in 1F, 2F and R oligonucleotides were resuspended in 10 mM Tris pH = 7.5 to 20 μM. Equimolar amount of them were mixed, annealed and ligated with Quick Ligation™ Kit (NEB). The reactions were purified by TBE PAGE gel, followed by final purification with Sep-Pak C18 cartridge (Waters). The single stranded spike-in oligoes were annealed with their complementary stands, ready for Jump-seq validation.

For the 'Jump' pattern test and method validation, 0.5% 5hmC modified single site or two sites spike-in was added into mESC genomic DNA (Figure 2). For the 'Jump' false positive and effectiveness evaluation, gradient of 5hmC modified spike-in was added into 48 ng mESC genomic DNA: 0% (0 pg), 1% (2.4 pg), 5% (12 pg), 10% (24 pg), 25% (60 pg), 50% (120 pg), 75% (180 pg) and 100% (240 pg); spike-in without 5hmC modification was added to make the final spike-in amount to be 100% (240 pg).

False positive and enrichment validation with 5hmC Spike-in model

6 ng each biotin-P7 adapter modified background oligoes (7 background oligoes in total) were combined with 6 ng of 5hmC-Jump single site spike-in (Nextera DNA kit compatible) (Table S3) and subjected to 5hmC Jump-seq. The false positive rate was calculated by (negative background mapped reads) / (5hmC spike-in model mapped reads). The 5hmC enrichment was calculated by (5hmC spike-in hmCpG site mapped reads / total hmCpG site number) / (negative background CpG sites mapped reads / total CpG site number). The enrichment fold is: (22 / 1) / (0.78/ (72)) =2030.8 and false positive rate is: 0.78 / 22 = 3.6% (Figure 2c).

Reads Mapping

For each 50 bp Illumina sequencing read, umi_tools were used to trim the first 8 bases which constitute a unique molecular identifier (UMI) and attach the UMI information to the name of each read. The UMI sequence of each read was used later to remove PCR duplicates (reads starting at the same genomic location and sharing a same UMI sequence are likely to arise from one DNA fragment with a hydroxymethylated site, thus need to be collapsed and counted as one read). After extracting UMI, cutadapt was used to retain reads with a Jump-Seq barcode "NNNNNNN" (The Jump barcode used for each library used in this study could be found in Table S4) and to trim the barcode from each of these retained reads. We also used cutadapt to remove the bases that match the invariant adaptor sequence at the 3'end of each read. Then bowtie was used to map the reads to the mouse genome (mm9) not allowing any mismatch in the first 35 bp of each read. Only uniquely mapped reads were kept and processed with umi_tools to remove PCR duplicates based on the UMI sequence of each read.

5hmC sites in mouse ES cells identified by TAB-Seq [5]were used as references to study the distribution of distance between Jump-Seq read 5'ends and bona fide 5hmC sites. For plus-strand 5hmC sites, the distance distribution was plotted using reads aligning to minus-strand. For minus-strand 5hmC sites, reads aligning to plus-strand were used. To disentangle Jump-Seq signals from single 5hmC sites, each 5hmc site was extended 100 bps both ways and only those extended intervals that don't overlap with others were used for calculating reads coverage. Reads coverage (5' position) for each 5hmC-containing 201 bp interval was calculated by bedtools and added up across all intervals.

Peak window calling

Since reads are generated from 5hmC's (unknown), it is more appropriate and reasonable to check the distribution of reads over C's (known), rather than over every possible base. To avoid multiple counting of one read, the 5' end of every reads was used when calling the coverage, instead of the entire length of the reads. Intuitively, more C's (not necessarily 5hmC) in a region, more reads will be falling in that region. So, we need to adjust the effect of non-5hmC. We formalize this problem to be a statistical testing problem. Specifically, in a region, we want to test if there is 5hmC sites by looking at the reads enrichment comparing to the null hypothesis of no 5hmC. Denote R by the total number of mapped reads, K by the total number of C's in the whole genome. Under null hypothesis that there is no 5hmC, each read is independently aligned into any regions with C with equal probability. In total, there are R Bernoulli trials by treating assigning each read as an independent trial. For a window with L C's, the success probability of having one reads is $\frac{L}{K}$, and thus the number of reads is essentially the number of successful Bernoulli trials in R trials, which is the Binomial random variable by definition as

$$\boldsymbol{X \sim Bin(R, \frac{L}{K})}$$

where $\frac{L}{K}$ is the probability of one read falling in the window. Let O be the number of observed reads in the window. The P value is calculated as

$$P(X > O | X \sim Bin\left(R, \frac{L}{K}\right))$$

For any region in the genome, we calculated its p value. Regions with small p values indicate there are more likely to contain 5hmC sites. Meanwhile, reads could be mistakenly falling into a window by random noise which could be reads alignment error or from other sources. To control false discovery rate, we applied Benjamini-Hochberg Procedure to p values for all windows at FDR 0.05 to call peaks.

Validation of 5hmC Jump-seq data

For the enrichment test, at different genomic concentrations, we called peak windows of 20 bp and selected "enriched" peak windows at FDR 0.05 (Figure 3a, b). The ratios of enriched peak windows were calculated when overlapping with Nano-seal and strong TAB-seq peak windows extended by 20 bp (10 bp upstream and 10 bp downstream) from strong TAB-seq base. The same number of peak windows were randomly picked from whole genome. For the sensitivity test, we combined twelve 48 ng samples of mESC Jump-seq data, called "enriched" peak windows of 20 bp at FDR 0.05.

Accession codes

Sequencing data reported in this paper has been deposited into the Gene Expression Omnibus (GEO) under accession number GSE127906. TAB-Seq data sets were obtained from GSE36173. 5hmC-Seal data sets of mESC genomic DNA were obtained from GSE77967.

Supplementary Figures

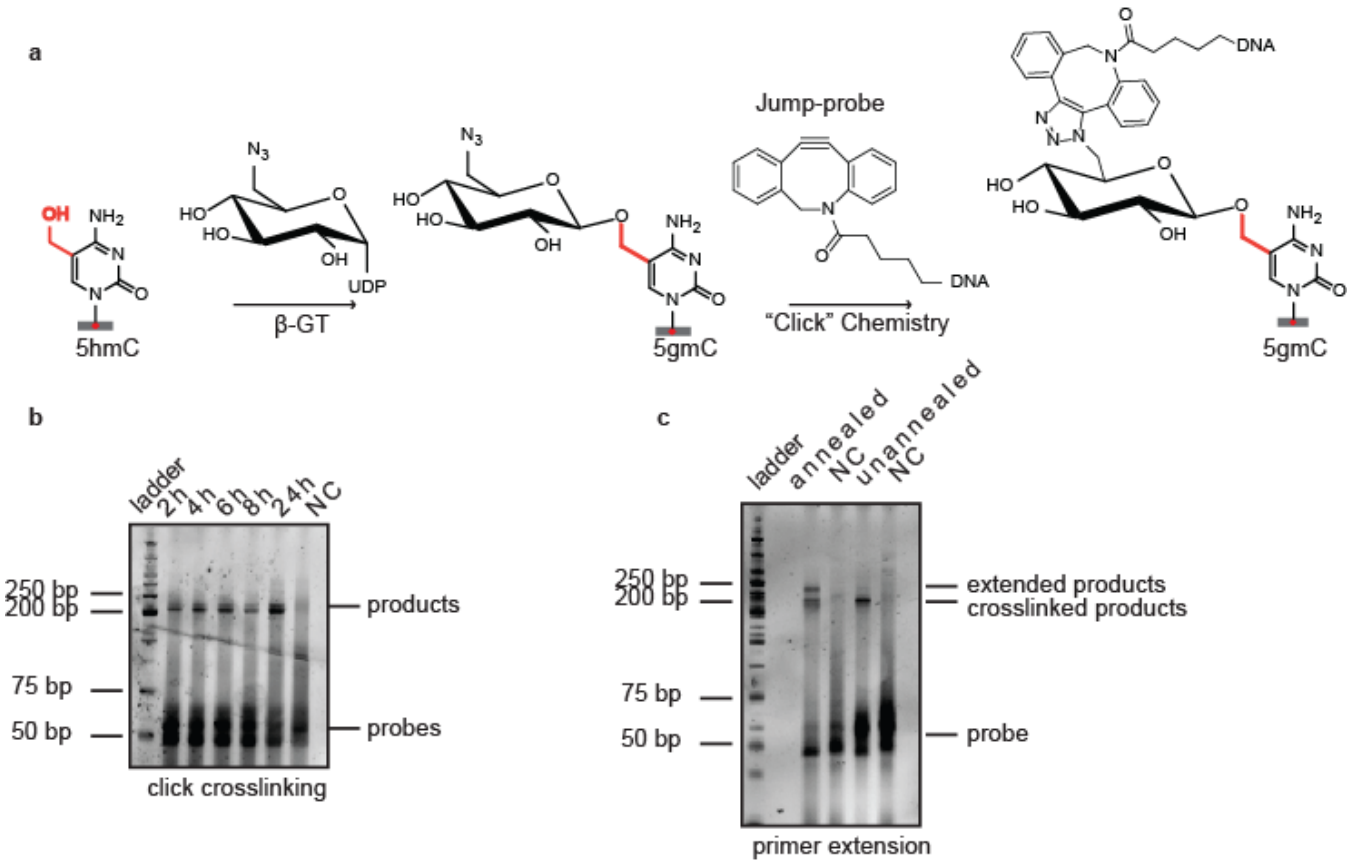

Figure S1. Efficiency of DBCO-azide mediated crosslinking and primer extension. (a) 5hmC in genomic DNA is labeled with an azide modified glucose using β-GT. A single-stranded DNA (with a unique sequence) carrying an DBCO is added covalently to the modified glucose. (b) Click reaction mediated by DBCO-modified jump probe and azide glucose-modified, 5hmC-containing 76 mer DNA model. Unmodified glucose was used as a negative control (NC). (c) Primer extension efficiency of annealed and unannealed hairpin jump probe. Unmodified glucose served as a negative control (NC).

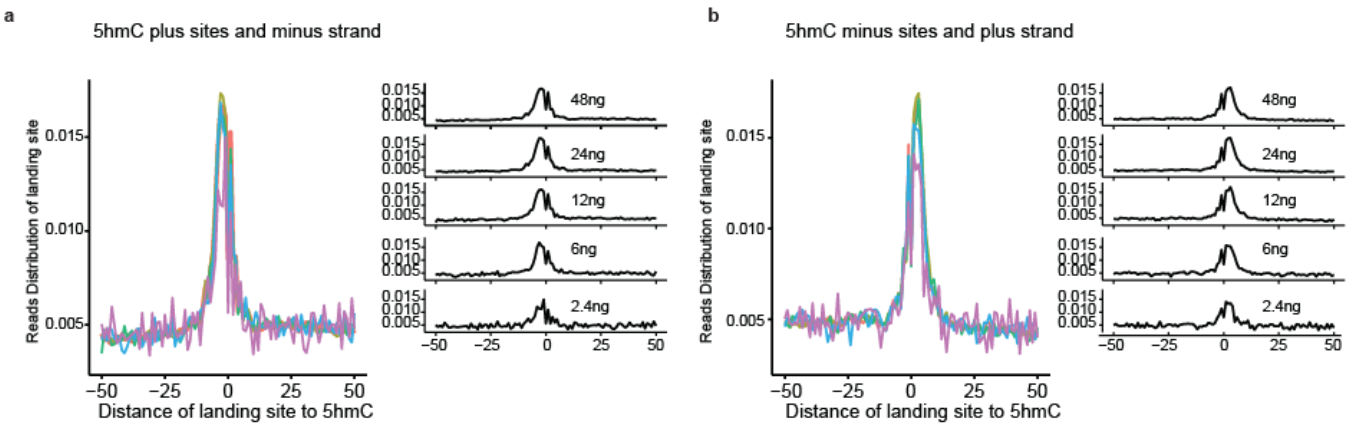

Figure S2. Reads distribution of Jump-seq strategy. Jump-seq results performed on genomic DNA isolated from 400 (2.4 ng), 1000 (6 ng), 2000 (12 ng), 4000 (24 ng), 8000 (48 ng) mouse ESCs showing a base-resolution "valley" overlaid on top of the 5hmC. "0" denotes the 5hmC site. (a) Jump-seq minus stand 5hmC sites (Jump-5hmC-). (b) Jump-seq plus stand 5hmC sites (Jump-5hmC+). Noting that the Jump-seq strategy has a complementary strand synthesis step, therefore the reads mapped to the plus stand actually represent the 5hmC sites in minus strand.

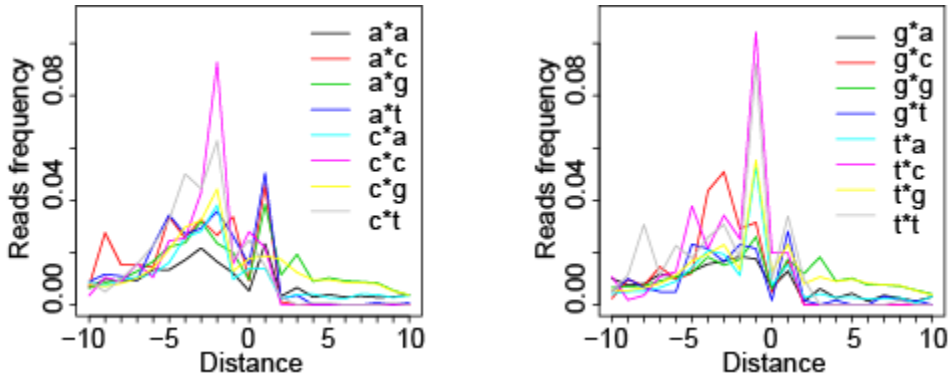

Figure S3. Jump-seq distribution pattern in different base composition context. To investigate the jump distribution of reads to ensure presence of just one 5hmC, most reads are concentrated within 20 bp window with the 5hmC at the center. Figure S3 visualizes the reads jump distribution of 16 kinds of base compositions. We performed two sample Kolmogorov-Smirnov test and found that ncn pair compositions with the same last base tend to have the same distribution, for instance, cca vs aca, etc.

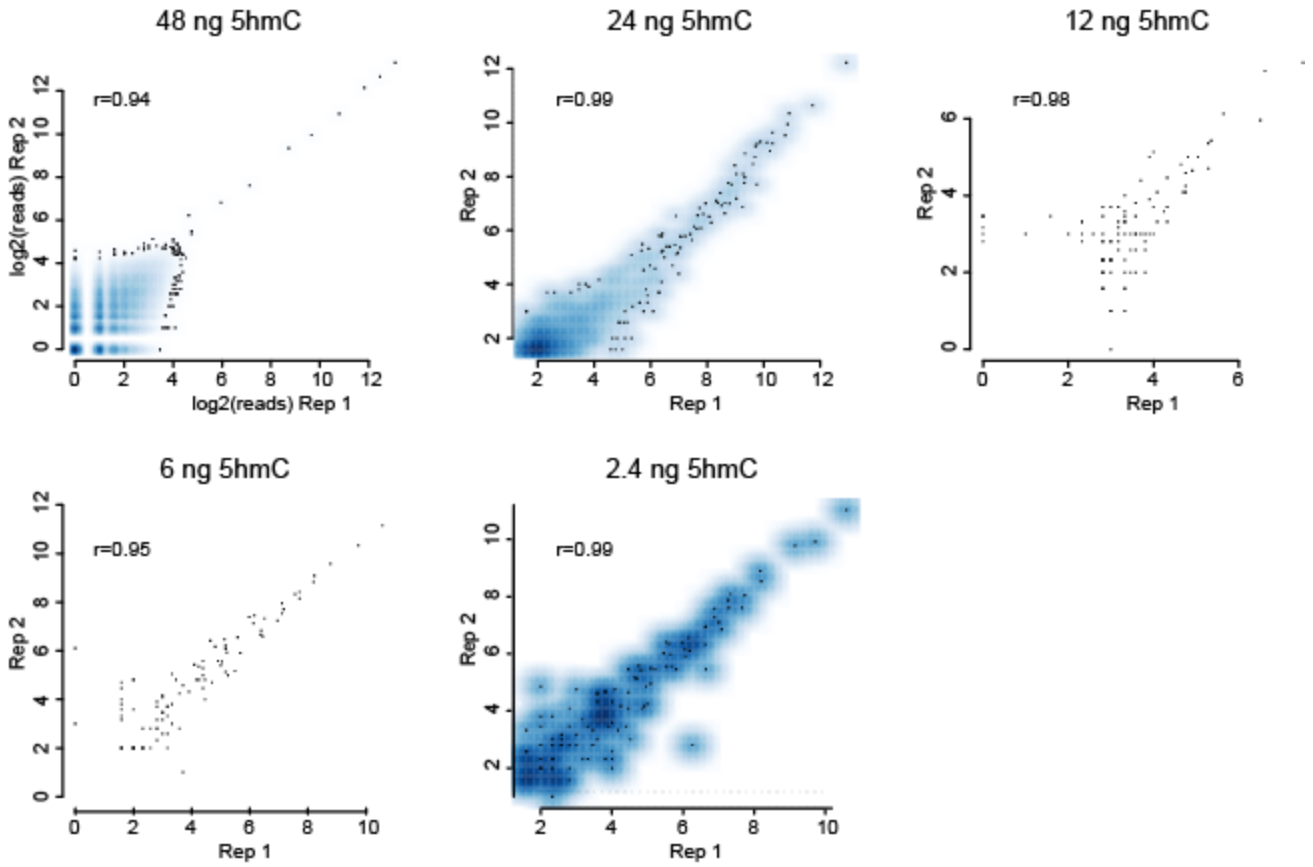

Figure S4. Correlation density plot of 5hmC signals between replicates at different gDNA levels. 5hmC Jump-seq results performed on genomic DNA isolated from 400 (2.4 ng), 1000 (6 ng), 2000 (12 ng), 4000 (24 ng), 8000 (48 ng) mouse ESCs showing good correlation between replicates.

Supplementary Tables

Table S1. Spike-in DNA model sequences

ID No.	Sequence (5'-3')	Annotation
1	CCTCACCATCTCAACCAATATTATATTACGCGTATATCGCGTATTTCGXGTTATAATATTGAGGGAGAAGTGGTGA (X = hmC or C)	76 mer model DNA for the crosslinking and primer extension efficiency validation
2	/Biotin/GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTATATTAGTCCATGCGGTGACCATCAGGTGCCAAAGTTGGCGTACACCCCGTGGTAGTACTGTXGCAGATCGAGAACTGTCGATAAGTCTAGGGTCGCCACCAGTAACGCTAGACGATCGGCTTCGAGCTACAGCTTGACCGCTCTAGTGACGGCATAGACTAGddC (X = C or 5hmC)	Jump single site spike-in (Nextera DNA kit compatible)
3	/Biotin/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGATGTGTATAAGAGACAGTGCCAAAGTTGGCGTACACCCCGTGGTAGTACTGTXGCAGATCGAGAACTGTCGATAAGTCTAGGGTCGCCACCAGTAACGCTAGACGATCGGCTTCGAGCTACAGCTTGACCGCTCTAGTGACGGddC(X = C or 5hmC)	Jump NEB spike-in (NEB next DNA kit compatible)
4	Biotin/GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTATATTAGTCCATGCGGTGACCATCAGGTGCCAAAGTTGGCGTACACCCCGTGGTXGTACTGTXGCAGATCGAGAACTGTCGATAAGTCTAGGGTCGCCACCAGTAACGCTAGACGATCGGCTTCGAGCTACAGCTTGACCGCTCTAGTGACGGCATAGACTAGddC (X = 5hmC)	Jump 2 sites spike-in (Nextera DNA kit compatible)

Table S2. The oligonucleotides synthesized at Integrated DNA Technologies (IDT) for the ligation of 200 bp long ds 5hmC Spike-in.

ID No.	Sequence (5'-3')	Annotation
1	/Biotin/GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTATATTAGTCCATGCGGTGACCATCAGGTGCCAAAGTTGGCGTACACCCCGTGGTAGTACTGTXG (X = C or 5hmC)	Jump single site spike-in 1F (Nextera DNA kit compatible)
2	/App/CAGATCGAGAACTGTCGATAAGTCTAGGGTCGCCACCAGTAACGCTAGACGATCGGCTTCGAGCTACAGCTTGACCGCTCTAGTGACGGCATAGACTAGddC	Jump single site spike-in 2F (Nextera DNA kit compatible)
3	TAGACTTATCGACAGTTCTCGATCTGCGACAGTACTACCACGGGGTGTACGCC	Jump single site spike-in R (Nextera DNA kit compatible)
4	/Biotin/GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTATATTAGTCCATGCGGTGACCATCAGGTGCCAAAGTTGGCGTACACCCCGTGGTXGTACTGTXG (X = 5hmC)	Jump 2 sites Spike-in 1F (Nextera DNA kit compatible)
5	/App/CAGATCGAGAACTGTCGATAAGTCTAGGGTCGCCACCAGTAACGCTAGACGATCGGCTTCGAGCTACAGCTTGACCGCTCTAGTGACGGCATAGACTAGddC	Jump 2 sites Spike-in 2F (Nextera DNA kit compatible)
6	TATCGACAGTTCTCGATCTGCGACAGTACGACCACGGGGTGTACGCC	Jump 2 sites Spike-in R (Nextera DNA kit compatible)
7	/Biotin/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGATGTGTATAAGAGACAG TGCCAAAGTTGGCGTACACCCCGTGGTAGTACTGTXG (X = 5hmC)	Jump NEB spike-in 1F
8	/App/CAGATCGAGAACTGTCGATAAGTCTAGGGTCGCCACCAGTAACGCTAGACGATCGGCTTCGAGCTACAGCTTGACCGCTCTAGTGACGGddC	Jump NEB spike-in 2F
9	GCTCGAAGCCGATCGTCTAGCGTTACTGGTGGCGACCCTAGACTTATCGACAGTTCTCGATCTGCGACAGTACTACCACGGGGTGTACGC	Jump NEB spike-in R

Table S3. hmC Spike-in DNA and background DNA sequences for the linear correlation of read counts and 5hmC amount test.

ID No.	Sequence (5'-3')	Annotation
1	/Bio-tin/GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTATATTAGTCCATGCGGTGACCATCAGGTGCCAAAGTTGGCGTACACCCCGTGGTAGTACTGTXGCAGATCGAGAACTGTCGATAAGTCTAGGGTCGCCACCAGTAACGCTAGACGATCGGCTTCGAGCTACAGCTTGACCGCTCTAGTGACGGCATAGACTAGddC (X = 5hmC)	hmC spike-in (Nextera DNA kit compatible)
2	/Biotin/ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATGTCGGACACGTGGAGCTCTATCCAG	Background-1
3	/Biotin/ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GCAGAAGTTTGCAGCTCAGGAGTTGATTG	Background-2
4	/Biotin/ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GTTGATCTGGAGATAGAGAGCCTTCTG	Background-3
5	/Biotin/ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GATGATGAGATGCGCAGGCTCAACATAC	Background-4
6	/Biotin/ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG ATGGATAGCCGCTTGCAGGAGATC	Background-5
7	/Biotin/ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG CCAACATTTTGTAGACACTGGACATAG	Background-6
8	/Biotin/ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG AGACGGGGACTTCATTCATGCTAATG	Background-7

Table S4. Jump adaptor and jump-probe sequences

The oligonucleotides for Jump-seq adaptors and jump-probes were synthesized at Integrated DNA Technologies (IDT).

ID No.	Sequence (5'-3')	Annotation
1	/phos/ CTGTCTCTTATACAddC,-3'	Tn5-bottom adaptor
2	/Biotin/GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	Jump P7 adaptor (Nextera DNA kit compatible)
3	/Biotin/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT ACCCCGTT AGATGTGTATAAGAGACAG	Jump P7 adaptor NEB barcode
4	/DBCO/CATTGATNNNNNNNN AGATCGGAAGAGCGTCGTGTAGGGAAA-GAdUdUdUACACTCTTTCCCTA	Jump P5 probe (NEB Next DNA kit compatible)

Table S5. Summary of mESC genomic DNA sequencing results

This excel file is submitted separately.

References

1. Picelli, S.; Bjorklund, A. K.; Reinius, B.; Sagasser, S.; Winberg, G.; Sandberg, R., Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome research* **2014,** *24* (12), 2033-40.
2. Wang, Q.; Gu, L.; Adey, A.; Radlwimmer, B.; Wang, W.; Hovestadt, V.; Bahr, M.; Wolf, S.; Shendure, J.; Eils, R.; Plass, C.; Weichenhan, D., Tagmentation-based whole-genome bisulfite sequencing. *Nature protocols* **2013,** *8* (10), 2022-32.
3. Song, C. X.; Szulwach, K. E.; Fu, Y.; Dai, Q.; Yi, C.; Li, X.; Li, Y.; Chen, C. H.; Zhang, W.; Jian, X.; Wang, J.; Zhang, L.; Looney, T. J.; Zhang, B.; Godley, L. A.; Hicks, L. M.; Lahn, B. T.; Jin, P.; He, C., Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature biotechnology* **2011,** *29* (1), 68-72.
4. 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.
5. Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C. X.; Jin, P.; Ren, B.; He, C., Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nature protocols* **2012,** *7* (12), 2159-70.