Tagmentation-based whole-genome bisulfite sequencing

Qi Wang[1,6], Lei Gu[1,2,6], Andrew Adey[3], Bernhard Radlwimmer[4], Wei Wang[4], Volker Hovestadt[4], Marion Bähr[2], Stephan Wolf[5], Jay Shendure[3], Roland Eils[1], Christoph Plass[2] & Dieter Weichenhan[2]

[1]Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [2]Division of Epigenomics and Cancer Risk Factors, DKFZ, Heidelberg, Germany. [3]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [4]Division of Molecular Genetics, DKFZ, Heidelberg, Germany. [5]Core Facility, High-Throughput Sequencing, DKFZ, Heidelberg, Germany. [6]These authors contributed equally to this work. Correspondence should be addressed to D.W. (d.weichenhan@dkfz.de).

Published online 26 September 2013; doi:10.1038/nprot.2013.118

Epigenetic modifications such as carbon 5 methylation of the cytosine base in a CpG dinucleotide context are involved in the onset and progression of human diseases. A comprehensive understanding of the role of genome-wide DNA methylation patterns, the methylome, requires quantitative determination of the methylation states of all CpG sites in a genome. So far, analyses of the complete methylome by whole-genome bisulfite sequencing (WGBS) are rare because of the required large DNA quantities, substantial bioinformatic resources and high sequencing costs. Here we describe a detailed protocol for tagmentation-based WGBS (T-WGBS) and demonstrate its reliability in comparison with conventional WGBS. In T-WGBS, a hyperactive Tn*5* transposase fragments the DNA and appends sequencing adapters in a single step. T-WGBS requires not more than 20 ng of input DNA; hence, the protocol allows the comprehensive methylome analysis of limited amounts of DNA isolated from precious biological specimens. The T-WGBS library preparation takes 2 d.

INTRODUCTION

WGBS is the gold-standard method for quantitative interrogation of the methylation state of all CpG dinucleotides, i.e., the methylome, in a genome. The most widely used protocol for WGBS library preparation requires several enzymatic and cleaning steps, necessitating a minimum of 200–500 ng and up to 5 µg of human input DNA, as described in the original protocol[1]. Low input DNA amounts down to 5 ng have been used as well, but at the expense of genomic coverage; moreover, the methodological details involved in using these low DNA quantities have not been clearly described[2–4]. As the large DNA amount required in the standard protocol is prohibitive for many methylome studies, alternative approaches such as reduced-representation bisulfite sequencing (RRBS) or HpaII tiny fragment enrichment by ligation-mediated PCR (HELP) are frequently applied. These methods require much less DNA (between 10 and 300 ng); however, they allow examination of only a minor fraction of the methylome[5–8]. Another widely used alternative that requires only 500 ng of input DNA is the HumanMethylation450 BeadChip technology, which addresses more than 485,000 CpG dinucleotides in the human genome, including gene promoters, gene bodies, microRNAs and CpG islands (http://www.illumina.com/products/methylation_450_beadchip_kits.ilmn). However, only a complete picture of the methylome will enable a complete assessment of how methylation states correlate with and possibly influence the phenotype in different developmental stages and diseases. Recently, new methods addressing the complete methylome of samples with a very limited amount of DNA, post-bisulfite adapter tagging (PBAT)[9] and T-WGBS[10], were published; moreover, a commercial kit became available in the case of the Ovation Methyl-seq platform (http://www.nugeninc.com/nugen/index.cfm/products/cs/ngs/methyl-seq/).

We present here a detailed protocol for T-WGBS. This method uses a hyperactive Tn*5* transposase that fragments genomic DNA and attaches sequencing adapters in a single step[11]. Compared with the much higher input DNA amount required in routine conventional WGBS, T-WGBS requires not more than 20 ng of human or mouse DNA. This input DNA amount is similar to that used in RRBS, HELP and PBAT, yet as in conventional WGBS and PBAT, it addresses the complete methylome.

Development of the protocol

Tagmentation-based next-generation sequencing was initially developed as a less-time-consuming and less-input-DNA–consuming alternative approach to the conventional generation of next-generation sequencing libraries (Nextera, Epicentre). Whereas conventional protocols require several steps for DNA fragmentation, end polishing, A-tailing and sequencing adapter ligation (e.g., ref. 12), tagmentation does this in a single reaction by using a hyperactive Tn*5* transposase in combination with a DNA adapter[11]. T-WGBS uses a methylated adapter to prevent its conversion by bisulfite treatment[10] (**Fig. 1**). An oligonucleotide replacement and gap repair step is required to covalently append methylated adapter sequences to each single strand of tagmented DNA fragments. Before the final PCR amplification, the DNA is bisulfite-treated (**Fig. 1**). Detailed comparative analyses of tagmentation-based and conventional sequencing libraries from humans, *Drosophila melanogaster* and *Escherichia coli* revealed similar coverage biases, demonstrating high utility of the tagmentation approach[11]. Comparison of sequencing data using various amounts of human input DNA indicated that as little as 10 ng provided results of similar quality to those obtained with 100-fold more input DNA in conventional WGBS[10]. We have now further optimized the protocol for routine T-WGBS library preparation starting with 10–30 ng of human or mouse input DNA.

Applications of the method

The method is particularly suited for the preparation of next-generation sequencing libraries (either without bisulfite treatment

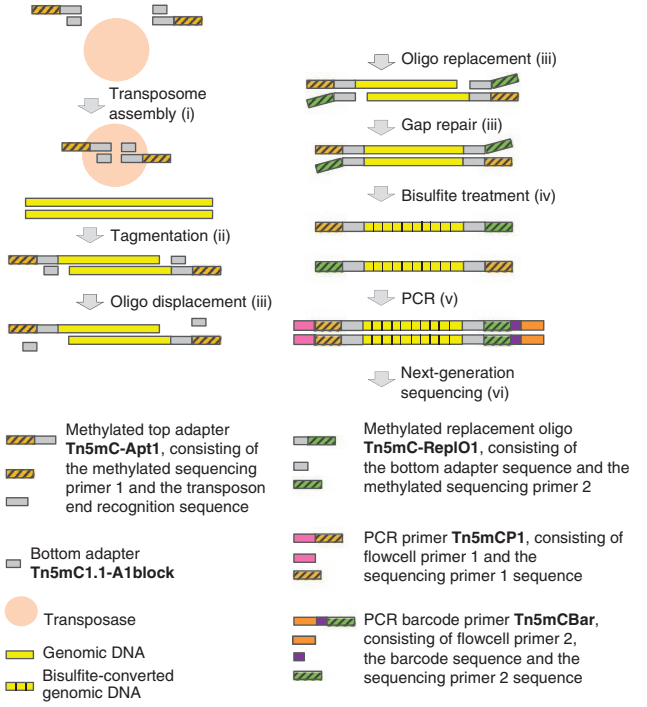

Figure 1 | Overview and components of T-WGBS library preparation. For a detailed description, refer to Experimental design in the INTRODUCTION. Roman numerals indicate stages described therein.

to determine the genomic sequence or with bisulfite treatment to determine the methylome sequence) of a limited number of cells of special cell types, such as high-purity solid tumor or embryonic mouse cells harvested by laser-capture microdissection, as well as healthy hematopoietic cells enriched by flow cytometry. Both library types can be prepared in parallel from a common DNA input amount of as little as 20 ng (human or mouse). Theoretically, as few as 1,700 human or mouse cells would be sufficient for tagmentation-based NGS library preparation; however, a more realistic number to yield the required minimum of 10 ng of genomic DNA is ~3,500 cells. The tagmentation protocol has also been extended to additional applications such as PCR-free and colony PCR library construction[11].

Comparison with other methods addressing the methylome

Methylated CpG-enrichment strategies such as MeDIP[13,14], MethylCap[15] and MCIp[16], followed by sequencing or microarray analysis, are widely used methods. Such strategies provide, by counting read numbers or assessing relative fluorescence ratios of regional sequences, enrichment values for diverse methylation states of genomic regions. These methods do not provide quantitative values of CpG methylation states, which can only be determined for regions of interest by additional approaches such as bisulfite sequencing of cloned fragments, pyrosequencing or mass spectrometric analysis. The restriction enzyme–based approaches RRBS[5,7] and HELP[6] provide quantitative methylation values, yet they cover only about 10% of all CpGs in the human or mouse genome. Depending on the biological question or the regions of interest (for example, promoters and CpG islands), enrichment- or enzyme-based methods or combinations can be considered valuable and inexpensive strategies for methylome analysis[17,18]. Similarly, the HumanMethylation450 BeadChip technology, which provides quantitative methylation values of selected sites, is cost-efficient and of increasing use in the community, as reflected by a plethora of recent publications (e.g., refs. 19 and 20) describing methylome changes in human disease. However, aside from its limited number of interrogated CpGs, a major disadvantage of this method is its current limitation to the human genome.

Conventional WGBS[1], PBAT[9], T-WGBS[10] and the commercially available Ovation Methyl-seq platform address all CpG sites in a genome, including those in regions of intermediate and low CpG density. Because these regions are almost always excluded from the analytical methods described above, their role in epigenetically regulated developmental and pathological processes is still largely elusive and thus of particular interest in current and future studies. With the exception of conventional WGBS, the other three methods claim to require only very small amounts of genomic input DNA, making these approaches highly attractive for the complete methylome analysis of limited cell numbers or tiny biopsy specimens. T-WGBS and the commercial approach require lower input amounts than PBAT; the latter, however, does not include global genomic PCR amplification, thus avoiding PCR-based read duplications.

As shown previously[10] and covered in this protocol, T-WGBS and conventional WGBS provide largely identical results with respect to methylation levels and coverage of the human methylome. Besides the small input DNA amount required, a higher methodological robustness, reproducibility and shortened procedural time owing to fewer manipulation steps[10] are further advantages of T-WGBS as compared with conventional WGBS.

Experimental design

Input DNA amounts. T-WGBS is particularly suited for low input DNA amounts and has been optimized for 10–30 ng of human or mouse input DNA. The use of only 5 ng of human input DNA requires a higher PCR cycle number (17 compared with 15 cycles for a 10-ng input) to reach the amplification plateau and may therefore lead to a higher sequencing read–duplication frequency. However, even with only 1 ng of input DNA, 78% unique reads can be obtained[10]. The use of more input DNA, e.g., 50 ng or 200 ng, may improve sequencing library complexity[10], but it did not decrease the PCR cycle number while the process of the protocol was being established in our laboratory. Additional protocol optimization, such as doubling the transposase activity, may be required if more than 30 ng of input DNA is used. It is recommended to generate T-WGBS libraries in parallel with the indicated input amount, in case the genomic sequencing coverage is too low with a single library.

Modified oligonucleotides. (Sequences and modifications of oligonucleotides are mandatory because of the transposon Tn5-based tagmentation, bisulfite treatment and the sequencing platform; however, other barcode sequences may be used.) The top adapter Tn5mC-Apt1 and the replacement oligonucleotide Tn5mC-ReplO1 must be methylated at the carbon 5 atom of all cytosine bases to prevent conversion to uracil upon bisulfite treatment (**Table 1**). A phosphate at the 5′ end of Tn5mC-ReplO1 is essential for covalent linkage to nucleotides upon gap repair (**Fig. 1**); the phosphate at the 5′ end of the bottom adapter Tn5mC1.1-A1block may be

TABLE 1 | Oligonucleotides and their modifications.

Name	Sequence (5′ to 3′)	Modifications
Methylated top adaptor: Tn5mC-Apt1	TcGTcGGcAGcGTcAGATGTGTATAAGAGAcAG	c: 5C-methylated
Bottom adaptor: Tn5mC1.1-A1block	pCTGTCTCTTATACAddC	p: phosphate, ddC: dideoxycytidylate
Replacement oligo: Tn5mC-ReplO1	pcTGTcTcTTATAcAcATcTccGAGccCAcGAGAcinvT	p: phosphate, c: 5C-methylated, invT: inverted deoxythymidylate
PCR primer: Tn5mCP1	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTC	
PCR barcode primer: Tn5mCBar	CAAGCAGAAGACGGCATACGAGAT-NNNNNNNNNGTCTCGTGGGCTCGG	

Barcodes for PCR primer Tn5mCBar: Bar1, GGATGTTCT; Bar2, CTTATCCAG; Bar3, GTAAGTCAC; Bar4, TTCAGTGAG; Bar5, CTCGTAATG; Bar6, CATGTCTCA; Bar7, AATCGTGGA; Bar8, GTATCAGTC.

omitted, yet the bottom adapter without the phosphorylated 5′ end was not tested in our laboratory. The Tn5mC1.1-A1block has a dideoxycytidylate at the 3′ end to prevent unwanted extension during the PCR if residual amounts of this oligonucleotide are still present in the PCR reaction mix. Replacement oligonucleotide Tn5mC-ReplO1 bears an inverted deoxythymidylate at the 3′ end that leads to a 3′-3′ linkage, which inhibits both degradation by 3′ exonucleases and extension by DNA polymerases.

Workflow. T-WGBS can be subdivided into six consecutive stages (**Fig. 1**): (i) assembly of the transposome consisting of transposase and adapter; (ii) tagmentation of genomic DNA; (iii) oligonucleotide replacement and gap repair; (iv) bisulfite treatment; (v) real-time PCR; and (vi) next-generation sequencing. Solid-phase reversible immobilization (SPRI) bead purification is obligatory between stages ii and iii, iii and iv, and v and vi. After stage v, the bisulfite-treated DNA is column-purified.

Tagmentation. In the second stage, the transposome fragments the genomic DNA and appends the adapter containing the Tn5 recognition sequence and a methylated primer sequence to both ends of the fragments. At this early stage, testing of a small aliquot as a control for fragmentation (by Bioanalyzer analysis) is recommended to prevent unnecessary time and material loss in case of failure (**Supplementary Fig. 1**).

Oligonucleotide replacement and gap repair. In the third stage, the shorter adapter oligonucleotide, which is not covalently linked to the genomic DNA, is replaced by a methylated oligonucleotide, and a gap of nine bases is repaired by the combined action of DNA polymerase and DNA ligase. Consequently, the tagmentation fragments are flanked by identical methylated adapters at both ends. At this point, a small sample aliquot (e.g., 10% of the sample volume) may be saved for use as a control in the later real-time PCR to assess extensive DNA damage by the subsequent bisulfite treatment. This sample aliquot can then even be used for genomic sequencing of a non-bisulfite-treated library (see below).

Bisulfite conversion. In the fourth stage, bisulfite treatment of the tagmented DNA converts unmethylated cytosines to uracil, whereas methylated cytosines including those of the two adapter oligonucleotides remain unchanged. Alternative bisulfite chemistry other than that indicated might be used as well; however, we have not tested such alternatives because of our long-lasting positive experience with the EZ DNA methylation kit (Zymo). To control for the efficiency of bisulfite conversion, unmethylated λ- phage DNA is spiked into the genomic input DNA. In T-WGBS, a bisulfite conversion control can be also performed by assessing the sequences of the nine base gaps into which unmethylated nucleotides are incorporated during the repair step[10] (**Fig. 1** and **Supplementary Table 1**). Spiking the input DNA with *in vitro*–methylated DNA may also be considered to control for overconversion, yet this was not part of the development of the presented protocol.

PCR amplification. Real-time PCR (the fifth stage) with adapter-specific primers that are compatible with the sequencing oligonucleotides used (from the TruSeq dual-index sequencing primer box, paired end; Illumina) is applied in the final amplification step, whether the DNA was bisulfite treated or not, until the amplification curve indicates transition to the plateau phase, usually between cycles 10 and 15 (**Fig. 2** and **Supplementary Fig. 2a**). The KAPA 2G robust hot start ready mix (Peqlab) is commonly used here, but the high-fidelity system KAPA HiFi uracil$^+$ (Peqlab) is also recommended. As outlined in the protocol, libraries are produced with only half of the sample from the fourth stage, thus providing the option to prepare two libraries from one input sample by independent PCR reactions. Moreover, this offers the opportunity to repeat the real-time PCR in case of failure. To control for extensive DNA damage by bisulfite treatment as a possible cause of PCR failure or too high a cycle number, half of the non-bisulfite-treated aliquot saved in stage three is routinely amplified in parallel. Usually, the bisulfite-treated and non-bisulfite-treated aliquots reach the amplification plateau at the identical cycle number, indicating that about 90% of the bisulfite-treated DNA

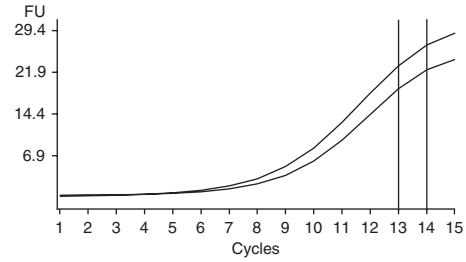

Figure 2 | PCR amplification curves of T-WGBS libraries. Curves move into the plateau phase between cycles 13 and 14. FU, fluorescence units.

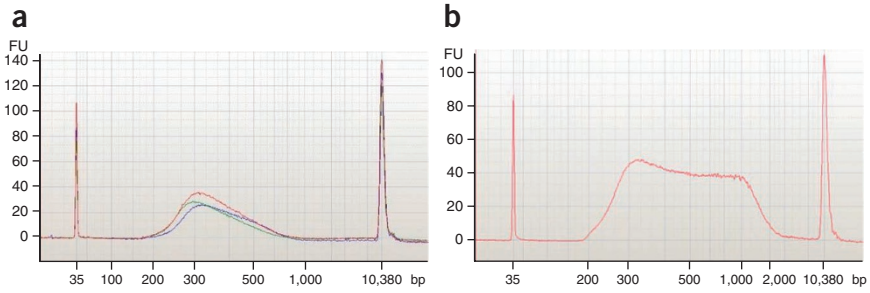

Figure 3 | Size distribution of T-WGBS libraries. (**a**) Three libraries prepared with bisulfite treatment. (**b**) Single library prepared without bisulfite treatment. The spikes flanking the curves represent size markers loaded together with each sample. FU, fluorescence units.

has either been lost between the third and fifth stages or has been damaged in such a way that PCR amplification is prevented.

Sequencing. Paired-end sequencing (stage six) of the amplified DNA is performed on a next-generation sequencing platform. Efficient 101-bp paired-end sequencing on an Illumina HiSeq 2000 requires an average fragment size of about 300–350 bp. Although amplified fragments from bisulfite-treated DNA are in this range (**Fig. 3a** and **Supplementary Fig. 2b**), those from non-bisulfite-treated DNA may have a larger size range (**Fig. 3b**), thus requiring an additional gel sizing step.

Limitations

T-WGBS routinely works with a starting amount of 10–30 ng of human or mouse DNA, equivalent to about 1,700–5,100 cells. This cell number may be still too high for studies of difficult-to-enrich cell types such as cancer stem cells or special progenitors in the hematopoietic lineage. Further technical optimization may improve the efficiency of the tagmentation process, the bisulfite treatment, the PCR and the DNA recovery in the purification steps. As bisulfite treatment does not distinguish among different cytosine modifications such as 5-methyl, 5-hydroxymethyl, 5-carboxyl or 5-formyl, these epigenetic modifications will be detected as one, and their differences will escape detection in T-WGBS. During the gap-repair step, unmethylated nucleotides are incorporated so that the first nine bases of the second read and the last nine bases before the adapter on the first read must be excluded from methylation analysis[10] (**Fig. 1**). These bases, however, may be used for controlling proper bisulfite conversion rate. It was reported that the Tn*5* transposase has a slightly higher sequence bias than fragmentation by sonication[11]. However, compared with conventional WGBS, we observed no obvious limitation of T-WGBS in this respect on the basis of genomic coverage or base composition of sequencing reads (**Supplementary Fig. 3**), with either human or mouse DNA. DNA isolated from formalin-fixed, paraffin-embedded tissue specimens required PCR cycle numbers around 20 and led to sequencing read–duplication frequencies of more than 80%; hence, it has so far proven to be unsuitable for T-WGBS. Further systematic analyses of various parameters such as the ratio between input DNA amount and transposase activity, the initial fragment size and the effect of residual contamination from the DNA extraction process are necessary to understand whether T-WGBS can be optimized so that it is suitable for archived biological specimens such as formalin-fixed, paraffin-embedded samples or ancient forensic material.

MATERIALS

REAGENTS

▲ **CRITICAL** All experiments that use human and animal samples must comply with the relevant institutional and governmental guidelines. Human samples must be obtained in accordance with informed consent policies.

- Ethanol, 100% (Sigma-Aldrich, cat. no. 32205) ! **CAUTION** Ethanol is flammable. Handle it properly during experiments.
- Acetic acid (100%, Merck, cat. no. 1.00063) ! **CAUTION** Acetic acid is flammable. It causes severe skin burns and eye damage. Work in a fume hood.
- Dimethylformamide (Merck, cat. no. 1.03053) ! **CAUTION** Dimethylformamide is flammable. It is harmful on contact with skin or if inhaled. It causes serious eye irritation. Work in a fume hood.
- $MgCl_2$ (Sigma, cat. no. M8266)
- Tris(hydroxymethyl)aminomethane (Sigma-Aldrich, cat. no. T1503) ! **CAUTION** This compound is harmful if inhaled. Work in a hood if the compound is not dissolved in aqueous solution.
- Glycerol (≥99.5%, Roth, cat. no. 3783)
- EZ-Tn5 ⟨Kan-2⟩ insertion kit (Biozym, cat. no. 190009; Epicentre cat. no. EZ1982K)
- Tagmentation DNA buffer (2× TD; the buffer is a component of the Nextera DNA sample preparation kit; Illumina, cat. no. FC-121-1030) ▲ **CRITICAL** Alternatively, a tagmentation buffer (2× tagmentation buffer) can be prepared as follows: 20 mM Tris(hydroxymethyl)aminomethane; 10 mM $MgCl_2$; 20% (vol/vol) dimethylformamide, according to ref. 21. Before the addition of dimethylformamide, adjust the pH to 7.6 with 100% acetic acid.
- Sterilize the buffer by filtration. This solution can be stored at −20 °C for at least 6 months.
- Phage-λ DNA, unmethylated (Promega, order no. D1521)
- T4 DNA polymerase (New England BioLabs, cat. no. M0203S)
- Ampligase, thermostable DNA ligase, with 10× buffer (Biozym, cat. no. 111025)
- KAPA 2G robust hot start ready mix (Peqlab, cat. no. 07-KK5701-01)
- SYBR Green I nucleic acid gel stain, 10,000× (Life Technologies, cat. no. S7563) ! **CAUTION** SYBR Green I is a potential mutagen. Handle it properly during experiments.
- dNTP mix (100 mM of each dNTP; Fermentas, cat. no. R0181)
- Guanidinium thiocyanate (Sigma-Aldrich, cat. no. 50990) ! **CAUTION** Guanidinium thiocyanate is hazardous in case of skin contact (irritant), eye contact (irritant), ingestion and inhalation. Work in a fume hood.
- PEG 8000 (Roth, cat. no. 0263.1)
- NaCl (Sigma-Aldrich, cat. no. 31434)
- EZ DNA methylation kit (Zymo Research, cat. no. D5002) with the following components: CT conversion reagent, M-dilution buffer, M-binding buffer, M-desulfonation buffer, M-wash buffer and M-elution buffer ! **CAUTION** This kit contains sodium metabisulfite, sodium hydroxide, guanidine hydrochloride, ethanol, Tris(hydroxymethyl)amino methane and EDTA. Many kit components are hazardous in case of skin contact (irritant), eye contact (irritant), ingestion and inhalation. Ethanol is flammable. Work in a fume hood.
- Agencourt AMPure XP beads (Beckman Coulter, cat. no. A63881)
- EB elution buffer (Qiagen, cat. no. 19086)
- Quant-iT dsDNA high-sensitivity (HS) assay kit (Agilent, cat. no. 5067-4626)
- Qubit dsDNA HS assay kit (Invitrogen, cat. no. Q32854)
- Oligonucleotides: all modified and unmodified oligonucleotides (**Table 1**) were custom-made and purchased from Sigma-Aldrich. The PCR primers Tn5mCP1 and Tn5mCBar (Tn5mCBar1–Tn5mCBar8) were synthesized at a 25-nmol scale and desalted. The modified oligonucleotides Tn5mC-Apt1,

Tn5mC1.1-A1block and Tn5mC-ReplO1 were synthesized at a 50-nmol scale and HPLC-purified
• QIAmp DNA mini or micro kits (Qiagen, cat. no. 51304 or 56304)

EQUIPMENT
• Mastercycler Epgradient S with a 96-well block (Eppendorf, cat. no. 950010151)
• Lightcycler 480 with a 96-well block (Roche, cat. no. 05015278001)
• Lightcycler 480 96-well plates (Roche, cat. no. 04729692001)
• Zymo-Spin IC columns and corresponding collection tubes (components of the EZ DNA methylation kit; Zymo Research, cat. no. D5002)
• DNA LoBind reaction tubes (Eppendorf, cat. no. 022431021)
• Qubit 2.0 fluorometer (Invitrogen, cat. no. Q32866)
• Qubit assay tubes (Invitrogen, cat. no. Q32856)
• Bioanalyzer 2100 with electrophoresis set (Agilent, cat. nos. G2940CA and G2947CA)
• Magnetic separator MagnetoPure-Micro for eight-well strips (Chemicell, cat. no. MP-50)
• TruSeq dual-index sequencing primer box, paired end (Illumina, cat. no. PE-121-1003)
• Illumina HiSeq2000 (Illumina)

REAGENT SETUP

Ultrapure Milli-Q-filtered and autoclaved H_2O Milli-Q-filtered and autoclaved H_2O is used throughout the protocol.

AMPure buffer AMPure buffer is 2.5 M NaCl and 20% (wt/vol) PEG 8000; filter-sterilize the buffer. This solution can be stored at room temperature (RT, 20–25 °C) for at least 6 months.

5 M guanidium thiocyanate Guanidium thiocyanate (5 M) is dissolved in H_2O. This solution can be stored protected from light at RT for at least 6 months.

Ethanol, 80% (vol/vol) Dissolve 40 ml of 100% ethanol in 10 ml of H_2O. This solution can be stored at RT for at least 1 week.

CT conversion reagent CT conversion reagent is supplied with the EZ DNA methylation kit as a solid mixture. Before use, add 750 µl of H_2O and 210 µl of M-dilution buffer, which is also supplied ready-to-use in the EZ DNA methylation kit. Mix the liquid CT conversion reagent with frequent vortexing for 10 min. This solution can be stored protected from light at −20 °C for up to 1 month.

M-wash buffer M-wash buffer is supplied with the EZ DNA methylation kit. Before use, add 96 ml of 100% ethanol to 24 ml of M-wash buffer and mix it thoroughly. This mixture can be stored at RT for at least 1 month.

SYBR Green, 100× Dissolve 1 µl of SYBR Green I nucleic acid gel stain, 10,000×, in 99 µl of H_2O. This solution can be stored protected from light at −20 °C for at least 6 months.

PROCEDURE

Isolation of human or mouse genomic DNA ● TIMING 1–2 h (plus incubation, if necessary)

1| Isolate genomic DNA from cell pellets, whole blood, frozen tissue slices or microdissected tissue specimen by using commercial DNA isolation kits such as the QIAmp DNA mini or micro kits. DNA isolated from tissue by standard phenol/chloroform extraction is suitable as well.

▲ CRITICAL STEP All experiments that use human and animal samples must comply with the relevant institutional and governmental guidelines.

Preannealing of the adapter Tn5mC-Adapt ● TIMING 1 h

2| In a 200-µl PCR tube, mix 10 µl each of oligonucleotides Tn5mC-Apt1 and Tn5mC1.1-A1block (100 µM each; **Table 1**) and 80 µl of H_2O.

3| Set the thermocycler conditions for adapter assembly as follows:

Cycle number	Denature	Anneal	Ramp to 26 °C	Hold
1	95 °C, 3 min	70 °C, 3 min		
2–46		70 °C, 30 s	−1 °C per cycle, 30 s	
47				25 °C, infinite

Assembly of the transposome ● TIMING 1 h

4| Heat 100 µl glycerol (≥99.5%) to 90 °C in a thermocycler. Hot glycerol can be readily and exactly pipetted.

5| Transfer 50 µl of hot glycerol to a 1.5-ml reaction tube and cool it down to RT by incubation on ice for 3 min.

6| Add 50 µl of adapter Tn5mC-Adapt (10 µM) from Step 2 and mix it by repeated pipetting.

7| Transfer 10 µl of the glycerol-adapter mixture (stable at −20 °C for at least 6 months) to a new 1.5-ml reaction tube, add 10 µl of Ez-Tn*5* transposase (from the EZ-Tn5 ⟨Kan-2⟩ insertion kit); mix by repeated pipetting.

8| Maintain the adapter-transposase mixture, the transposome, at RT for 30 min, and then place it on ice. The 20-µl mixture is sufficient for eight tagmentation reactions.

■ PAUSE POINT The transposome can be stored at −20 °C for at least 1 month.

Tagmentation of genomic DNA ● TIMING 15 min

9| Set up eight tagmentation reactions, one in each well of an eight-well PCR strip. Per well, mix on ice 10 µl of 2× TD and 7.5 µl of aqueous DNA solution containing 20 ng of genomic DNA and 10 pg of phage-λ DNA.

10| Add 2.5 µl of the assembled transposome (from Step 8) to each well and mix by repeated pipetting.

11| Set the thermocycler conditions for tagmentation as follows:

Cycle number	Tagmentation	Cool
1	55 °C, 8 min	4 °C, infinite

Post-tagmentation SPRI bead purification ● TIMING 30 min

12| To each well, add 15 µl of 5 M guanidinium thiocyanate (total volume of DNA solution 35 µl), 10 µl of AMPure beads and 36 µl of AMPure buffer (total volume of bead solution 46 µl) and mix to homogeneity by repeated pipetting.

13| Keep reactions at RT for 10 min.

14| Transfer the eight-well strip to a magnetic separator, wait for 1 min and remove the supernatant thoroughly by pipetting without disturbing the bead pellet in each well; discard the supernatant.
▲ **CRITICAL STEP** To avoid bead carryover or loss of beads, removal of the supernatant in several small volume steps rather than in a single large volume may be preferred.

15 While the strip is still on the magnetic separator, wash the beads in each well with 50 µl of 80% ethanol by repeated (10×) pipetting without disturbing the pellets. Remove the supernatant in each well completely by pipetting without disturbing the bead pellet; discard the supernatant. Keep the eight-well strip open on the magnetic separator for 10 min.
▲ **CRITICAL STEP** Keep the washing step as short as possible. Complete removal of any liquid and droplets is essential; the beads must be completely dry before the next step.

16| Remove the eight-well strip from the magnetic separator and thoroughly resuspend the magnetic beads in 12 µl of H_2O per well.

17| Transfer the strip to the magnetic separator, wait for 1 min and transfer the supernatant (the eluate containing the tagmented DNA) from each well into a new well of a new eight-well strip, without disturbing the bead pellets. Transfer 1 µl of the eluate from each well to a DNA LoBind reaction tube for troubleshooting by monitoring the fragment size using the Quant-iT dsDNA HS assay kit and the Bioanalyzer 2100 (**Supplementary Fig. 1a**).
? TROUBLESHOOTING
■ **PAUSE POINT** The purified tagmented DNA can be stored at –20 °C for at least 1 month.

Oligonucelotide replacement and gap repair ● TIMING 1 h

18| To the 11 µl of eluate in each well, add 2 µl of dNTP mix (2.5 mM each, 10 mM), 2 µl of 10× Ampligase buffer and 2 µl of replacement oligo Tn5mC-Repl01 (10 µM; **Table 1**); mix by repeated pipetting.

19| Set the thermocycler conditions for replacement and annealing as follows:

Cycle number	Denature	Anneal	Ramp to 37 °C	Hold
1	50 °C, 1 min	45 °C, 10 min		
2		45 °C, 0.1 s	–0.1 °C s^{-1}	37 °C, infinite

20| While the strip remains in the thermocycler, add 1 µl of T4 DNA polymerase and 2.5 µl of Ampligase per well; mix by repeated pipetting.

21| Continue the reaction in the thermocycler with the following conditions for gap repair:

Cycle number	Gap repair	Hold
1	37 °C, 30 min	4 °C, infinite

SPRI bead purification after oligonucleotide replacement ● TIMING 30 min

22| To the samples (20.5 µl per well) in the eight-well strip from Step 21, add 10 µl of AMPure beads and 26 µl of AMPure buffer (total volume of bead solution 36 µl) per well; mix to homogeneity by repeated pipetting.

23| Repeat Steps 13–16, resuspending the beads in 50 µl of H_2O per well.

24| Transfer the strip to the magnetic separator. Wait for 1 min and transfer the supernatant (the eluate containing the tagmented and gap-repaired DNA) from each well into a new well of a new eight-well strip, without disturbing the bead pellets. Transfer 5 µl of the eluate from each well to a LoBind reaction tube for troubleshooting by real-time PCR or for preparation of a sequencing library for genome analysis without bisulfite treatment.

? TROUBLESHOOTING

■ **PAUSE POINT** The purified gap-repaired DNA can be stored at –20 °C for at least 1 month.

Bisulfite treatment with the EZ DNA methylation kit ● TIMING 1 d

25| To the 45 µl of eluate in each well of the eight-well strip from Step 24, add 5 µl of M-dilution buffer and mix by repeated pipetting.

26| Set the thermocycler conditions as follows:

Cycle number	Gap repair	Hold
1	37 °C, 15 min	25 °C, infinite

27| Add 100 µl of liquid CT conversion reagent to each well and mix by repeated pipetting; split the mixture from each well into two new wells, each containing an equal aliquot of 75 µl, and continue the reaction in the thermocycler as follows:

Cycle number	Denature	Conversion	Hold
1–16	95 °C, 15 s	50 °C, 1 h	
17			4 °C, infinite

Column-based DNA purification after bisulfite treatment ● TIMING 30 min

28| Transfer 400 µl of M-binding buffer to each of eight Zymo-Spin IC Columns inserted in collection tubes. Load each cooled sample, consisting of 2 × 75 µl, into a Zymo-Spin IC column containing M-binding buffer. Close the column caps and mix by inverting several times. Centrifuge at 11,000*g* for 30 s at RT and discard the flow-through.

29| Add 100 µl of M-wash buffer to each column, centrifuge at 11,000*g* for 30 s at RT and discard the flow-through.

30| Add 200 µl of M-desulfonation buffer to each column and keep at RT for 15–20 min; centrifuge at 11,000*g* for 30 s at RT and discard the flow-through.

31| Add 200 µl of M-wash buffer to each column, centrifuge at 11,000*g* for 30 s at RT and discard the flow-through. Repeat this wash step once, and then finally centrifuge at 17,000*g* for 3 min at RT.

32| Transfer each column to a LoBind reaction tube and add 12 µl of M-elution buffer directly to each column matrix. Centrifuge the tubes at 16,000*g* for 30 s at RT to elute the converted DNA. Repeat this elution step once with another 12 µl of M-elution buffer.

■ **PAUSE POINT** The converted DNA can be stored at –80 °C for at least 1 month.

Real-time PCR for the preparation of the T-WGBS library ● TIMING 1.5 h

33| For each sample, prepare the following mixture on ice in a well of a 96-well real-time PCR plate:

Component	Amount	Final
KAPA 2G robust hot start ready mix (2×)	12.5 μl	1×
Primer Tn5mCP1n (10 μM; **Table 1**)	0.75 μl	0.3 μM
Primer Tn5mCBar (10 μM; **Table 1**)	0.75 μl	0.3 μM
SYBR Green (100×)	0.25 μl	1×
Converted template DNA (from Step 32)	10.75 μl	
Total	25 μl	

▲ **CRITICAL STEP** Prepare a master mixture without DNA templates and barcode primer Tn5mCBar if multiple samples are to be analyzed. Use individual barcode primers for different template DNAs. Use tagmented and gap-repaired, but not bisulfite-treated, template DNA from Step 24 for troubleshooting or sequencing library preparation. Use nontagmented genomic DNA as a negative control.

34| Set the real-time thermocycler conditions as follows:

Cycle number	Denature	Annealing	Read	Extension
1	95 °C, 3 min			
1–99	95 °C, 20 s	62 °C, 15 s	72, 30 s	72, 10 s

35| Stop the PCR when the amplification curve indicates transition from the exponential phase to the plateau phase, usually between cycles 10 and 15 (**Fig. 2** and **Supplementary Fig. 2a**).

? TROUBLESHOOTING

SPRI bead purification and library quality control ● TIMING 1.5 h

36| Transfer 25 μl of each PCR mix to a well of an eight-well strip, add 45 μl (1.8× volume) of AMPure beads per well and mix to homogeneity by repeated pipetting.

37| Repeat Steps 13–15 by washing the beads with 200 μl of 80% (vol/vol) ethanol and resuspending the magnetic beads in 21 μl of EB elution buffer per well.

38| Transfer the strip to the magnetic separator and wait for 1 min. Next, without disturbing the bead pellets, transfer the supernatant (the eluate containing the T-WGBS library) from each well into an individual 1.5-ml reaction tube. Transfer 1 μl of each eluate to a LoBind reaction tube and dissolve it with 9 μl of H_2O (1:10 dilution) for subsequent quantification (Qubit) and size distribution (Bioanalyzer) analysis.

39| Quantify the 1:10 diluted T-WGBS library with a Qubit fluorometer and determine the size distribution with a Quant-iT dsDNA HS assay kit.

? TROUBLESHOOTING

Preparation of a second T-WGBS library ● TIMING 3 h

40| To prepare an additional T-WGBS library from the same input DNA sample, repeat Steps 33–39 with the second half (10.75 μl) of the converted DNA from Step 32 once the T-WGBS library from Step 38 has passed the quantity and the size distribution test in Step 39.

Next-generation sequencing ● TIMING 12 d

41| Analyze each T-WGBS library on a different lane in a 101-bp paired-end sequencing run on an Illumina HiSeq2000 by using the appropriate sequencing primers (TruSeq dual-index sequencing primer box, paired end) compatible with the T-WGBS library primers Tn5mCP1 and Tn5mCBar (**Table 1**).

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Step	Problem	Possible reason	Solution
17	No or insufficient fragmentation	Defective transposome or impure DNA	Test the transposome with pure DNA, e.g., commercial phage-λ DNA; replace the transposome if it fails. Impure DNA may be cleaned by mini-dialysis, re-precipitation or gel purification, provided a sufficient amount is available
24	Too high a PCR cycle number (>15) to reach amplification plateau in test real-time PCR	Inefficient oligonucleotide replacement or gap repair	Use successfully tagmented phage-λ DNA in a test reaction. Replace the polymerase, Ampligase and replacement oligo if PCR cycle number is still too high
35	Too high a PCR cycle number (>15) after bisulfite treatment	Bisulfite treatment has destroyed the DNA	Lower the conversion cycle number (Step 30) to improve DNA integrity. Bisulfite conversion frequency should still be 99.5% as determined with spiked phage-λ DNA
39	Inappropriate size range (expected is 200–600 bp). Low DNA amount (<40 ng) in library	Unbalanced stoichiometry between input DNA and transposase; DNA loss during SPRI bead purification	Use proper amount of input DNA (10–30 ng) and transposase as indicated in protocol. Avoid the loss of SPRI beads and extensive (>30 s) bead washing. Replace beads, 80% (vol/vol) ethanol and EB buffer

● TIMING

Day 1

Step 1, isolation of human or mouse genomic DNA: 1–2 h (plus overnight incubation for tissue lysis, if necessary)
Steps 2 and 3, preannealing of the adapter Tn5mC-Adapt: 1 h
Steps 4–8, assembly of the transposome: 1 h
Steps 9–11, tagmentation of genomic DNA: 15 min
Steps 12–17, post-tagmentation SPRI bead purification: 30 min
Steps 18–21, oligonucleotide replacement and gap repair: 1 h
Steps 22–24, SPRI bead purification after oligonucleotide replacement: 30 min
Steps 25–27, bisulfite treatment with the EZ DNA methylation kit: 1 d

Day 2

Steps 28–32, column-based DNA purification after bisulfite treatment: 30 min
Steps 33–35, real-time PCR for the preparation of the T-WGBS library: 1.5 h
Steps 36–39, SPRI bead purification and library quality test: 1.5 h
Step 40, preparation of a second T-WGBS library (from the other half of the tagmented, bisulfite-treated DNA): 3 h
Step 41, next-generation sequencing: 12 d

ANTICIPATED RESULTS

Starting with a fragment size of the input DNA of about 20 kb, after tagmentation and the first SPRI bead purification (Step 17) the DNA fragments should range in size from smaller than 1 kb to about 10 kb (**Supplementary Fig. 1a**). The number of PCR cycles required to reach the plateau phase usually ranges from 10 to 15 (**Fig. 2** and **Supplementary Fig. 2a**). Higher cycle numbers may bear the risk of sequencing read duplication frequencies >40%, which may lead to low genomic coverage per sequencing lane (i.e., less than twofold coverage per DNA strand). After bisulfite treatment, fragments of the sequencing library should have a size range of 200–600 bp with a peak at about 300–400 bp (**Fig. 3a** and **Supplementary Fig. 2b**); no or only a very low primer peak should be visible. The DNA concentration of a T-WGBS library usually ranges from 2 to 15 ng μl^{-1} in a 20-µl SPRI bead eluate, sufficient to load more than one lane on a HiSeq2000. As the presented protocol produces two T-WGBS libraries per 10–30 ng of input DNA, one can usually load at least three HiSeq2000 lanes per input DNA and can expect a total genomic coverage of about 20-fold (tenfold per DNA strand), which is considered suitable to properly assess the methylation percentage at the single-CpG level[22].

For comparison, a single DNA isolate from a human glioblastoma multiforme tumor biopsy was subjected to T-WGBS (30 ng of input DNA each for two independent tagmentations with two libraries per tagmentation) and conventional WGBS

(performed as described previously with 5 μg of input DNA for a single library[1]). The conventional WGBS library was loaded onto three HiSeq2000 lanes, whereas the four T-WGBS libraries were sequenced one lane per library. Sequencing data have been deposited at the European Genome-phenome Archive (EGA; https://www.ebi.ac.uk/ega/datasets/) hosted by the European Bioinformatics Institute, under accession number EGAD00001000601. Sequencing data were aligned and methylation calling was carried out as described in the **Supplementary Methods**. The four T-WGBS libraries were almost identical to each other, and they performed similarly well compared with the conventional WGBS library with respect to the percentage of mapped reads, the overall methylation level assessment and the conversion frequency (**Supplementary Table 1**). The higher duplication frequency in T-WGBS probably reflects the higher PCR cycle number used (ten or eleven cycles in T-WGBS versus eight in conventional WGBS), yet it proved sufficiently low to attain an only slightly lower CpG coverage in T-WGBS, 12.1 versus 13.7, when reads from three corresponding T-WGBS libraries were merged (three T-WGBS libraries with one lane per library versus one conventional WGBS library loaded onto three lanes; **Supplementary Table 1**). The bisulfite conversion frequency of the T-WGBS libraries was marginally lower than that of the conventional WGBS library, 99.5 versus 99.9%, and, in line with this, the average CpG methylation level in T-WGBS was slightly higher, 77.2 versus 75.8%; both differences probably reflect a better bisulfite treatment performance in the conventional WGBS rather than a better overall performance of the conventional method. High similarity in the performance between the two protocols was further supported by the high correlation of the methylation levels (Pearson's correlation, 0.95; **Fig. 4** and **Supplementary Fig. 4a**; calculated based on about 3.7 million CpGs that are covered by at least 30 reads in both protocols). To further quantify the consistency between the two WGBS protocols, we defined a concordance metric as the percentage of CpG sites (at least 30-fold coverage) with a <0.2 (20%) difference in methylation level. The concordance (Concordance value was calculated by subtracting from 100% the respective numbers in the corners of the density plot; these numbers indicate the percentage of data points that fall above or below the envelope marked with dotted lines.) between the two protocols was 97.3% (**Fig. 4a**). Such reliability was further supported by the high level of agreement in the methylation level assessment between the two protocols applied on two human blood samples (**Supplementary Fig. 5**). Moreover, the observed concordance between conventional WGBS and T-WGBS libraries (98.5%; **Supplementary Fig. 6a**) was comparable to that between the Watson and Crick strands in the conventional WGBS (98.1%; **Supplementary Fig. 6b**). As determined in the same manner, the concordance between two independent T-WGBS experiments was 97.8% (r = 0.92; **Fig. 4b** and **Supplementary Fig. 4b**), indicating high robustness and reliability of the T-WGBS protocol. T-WGBS and conventional WGBS also show similar sequencing coverage at CpG sites as a function of CpG density (**Fig. 4c**). Comparative analysis of sequencing coverage versus density of cytosines in CpG, CHG and CHH context (H can be A, C or T; **Supplementary Fig. 7**) or versus local GC content (**Supplementary Fig. 8**) revealed similar patterns from T-WGBS and conventional WGBS. For methylome characterization, genomic features such as promoters, CpG islands, exons, introns and intergenic regions are of particular interest. The proportion of CpGs that were covered at least tenfold in these features was about 90% or higher and nearly identical between T-WGBS and conventional WGBS (**Fig. 4d**; **Supplementary Table 2**).

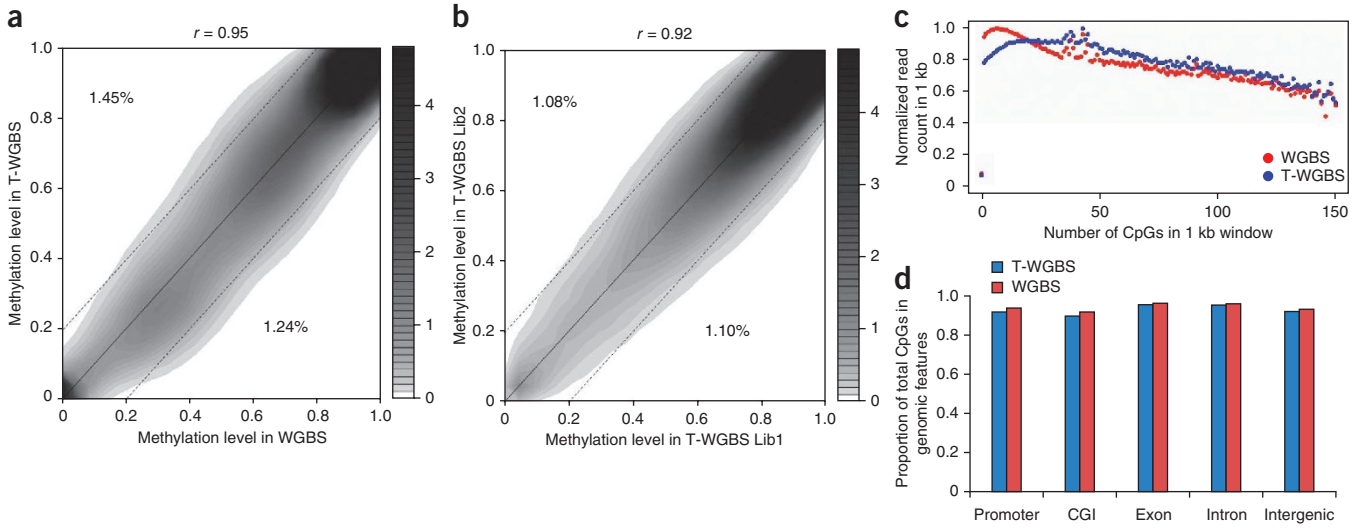

Figure 4 | Reliability and reproducibility of T-WGBS. (**a**) High consistency is shown with a Pearson's correlation of r = 0.95 between the methylation levels of corresponding single CpGs covered at least 30-fold in T-WGBS and in conventional WGBS. (**b**) High reproducibility of T-WGBS is indicated by strong agreement of the methylation levels (r = 0.92) in windows of five CpGs (read numbers too low for single-CpG analysis) in libraries from two independent tagmentations analyzed on a single HiSeq2000 lane each. The numbers in the corners of **a** and **b** indicate the percentage of data points that fall above or below the envelope marked with dotted lines (marking the 0.2 difference between the x and y axes). (**c**) Nearly identical sequencing coverage of CpGs as a function of CpG density between T-WGBS and conventional WGBS. (**d**) Ninety percent or higher and almost identical proportions of CpGs are covered at least tenfold in five genomic features. T-WGBS and conventional WGBS reads from three lanes each were compared.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS We gratefully acknowledge excellent technical support by M. Helf and helpful discussions with D. Lipka. We also acknowledge the excellent support by the sequencing core facility at DKFZ. Work in the Plass laboratory was supported by the Helmholtz Foundation and the German Federal Ministry of Education and Science in the program for medical genome research (FKZ; no. 01KU1001A). Q.W. obtained support by the Humboldt Research Fellowship for Postdoctoral Researchers. A.A. is funded by a National Science Foundation Graduate Research Fellowship.

AUTHOR CONTRIBUTIONS A.A., J.S. and D.W. conceived the study. B.R., W.W. and V.H. contributed data. B.R., R.E. and C.P. contributed materials. W.W., M.B., S.W. and D.W. performed the experiments and analyzed data. Q.W., L.G. and V.H. performed the bioinformatic analysis. Q.W., L.G. and D.W. wrote the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare competing financial interests: details are available in the online version of the paper.

1. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
2. Kobayashi, H. *et al.* Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* **8**, e1002440 (2012).
3. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–1105 (2010).
4. Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48**, 849–862 (2012).
5. Gu, H. *et al.* Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat. Protoc.* **6**, 468–481 (2011).
6. Khulan, B. *et al.* Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res.* **16**, 1046–1055 (2006).
7. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877 (2005).
8. Oda, M. *et al.* High-resolution genome-wide cytosine methylation profiling with simultaneous copy number analysis and optimization for limited cell numbers. *Nucleic Acids Res.* **37**, 3829–3839 (2009).
9. Miura, F., Enomoto, Y., Dairiki, R. & Ito, T. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.* **40**, e136 (2012).
10. Adey, A. & Shendure, J. Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. *Genome Res.* **22**, 1139–1143 (2012).
11. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol.* **11**, R119 (2010).
12. Quail, M.A. *et al.* A large genome center's improvements to the Illumina sequencing system. *Nat. Methods* **5**, 1005–1010 (2008).
13. Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* **37**, 853–862 (2005).
14. Taiwo, O. *et al.* Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat. Protoc.* **7**, 617–636 (2012).
15. Brinkman, A.B. *et al.* Whole-genome DNA methylation profiling using MethylCap-seq. *Methods* **52**, 232–236 (2010).
16. Gebhard, C. *et al.* Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res.* **66**, 6118–6128 (2006).
17. Bock, C. *et al.* Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat. Biotechnol.* **28**, 1106–1114 (2010).
18. Harris, R.A. *et al.* Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat. Biotechnol.* **28**, 1097–1105 (2010).
19. Hovestadt, V. *et al.* Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol.* **125**, 913–916 (2013).
20. Sturm, D. *et al.* Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* **22**, 425–437 (2012).
21. Grunenwald, H.L, Caruccio, N., Jendrisak, J. & Dahl, G. Transposon end compositions and methods for modifying nucleic acids. US patent 20100120098A1 (2010).
22. Hansen, K.D., Langmead, B. & Irizarry, R.A. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol.* **13**, R83 (2012).