A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA

Adam B. Robertson[1], John A. Dahl[1], Cathrine B. Vågbø[2], Pankaj Tripathi[3], Hans E. Krokan[2] and Arne Klungland[1,4,*]

[1]Centre for Molecular Biology and Neuroscience and Institute of Medical Microbiology, Oslo University Hospital, Rikshospitalet, [2]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, N-7489 Trondheim, Norway, [3]Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam, The Netherlands and [4]Institute of Basic Medical Sciences, University of Oslo, PO Box 1018 Blindern, N-0315 Oslo, Norway

ABSTRACT

Recently, 5-hydroxymethylcytosine (5hmC) was identified in mammalian genomic DNA. The biological role of this modification remains unclear; however, identifying the genomic location of this modified base will assist in elucidating its function. We describe a method for the rapid and inexpensive identification of genomic regions containing 5hmC. This method involves the selective glucosylation of 5hmC residues by the β-glucosyltransferase from T4 bacteriophage creating β-glucosyl-5-hydroxymethylcytosine (β-glu-5hmC). The β-glu-5hmC modification provides a target that can be efficiently and selectively pulled down by J-binding protein 1 coupled to magnetic beads. DNA that is precipitated is suitable for analysis by quantitative PCR, microarray or sequencing. Furthermore, we demonstrate that the J-binding protein 1 pull down assay identifies 5hmC at the promoters of developmentally regulated genes in human embryonic stem cells. The method described here will allow for a greater understanding of the temporal and spatial effects that 5hmC may have on epigenetic regulation at the single gene level.

INTRODUCTION

DNA methylation at cytosine residues in mammalian cells is a well known and well described modification affecting gene expression. This DNA modification consists of a cytosine that is modified by a methyl group at the N5 position (5meC). The 5meC modification generally occurs at the CpG dinucleotide sequence; however, the 5meC modification has been identified elsewhere in the genome [For a recent review see ref. (1)].

Two independent groups recently discovered another type of DNA modification that is speculated to be involved in gene regulation, 5-hydroxymethylcytosine (5hmC) (2,3). Tahiliani *et al.* (2) demonstrated that the enzyme Tet1, an iron-dependent α-ketoglutarate dioxygenase, catalyzes the formation of 5hmC from 5meC. Furthermore, this group suggested that the 5hmC base may be an intermediate in the conversion of 5meC to cytosine, thus identifying an enzyme that can potentially demethylate DNA (2). Kriaucionis and Heintz demonstrated that 5hmC is a stable DNA modification found in specialized non-dividing neurons and was likely present in the animal tissues they studied. Further studies have confirmed the presence of 5hmC in these mammalian tissues (4,5). Intriguingly, 5hmC was not detected in cancerous cell lines. The inability to detect 5hmC in cancerous cell lines suggests that the lack of 5hmC may be involved in tumorigenesis (3). This result also allows for speculation that 5hmC is involved in epigenetic gene regulation. This alleged involvement of 5hmC in epigenetic regulation has been experimentally verified by Jin *et al.* who show that 5hmC in DNA inhibits the binding of several methyl-CpG-binding domain proteins (6)—proteins that are known to regulate transcription by interaction with 5meC (7). Further research by Ito *et al.* has shown that Tet2 and Tet3 also catalyze the formation of 5hmC (8). Additionally, two independent groups have been able to precisely quantify the amount of 5hmC present in several mammalian tissues (9–11); however, the specific genomic location of 5hmC remains unknown.

*To whom correspondence should be addressed. Tel: +47 23074072/+47 47 840 305; Fax: +47 23074061; Email: arne.klungland@rr-research.no

The identification of specific genomic regions containing 5hmC has been shown to be technically challenging. The most frequent method of identifying 5meC, bisulfite sequencing, cannot distinguish 5meC from 5hmC (11–13). Furthermore, it has been questioned whether commercially available antibodies raised against 5hmC can distinguish between 5meC and 5hmC (8). We note that Ko *et al.* have developed an antiserum that recognizes 5-methylenesulfonate, the product of bisulfite conversion of 5hmC and therefore this antiserum may prove useful for identifying the location of 5hmC (14). Using polymerase kinetics one group has been able to differentiate between 5meC and 5hmC using model substrates (15). This sequencing method may prove useful for genome wide analysis; however, it requires specialized and expensive equipment making this method impractical for identification of the 5hmC status of individual genes and has yet to be demonstrated in *in vivo* studies (11). With these difficulties in mind we tried to develop a simple and cost-effective means of identifying the 5hmC-containing genomic regions.

The DNA of wild-type bacteriophage T4 is nearly devoid of cytosine residues, which are replaced by 5hmC. Furthermore, these 5hmC residues are glucosylated by the T4-encoded α-glucosyltransferase or β-glucosyltransferase (16), the latter proving more efficient when used for *in vitro* glucosylation assays (9,17,18). We have exploited the glucosylation of 5hmC residues to mark 5hmC residues in mammalian DNA. Glucosylated 5hmC is chemically similar to β-glucosyl-5-hydroxymethyluracil, which is specifically recognized by DNA binding proteins from certain protozoa.

African trypanosomes and related kinetoplastids contain the highly modified base β-glucosyl-5-hydroxymethyluracil (often referred to as the J-base) in their DNA (19–21). These organisms contain J-binding proteins (JBP1 and JBP2) and notably JBP1 specifically binds to DNA containing the J-base (22–24). Given that JBP1 is able to bind J-containing DNA we reasoned that JBP1 may be able to cross-react significantly with β-glucosyl-5-hydroxymethylcytosine (β-glu-5hmC) containing DNA.

We show here that 5hmC can be selectively identified in genomic regions (Figure 1) by modifying 5hmC residues in genomic DNA using the T4 β-glucosyltransferase (β-gt) to create β-glu-5hmC residues. DNA containing these residues can be pulled down with J-binding protein 1 (JBP1) from *Crithidia fasciculate,* allowing for the efficient identification of genomic regions containing 5hmC. Importantly, the JBP1 pull down assay was applied to identify 5hmC at promoters of developmentally regulated genes in human embryonic stem (hES) cells. This report is, to our knowledge, the first to identify the location of 5hmC in any genome.

MATERIALS AND METHODS

Protein purifications

β-Glucosyltransferase. The *bgt* gene was amplified from T4 bacteriophage DNA and was cloned into pET28a. Cultures of Rosetta(DE3)pLysS harboring pET28a-bgt were grown in 500 ml Studier auto-inducing media (25) to an A_{600} of 0.6 at 37°C followed by a shift to 18°C for 20 h. The cells were harvested by centrifugation and suspended in 10 ml β-gt lysis buffer [500 mM NaCl, 25 mM Hepes KOH (pH 7.9), 5 mM imidazole, 10% (v/v) glycerol]. All subsequent steps were carried out at 4°C unless otherwise specified.

After suspension in lysis buffer, cells were incubated for 1 h with lysozyme added to a final concentration of 200 μg/ml. After 1 h Triton-X100 was added to the lysate to a final concentration of 0.1% (v/v) and the lysate was heated briefly to 20°C. The lysate viscosity was reduced by sonication on ice. The lysate was clarified by centrifugation for 60 min at 20 000 rpm in a SS34 rotor (Sorvall). After centrifugation the supernatant was applied and batch bound for 1 h to 2 ml of Talon Resin (GE Healthcare) equilibrated in β-gt lysis buffer. The lysate was applied to a column and washed with β-gt lysis buffer until the flow through had had an A_{280} of 0. β-gt was eluted with 20 ml β-gt lysis containing 200 mM imidazole.

Fractions containing the highest concentration of protein were pooled and dialyzed 3 times against 100 volumes of β-gt low salt buffer [50 mM NaCl, 25 mM Hepes KOH (pH 7.9), 0.5 mM EDTA (pH 8.0), 10% (v/v) glycerol]. The pool was then applied to a Resource S cation exchange column. The column was eluted with a 50–500 mM linear NaCl gradient. Relevant fractions were pooled and dialyzed against 100 volumes β-gt storage buffer [250 mM NaCl, 25 mM Hepes KOH (pH 7.9), 1 mM EDTA (pH 8.0), 50% (v/v) glycerol].

J-binding protein 1. The gene coding for J-binding protein 1 (JBP1) from *C. fasciculata* was synthesized by GeneArt (Germany). After synthesis the gene was cloned into pET28a. Cultures of Rosetta(DE3)pLysS harboring pET28a-JBP1 were grown in 500 ml Studier auto-inducing media to an OD A_{600} of 0.6 at 37°C followed by a shift to 18°C for 20 h. The cells were harvested by centrifugation and suspended in 10 ml JBP1 lysis buffer [500 mM NaCl, 25 mM Tris-HCl (pH 7.5), 5 mM imidazole, 10% (v/v) glycerol]. All subsequent steps were carried out at 4°C unless otherwise specified.

After suspension in JBP1 lysis buffer, cells were incubated for 1 h with lysozyme added to a final concentration of 200 μg/ml. After 1 h Triton-X100 was added to the lysate to a final concentration of 0.1% (v/v) and the lysate was heated briefly to 20°C. The lysate viscosity was reduced by sonication on ice. The lysate was clarified by centrifugation for 10 min at 10 000 rpm in a SS34 rotor (Sorvall). After centrifugation the supernatant was applied and batch bound for 1 h to 2 ml of Talon Resin (GE Healthcare) equilibrated in JBP1 lysis buffer. The lysate was applied to a column and washed with JBP1 lysis buffer until the flow through had an OD A_{280} of 0. JBP1 was eluted with 20 ml JBP1 lysis buffer containing 200 mM imidazole.

Fractions containing the highest concentration of protein were pooled and concentrated to 500 μl using a centricon MWCO 30000 according to the manufacturer's
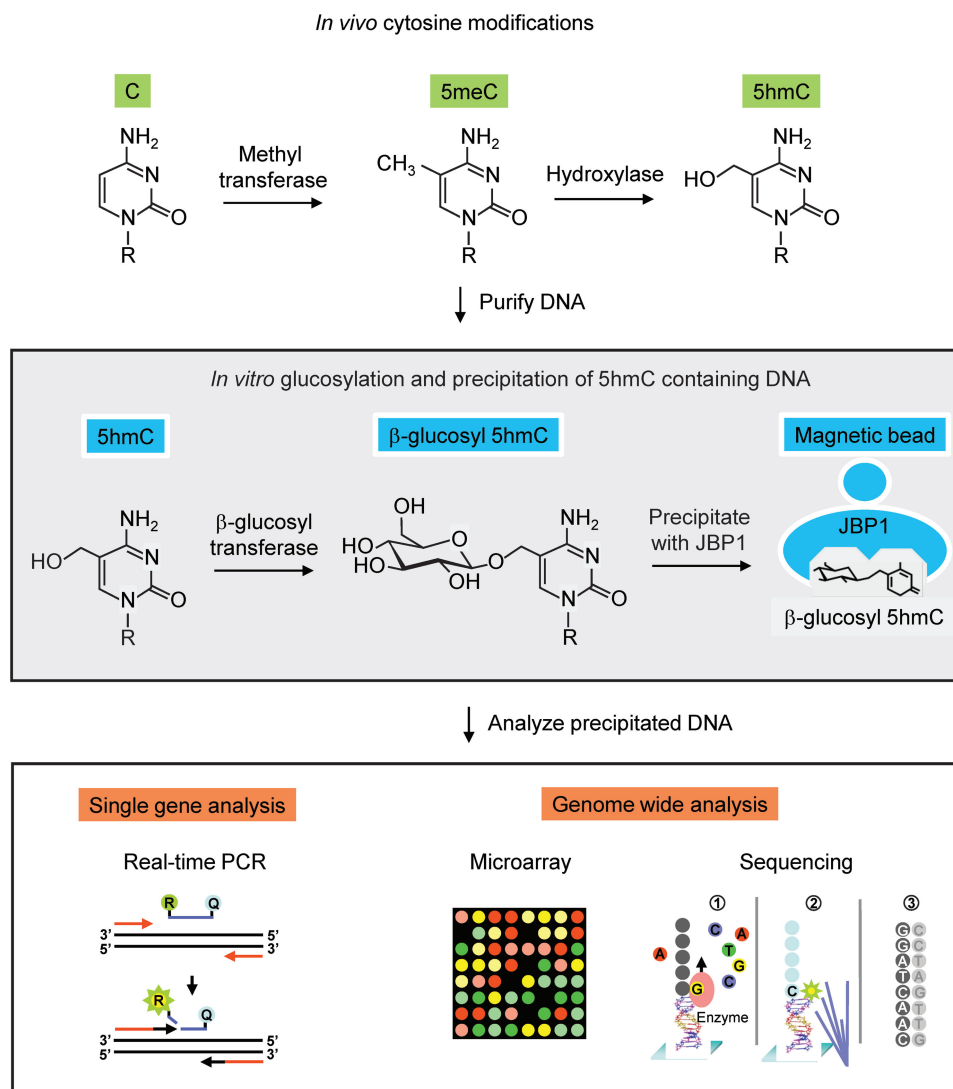
Figure 1. Scheme for the selective pull down of 5hmC. The top panel shows the known naturally occurring cytosine modifications in mammalian cells. The middle panel gives an overview of the method used to specifically identify 5hmC bases in genomic DNA. The bottom panel shows some applications that can be employed using this method of identifying 5hmC regions.

instructions. The concentrated protein was applied to a Superdex75 size exclusion column and eluted with JBP1 Superdex Buffer [250 mM NaCl, 25 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0), 1 mM DTT, 10% (v/v) glycerol]. Fractions containing no detectable contaminants (as judged by SDS–PAGE) were pooled and 3 times dialyzed against 100 volumes PBS [150 mM NaCl, 50 mM KPO$_4$ (pH 7.2)].

Coupling JBP1 to magnetic beads

Five milligrams of Epoxy modified magnetic beads (Dynal, Oslo, Norway) were equilibrated according to the manufacturer's instructions and suspended in 60 µl PBS. About 100 µg of JBP1 in PBS was added to the beads and the PBS was added until the volume equaled 120 µl, followed by the addition of 40 µl 4 M (NH$_4$)$_2$SO$_4$. The bead/JBP1 solution was slowly rotated at 4°C for 48 h. After incubation, the protein not bound to the

beads was removed and binding efficiencies were calculated (typical binding efficiencies were between 70 and 80%). The beads were then blocked with 300 µl binding buffer [2 mM EDTA (pH 8.0), 10 mM Tris (pH 8.0), 150 mM NaCl, 0.02% (v/v) Tween-20, 1 mg/ml BSA]. After blocking beads were washed three times with 300 µl binding buffer and finally suspended in 300 µl binding buffer. Beads were prepared freshly for each experiment as it was found that several freeze thaw cycles dramatically reduced binding capacity.

Substrates

Oligonucleotide substrates. Annealing reactions (50 µl) containing 100 pmol of each complementary oligonucleotide (Supplementary Table S1), 40 mm Tris–HCl (pH 7.5), 10 mM MgCl$_2$, 1 mM dithiothreitol, 50 mM NaCl were heated to 95°C in a thermocycler for 5 min followed decrease in temperature of 1°C/min to 25°C. Duplex

oligonucleotides were then purified from a 15% non-denaturing PAGE according to the protocol established by Sambrook *et al.* (26).

PCR amplified substrates. Substrates containing cytosine residues or 5meC residues were amplified from pUC18 or from specific mouse genomic regions (Supplementary Table S2) using a PCR reaction containing unmodified dNTPs with Pfu Turbo (Stratagene) according to the manufacturer's instructions. 5meC residues were created by incubating the PCR product amplified using unmodified dNTPs with M. SssI methyltransferase (NEB) according to the manufacturer's instructions. Substrates containing 5hmC residues or β-glu-5hmC residues were amplified as other substrates except that d5hmCTP (Bioline) was used in place of dCTP. β-glu-5hmC substrates were generated by incubating the PCR product created using d5hmCTP with 10 μg β-gt, 20 μM UDP-glucose (Sigma) in β-gt reaction buffer [20 mM KPO_4 (pH 8.0), 25 mM $MgCl_2$]. Substrates were purified on a 0.8% agarose gel and excised using a gel extraction kit (Qiagen).

Radiolabeled substrates. Reactions (50 μl) containing 1 μg DNA substrate, 10 μCi γ-[^{32}P]ATP, 2.5 U T4 polynucleotide kinase (NEB) and buffer supplied by the manufacturer were incubated at 37°C for 30 min. After 30 min 100 μmol ATP was added and the reaction was allowed to proceed for 10 min. Enzyme and free nucleotide was removed using a nucleotide removal kit (Qiagen).

β-gt specificity and activity assay

Reactions (50 μl) containing 1 pmol double stranded oligonucleotide substrate, 20 μM UDP-glucose, 10 μg β-gt and β-gt reaction buffer were incubated at 37°C for 60 min. Reactions were terminated by the addition of 10 μl Stop Buffer [20 μg Protenase K, 50 mM Tris–HCl (pH 8.0), 100 mM EDTA (pH 8.0), 2% (w/v) SDS] followed by incubation at 42°C for 60 min. Oligonucleotides were purified by phenol:CHCl$_3$:IAA extraction followed by an ethanol precipitation. Pellets were suspended in 50 μl NEB buffer 4 and 10 U *Taq*I (NEB) and were allowed to incubate at 65°C for 16 h. *Taq*I was inactivated by heating to 80°C for 20 min followed by an incubation with 2.5 U of shrimp alkaline phosphatase (NEB) for 30 min. The alkaline phosphatase was inactivated by heating to 65°C for 5 min. Reactions were radiolabeled by adding 2.5 U T4 polynucleotide kinase, 10 μCi γ-[^{32}P]-ATP and the volume was raised to 70 μl with T4 polynucleotide kinase buffer. DNA was cleaned by phenol:CHCl$_3$:IAA (isoamyl alcohol) extraction followed by an ethanol precipitation. Pellets were suspended in 5 μl DNase I buffer and 0.2 U DNase I (NEB) and 0.2 U snake venom phosphodiesterase (Worthington) was added to the reactions. Reactions were incubated at 25°C for at least 4 h. About 0.5 μl of each reaction was spotted onto a $100 \times 20 \times 20$ cm cellulose thin layer plate. The nucleotides were resolved in two phases: the first phase contained isobutyric acid:water:NH$_3$ (66:20:1) the second phase contained 4 M ammonium sulfate: 1 M acetic acid:

isopropanol (80:17:2). Radioactive spots were identified using a phosphor screen.

LC/MS/MS analysis

DNA substrates were composed of 2.7 kb linear pUC18 PCR products that contained only cytosines, 5meC at CpG regions, only 5hmC, or the same three DNA substrates treated with 100 ng β-gt in the presence of 20 μM UDP-glucose. Substrates were hydrolyzed to nucleosides by incubation with nuclease P1, snake venom phosphodiesterase and alkaline phosphatase (Sigma-Aldrich, St Louis, MO, USA) as described (27). Three volumes of methanol were added to the reactions after digestion was completed and the reactions were centrifuged at 16 000*g* for 30 min. The supernatants were dried under vacuum and the resulting residues were dissolved in 50 μl 5% (v/v) methanol for analysis by LC/MS/MS. Chromatographic separation of nucleosides was performed using a Shimadzu Prominence HPLC system with a Zorbax SB-C18 2.1×150 mm i.d. (3.5 μm) reverse phase column equipped with a Eclipse XDB-C8 2.1×12.5 mm i.d. (5 μm) guard column (Agilent Technologies, USA), with a flow rate of 0.2 ml/min at ambient temperature. The mobile phase consisted of A (0.1% formic acid in water) and B (0.1% formic acid in methanol), starting with 95% A/5% B for 0.5 min, followed by a 6.5-min linear gradient of 5–50% B, 2 min with 50% B and 6 min re-equilibration with the initial mobile phase conditions. Online mass spectrometry detection was performed using an Applied Biosystems/MDS Sciex 5000 triple quadrupole mass spectrometer (Applied Biosystems Sciex, USA) with TurboIonSpray probe operating in positive electrospray ionization mode. The deoxyribonucleosides were monitored by multiple reaction monitoring using the mass transitions 228.2→112.1 (dC), 242.2→126.1 [5-me(dC)], 258.2→142.1 [5-hm(dC)] and 420.2→304.1 [5-Gly-hm(dC)].

JBP1 pull down reactions

Pull down reactions (200 μl) containing 132 ng JBP1 coated beads and 10 ng radiolabeled DNA when using PCR amplified substrates or 208 fmol 37mer, binding buffer [2 mM EDTA (pH 8.0), 10 mM Tris–HCl (pH 8.0) 150 mM NaCl, 0.02% (v/v) Tween-20, 1 mg/ml BSA] were incubated for 60 min at room temperature with gentle rotation. After incubation the supernatant was collected and counted by scintillation counting. The JBP1 coated beads were washed three times with 300 μl binding buffer without BSA. Radioactivity bound to the beads was counted using scintillation counting.

qPCR varying bead concentrations

Reactions (200 μl) contained 10 ng of each DNA substrate combined in one vial, with binding buffer and varying amounts of JBP1 modified magnetic beads were incubated for 60 min at room temperature with gentle rotation. After washing three times with 300 μl binding buffer without BSA and the beads were suspended in 90 μl binding buffer and 10 μl stop buffer was added [20 μg/ml Proteinase K, 50 mM Tris–HCl (pH 8.0), 100 mM

EDTA (pH 8.0), 2% (w/v) SDS] and allowed to incubate at 42°C for 60 min. After incubation the DNA was cleaned using a PCR clean kit (Qiagen) and the amount of each DNA pulled down relative to the input was measured by quantitative real-time PCR.

Complete glucosylation and pull down assay

Reactions (50 µl) containing 20 ng of each of three DNA substrates (unmodified cytosine, 5meC, 5hmC genomic regions combined in one vial), 20 µM UDP-Glucose and β-gt reaction buffer, and either 10 µg β-gt or β-gt storage buffer were incubated at 37°C for 30 min. Reactions were stopped by the addition of 10 µl stop buffer. DNA was purified using a PCR clean kit (Qiagen). The cleaned DNA was incubated with 20 µg JBP1 coated magnetic beads in 200 µl binding buffer for 60 min with gentle rotation. Beads were washed thre times with 300 µl binding buffer without BSA and the beads were suspended in 90 µl binding buffer and incubated at 42°C for 60 min with 10 µl stop buffer. DNA was then cleaned using a PCR clean kit. The amount of each DNA pulled down relative to the input was quantified using real-time PCR. The fold enrichment above cytosine was calculated by dividing each pulled down amount by the amount of cytosine containing DNA pulled down.

JBP1 pull down of 5hmC from human ES cell genomic DNA

Purified human ES cell DNA was sonicated on ice using a probe sonicator (Labsonic-M, 3-mm probe; cycle 0.5, 30% power; Sartorius AG, Göttingen, Germany) to produce an average fragment length of 300 bp. One microgram of this DNA was treated with β-gt and subjected to JBP1 pull down as described above except the reactions contained 264 ng JPB1 coated magnetic beads. Negative controls were performed without the addition of β-gt. JBP1-bound DNA was purified and used as template in qPCR. Quantitative PCR primers used in this study were as described (28,29) or are shown in Supplementary Table S1.

RESULTS

A brief overview of our scheme to specifically identify genomic regions containing 5hmC residues is shown (Figure 1). Briefly, we use the β-gt to specifically modify 5hmC residues creating β-glu-5hmC residues. Following the complete conversion of 5hmC to β-glu-5hmC we incubate the DNA with JBP1-coated magnetic beads. JBP1 specifically binds β-glu-5hmC containing DNA allowing for efficient and selective identification of DNA containing 5hmC. The resulting DNA is purified and ready for analysis by various methods.

The β-glucosyltransferase specifically glucosylates 5hmC

Recombinant β-gt from phage T4 was expressed and purified to >90% purity (Supplementary Figure S1; lane 1) by Cobalt affinity chromatography, followed by cation exchange chromatography. The β-gt has been shown by many groups to specifically glucosylate 5hmC

residues(16–18); yet, we tested the efficiency and specificity of our purified enzyme to ensure that our purification scheme resulted in active T4 β-gt. Three double-stranded 37-bp substrates were created: one contained unmodified cytosine at a *Taq*I restriction site, the second contained 5meC at the *Taq*I site and the third contained a 5hmC at the *Taq*I site. Each of the three substrates was treated with purified β-gt in the presence of UDP-glucose or incubated in the absence of β-gt. The resulting products were digested with *Taq*I, treated with alkaline phosphatase, 5′-end labeled using T4 polynucleotide kinase and digested to 5′-mononucleotides using DNase I and Snake Venom Phosphodiesterase. The resulting nucleotides were then resolved using two-dimensional thin-layer chromatography (TLC). From the TLC analysis we deduced that the β-gt has no effect on the substrates that lack 5hmC (Figure 2a; left and center); however, upon treatment with β-gt the 5hmC spot is absent from the TLC plate (Figure 2a; right); this result strongly suggests that the 5hmC has been specifically modified by the β-gt, demonstrating that β-gt is active. The cytosine and thymidine spots are good reference markers and are present because the substrates terminate with 5′-cytosine and 5′-thymidine residues. Additionally, we performed mass spectrometry on each of the substrates to reinforce that the β-gt can convert 5hmC residues to β-glu-5hmC at nearly 100% efficiency (Figure 2b and Supplementary Figure S2).

JPB1 from *C. fasciculata* can specifically pull down β-glucosyl-5hmC

Recombinant *C. fasciculata* JBP1 containing a his-tag was purified to >90% homogeneity by cobalt affinity chromatography followed by size exclusion chromatography. Fractions included in the final pool of purified JBP1 were determined to be free of detectable nucleases (Supplementary Figure S1; lane 2). Following purification, JBP1 was covalently linked to epoxy modified magnetic beads ('Materials and Methods' section).

JBP1 coated magnetic beads were incubated with four different radiolabeled substrates and precipitated using a magnet. Each substrate was a 2.7 kb linear PCR product created using pUC18 as a template (Supplementary Table S1). The substrates contained normal adenines (A), guanines (G) and thymidines (T) and either: only cytosine residues, 5meC residues in CpG sequences, only 5hmC residues or β-glu-5hmC residues produced by incubating the 5hmC substrate with the β-gt. JBP1 coated magnetic beads (132 ng) could effectively pull down up to 6% of the DNA that contained the β-glu-5hmC modification (Figure 3a). Additionally, JBP1 had little affinity for cytosine, 5meC or 5hmC containing substrates. Furthermore, BSA coated magnetic beads were unable to pull down any of the substrates, suggesting that the binding of β-glu-5hmC is due to JBP1 bound to the magnetic beads.

The efficiency of β-gly-5hmC pull down is dependent on the amount of JBP1 coated magnetic beads

Four different DNA substrates were created by PCR amplification of four different 2 kb mouse genomic
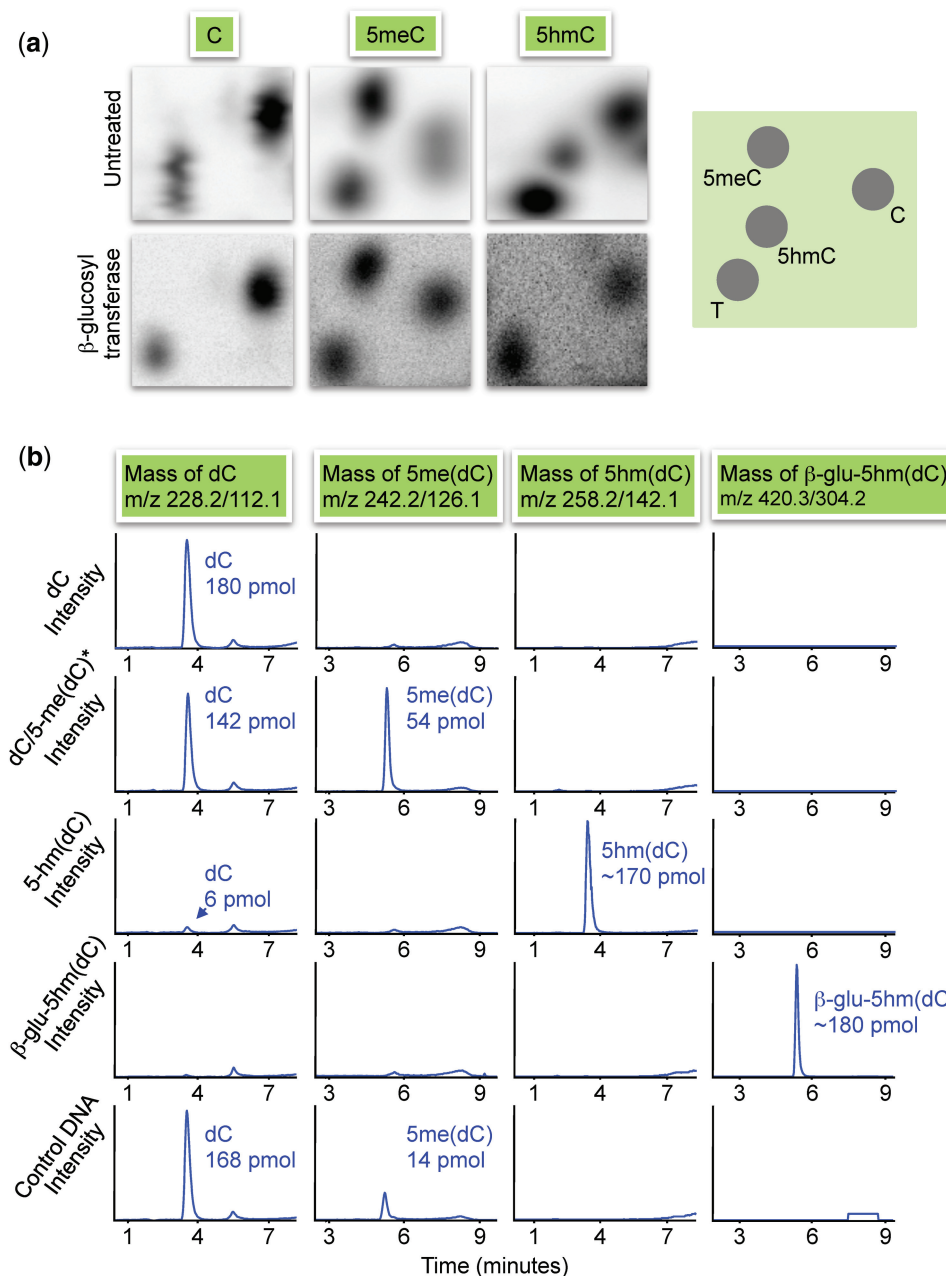
Figure 2. The β-gt can specifically modify 5hmC residues at a high efficiency. (a) Oligonucleotides that were either incubated in the presence or absence of the β-gt were digested with *Taq*I, treated with alkaline phosphatase, 5′-end labeled using T4 polynucleotide kinase and digested to 5′-mononucleotides using DNase I and Snake Venom Phosphodiesterase. Radiolabeled mononucleotides were analyzed by two-dimensional TLC. C, 3′-deoxyribocytosine-5′-monophosphate; T, 3′-deoxyribothymidine-5′-monophosphate; 5meC, 3′-deoxyribo-N5-methylcytosine-5′-monophosphate; 5hmC, 3′-deoxyribo-N5-hydroxymethylcytosine-5′-monophosphate. (b) HPLC coupled to tandem mass spectrometry was used to measure the efficiency of the β-gt reaction. Substrates analyzed were 2.7 kb linear PCR products of pUC18: the dC substrate contained only cytosine residues; the 5meC substrate was created by methylating the CpG dinucleotide of the cytosine substrate; the 5hmC substrate was created by using d5hmC in place of dCTP in the PCR reactions; the β-glu-5hmC substrate was created by incubating the 5hmC substrate with the β-gt in the presence of UDP-glucose. Control DNA was prepared from salmon sperm. LC/MS/MS chromatograms of the cytosine residues from each of the substrates are presented. Abbreviations: dC, 3′-deoxyribocytosine; 5me(dC), 3′-deoxyribo-N5-methylcytosine; 5hm(dC), 3′-deoxyribo-N5-hydroxymethylcytosine; 5-glu-hm(dC), 3′-deoxyribo-N5-(β-D-glucosyl(hydroxymethyl))cytosine. Asterisks indictes that cytosines are only 5meC modified at CpG sequences.

regions (Sub1: contained only cytosines, Sub2 5me: contained 5meC at CpG dinucleotides, Sub3 5hmC: contained 5hmC instead of cytosines, Sub4 β-glu-5hmC: contained β-glu-5hmC instead of cytosines—sequences are provided in Supplementary Table S2). These four substrates were incubated together with various concentrations of JBP1 coated magnetic beads. The beads were pulled down and the amount of each substrate captured was quantified using real-time quantitative PCR. There is a clear dependence on the amount of JBP1 coated magnetic beads and
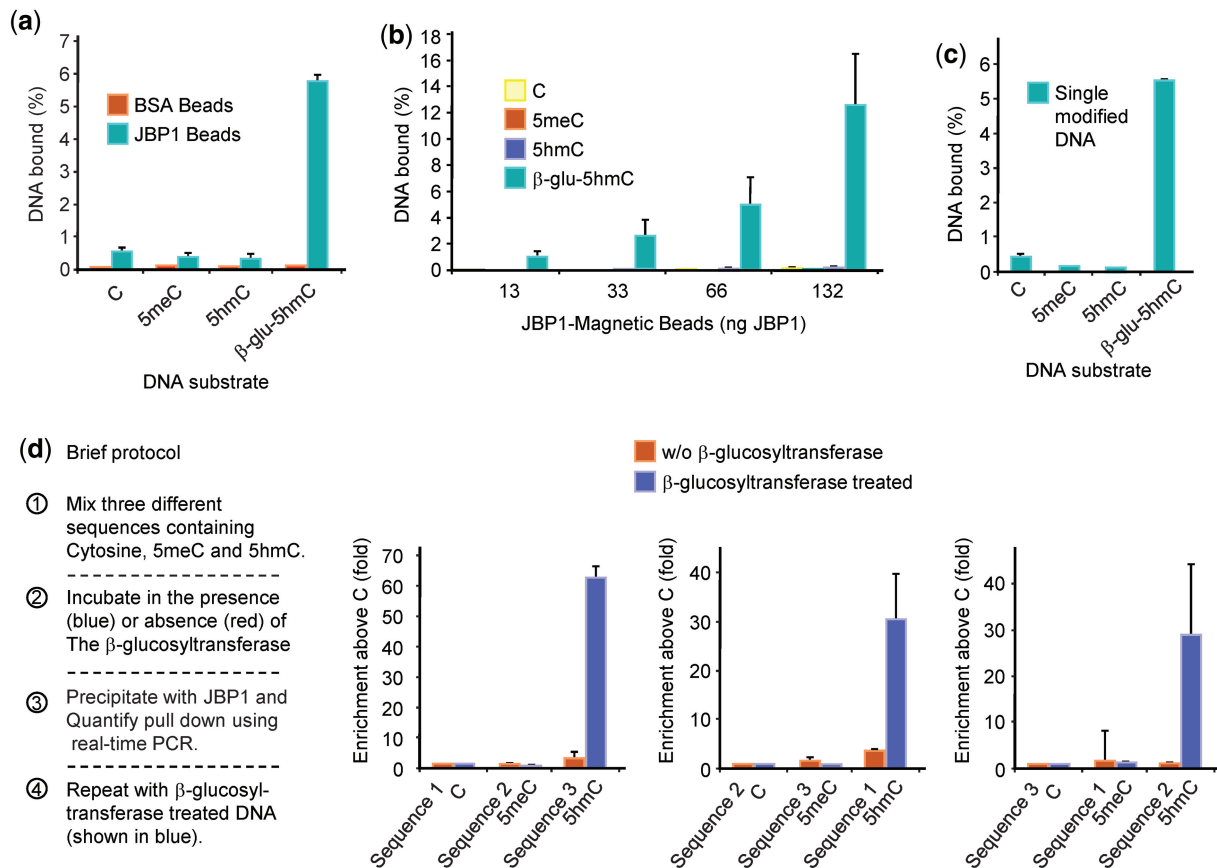
Figure 3. JBP1 can selectively pull down DNA that contains β-glu-5hmC resulting from β-gt treatment (**a**) Ten ng of each radiolabeled PCR product ('Materials and Methods' section) was incubated with magnetic beads coated with 132 ng JBP1 or 132 ng BSA. The beads were pulled down with a magnet and the percent bound to the beads was measured using scintillation counting. (**b**) Ten ng of each substrate were incubated in the same tube with various concentrations of JBP1 coated magnetic beads. The beads were pulled down using a magnet and the DNA that was bound to the magnetic beads was measured by quantitative real-time PCR. (**c**) 208 fmol of 37 bp double stranded oligonuleotides containing unmodified cytosines, a single 5meC, 5hmC or β-glu-5hmC on each strand were incubated with 132 ng JBP1 coated magnetic beads. Beads were pulled down as described and the percent DNA bound to the JBP1 coated magnetic beads were quantified using scintillation counting. (**d**) Three different combinations of three different PCR products composed of only cytosines, only 5meC, or 5hmC (20 ng of each) were combined and incubated with purified β-gt in the presence of UDP-Glucose. The DNA was purified and incubated with 132 ng JBP1 coated magnetic beads. The JBP1 magnetic beads were pulled down as above and the amount of DNA pulled down was quantified using quantitative real-time PCR. Data is presented as the fold enrichment over cytosine-containing DNA.

the pull down efficiency (Figure 3b). Furthermore we demonstrate that the highest concentration of JBP1 coated magnetic beads can pull down 14% of the β-glu-5hmC DNA and provide an 87-fold enrichment of β-glu-5hmC modified DNA over cytosine containing DNA and a 319-fold enrichment over 5meC containing DNA.

JBP1 can pull down DNA containing a symmetrical β-glu-5hmC modification

The previous experiments were conducted using substrates that contained many β-glu-5hmC modifications. We wanted to determine if a symmetrical β-glu-5hmC residue is sufficient for recognition by JBP1 coated magnetic beads. Therefore, we created four double stranded 37 bp substrates (Supplementary Table S1): the first contained unmodified cytosines, the second contained a symmetrical 5meC modification on each strand, the third contained a symmetrical 5hmC modification on each strand and the fourth contained a symmetrical

β-glu-5hmC on each strand. Each of these substrates was ^{32}P end labeled and incubated with 132 ng JBP1 coated beads. After the substrate was incubated with JBP1 the bound fraction was measured by scintillation counting. JBP1 coated magnetic beads pull down the substrate containing a symmetrically modified β-glu-5hmC efficiently while showing little affinity for the other substrates composed of an identical sequence (Figure 3c).

JBP1 preferentially pulls down β-glu-5hmC containing DNA from a mixture of competing DNA substrates

We performed an assay that mimicked the JBP1 pull down that will be used on mammalian DNA. We combined several different mixtures of three DNA substrates (Supplementary Table S2) with differing sequences, with each substrate mixture containing unmodified cytosine substrate, 5meC modified substrate and 5hmC modified substrate. We then treated the substrate mixes with purified β-gt in the presence of UDP-glucose. The
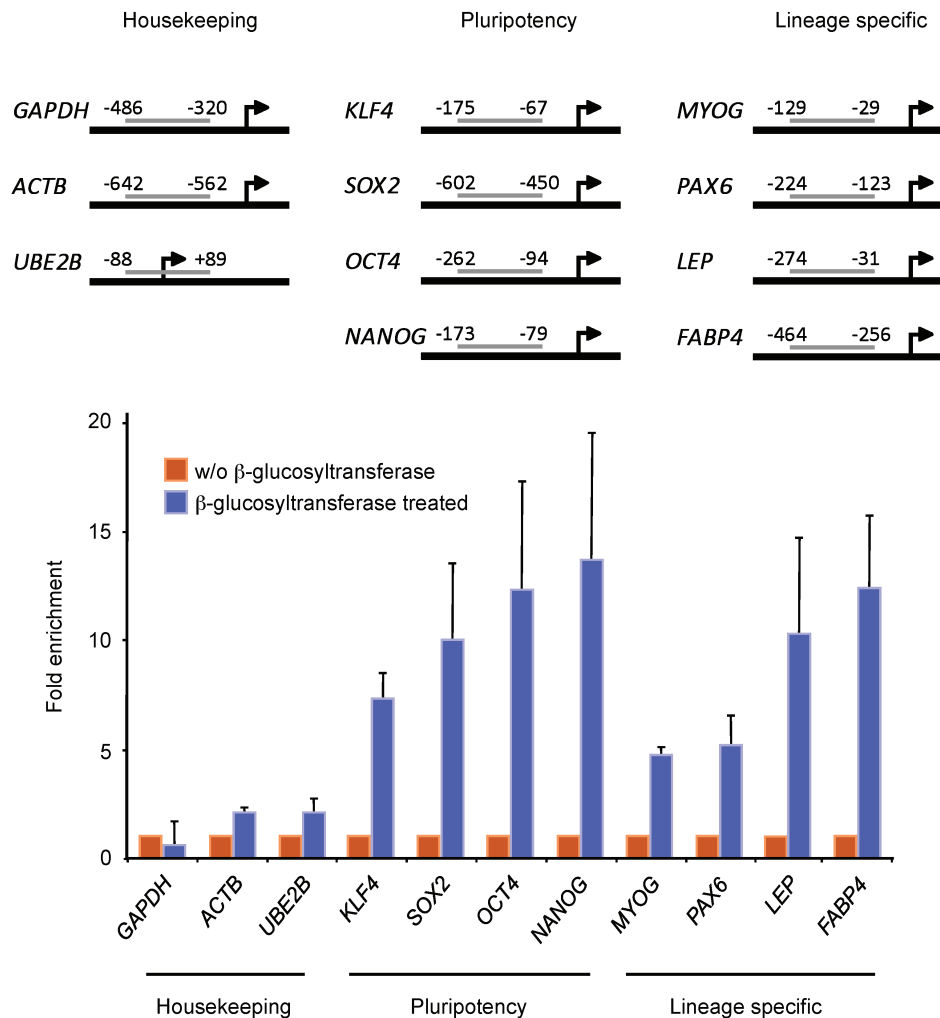
Figure 4. JPB1 can specifically pull down genomic DNA from human embryonic stem cells that has been incubated with the β-glucosyltransferase. One microgram of genomic DNA from hES cells was incubated in the presence (blue bars) or in the absence (red bars) of the β-glucosyltransferase. After incubation DNA was incubated and pulled down with JBP1 coated magnetic beads. The amount of DNA pulled down was quantified by real-time PCR. The genomic location of the PCR amplified region relative to the transcription start site is shown for all loci in the top panel. Data is represented as fold enrichment above background pull down. Error bars represent the standard deviation from the mean of at least three experiments.

resulting DNA products were purified and incubated with JBP1 coated magnetic beads. In all three mixtures tested, the JBP1 coated magnetic beads were able to significantly enrich for the DNA substrate that contained the 5hmC modification after treatment with the β-gt (Figure 3d). The data presented does not suggest an absolute sequence independence for β-glu-5hmC modified DNA; however, JBP1 has been previously shown to bind DNA containing the J-base in a sequence independent manner (22,23). Taken together, these results suggest that JBP1 binding to DNA is not dependent on sequence. Further study on the binding properties of JBP1 to β-glu-5hmC will be necessary to confirm JBP1 binding in a sequence independent manner.

The JBP1 pull down assay enables identification of 5hmC on the promoters of developmentally regulated genes in human embryonic stem cells

The *in vitro* validated JBP1 pull down assay was applied to assess the location of 5hmC in genomic DNA from hES cells. One microgram of purified DNA sonicated to an average fragment length of 300 bp was treated with β-gt and subjected to JBP1 pull down as described above. Negative controls were performed without the addition of the β-gt. This unglucosylated DNA will not be specifically pulled down by JBP1, providing a background pull down level. JBP1 bound DNA was purified and used as template for qPCR. Quantitative PCR analysis (Figure 4) revealed that the 5hmC levels at the promoters of the housekeeping genes *GAPDH*, *ACTB* and *UBE2B* was at or slightly above the background pull down seen in the unglucosylated samples, suggesting that these promoters do not contain 5hmC. In contrast, promoters of the pluripotency marker genes *KLF4*, *SOX2*, *OCT4* (*POU5F1*) and *NANOG* were enriched for 5hmC 7- to 14-fold above the unglucosylated pull down level, showing that 5hmC is present at the promoters of these genes. Moreover, genes which are expressed upon lineage specific differentiation: *MYOG*, myogenic lineage; *PAX6*, neurogenic lineage; *LEP* and *FABP4*, adipogenic lineage,

showed relatively high levels of 5hmC, ∼5- to 12-fold enrichment over background, at their promoters. In summary, all developmentally regulated loci assessed harbored 5hmC in human ES cells whereas the promoters of cellular housekeeping genes lacked the 5hmC modification.

DISCUSSION

In this study we have developed a novel method for efficient and selective identification of 5hmC modified nucleotides in genomic DNA. At the present time the only other method available to recognize the location of 5hmC modifications—using polymerase kinetics—is prohibitively expensive and has yet to be demonstrated in *in vivo* studies(11). Currently, there is no antibody that has been shown to specifically react with 5hmC; rather the current literature shows that the commercially available antisera raised against 5hmC cross-reacts significantly with 5meC (8). Furthermore, a 5hmC residue produces similar results as a 5meC residue when bisulfite sequencing is employed to identify these residues (12), rendering these approaches to identifying 5hmC unusable without significant improvements. However, an antiserum has been developed that recognizes 5-methylenesulfonate, the product of bisulfite conversion of 5hmC, this antiserum may prove useful for identifying the location of 5hmC (14).

The method presented in this article ensures that only 5hmC-containing genomic regions will be precipitated and identified in subsequent assays. Given that the β-gt from T4 specifically catalyzes the glucosylation of 5hmC [Figure 2 (16)] and JBP1 specifically recognizes the resulting β-glu-5hmC base (Figure 3a–d), the DNA pulled down by JBP1 is highly enriched in the 5hmC modification. The pulled down DNA is then ready for analysis by real-time quantitative PCR, microarray analysis, or sequencing by any method, including high-throughput sequencing.

This report is, to our knowledge, the first assessment of the location of 5hmC in any genome. We demonstrate that the JBP1 pull down assay can identify 5hmC at the promoters (Figure 4) of developmentally regulated genes in human embryonic stem cells, whereas the promoters of housekeeping genes assayed do not appear to contain the 5hmC modification. Interestingly, the absence of 5hmC at promoters of the studied housekeeping genes suggests that 5hmC does not directly correlate with gene expression. Moreover, 5hmC is observed at the promoters of genes that are not expressed in hES cells but that are expressed during lineage specific differentiation. However, pluripotency related genes are both expressed and show promoter enrichment for 5hmC.

More simplistically, one might have envisaged a generic role for 5hmC in transcriptional activation. Given that promoter methylation generally has a transcriptionally repressive role facilitated by the recruitment of methyl-CpG-binding domain proteins and given that 5hmC inhibits the binding of these proteins it was speculated that 5hmC may enhance transcription (1). Some of the genes analyzed for 5hmC content are transcriptionally silent in hES cells yet primed for transcriptional activation; therefore, it is tempting to speculate that the 5hmC modification may prime these genes for activation during differentiation. Indeed, our data does not exclude the possibility that 5hmC may have a more direct role in the activation of some genes or particular groups of genes. An alternative scenario posits that 5hmC, rather than being an activating mark by itself, sets up a transcriptional permissive environment that allows for further transcriptional directions to be provided by other cellular factors or modifications. More study into the relationship between 5hmC at promoters and gene expression is necessary to validate either of these hypotheses.

The simple and cost-effective method for identifying genomic regions containing 5hmC presented in this report will allow further investigation of the scenarios outlined above and will significantly increase the ability of researchers to study the temporal and spatial patterns of 5hmC residues in any genomic DNA.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Illingworth,R.S. and Bird,A.P. (2009) CpG islands–'a rough guide'. *FEBS Lett.*, **583**, 1713–1720.
2. 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.
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. Munzel,M., Globisch,D., Bruckl,T., Wagner,M., Welzmiller,V., Michalakis,S., Muller,M., Biel,M. and Carell,T. (2010) Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.*, **49**, 5375–5377.
5. Szwagierczak,A., Bultmann,S., Schmidt,C.S., Spada,F. and Leonhardt,H. (2010) Sensitive enzymatic quantification of

5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.*, **38**, e181.

6. Jin,S.G., Kadam,S. and Pfeifer,G.P. (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **38**, e125.

7. Hendrich,B. and Bird,A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.*, **18**, 6538–6547.

8. 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**, 1129–1133.

9. Szwagierczak,A., Bultmann,S., Schmidt,C.S., Spada,F. and Leonhardt,H. (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.*, **38**, e181.

10. Munzel,M., Globisch,D., Bruckl,T., Wagner,M., Welzmiller,V., Michalakis,S., Muller,M., Biel,M. and Carell,T. (2010) Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed.*, **49**, 5375–5377.

11. Flusberg,B.A., Webster,D.R., Lee,J.H., Travers,K.J., Olivares,E.C., Clark,T.A., Korlach,J. and Turner,S.W. (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods*, **7**, 461–465.

12. Nestor,C., Ruzov,A., Meehan,R. and Dunican,D. (2010) Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *BioTechniques*, **48**, 317–319.

13. Jin,S.G., Kadam,S. and Pfeifer,G.P. (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **38**, e125.

14. Ko,M., Huang,Y., Jankowska,A.M., Pape,U.J., Tahiliani,M., Bandukwala,H.S., An,J., Lamperti,E.D., Koh,K.P., Ganetzky,R. *et al.* (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, **468**, 839–843.

15. Flusberg,B.A., Webster,D.R., Lee,J.H., Travers,K.J., Olivares,E.C., Clark,T.A., Korlach,J. and Turner,S.W. (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods*, **7**, 461–465.

16. Tomaschewski,J., Gram,H., Crabb,J.W. and Ruger,W. (1985) T4-induced alpha- and beta-glucosyltransferase: cloning of the genes and a comparison of their products based on sequencing data. *Nucleic Acids Res.*, **13**, 7551–7568.

17. Georgopoulos,C.P. and Revel,H.R. (1971) Studies with glucosyl transferase mutants of the T-even bacteriophages. *Virology*, **44**, 271–285.

18. Kornberg,S.R., Zimmerman,S.B. and Kornberg,A. (1961) Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected Escherichia coli. *J. Biol. Chem.*, **236**, 1487–1493.

19. Gommers-Ampt,J.H., Van Leeuwen,F., de Beer,A.L., Vliegenthart,J.F., Dizdaroglu,M., Kowalak,J.A., Crain,P.F. and Borst,P. (1993) beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan T. brucei. *Cell*, **75**, 1129–1136.

20. Borst,P. and Sabatini,R. (2008) Base J: discovery, biosynthesis, and possible functions. *Ann. Rev. Microbiol.*, **62**, 235–251.

21. van Leeuwen,F., Taylor,M.C., Mondragon,A., Moreau,H., Gibson,W., Kieft,R. and Borst,P. (1998) beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Procd. Natl Acad. Sci. USA*, **95**, 2366–2371.

22. Sabatini,R., Meeuwenoord,N., van Boom,J.H. and Borst,P. (2002) Recognition of base J in duplex DNA by J-binding protein. *J. Biol. Chem.*, **277**, 958–966.

23. Cross,M., Kieft,R., Sabatini,R., Wilm,M., de Kort,M., van der Marel,G.A., van Boom,J.H., van Leeuwen,F. and Borst,P. (1999) The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO J.*, **18**, 6573–6581.

24. Grover,R.K., Pond,S.J., Cui,Q., Subramaniam,P., Case,D.A., Millar,D.P. and Wentworth,P. Jr (2007) O-glycoside orientation is an essential aspect of base J recognition by the kinetoplastid DNA-binding protein JBP1. *Angew. Chem. Int. Ed.*, **46**, 2839–2843.

25. Studier,F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Express. Purif.*, **41**, 207–234.

26. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning, A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press.

27. Crain,P.F. (1990) Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. *Methods Enzymol.*, **193**, 782–790.

28. Noer,A., Lindeman,L.C. and Collas,P. (2009) Histone H3 modifications associated with differentiation and long-term culture of mesenchymal adipose stem cells. *Stem Cells Dev.*, **18**, 725–736.

29. Dahl,J.A., Reiner,A.H. and Collas,P. (2009) Fast genomic muChIP-chip from 1,000 cells. *Genome Biol.*, **10**, R13.