Published online 27 March 2014

Nucleic Acids Research, 2014, Vol. 42, No. 9 ***e81***
doi: 10.1093/nar/gku216

Boronic acid-mediated polymerase chain reaction for gene- and fragment-specific detection of 5-hydroxymethylcytosine

Chao Zhao[1], Hailin Wang[1,*], Bailin Zhao[1], Cuiping Li[1], Ruichuan Yin[1], Maoyong Song[1], Baodong Liu[1], Zhen Liu[2] and Guibin Jiang[1]

[1]State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China and [2]Department of Chemistry, Nanjing University, Nanjing 210093, China

Received July 1, 2013; Revised February 20, 2014; Accepted March 3, 2014

ABSTRACT

The gene- or fragment-specific detection of newly recognized deoxyribonucleic acid (DNA) base 5-hydroxymethylcytosine (5hmC) will provide insights into its critical functions in development and diseases, and is also important for screening 5hmC-rich genes as an indicator of epigenetic states, pathogenic processes and pharmacological responses. Current analytical technologies for gene-specific detection of 5hmC are heavily dependent on glucosylated 5hmC-resistant restriction endonuclease cleavage. Here, we find that boronic acid (BA) can inhibit the amplification activity of Taq DNA polymerase for replicating glucosylated 5hmC bases in template DNA by interacting with their glucose moiety. On the basis of this finding, we propose for the first time a BA-mediated polymerase chain reaction (PCR) assay for rapid and sensitive detection of gene- or fragment-specific 5hmC without restriction-assay-like sequence limitations. To optimize the BA-mediated PCR assay, we further tested BA derivatives and show that one BA derivative, 2-(2′-chlorobenzyloxy) phenylboronic acid, displays the highest inhibitory efficiency. Using the optimized assay, we demonstrate the enrichment of 5hmC in an intron region of *Pax5* gene (a member of the paired box family of transcription factors) in mouse embryonic stem cells. Our work potentially opens a new way for the screening and identification of 5hmC-rich genes and for high throughput analysis of 5hmC in mammalian cells.

INTRODUCTION

5-hydroxymethylcytosine (5hmC) is a recently re-discovered deoxyribonucleic acid (DNA) base that is converted from the well-characterized epigenetic mark 5-methylcytosine (5mC) (1–3). The conversion of 5mC to 5hmC is catalyzed by Tet family dioxygenases in mammalian cells (2). 5hmC functions critically in nuclear reprogramming, development and diseases (4–13). Moreover, the level of 5hmC has been observed to be altered in several diseases, including hematopoietic malignancies and a broad range of solid tumours (12,13). A lack of 5hmC was implicated as a useful biomarker for cancer diagnosis (13). Interestingly, both nutritional [e.g. vitamin C (14)] and environmental factors [e.g. redox-active quinones (15)] affect the activity of Tet family dioxygenases mediating 5mC oxidation and the cellular level of 5hmC. These observations may indicate that 5hmC-varying or -rich genes or sequences could be exploited as indicators or biomarkers of epigenetic states, pathogenic processes, pharmacological responses and environmental exposures. Therefore, the development of analytical technologies is critical for the detection of 5hmC distribution in the context of sequences or genes.

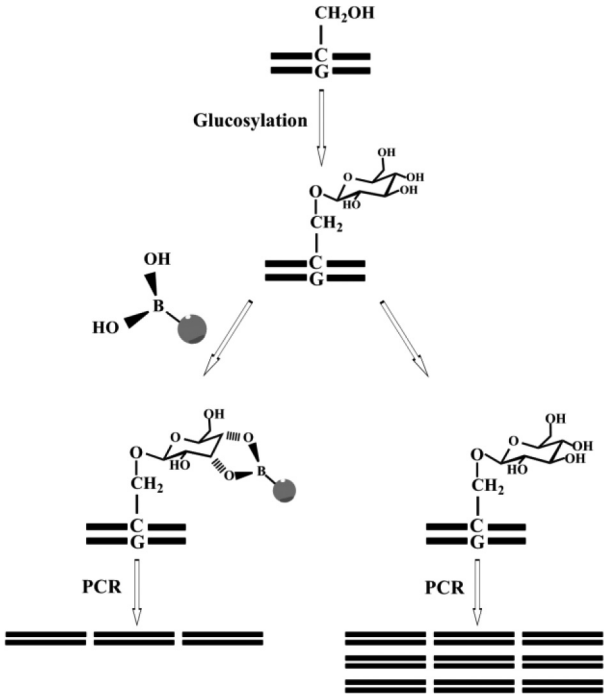
It is very difficult to discriminate the rare 5hmC from the abundant 5mC due to their similar chemical structures. For example, 5hmC is indiscernible from 5mC in bisulfite sequencing (16) and in most of 5mC-sensitive restriction assays (17,18). Interestingly, 5hmC can be selectively glucosylated by T4 phage β-glucosyltransferase (β-GT) to form β-glucosyl-5-hydroxymethylcytosine (5ghmC) (19), which is resistant to the cleaving activities of some methylation-insensitive restriction enzymes, e.g. MspI (https://www.neb.com/nebecomm/tech_reference/epigenetics/epimark.asp). Therefore, 5hmC (in its glucosylated state) but not 5mC can be retained at short and cleavable sites of restriction endonucleases for the specific genes or sequences (20–23). The fragments obtained by the digestion of glucosylated genomic DNA with combined restriction enzymes

*To whom correspondence should be addressed. Tel: +86 10 62849600; Fax: +86 10 62849600; Email: hlwang@rcees.ac.cn

(MspI/HpaII) could be enriched by ligation-mediated polymerase chain reaction (PCR) for identification of 5hmC loci (24). However, the applications of these assays are conditional: (i) the target DNA region must contain the specific cleavable sites (e.g. CCGG for MspI) and (ii) the 5hmC must be located precisely at the cleavable sites. Therefore, the restriction endonuclease assays are only applicable to certain loci or sequences and cannot identify 5hmC in a number of targeted but non-cleavable DNA regions, where it may be a critical factor affecting gene activity.

A number of affinity trapping approaches have also been developed in the 5hmC analysis of genomic DNA using antibodies (25,26), JBP1 (27) and biotinylation (28,29). The diol of the gluocosylated 5hmC could be used to develop oxidation chemistry for 5hmC capture (29). Meanwhile, several techniques for base-resolution analysis of 5-hmC were developed, including Tet-assisted bisulfite sequencing (30) and oxidative bisulfite sequencing (31).

In this work, we attempted to develop a novel and generic PCR approach to perform a fast and high throughput analysis of 5hmC at gene-specific or fragment-specific levels. Among currently known DNA bases in mammalian cells (32,33), only 5hmC can be specifically modified by β-GT to glucosylated 5hmC. However, the glucosylation of 5hmC in genomic DNA cannot block the replicating action of DNA polymerses (see the Results section). Thereby, we explored the further modification of glucosylated 5hmC in genomic DNA without introducing any DNA damages and other byproducts. We hypothesized that an appropriately structured boronic acid (BA) can selectively bond with the vicinal *cis*-diol group in glucosylated 5hmC and increase the size of 5ghmC, leading to specific stalling of DNA polymerases on 5ghmC in template DNA (Scheme 1). Although BA and its derivatives have been exploited to separate and capture *cis*-diol-containing biomolecules (34–38), it remains unknown whether this interaction can block the replication of DNA polymerase on 5ghmC (glucosylated 5hmC). Therefore, we tested the possible BA derivatives blocking of DNA polymerase amplification activity linking to 5hmC loci.



Scheme 1. Illustration of the BA-mediated inhibition on amplification activity of DNA polymerase. First, the 5hmC in dsDNA is exclusively glucosylated; then, one boronic acid molecule is selectively bonded with the vicinal diol in the conjugated glucose of 5hmC, inhibiting the replication of the 5hmC-containing DNA and leading to a reduced PCR yield (left). The glucosylated 5hmC-containing DNA can be normally amplified by PCR without boronic acid (right). The symbol '📍' Indicates the substituted group for boronic acid.

MATERIALS AND METHODS

Materials and chemicals

All of the unmodified oligos were synthesized and purified [by high performance liquid chromatography (HPLC)] by Sangon Biological Engineering Technology and Services (Shanghai, China). T4 DNA ligase, T4 polynucleotide kinase (T4 PNK), calf intestinal alkaline phosphatase, deoxyribonuclease I and adenosine triphosphate (ATP) were purchased from New England BioLabs (Ipswich, MA, USA). Snake venom phosphodiesterase, boronic acid (BA), phenylboronic acid (PBA), 3-chlorophenylboronic acid (3-CPBA), 2-(2′-chlorobenzyloxy) phenylboronic acid (2-CBPBA) and 3-(Dansylamino) phenylboronic acid (3-D-PBA) were purchased from Sigma-Aldrich (St. Louis, MO, USA). 5mdCTP and 5hmdCTP were purchased from Zymo Research (Irvine, CA, USA), and deoxyguanosine triphosphate (dGTP), thymine triphosphate (TTP), deoxycytidine (dCTP), GO Taq hot start polymerase were ordered from Promega (Madison, WI, USA). Other biochemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA) or Fisher Scientific (Pittsburgh, PA, USA). The 5hmC-ds83mers with a sequence of 5′-TTTCCTACCT TAAGATCCTT CCAGTCTC CGCCGCG CAGTG TTACCCTTAG AGCTCATACC ATTCGCCAAT TTCTTCGCAC GTT-3′ (only one strand shown) were synthesized and purified by TaKaRa Biotechnology Co., Ltd. (Dalian, China).

Synthesis, purification and characterization of 5hmC/5mC-containing oligos

To synthesize 5hmC-containing oligodeoxynucleotide 5′-TTTTCGAATTCCTCCCTGTA (5hmC) GTTTT-3′ (5hmC-Oligo 1, 26mer), a 41mer template (T, 5′-TTACTCATCATTTTTAAAACGT ACAGGGAGGAA TTCGAAAA-3′) and a 20mer primer (P, 5′-TTTTCGAATTCCTCCCTGTA-3′) were used (Supplementary Figure S1). With the assistance of the template 41mer, the 20mer primer (25 μM) was extended by six nucleotides using Taq DNA polymerase and limited dNTP (deoxynucleotide triphosphate) substrates. The extension reaction solution was prepared by mixing 100-μM 41mer template T (4.0 μl), 100-μM 20mer primer P (5.0 μl), 5.0 units/μl GO Taq hot start polymerase (0.5 μl), 5 × GO Taq hot start polymerase reaction buffer (4.0 μl) and 25-mM $MgCl_2$ (1.6 μl) with 50-mM 5hmdCTP (0.4 μl), 50-mM

dGTP (0.4 μl), 100-mM TTP (0.8 μl) and nuclease-free water (3.3 μl). The total volume was about 20 μl. No dCTP and dATP (deoxyadenosine triphosphate) were used in the extension reaction. The lack of dATP ensured the extension of the primer P to stop exactly after six nucleotides. The reaction was performed using a MyCycler Thermal Cycler (Bio-Rad Laboratories, Hercules, CA, USA). The extension program was set as: initial heating at 94°C for 5 min, degeneration at 94°C for 1 min, annealing at 53°C for 30 s, extension at 72°C for 90 min and final chilling at 4°C for 5 min.

Two control versions of oligo 1 (5C-oligo 1 and 5mC-oligo 1) were synthesized following the same procedure except that 5hmdCTP was replaced with dCTP and 5mdCTP, respectively.

The three synthesized oligo 1s were purified using 16% denaturing polyacrylamide gel electrophoresis (PAGE) at 200 V for 80 min. The denaturing gel was prepared with 7.0-M urea. 1 × TBE buffer [90-mM Tris-borate, 2-mM ethylenediaminetetraacetic acid (EDTA), pH 8.3] was used for PAGE separation. The bands (stained by ethidium bromide) were visualized using a GBOX/HR-E-M gel documentation system (Syngene, Cambridge, UK) and cut from the gel for recovery. The recovered gel slices were crushed in collection vials and soaked in 2.0 volume of polyacrylamide gel elution buffer (0.5-mol/l ammonium acetate, 10-mmol/l magnesium acetate, 1.0-mmol/l EDTA, pH 8.0), followed by shaking in a rotary shaker in the dark overnight at 37°C. The supernatants were aspirated out and transferred into other vials, followed by the addition of 2.0 volume of ice-cold ethanol and 0.3 volume of 100-mM $MgCl_2$. The solutions were placed at −20°C for 1.5 h (to facilitate the precipitation of the oligodeoxynucleotides) and then centrifuged by 12 000 rpm × 30 min at 4°C. The pellets were collected and washed once or twice with 70% ethanol. The collected pellets were air-dried, re-suspended in ddH_2O and quantified using a Nano Drop 2000 (Thermo Fisher Scientific Inc., Waltham, MA, USA) at 260 nm.

Design, synthesis and purification of X-ds100mers

The DNA probes containing a single modified site (C, 5mC or 5hmC) were designed and synthesized (Supplementary Figure S1). Briefly, oligo 1 (with one modified site of C, 5mC or 5hmC), oligo 3, oligo 4 and oligo 5 were phosphorylated at the 5′ end and then annealed with unphosphorylated oligos 2 and 6. The annealed oligos were ligated by T4 ligase to form an intact X-ds100mer. In this design, the oligo 5 was used as the complementary strand to assist with the ligation of oligo 1, oligo 2 and oligo 3 into one strand.

Firstly, oligo 1, oligo 3, oligo 4 and oligo 5 (150 pmol each) were phosphorylated by 10 units T4 PNK at 37°C for 2 h. The solution was buffer with 1 × ligation buffer (New England Biolabs, Ipswich, MA, USA). The total volume was about 35 μl. Next, the T4 PNK was denatured by heating at 70°C for 10 min. Then, two additional oligos (oligos 2 and 6; 150 pmol each) and 1-mM ATP were added to the solution, and the solution was heated at 70°C for 10 min, followed by a cooling at room temperature to anneal the six oligos. Lastly, ligation was initiated by adding 8.5 Weiss units of T4 DNA ligase and was carried out overnight at 16°C. The control probe was synthesized according to the same procedure, but 5hmC-oligo 1 was replaced with the oligo 1 containing a C or 5mC at the same position of 5hmC. The ligation products were purified using 16% native PAGE. The bands were visualized by the fluorescence staining and cut from the gel for recovery. The other procedures were the same as that for the purification of oligo 1.

Glucosylation of 5hmC-dsDNA probes or genomic DNA

All the glucosylation reactions were performed using β-GT provided with the 5hmC and 5mC Analysis Kit (NEB EpiMark™).

The 5hmC-double-stranded DNA (dsDNA) probes (2.5 μg each) were glucosylated using 30 units of β-GT for 18 h at 37°C in a reaction solution with a total volume of 40 μl. This solution contained 4.0 μl of 10 × NEBuffer 4, 1.6 μl of 80-μM UDP-glucose (uridine-5'-diphosphoglucose), and supplemented with 24.4-μl nuclease-free water. After the glucosylation, the enzyme β-GT was digested at 40°C for 30 min by adding 1.0-μl proteinase K (20 mg/ml) to the solution. Then, the added proteinase K was inactivated by heating at 95°C for 10 min. The glucosylated DNA was then mixed with 2 volume of ice-cold ethanol and 0.3 volume of 100-mM $MgCl_2$, and placed at −20°C for 1.5 h. Then, DNA was precipitated by centrifugation of 12 000 rpm × 10 min at 4°C. The supernatant was casted out, meanwhile the DNA pellet was washed once or twice with 70% ethanol. The DNA pellets were collected, air-dried and re-suspended in a phosphate buffer (100-mM Na_2HPO_4 buffer, pH 8.5, plus 50-mM NaCl). The final DNA solutions were quantified using a Nano Drop 2000 at 260 nm.

The genomic DNA (5.0 μg) was glucosylated using 40 units of β-GT for 18 h at 37°C in a reaction solution with a total volume of 100 μl. The solution comprised 10.0 μl of 10 × NEBuffer 4, 4.0 μl of 80-μM UDP-glucose, and supplemented with 86.0-μl nuclease-free water. The other steps were same as described above.

The characterization of the 5ghmC-dsDNA probe

The DNA products were digested into mononucleosides and subjected to ultra-performance liquid chromatograph-tandem mass spectrometry (UHPLC-MS/MS) analysis as described previously (14,39). Briefly, the UHPLC-MS/MS analysis was performed on Agilent 1290 UHPLC system coupled with a G6410B triple quadrupole mass spectrometer with an electrospray ionization source (Agilent Technologies, Santa Clara, CA, USA). A reversed-phase Zorbax Eclipse Plus C18 column (100 mm × 2.1 mm i.d., 1.8-μm particle size, Agilent Technologies) was used for separation. A MassHunter workstation software version B.01.03 was used for data acquisition. The mobile phase consisted of 5.0% methanol and 95% water (plus 0.1% formic Acid) and was used for UHPLC separation of the mononucleosides at a flow rate of 0.3 ml/min.

BA-mediated PCR assay

To develop BA-mediated quantitative PCR (qPCR) analysis, we tested BA and four BA derivatives using four dsDNA

probes (c-ds100mer, 5mC-ds100mer, 5hmC-ds100mer and 5ghmC-ds100mer).

Firstly, an aqueous solution of double-stranded oligodeoxynucleotides or their mixture (1.0 μl, 0–125 nM) was mixed with the stock solution of BA or BA derivaties (3.1 μl, 40 mM), 2 × GoTaq® SYBR Green qPCR Master Mix (12.5 μl; Promega, Madison, WI, USA), 10-μM forward and reverse primers (0.5 μl each), 200-mM Na_2HPO_4 buffer [pH 8.5 (5.0 μl)] and 400-mM NaCl (3.0 μl). As a control without any BA, 1 × BA stock buffer of the same volume (3.1 μl) was added to the solution instead. The solutions were supplemented with nuclease-free water to the total volume of 25.0 μl. The qPCR was performed using a Mx3005P Real-Time qPCR system (Stratagene, La Jolla). The cycling conditions were set as: initial denaturation at 95°C for 2.0 min and then 40 cycles of PCR including denaturation at 95°C for 15 s, annealing 57°C for 15 s and extension 25°C for 60 s. Each sample was analyzed in triplicate. The final PCR products were also separated by 16% native PAGE. Another was mixed with the phosphate buffer (1:1, volume).

The abundance of gene-specific 5hmC in genomic DNA was exclusively evaluated using 2-CB-PBA-mediated qPCR analysis. The glucosylated or un-glucosylated dsDNA samples were prepared for qPCR analysis following the same protocol as described above, except that the total amount of genomic DNA was about 10 ng.

Note: the stock solutions of BA or BA derivatives (40 mM) were prepared in prior by dissolving BA or BA derivatives in 200-mM Na_2HPO_4 buffer, pH 8.5 and 10% (v/v) methanol solution (1 × BA stock buffer). The dissolution of BA or BA derivatives was assisted transiently with ultrasonication.

Cell culture and DNA extraction

The cells were cultured and genomic DNA was extracted as described previously (14,15). Mouse embryonic stem (ES) cells of wild type (WT; 129 SvEv) and of Tet1/Tet2 double knockout (Tet1/Tet2$^{-/-}$; a gift from Dr Guoliang Xu at the State Key Laboratory of Molecular Biology, Shanghai, China) were maintained in Dulbecco's modified Eagle's medium (HyClone, Thermo Fisher Scientific Inc., MA, USA) supplemented with 20% ES FBS (Gibco, Life Technologies Corporation, Grand Island, NY, USA), 0.1-mM non-essential amino acids, 2-mM L-glutamine, 0.1-mM β-mercaptoethanol, 1000 units/ml leukemia inhibitory factor (Millipore, Billerica, MA, USA), 1.0-μM PD 0325901 (Stemgent, Cambridge, MA, USA) and 3.0-μM CHIR 99021 (Stemgent, Cambridge, MA, USA). Cells were trypsinized and plated in culture dishes pre-treated with 0.1% gelatin, then incubated in a humidified 37°C incubator supplied with 5.0% CO_2. After 24-h treatment, the cells were collected for DNA extraction. Genomic DNA (20 μg) was extracted from the harvested mouse ES cells by a Genomic DNA Purification Kit (Promega, Madison, WI, USA) following the manufacturer's instructions. The concentration and quality of the extracted genomic DNA were evaluated by measuring the absorbance at 260 nm and 280 nm. To assure the quality of the extracted genomic DNA, the ratio of absorbance at 260 to that at 280 nm is required to be about 1.80.

MspI-qPCR assay

We performed this assay following the instruction of manufacturer (https://www.neb.com/nebecomm/tech_reference/epigenetics/epimark.asp). Briefly, 5ghmC-ds100mer or glucosylated genomic DNA (100 ng) was incubated with 10 units MspI at 37°C. After 12-h enzymatic digestion, 1.0-μl proteinase K was added to each tube and incubated at 40°C for 30 min. Then, proteinase K was inactivated by heating at 95°C for 10 min. The digested products were purified by ice-cold ethanol precipitation, and then re-suspended, quantified and analyzed by qPCR. An aliquot without undergoing MspI digestion was used as a control. We would obtain a value of ΔC_t by subtracting the C_t values from MspI-digested aliquot from that of the control.

RESULTS

Design and synthesis of 5hmC-containing oligonucleotide probes

We first designed and synthesized four X-ds100mer probes (X = C, 5mC, 5hmC or 5ghmC; Supplementary Figures S1 and S2 and Supplementary Table S1). These probes comprised a double-stranded oligonucleotide of 100 bp containing a C, 5mC, 5hmC or 5ghmC located at the 58th nucleotide from the 5′ end (Figure 1a). The 5mC or 5hmC was incorporated precisely into the 58th position of the probes using well-designed primer extension reaction (Supplementary Figure S1), and no modified C was incorporated at any other positions. Therefore, both 5mC- and 5hmC-ds100mers only contained a single 5mC and 5hmC, respectively. The C-ds100mer did not contain any 5mC or 5hmC and could be obtained simply by commercial synthesis. UHPLC-MS/MS analysis validated the presence of 5mC and 5hmC in corresponding ds100mers (data not shown). The 5ghmC-ds100mer was obtained by glucosylation of the 5hmC-ds100mer using T4 phage β-GT.

Inhibition of BA on replication action of DNA polymerases

We next tested the effect of BA on replication action of DNA polymerases on 5ghmC-ds100mer using Taq DNA polymerase. qPCR was used to evaluate the replication activity. The cycle threshold (C_t) is the cycle number with a fluorescence signal equal to the defined threshold (40). We observed that BA increased the C_t values of 5ghmC-ds100mer from the BA-absent level (18.0) to 20.3 (Figure 1b). Because C_t levels are inversely proportional to the amount of amplifiable DNA template in the sample, the observed increase in C_t values suggests that the BA could reduce the initial amount of template glucosylated 5hmC to be amplified by Taq DNA polymerase. This result indicates that the BA could partially block the replication of glucosylated 5hmC in the ds100mer. In contrast, 5ghmC-ds100mer displayed the same qPCR amplification curve as 5hmC-ds100mer (unglucosylated), 5mC-ds100mer and C-ds100mer in the absence of BA (Figure 1c), suggesting that the glucosylation of 5hmC in DNA itself cannot affect the replication

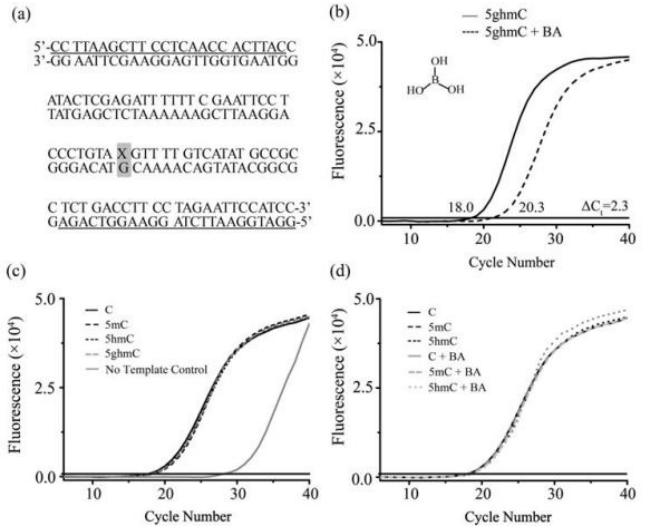


Figure 1. Boronic acid (BA) specifically inhibit the amplification activity of Taq DNA polymerase on 5ghmC-containing dsDNA. (**a**) The sequences of oligonucleotides containing modified cytosines, where X indicates C, 5mC 5hmC or 5ghmC, and the underlined sequences correspond to the forward and reverse primers that were used for PCR amplification. (**b**) The qPCR curves of 5ghmC-ds100mer in the presence (dashed line) or absence (solid line) of BA. (**c**) The qPCR curves of ds100mers containing a C (black solid line), 5mC (black dashed line), 5hmC (black dense dashed line) or 5ghmC (gray dashed line) at the 58th nucleotide in the absence of any BAs; gray solid lines indicate the PCR products of NTC (No Template Control). The solid straight lines indicate the defined threshold of qPCR. (**d**) The qPCR curves of C-ds100mer, 5mC-ds100mer and unglucosylated 5hmC-ds100mer in the presence of BA (followed by the gray solid line, gray dashed line and gray dense dashed line) or absence of BA (followed by the black solid line, black dashed line and black dense dashed line).

activity of Taq DNA polymerase and further supporting a critical role for BA to inhibit the replication activity of Taq DNA polymerase on 5ghmC-containing DNA. Moreover, BA cannot change the qPCR amplification curves of 5mC-ds100mer, C-ds100mer and unglucosylated 5hmC-ds100mer (Figure 1d), showing high specificity of BA blocking the replication of Taq DNA polymerase of the glucosylated 5hmC. This is the first report on BA blocking the replication activity of DNA polymerase by interacting with modified bases.

On the basis of above finding, we proposed a novel assay for rapid and sensitive detection of gene- or fragment-specific 5hmC, named as BA-mediated PCR assay. This assay does not require any restriction endonucleases and should be applicable to the detection of any sequences in genomic DNA.

Testing of BA derivatives for developing 5ghmC-sensitive qPCR assay

As described above, BA can inhibit the amplification of 5ghmC-containing DNA by Taq DNA polymerase. However, the inhibition is limited as shown by qPCR analysis ($\Delta C_t = 2.3$). To improve BA-mediated PCR assay, we further tested four BA derivatives, including PBAs, 3-CPBA, 2-CB-PBA and 3-D-PBA. Remarkably, we observed that selected BA derivatives (PBA, 3-CPBA, 2-CB-PBA and 3-D-PBA) could increase the C_t values of 5ghmC-ds100mer

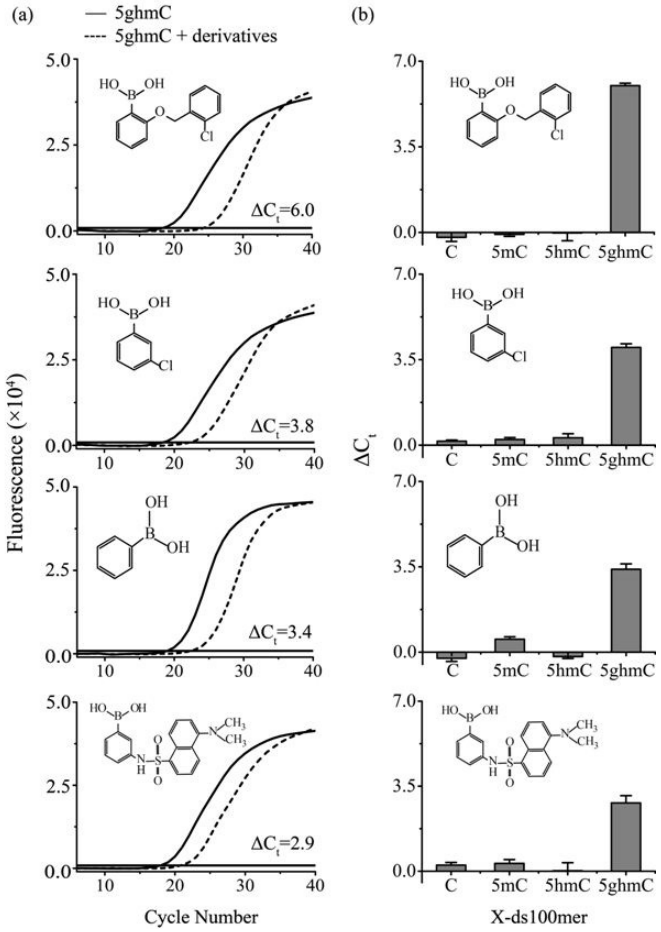


Figure 2. The effects of four BA derivatives on Taq DNA polymerase replicating 5ghmC-ds100mer. (**a**) qPCR curves in the presence (dashed line) or absence (solid line) of BA derivatives. (**b**) ΔC_t values for C-, 5mC-, 5hmC- and 5ghmC-containing DNA in the presence and absence of BA derivations. ΔC_t is the difference between the C_t value with BA derivatives and without BA derivatives. Four BA derivatives are shown 2-CB-PBA (2-(2'-chlorobenzyloxy) phenylboronic acid), 3-CPBA (3-chlorophenylboronic acid), PBA (phenyl-boronic acid) and 3-D-PBA (3-(Dansylamino) phenylboronic acid) from top to bottom. Error bars represent the standard deviation from the mean of three independent experiments.

from the BA-absent level (18.0) to 21.9–24.0 ($\Delta C_t = 2.9$–6.0), displaying higher blocking efficiency than BA.

Similarly, among the four synthesized probes (i.e. the C-, 5mC-, 5hmC- and 5ghmC-ds100mers), only the 5ghmC-ds100mer can be specifically discriminated by qPCR in the presence of one of the four BA derivatives (Figure 2b).

As shown in Figure 2a and b, 2-CB-PBA displays the highest inhibition efficiency ($\Delta C_t = 6.0$) among the five tested BAs. Moreover, non-specific blocking by other three probes (i.e. the C-, 5mC- and 5hmC-ds100mers) was not observed for 2-CB-PBA. In the following work, only 2-CB-PBA was used in BA-mediated qPCR assay to detect 5hmC-containing DNA.

It is not known whether the relative placement of 5hmC in DNA affects 2-CB-PBA-mediated qPCR efficiency. To test this possible effect, we synthesized three additional

5ghmC-ds83mer probes by placing a single 5ghmC at the 29th nucleotide, 32th nucleotide and 64th nucleotide from 5′ end (Figure 3a), respectively. It showed that all three ds83mers with the different placement of 5ghmC generated identical ΔC_t values ($\Delta C_t = 5.7$; Figure 3b) in response to the 2-CB-PBA. Therefore, the relative placements of the 5ghmC in DNA have no effect on the replication activity of Taq DNA polymerase, suggesting that the efficiency of 2-CB-PBA-mediated qPCR is mainly dependent on the presence of 5ghmC in DNA.

The next question we asked is whether the inhibition of Taq DNA polymerase by 2-CB-PBA remains to be specific on 5ghmC-containing DNA when co-existed with non-5ghmC DNA. To this purpose, we mixed the 5ghmC-ds100mer probe with a ds160mer without any modified site (Supplementary Table S1) at the molar ratio of 1:1. The total concentration of dsDNA was kept about 10.0 nM. As shown by melting curve analysis, the 5ghmC-ds100mer and the ds160mer have very distinct melting temperature of 84.4°C and 86.8°C, respectively. In the presence of 2-CB-PBA, only the amplification of 5ghmC-ds100mer probe was partially inhibited ($\Delta C_t = 6.0$, $T_m = 84.4$°C; Supplementary Figure S3a). In contrary, non-5ghmC-ds160mer can be amplified without any inhibition ($\Delta C_t = 0.1$, Tm = 86.8°C; Supplementary Figure S3b).

We further mixed the 5ghmC-ds100mer with genomic DNA from mouse Tet1/Tet2-knockout ES cells. The double knockout of Tet1 and Tet2 in mouse ES cells indeed eliminates the 5hmC in genomic DNA (15). Similarly, the amplification of 5ghmC-ds100mer was partially inhibited by 2-CB-PBA (Supplementary Figure S3c). This suggests the presence of genomic DNA would not interfere with the partial inhibition on the amplification of 5ghmC-ds100mer being caused by 2-CB-PBA.

In some regions of genomic DNA where 5hmCs are clustered, thereby it is helpful to test oligonucleotides containing multiple 5hmCs. For this purpose, we tested six ds83mers containing multiple hemi-hydroxymethylated CpG sites or symmetrically hydroxymethylated CpG sites (Figure 4a and c). All these ds83mers were glucosylated before PCR analysis. For hemi-hydroxymethylated ds83mers, the ΔC_t value increases from 6.0 to 6.99 with increasing the 5hmC number (1–3) (Figure 4b). For symmetrically hydroxymethylated ds83mers, the ΔC_t value increases from 6.5 to 9.0 with increasing the 5hmC pair number (1–3) (Figure 4d). The results clearly show that the blocking efficiency is positively correlated with 5hmC amounts within the tested regions.

Quantitative evaluation of 5hmC

To demonstrate that our BA-mediated PCR assay had the possibility of quantifying 5hmC levels, we diluted the 5ghmC-ds100mer with 5hmC-ds100mer at molar ratios of 0:1, 1:16, 1:8, 1:4, 1:2, 1:1, 2:1, 4:1 and 1:0 (5ghmC-ds100mer:5hmC-ds100mer). In this case, the 5ghmC-ds100mer probe was obtained from the complete glucosylation of 5hmC-ds100mer by T4 phage β-GT (Supplementary Figure S4). The total concentration of the two ds100mers was kept at 5.0 nM. As shown in Figure 5a and b, the ΔC_t value increased (from 0 to 6.0) with the increasing content of 5ghmC in the ds100mers. A linear relationship was observed between the ΔC_t values and log [5ghmC/ds100mer] ($\Delta C_t = 6.14 + 4.87 \log$ [5ghmC/ds100mer], $R^2 = 0.97$; Figure 5b). Interestingly, the ghmC-containing oligo could be sensitively detected even when mixing with 16-fold more unglucosylated 5hmC-containing oligo (Figure 5a). Native gel electrophoresis analysis of the final qPCR products was not as sensitive as qPCR, but clearly indicated that 2-CB-PBA reduced the product yield in the presence of 5ghmC (Supplementary Figure S5).

As described above, our 2-CB-PBA-mediated PCR assay can distinguish the 5ghmC-ds100mer in the presence of 16-fold more unglucosylated 5hmC-ds100mer (Figure 5). In contrast, the restriction enzyme MspI-based approach failed to detect glucosylated 5hmC-ds100mer ($\Delta C_t = 0$; Figure 5c), even without any dilution by the other ds100mers. This is reasonable. The recognition of 5hmC is solely dependent on MspI, which can cleave both 5mC and 5hmC but not glucosylated 5hmC. Given that 5hmC cleavage occurs only in the context of CCGG sequences (https://www.neb.com/nebecomm/tech_reference/epigenetics/epimark.asp), glucosylated 5hmC-ds100mers without any 5hmC in a CCGG sequence cannot be detected by the MspI-based approach. Noteworthy, early work showed that a duplex containing either 5hmC or 5mC at the inner position of CCGG is cleaved by MspI, however, if the 5hmC/5mC is placed at the outer cytosine position, the modified strand is protected from MspI whereas the unmodified strand is cleaved (18). Therefore, it is also possible that the sites containing 5mC and/or unglucosylated 5hmC cannot be cleaved by MspI.

Detection of gene-specific 5hmC in genomic DNA by 2-CB-PBA-mediated qPCR assay

We then asked whether BA-mediated PCR assay could be applied to the 5hmC detection of some specific genes or regions in genomic DNA. To this end, we tested three intron regions from a B cell transcription factor gene *Pax5* (marked as intron_*Pax5_1*, intron_*Pax5_2* and intron_*Pax5_3*) in mouse ES cell genomic DNA. As measured by 5hmC-deep sequencing (15, GSE43262) in our recent work, these regions display differential 5hmC abundance (evaluated by peak values: 246, 798 and 1534) in mouse ES cell genomic DNA. We also chose a region from UTR-5 of *Srr* (marked as UTR_*Srr*) that displays negligible 5hmC (peak value: 13) as an inner control. In this experiment, the optimized approach (2-CB-PBA-mediated qPCR assay) with higher sensitivity was used to test the 5hmC abundance in the chosen regions of genomic DNA of mouse ES cells. Our assay of the glucosylated genomic DNA from mouse ES cells showed an increased C_t value for all three intron regions of *Pax5* ($\Delta C_t = 1.5$, 3.6 and 7.5), indicating the presence of 5hmC (Figure 6a). In contrast, no significant change in the C_t value ($\Delta C_t = 0.1$) was observed for the chosen region of UTR-5 of *Srr*. As measured by our approach, the three intron regions of *Pax5* in glucosylated genomic DNA display differential 5hmC abundance in 5hmC. The obtained ΔC_t values are well correlated with that of chemical labeling sequencing analysis ($R^2 = 0.99$; Figure 6b). The amplification curve and melting curve suggest that

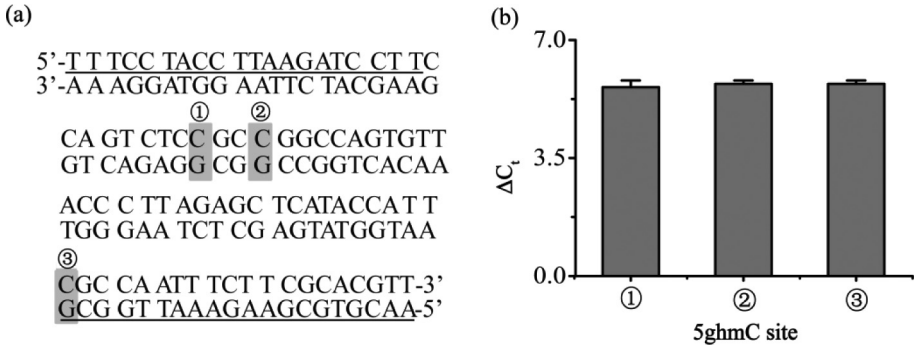


Figure 3. The effect of relative placement of the 5ghmC site for the replication activity of Taq DNA polymerase. (**a**) The sequences of the 5hmC-ds 83mer probes used for 2-CB-PBA-mediated PCR assay. We synthesized three 5ghmC-ds83mer probes by placing a single 5ghmC at the 29th nucleotide (?), 32th nucleotide (?) or 64th nucleotide (?) counted from 5′ end (Figure 3a). The underlined sequences correspond to the forward and reverse primers that were used for PCR amplification. (**b**) ΔC_t values for the different placements of the 5ghmC-ds 83mer DNA probes in the presence and absence of 2-CB-PBA.

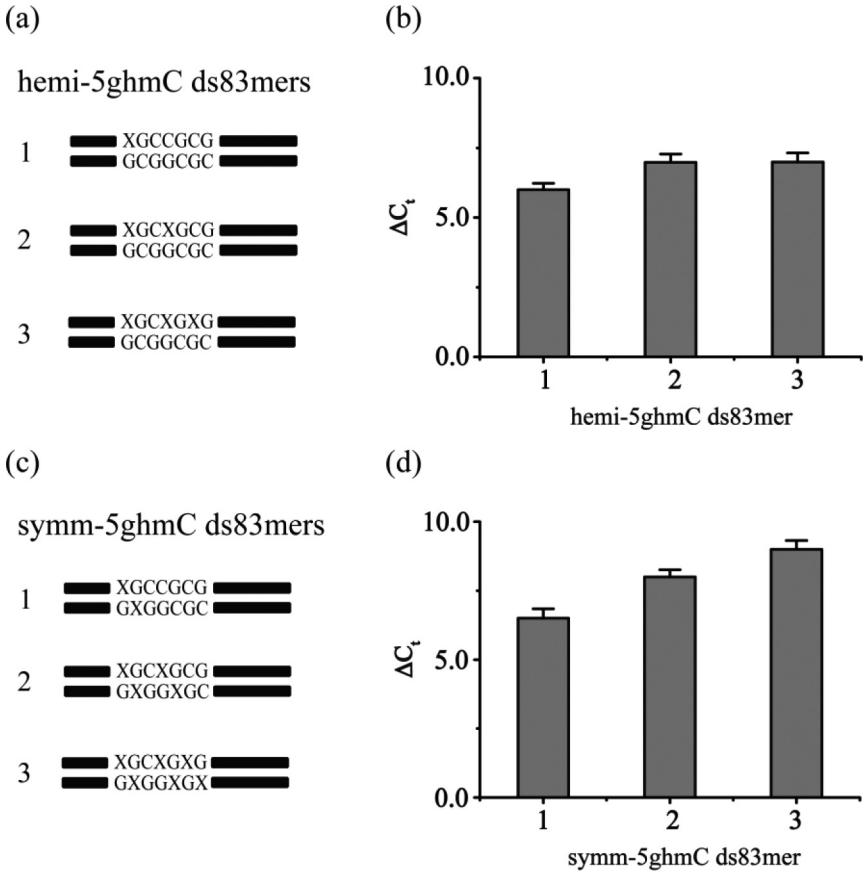


Figure 4. The effect of placement of the multiple 5hmC sites for the replication activity of Taq DNA polymerase using 2-CB-PBA-mediated PCR assay. (**a**) and (**c**): Six glucosylated ds83mers containing multiple hemi-hydroxymethylated CGs (a) or symmetrically hydroxymethylated CGs (c) at the 29th, 32th and 34th nucleotides from 5′ end were used as indicated. (**b**) and (**d**): ΔC_t values for the hemi-5ghmC (b) or symmetry-5ghmC-dsDNA probes (d) in the presence and absence of 2-CB-PBA. Error bars represent the standard deviation from the mean of three independent experiments. X indicates 5ghmC site. The sequence of 5hmC-ds83mers was the same as listed in Figure 3a. The 5hmC-ds83mers were subjected to be glucosylated as described in the Materials and Methods section.

the amplification of intron-*Pax5_3* is completely inhibited in the presence of 2-CB-PBA (Supplementary Figure S6a and b). The gel electrophoresis of final PCR products also confirmed the predominant inhibition of the amplification of glucosylated intron-*Pax5_3* by 2-CB-PBA (Supplementary Figure S6c). Consistently, both our current approach and chemical labeling-Seq analysis show that the intron-*Pax5_3* region (110 bp) displays the highest 5hmC abundance.

By mixing the genomic DNA of WT with the genomic DNA from the mutant cells depleting Tet1 and Tet2 in varying ratios, we observed that 5hmC-abundant *PAX5_3* could be detected at the ratio of 1:64 (Supplementary Figure S7).

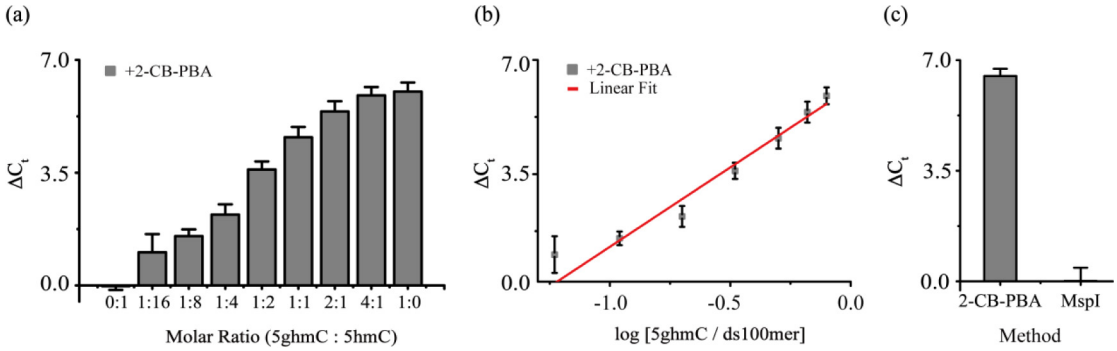


Figure 5. Quantitative evaluation of 5ghmC-ds100mer by 2-CB-PBA-mediated PCR. (**a**) The ΔC_t values of mixed dsDNA templates (5ghmC-ds100mer and 5hmC-ds100mer) with the varying molar ratios of 0:1, 1:16, 1:8, 1:4, 1:2, 1:1, 2:1, 4:1 and 1:0. (**b**) The linear relationship between the ΔC_t value and the log [5ghmC/ds100mer]. ($\Delta C_t = 6.14 + 4.87$ log [5ghmC/ds100mer], $R^2 = 0.97$). (**c**) The comparison of 2-CB-BA-mediated PCR assay and MspI-qPCR analysis of 5ghmC-ds100mer.

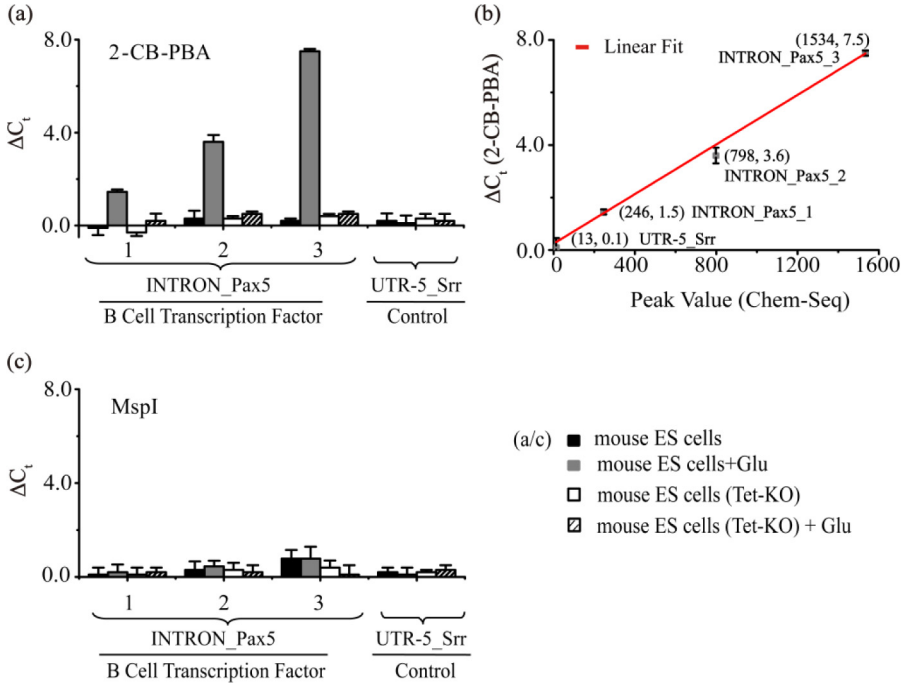


Figure 6. The 2-CB-PBA-mediated qPCR assay for fragment-specific detection of 5hmC in the indicated INTRON-*Pax5* regions of genomic DNA of mouse embryonic stem (ES) cells. (**a**) The ΔC_t values obtained by the 2-CB-PBA-mediated qPCR assay. (**b**) The correlation of the 2-CB-PBA-mediated qPCR with Chem-Seq analysis of 5hmC. (**c**) The ΔC_t values obtained by the MspI-qPCR assay. The primers designed for amplification of the target DNA regions were listed in Supplementary Table S2. UTR-5_*Srr* was included as inner control from qPCR analysis. Error bars represent the standard deviation from the mean of at least three experiments. 'Glu' indicates the glucosylation of DNA by β-GT. Tet-KO indicates the double knockout of Tet1 and Tet2.

The results hint that our approach is sensitive for detection of 5hmC-abundant genes. The observed better sensitivity for detection of 5hmC in *PAX_3* (Supplementary Figure S7 versus Figure 5) is attributed to the presence of multiple 5hmC in the region of *PAX5_3*.

Moreover, the C_t values of unglucosylated DNA are basically unchanged ($\Delta C_t = 0.1$–0.3; Figure 6a), suggesting that the selected PCR amplification regions do not contain any spontaneous 5ghmC in these cells. Therefore, the detected 5ghmC by 2-CB-PBA-mediated qPCR assay results exclusively from the β-GT-catalyzed glucosylation of 5hmC in the extracted genomic DNA of mouse ES cells, confirming the specificity of the BA-mediated PCR approach.

By knocking out Tet1/Tet2 mediating 5mC oxidation and 5hmC formation, no 5hmC was detected in all three intron regions of *Pax5* (Figure 6a), confirming the reliability of our approach.

To verify the depletion of Tet1 and Tet2, we measured the expression of Tet1 and Tet2 at mRNA levels in the Tet1/Tet2 double knockout ES cells using qPCR. Indeed, the levels of Tet1 and Tet2 expression in Tet1/Tet2 double knockout ES cells decrease significantly by two or three orders of magnitude (Supplementary Figure S8) when compared with that in ES cells of WT.

We further tested the three intron regions of *Pax5* using the MspI-based restriction–glucosylation assay from ge-

nomic DNA. 5hmC could not be detected in all three regions (Figure 6c). These regions do not have any CCGG site, and their C_t values for unglucosylated and glucosylated DNA are essentially unchanged ($\Delta C_t < 0.7$; Figure 6c). These results suggest the limitation of the restriction endonuclease approach, which requires cleavable sites. In contrast, our assay shows the differential distribution of 5hmC in the three regions, which are consistent with 5hmC-deep sequencing. Therefore, our assay does not exhibit such limitation and can be used to detect 5hmC in all of the selected regions of genomic DNA.

To evaluate the sensitivity of our approach and show the biologically relevant applications, we measured 5hmC distribution of 10 regions in human MRC-5 cells, which have naturally low expression of all three TET proteins. The 5hmC abundance in human MRC-5 cells is comparable with that of most of human tissues. The tested regions and the sequences of the designed primers were listed in Supplementary Table S3. By the 2-CB-PBA-mediated PCR method, we observed that three regions (one CDS_UBC, one INTRON_GOSR2 and one INTRON_FBXO32) exhibit the highest 5hmC abundance ($\Delta C_t = 3.2$) among 10 tested regions (Supplementary Figure S9). Consistently, 5hmC DIP-Seq analysis [GSE44457, (15)] showed that the three regions also have the highest 5hmC abundance (peak values: 850–1158). Together, we conclude that our approach is sensitive for the detection of gene-specific 5hmC in biologically relevant samples.

DISCUSSION

Here, we demonstrate that all unglucosylated ds100mers are normally amplified using Taq DNA polymerase regardless of whether any of BA and its derivatives is present. This result indicates that BA and its derivatives cannot directly inhibit or block the amplification activity of Taq DNA polymerase. Such amplification activity is partly blocked only when the 5hmC in template DNA is modified by one glucose molecule and when BA or one of its derivatives is present. Therefore, the observed blocking of Taq DNA polymerase by glucosylated 5hmC can be attributed to the specific interaction of BA (or its derivatives) with the glucosylated 5hmC, which most likely leads to the formation of a bulky covalent but dynamic complex of BA–glucose–5hmC (41–45).

Here, we also demonstrate the application of the BA-mediated PCR assay for the detection of 5hmC in *Pax5* gene of mouse ES cell genomic DNA. Our results reveal the highest abundance of 5hmC in intron-*Pax5* regions among tested 23 regions of genomic DNA of mouse ES cells (Figure 6 and Supplementary Figure S10). Therefore, we identified *PAX5_3* as a 5hmC-rich fragment. *Pax5* is a transcription factor to regulate the development and function of mature B cell (46). *Pax5* is also implicated in chromosome translocation, aberrant high frequency mutation and aberrant methylation, which were found in hematopoietic malignancies (lymphoblastic leukemia, myeloid leukaemia) (46). The role of the abundant 5hmC in *Pax5* remains unclear.

By our approach, the 23 regions of genomic DNA of mouse ES cells were simultaneously measured in one batch of tests using PCR instrument with 96 channels (Figure 6 and Supplementary Figure S10). This indicates that our assays are potentially high throughput.

In summary, we describe a novel DNA polymerase amplification assay for distinguishing 5hmC from 5mC without the need for restriction enzymes. This was achieved by introducing BAs to block the replication of DNA polymerase on glucosylated 5hmC. Theoretically, the assay can be applied to detect 5hmC in any DNA sequence, eliminating the sequence limitation of recently developed restriction–glucosylation assays. Essentially, our study open a new way of detecting 5hmC for chemical, biological, medical and pharmacological studies.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

Dr H. Wang acknowledges the support from the National Basic Research Program of China and the National Natural Science Foundation of China.

FUNDING

Basic Research Program of China [2009CB421605 and 2010CB933502 to H.W.]; National Natural Science Foundation of China [21077129 and 21125523 to H.W.]. Funding for open access charge: Natural Science Foundation of China [21125523].

Conflict of interest statement. None declared.

REFERENCES

1. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
2. Iyer,L.M., Tahiliani,M., Rao,A. and Aravind,L. (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, **8**, 1698–1710.
3. Kriaucionis,S. and Heintz,N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science*, **324**, 929–930.
4. Ito,S., D'Alessio,A.C., Taranova,O.V., Hong,K., Sowers,L.C. and Zhang,Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129–1135.
5. Hackett,J.A., Sengupta,R., Zylicz,J.J., Murakami,K., Lee,C., Down,T.A. and Surani,M.A. (2013) Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science*, **339**, 448–452.
6. Hajkova,P., Jeffries,S.J., Lee,C., Miller,N., Jackson,S.P. and Surani,M.A. (2010) Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science*, **329**, 78–82.
7. Wossidlo,M., Nakamura,T., Lepikhov,K., Marques,C.J., Zakhartchenko,V., Boiani,M., Arand,J., Nakano,T., Reik,W. and Walter,J. (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.*, **2**, 1–8.
8. Yamaguchi,S., Hong,K., Liu,R., Inoue,A., Shen,L., Zhang,K. and Zhang,Y. (2013) Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Res.*, **23**, 329–339.
9. Valinluck,V. and Sowers,L.C. (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.*, **67**, 946–950.

10. Pfaffeneder,T., Hackner,B., Truss,M., Muenzel,M., Mueller,M., Deiml,C.A., Hagemeier,C. and Carell,T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.*, **50**, 7008–7012.
11. Li,W. and Liu,M.J. (2011) Distribution of 5-hydroxymethylcytosine in different human tissues. *J. Nucleic Acids*, **2011**,870726, 1–5. http://dx.doi.org/10.4061/2011/870726 (18 March 2014, date last accessed).
12. Tan,L. and Shi,Y.G. (2012) Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development*, **139**, 1895–1902.
13. Jin,S.G., Jiang,Y., Qiu,R., Rauch,T.A., Wang,Y., Schackert,G., Krex,D., Lu,Q. and Pfeifer,G.P. (2011) 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res.*, **71**, 7360–7365.
14. Yin,R., Mao,S, Zhao,B., Chong,Z., Yang,Y., Zhao,C., Zhang,D., Huang,H., Gao,J., Li,Z. *et al.* (2013) Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J. Am. Chem. Soc.*, **135**, 10396–10403.
15. Zhao,B., Yang,Y., Wang,X., Chong,Z., Yin,R., Song,S. H., Zhao,C., Li,C., Huang,H., Sun,B. F. *et al.* (2014) Redox-active quinones induces genome-wide DNA methylation changes by an iron-mediated and Tet-dependent mechanism. *Nucleic Acids Res.*, **42**, 1593–1605.
16. Huang,Y., Pastor,W.A., Shen,Y., Tahiliani,M., Liu,D.R. and Rao,A. (2010) The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE*, **5**, e8888.
17. Nestor,C., Ruzov,A., Meehan,R.R. and Dunican,D.S. (2010) Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *Biotechniques*, **48**, 317–319.
18. Tardy-Planechaud,S., Fujimoto,J., Lin,S.S. and Sowers,L.C. (1997) Solid phase synthesis and restriction endonuclease cleavage of oligodeoxynucleotides containing 5-(hydroxymethyl)-cytosine. *Nucleic Acids Res.*, **25**, 553–558.
19. Szwagierczak,A., Bultmann,S., Schmidt,C.S., Leonhardt,H. and Spada,F. (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.*, **38**, e181.
20. Zheng,Y., Cohen-Karni,D., Xu,D., Chin,H.G., Wilson,G., Pradhan,S. and Roberts,R.J. (2010) A unique family of Mrr-like modification-dependent restriction endonucleases. *Nucleic Acids Res.*, **38**, 5527–5534.
21. Song,C.-X., Yu,M., Dai,Q. and He,C. (2011) Detection of 5-hydroxymethylcytosine in a combined glycosylation restriction analysis (CGRA) using restriction enzyme Taq$^{\alpha}$I. *Bioorg. Med. Chem. Lett.*, **21**, 5075–5077.
22. Wang,H., Guan,S., Quimby,A., Cohen-Karni,D., Pradhan,S., Wilson,G., Roberts,R.J., Zhu,Z. and Zheng,Y. (2011) Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **39**, 9294–9305.
23. Szwagierczak,A., Brachmann,A., Schmidt,C.S., Bultmann,S., Leonhardt,H. and Spada,F. (2011) Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **39**, 5149–5156.
24. Bhattacharyya,S., Yu,Y.T., Suzuki,M., Campbell,N., Jozef Mazdo,J., Vasanthakumar,A., Bhagat,T.D., Nischal,S., Christopeit,M., Parekh,S. *et al.* (2013) Genome-wide hydroxymethylation tested using the HELP-GT assay shows redistribution in cancer. *Nucleic Acids Res.*, **41**, e157.
25. Jin,S.G, Wu,X. and Li,A.X. (2011) Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Res.*, **39**, 5015–5024.
26. Ko,M., Huang,Y. and Jankowska,A.M. (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, **468**, 839–843.
27. Robertson,A.B., Dahl,J.A., Vagbo,C.B., Tripathi,P., Krokan,H.E. and Klungland,A. (2011) A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.*, **39**, e55.
28. Song,C.X., Szulwach,K.E., Fu,Y., Dai,Q., Yi,C., Li,X., Li,Y., Chen,C.H., Zhang,W., Jian,X. *et al.* (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.*, **29**, 68–72.
29. Pastor,W.A., Pape,U.J., Huang,Y., Henderson,H.R., Lister,R., Ko,M., McLoughlin,E.M., Brudno,Y., Mahapatra,S., Kapranov,P. *et al.* (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, **473**, 394–397.
30. Yu,M., Hon,G.C., Szulwach,K.E., Song,C.-X., Zhang,L., Kim,A., Li,X.K., Dai,Q., Shen,Y., Park,B. *et al.* (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian Genome. *Cell*, **149**,1368–1380.
31. Booth,M.J., Branco,M.R., Ficz,G., Oxley,D., Krueger,F., Reik,W. and Balasubramanian,S. (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*, **336**, 934–937.
32. Ito,S., Shen,L., Dai,Q., Wu,S.C., Collins,L.B., Swenberg,J.A., He,C. and Zhang,Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
33. He,Y.F., Li,B.Z., Li,Z., Liu,P., Wang,Y., Tang,Q.Y., Ding,J.P., Jia,Y.Y., Chen,Z.C., Li,L. *et al.* (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303–1307.
34. Le,X.C., Zhang,Y., Dovichi,N.J., Compston,C.A., Palcic,M.M., Beever,R.J. and Hindsgaul,O. (1997) Study of the enzymatic transformation of fluorescently labeled oligosaccharides in human epidermoid cells using capillary electrophoresis with laser-induced fluorescence detection. *J. Chromatogr. A*, **781**, 515–522.
35. Ren,L., Liu,Z., Liu,Y., Dou,P. and Chen,H.Y. (2009) Ring-opening polymerization with synergistic co-monomers: access to a boronate-functionalized polymeric monolith for the specific capture of cis-diol-containing biomolecules under neutral conditions. *Angew. Chem. Int. Ed. Engl.*, **48**, 6704–6707.
36. Pal,A., Berube,M. and Hall,D.G. (2010) Design, synthesis, and screening of a library of peptidyl bis (boroxoles) as oligosaccharide receptors in water: identification of a receptor for the tumor marker TF-antigen disaccharide. *Angew. Chem. Int. Ed.*, **49**, 1492–1495.
37. Li,H., Wang,H., Liu,Y. and Liu,Z. (2012) A benzoboroxole-functionalized monolithic column for the selective enrichment and separation of cis-diol containing biomolecules. *Chem. Commun.*, **48**, 4115–4117.
38. Lü,C., Li,H., Wang,H. and Liu,Z. (2013) Probing the interactions between boronic acids and cis-diol-containing biomolecules by affinity capillary electrophoresis. *Anal. Chem.*, **85**, 2361–2369.
39. Wang,X.L., Suo,Y.S., Yin,R.C., Shen,H.Q. and Wang,H.L. (2011) Ultra-performance liquid chromatography/tandem mass spectrometry for accurate quantification of global DNA methylation in human sperms. *J. Chromatogr. B*, **879**, 1647–1652.
40. Heid,C.A., Stevens,J., Livak,K.J. and Williams,P.M. (1996) Real time quantitative PCR. *Genome Res.*, **6**, 986–994.
41. Yan,J., Springsteen,G., Deeter,S. and Wang,B.H. (2004) The relationship among pK_a, pH, and binding constants in the interactions between boronic acids and diols—it is not as simple as it appears. *Tetrahedron*, **60**, 11205–11209.
42. Norrild,J.C. and Eggert,H. (1995) Evidence for mono-and bisdentate boronate complexes of glucose in the furanose form. *J. Am. Chem. Soc.*, **117**, 1479–1484.
43. Bielecki,M., Eggert,H. and Norrild,J.C. (1999) A fluorescent glucose sensor binding covalently to all five hydroxyl groups of α-D-glucofuranose. *J. Chem. Soc. [Perkin 2]*, **3**, 449–455.
44. Norrild,J.C. and Eggert,H. (1996) Boronic acids as fructose sensors. Structure determination of the complexes involved using $^{1}J_{CC}$ coupling constants. *J. Chem. Soc. [Perkin 2]*, **12**, 2583–2588.
45. Hall,D.G. (ed.). (2005) *Boronic Acids: Preparation and Applications in Organic Synthesis and Medicine*. Wiley-VCH, Germany.
46. Medvedovic,J., Ebert,A., Tagoh,H. and Busslinger,M. (2011) *Pax5*: a master regulator of B cell development and leukemogenesis. *Adv. Immunol.*, **111**, 179–206.