

# Tissue-specific Distribution and Dynamic Changes of 5-Hydroxymethylcytosine in Mammalian Genomes<sup>\*[5]</sup>

Received for publication, December 29, 2010, and in revised form, May 23, 2011. Published, JBC Papers in Press, May 24, 2011, DOI 10.1074/jbc.M110.217083

Shannon Morey Kinney<sup>†1</sup>, Hang Gyeong Chin<sup>†1</sup>, Romualdas Vaisvila<sup>‡</sup>, Jurate Bitinaite<sup>‡</sup>, Yu Zheng<sup>‡</sup>, Pierre-Olivier Estève<sup>‡</sup>, Suhua Feng<sup>§</sup>, Hume Stroud<sup>§</sup>, Steven E. Jacobsen<sup>§2</sup>, and Sriharsa Pradhan<sup>‡3</sup>

From <sup>†</sup>New England Biolabs, Inc., Ipswich, Massachusetts 01938 and the <sup>§</sup>Howard Hughes Medical Institute, Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, California 90095-1606

Cytosine residues in the vertebrate genome are enzymatically modified to 5-methylcytosine, which participates in transcriptional repression of genes during development and disease progression. 5-Methylcytosine can be further enzymatically modified to 5-hydroxymethylcytosine by the TET family of methylcytosine dioxygenases. Analysis of 5-methylcytosine and 5-hydroxymethylcytosine is confounded, as these modifications are indistinguishable by traditional sequencing methods even when supplemented by bisulfite conversion. Here we demonstrate a simple enzymatic approach that involves cloning, identification, and quantification of 5-hydroxymethylcytosine in various CCGG loci within murine and human genomes. 5-Hydroxymethylcytosine was prevalent in human and murine brain and heart genomic DNAs at several regions. The cultured cell lines NIH3T3 and HeLa both displayed very low or undetectable amounts of 5-hydroxymethylcytosine at the examined loci. Interestingly, 5-hydroxymethylcytosine levels in mouse embryonic stem cell DNA first increased then slowly decreased upon differentiation to embryoid bodies, whereas 5-methylcytosine levels increased gradually over time. Finally, using a quantitative PCR approach, we established that a portion of *VANGL1* and *EGFR* gene body methylation in human tissue DNA samples is indeed hydroxymethylation.

In the vertebrate genome, DNA methylation is the predominant epigenetic modification. Cytosine residues undergo enzymatic modification at the carbon-5 position (5-mC)<sup>4</sup> by DNA (cytosine-5) methyltransferases (1). In mammals, methylation is primarily found in CpG dinucleotides (2). In certain mammalian cell types, including embryonic stem cells, and in plants,

non-CpG and asymmetric cytosine methylation has been observed (3–5). CpG methylation may directly disrupt interaction between certain transcription factors and their corresponding DNA binding sites (6, 7) or may recruit methyl DNA-binding proteins, such as MeCP2 or MBDs to create a repressive chromatin state (8, 9). Changes in DNA methylation patterns in the genome are correlated with altered gene expression, including selective inactivation of one X chromosome in female mammals (10). DNA methylation is dynamic, heritable, and reversible, providing important determinants for an array of epigenetic states regulating phenotype and gene expression.

In addition to the presence of 5mC, there are other DNA modifications, including DNA damage, that arise via exposure of cells to physical or chemical agents (11). Such damage is generally repaired before the next cell cycle starts *in vivo*, thereby maintaining genome integrity (12). Recently, two independent reports demonstrated the presence of another modification, 5-hydroxymethylcytosine (5-hmC), in murine neuronal and embryonic stem cell genomes (13, 14). In one report, Tahiliani *et al.* (14) performed a homology search using the sequences of trypanosoma enzymes JBP1 and JBP2 that sequentially modify the methyl group of thymine by two-step hydroxylation and glucosylation to create  $\beta$ -D-glucosylhydroxymethyluracil (base J) and found mammalian homologues Tet1, Tet2, and Tet3. Using biochemical and cell biology techniques, they established that Tet1 catalyzes the conversion of 5-mC to 5-hmC and that the ratio of 5-mC to 5-hmC in murine embryonic stem cells is ~10:1. Coinciding with this observation, Kriacucionis and Heintz (13) also reported the presence of 5-hmC in murine Purkinje and granule neurons. Analysis of the Purkinje and granule cell genomes displayed a significant portion, 0.6 and 0.2%, respectively, of the total nucleotide pool as 5-hmC. As a potentially stable base, 5-hmC may influence chromatin structure and local transcriptional activity by repelling 5-mC-binding proteins or recruiting 5-hmC-specific proteins. Indeed, in a previous study it was demonstrated that methyl-binding protein MeCP2 does not recognize or bind to 5-hmC (15). More recent reports using several other methyl-binding proteins, including MBD1, MBD2, and MBD4, support this hypothesis (16). Because 5-hmC is present in the mammalian genome and may mediate biological functions differently from 5-mC, there is a need for distinguishing between the various forms of modified cytosine residues dispersed throughout the genome. Here we report a simple enzymatic method of determining 5-hmC in CpG context, embedded in CCGG sites in the mammalian genome. Coupling the enzymatic method

\* This work was supported in part by an innovation award from the Eli & Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA (to S. E. J.).

[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Tables 1–3 and Figs. 1–5.

⌘ Author's Choice—Final version full access.

<sup>1</sup> Both authors contributed equally to this work.

<sup>2</sup> Investigator of the Howard Hughes Medical Institute.

<sup>3</sup> To whom correspondence should be addressed: New England Biolabs, Inc; 240 County Rd., Ipswich, Massachusetts 01938. Tel.: 978-380-7227; Fax: 978-921-1350; E-mail: pradhan@neb.com.

<sup>4</sup> The abbreviations used are: 5-mC, 5-methylcytosine; 5-hmC, 5-hydroxymethylcytosine; NEB, New England Biolabs;  $\beta$ -GT,  $\beta$ -glucosyltransferase; ES, embryonic stem; qPCR, quantitative-PCR; 5-ghmC, glucosylated 5-hmC; MBD, methyl binding domain; nt, nucleotide(s); FAM, 6-carboxyfluorescein; Tricine, N-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine; -hm-, hydroxymethyl; LIF, leukemia inhibitory factor.

## Tissue-specific Distribution of 5-Hydroxymethylcytosine

with quantitative PCR, we are able to determine the percentage of unmodified cytosine (C) and its modified form (5-mC and 5-hmC) at the internal cytosine residue of CCGG sites. Furthermore, using this method, we demonstrate gene- and tissue-specific distribution of 5-hmC as well as dynamic changes in 5-hmC distribution during embryoid body formation.

### EXPERIMENTAL PROCEDURES

*In Vitro Analysis of MspI and HpaII Sensitivity to Glucosylation of 5-Hydroxymethylcytosine*—Fluorescein-labeled double-stranded oligonucleotides containing a single 5-hmC residue either on one or both strands (within the MspI recognition site, *i.e.* ChmCGG) were synthesized as follows. 5  $\mu$ mol each of two fluorescein-labeled 45-nt-long oligonucleotides, 5'-FAM-CCAACTCTACATTCAACTCTTATCCGGTGTAATGTGATGGGTGT-3', and a 19-nt primer oligonucleotide, 5'-FAM-ACCCATCACATTTACAC-3', were combined in 25  $\mu$ l of NEBuffer 4 (New England Biolabs, NEB) and annealed by incubating at 95 °C for 5 min followed by slowly cooling to room temperature. The annealing reaction was then supplemented with 5  $\mu$ l (1 mM) each of hydroxymethyl-dCTP (hmdCTP) (Bio-line), dATP, dTTP, and dGTP (NEB) and 1  $\mu$ l (5 units) of Klenow (NEB), and the reaction volume was adjusted to 50  $\mu$ l with Milli-Q water. The annealed oligonucleotide was made fully double-stranded by incubating the reaction at room temperature for 30 min, resulting in hemi-5-hmC substrate (100 pmol/ $\mu$ l): top strand (5'-FAM-CCAACTCTACATTCAACTCTTATCCGGTGTAATGTGATGGGTGT-3') and bottom strand (3'-GGTTGAGATGTAAGTTGAGAATAGGhmCCACATTACTACTACCCACA-FAM-5').

To make a double-stranded oligonucleotide containing hmC residue on both strands, the 45-nt template oligonucleotide was synthesized with 8 uracil residues distributed uniformly through the sequence: 5'-FAM-CCAACTCTACAUTCAACUCTTAUCCGGUGTAAAUGTGAUGGGUGT-3'. This oligonucleotide was annealed with a 19-nt primer oligonucleotide and made fully double-stranded as was described earlier: top strand (5'-FAM-CCAACTCTACAUTCAACUCTTAUCCGGUGTAAAUGTGAUGGGUGT-3') and bottom strand (3'-GGTTGAGATGTAAGTTGAGAATAGGhmCCACATTACTACTACCCACA-FAM-5'). In the next step a single-stranded template with a 5-hmC residue within the MspI site was created by destroying the uracil-containing oligonucleotide strand. For this purpose, 5  $\mu$ l (500 pmol) of the above double-stranded oligonucleotide were combined with 5  $\mu$ l of 10 $\times$  T4 DNA ligase buffer and 5  $\mu$ l (5 units) of USER Enzyme, NEB, and the reaction volume was adjusted to 50  $\mu$ l with Milli-Q water. The reaction was incubated at 37 °C for 60 min to excise uracil residues and additionally incubated for 20 min at 65 °C to fully dissociate the leftover double-stranded regions. An equimolar amount (500 pmol) of complementary 24-nt primer oligonucleotide, 5'-FAM-CCAACTCTACATTCAACTCTTATC, was annealed to the newly obtained single-stranded template by incubating for 5 min at 95 °C and slowly cooling down to room temperature. The reaction was then supplemented with 5  $\mu$ l of NEBuffer 4, 5  $\mu$ l (1 mM) each of hmdCTP, dATP, dTTP, and dGTP, and 1  $\mu$ l (5 units) of Klenow fragment, and the reaction volume was adjusted to 100  $\mu$ l with Milli-Q water.

The reaction was incubated at room temperature for 30 min to yield fully double-stranded substrate containing a single 5-hmC residue on either strand: top strand (5'-FAM-CCAACTCTACATTCAACTCTTATChmC GGTGTAATGTGATGGGTGT-3') and bottom strand (3'-GGTTGAGATGTAAGTTGAGAATAG GhmCCACATTTACTACTACCCACA-FAM-5'). The control duplex DNA with C, 5-mC and 5-hmC were obtained from NEB.

*Glucosylation of 5-hmC-containing Oligonucleotide Substrates with T4  $\beta$ -Glucosyltransferase*—The hmC residues within the MspI site were glucosylated by incubating 200 pmol of DNA substrates with 1  $\mu$ l (10 units) of T4  $\beta$ -glucosyltransferase ( $\beta$ -GT) (NEB) for 1 h at 37 °C in a total 50- $\mu$ l reaction containing 1 $\times$  NEBuffer 4 supplemented with 0.1 mM UDP-Glc. After glucosylation, the  $\beta$ -GT enzyme was heat-inactivated by incubating for 20 min at 75 °C.

*Cleavage with MspI and HpaII Restriction Endonucleases*—Double-stranded oligonucleotide substrates (20 pmol) were cleaved with 20 units of either MspI or HpaII. The reactions were carried out at 37 °C for 4 h in 20  $\mu$ l of either NEBuffer 4 (MspI) or NEBuffer 1 (HpaII). Reactions containing hemi-hydroxymethylated (hemi-glucosylated) substrates were terminated by the addition of 10  $\mu$ l formamide gel buffer (50% formamide, 7 M urea, 12% Ficoll 400, 0.01% bromphenol blue, and 0.02% Xylene) and heated at 95 °C for 5 min. Reaction products were separated by electrophoresis under denaturing conditions on 20% acrylamide, 7 M urea gel. Reactions containing fully hydroxymethylated (fully glucosylated) substrates were terminated by the addition of 5  $\mu$ l of Gel Loading Dye (2.5% Ficoll 400, 11 mM EDTA, 3.3 mM Tris-HCl (pH 8.0), 0.017% SDS, and 0.015% bromphenol blue), and reaction products were separated by electrophoresis on the native 10% Tris borate-EDTA acrylamide gel and visualized under UV light.

*Cell Culture and E14 Embryonic Stem (ES) Cell Differentiation to Embryoid Bodies*—ES cells were cultured as described with GMEM (Invitrogen) media containing 10% FBS (Gemcell), 1% non-essential amino acids (NEAA) (Hyclone), 1% sodium pyruvate (Invitrogen), 50  $\mu$ M  $\beta$ -mercaptoethanol (Sigma), and 1 $\times$  LIF (Millipore). To maintain the undifferentiated ES cells, they were grown on 0.1% gelatin (Stem Cell Technologies)-coated culture dishes. For differentiation of ES cells to embryoid bodies, LIF was removed, and cells were seeded on low adherence plates (Corning) with no gelatin for 1–10 days.

*Screen for Genomic DNA for Hydroxymethylated Loci*—Genomic DNA was extracted from E14 ES cells and embryoid bodies using the Qiagen DNeasy Blood and Tissue kit. NIH 3T3 and HeLa cells were obtained from ATCC. DNA from human tissues was purchased from Biochain. Five  $\mu$ g of genomic DNA (E14 and human normal brain) was digested with MspI (NEB). Digested DNA was purified with phenol chloroform and then ligated with T4 DNA ligase (NEB) into pCpG-MspI plasmid with zeocin selection marker containing one MspI site. Plasmid DNA was glucosylated with  $\beta$ -GT (NEB) and cofactor UDP-Glc (NEB) and digested with MspI. The remaining circular DNA was used to transform GT115 or 2924E competent *Escherichia coli* cells. All colonies were picked, grown to 5-ml cultures, purified with Qiagen miniprep kit, and sequenced at the NEB

sequencing facility. Sequences were then aligned to the appropriate genome with the NCBI Blast software.

**Glucosylation and Digestion of Genomic DNA with *MspI* or *HpaII***—2–5- $\mu\text{g}$  aliquots of genomic DNA were either glucosylated with  $\beta\text{-GT}$  and UDP-Glc or mock-treated with  $\beta\text{-GT}$  and no UDP-Glc for at least 3 h. These reactions were then split in half and digested separately with *MspI* (100 units/1.5  $\mu\text{g}$ ) and *HpaII* (50 units/1.5  $\mu\text{g}$ ). Both digested and undigested DNAs were diluted to a final concentration of 16 ng/ $\mu\text{l}$  to be used for PCR analysis.

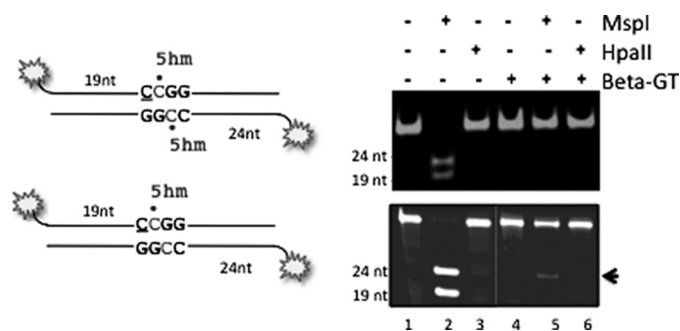
**End Point and Quantitative-PCR (qPCR) Analysis of  $\beta\text{-GT}$ -treated *MspI*- or *HpaII*-digested DNA**—End point PCR was completed with Phusion-GC (NEB) polymerase master mix. Two  $\mu\text{l}$  of diluted DNAs described above were used for each 50- $\mu\text{l}$  reaction. Half of each PCR reaction was separated on a 1.2% agarose gel (VWR) and stained with ethidium bromide (Sigma) to visualize. Typically, end point PCR amplifications were carried out for 30 cycles. qPCR was completed with Dynamo HS SYBR Green qPCR kit (NEB) using a Bio-Rad CFX384 Real-Time PCR Detection System. Background signal was subtracted from the copy number of each sample and then normalized to undigested control. Primers were designed with NCBI primer-blast software and purchased from IDT.

**Western Blot Analysis of E14 ES Cells**—Proteins were extracted from E14 cells with radioimmune precipitation assay buffer and quantified with the Coomassie Plus (Bradford) Protein Assay (Thermo Scientific). Equal amounts of protein were separated on 10% Tricine gels (Invitrogen) using SDS loading buffer (NEB). Proteins were then transferred to Whatman Protran nitrocellulose membrane, blocked with either 5% milk or BSA in PBST, probed with the appropriate antibodies, treated with Lumiglo chemiluminescent reagent (Cell Signaling Technology, CST), and exposed to Kodak Biomax MS film. Membranes were stripped with Restore Western blot stripping buffer (Thermo Scientific) and probed with *Gapdh* as a loading control. The following antibodies were used for Western blot analysis: Oct4 (1:1000; Stem Cell Technologies), Nanog (1:500; Abcam), *Gapdh* (CST), *Dnmt1* (1:1000; Abcam), *Dnmt3a* (1:5000; Abcam), *Dnmt3b* (1:2000; Novus), Tet1 (1:2500, rabbit bleed, CST).

**Liquid Chromatography Mass Spectrometry Analysis of E14 DNA for Genomic Deoxycytosine, 5-Methyldeoxycytosine, and 5-Hydroxymethyldeoxycytosine**—Two  $\mu\text{g}$  of genomic DNA was treated with 10 units of Antarctic phosphatase (NEB) and 2 milliunits of snake venom phosphodiesterase (Sigma) at 37 °C overnight. Eight microliters of hydrolyzed DNA was injected for analysis. Peak area was used to calculate percent of 5-mC (of total C) and percent 5-hmC (of total 5-mC).

## RESULTS

**Restriction Enzyme *MspI* Is Blocked by Glucosylated 5-Hydroxymethylcytosine**—The restriction enzyme *MspI* recognizes the sequence CCGG and is able to cleave when the internal cytosine is unmethylated or methylated at the C5 position (17). To determine the effect of internal cytosine hydroxymethylation on *MspI* cleavage, we made a synthetic 5'-FAM-labeled duplex DNA containing one *MspI* site with either the internal cytosine being hemi-hydroxymethylated or

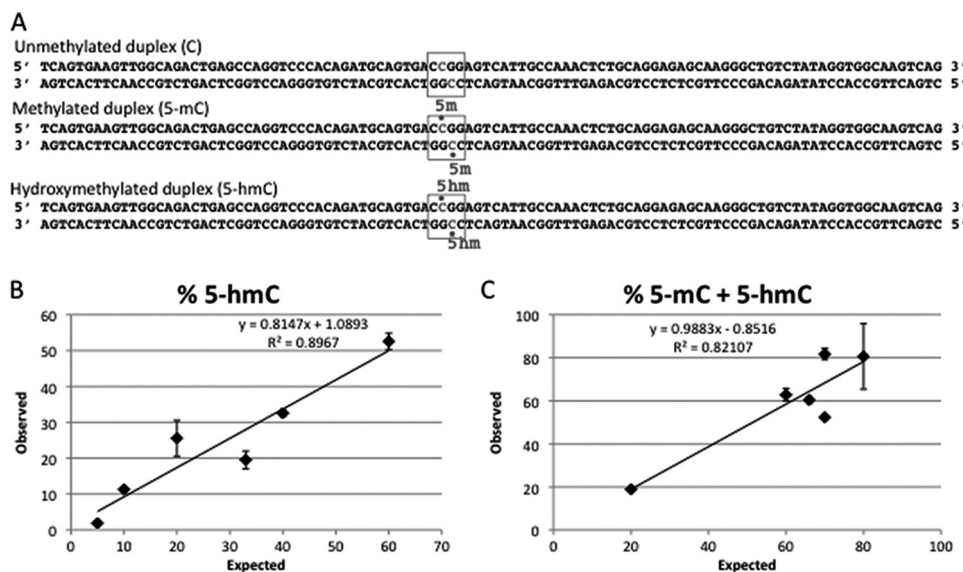


**FIGURE 1. *MspI* and *HpaII* isoschizomer can distinguish between 5-hmC and 5-ghmC at the internal cytosine residue.** Cleavage specificity is shown for *MspI* and *HpaII* on FAM-end-labeled oligonucleotide duplex with internal CG being symmetrically hydroxymethylated or hemihydroxymethylated in the presence or absence of  $\beta\text{-GT}$ . Complete cleavage products are labeled as 24 and 19 nt, respectively, on a non-denaturing acrylamide gel for double-hydroxymethylated (upper panel) and denaturing acrylamide gel for hemi-hydroxymethylated DNA (lower panel). The arrow at the right indicates small amounts of 24-nt-long product.

fully hydroxymethylated (Fig. 1, left panel). These oligonucleotide duplexes are either mock-glucosylated or glucosylated using  $\beta\text{-GT}$ .  $\beta\text{-GT}$  is a DNA-modifying enzyme encoded by bacteriophage T4 that catalyzes the transfer of glucose (Glc) from uridine diphosphoglucose (UDP-Glc) to 5-hmC in double-stranded DNA using a base flipping mechanism (18). After the reaction, duplex DNAs were subjected to *MspI* digestion, and products were separated on a polyacrylamide gel. *MspI* cleaved the internal 5-hmC containing DNA (Fig. 1, right upper panel, lanes 1 versus 2) but failed to digest the glucosylated 5-hmC (5-ghmC) containing DNA (Fig. 1, right upper panel, lanes 2 versus 5). *HpaII*, an isoschizomer of *MspI*, did not cleave either 5-hmC or 5-ghmC containing DNA (Fig. 1, upper panel, lanes 3 versus 6). We also performed the same experiment to determine the specificity of *MspI* on a hemi-5-hmC oligonucleotide duplex. In this experiment, the digested DNA products were separated on a denaturing polyacrylamide gel. As expected, *HpaII* digestion was blocked by hemi-5-hmC and hemi-5-ghmC (Fig. 1, lower panel, lanes 3 and 6). *MspI* fully cleaved the hemi-hmC DNA (Fig. 1, lower panel, lane 2). Once the 5-hmC was glucosylated, the strand containing 5-ghmC was protected from *MspI*, and the unmodified strand was poorly cleaved (Fig. 1, lower panel, lane 5). Overnight incubation yielded small amounts of 24-nt-long product, suggesting hemi-hydroxymethylated sequences on *MspI* sites cannot be reliably distinguished from fully hydroxymethylated DNA using this assay alone. However, coupling this technique with strand-specific bisulfite sequencing could result in the identification of hemi-hydroxymethylated sites. From these results we concluded that glucosylation along with *MspI* and *HpaII* might be used for distinguishing between C, 5-mC, and 5-hmC.

**Differential  $\beta$ -Glucosylation and Quantification of Methylated and Hydroxymethylated Cytosine**—Because *MspI* cleaves internal 5-hydroxymethylcytosine and is selectively blocked by glucosylation of this site, we used this property of the enzyme to establish a robust protocol for identification and quantification of C, 5-mC, and 5-hmC on a defined sequence of synthetic DNA. However, restriction enzymes display reduced or impaired cleavage by target site methylation. Therefore, we first determined the amount of *MspI* enzyme required for complete

## Tissue-specific Distribution of 5-Hydroxymethylcytosine



**FIGURE 2. Validation of MspI and HpaII isoschizomer- and  $\beta$ -glucosyltransferase-mediated 5-hmC detection and quantitation.** A, shown is unmethylated, methylated, and hydroxymethylated duplex DNAs with centrally located CCGG (boxed), where the internal C is either unmethylated C or 5-mC or 5-hmC (shown in gray). B, shown is validation for locus-specific 5-hmC detection in a fixed amount of pre-mixed DNAs, as shown in A, using MspI in the presence of  $\beta$ -GT and cofactor UDP-Glc. The reaction products were subjected to qPCR, and the ratio between observed and expected 5-hmC is plotted. C, shown is similar validation for locus-specific methylcytosine (5-mC + 5-hmC) detection in fixed amount of pre-mixed DNAs, as shown in A, using HpaII in the presence of  $\beta$ -GT and cofactor UDP-Glc. The reaction products were subjected to qPCR, and the ratio between observed and expected total methyl cytosine (5-mC + 5-hmC) is plotted.

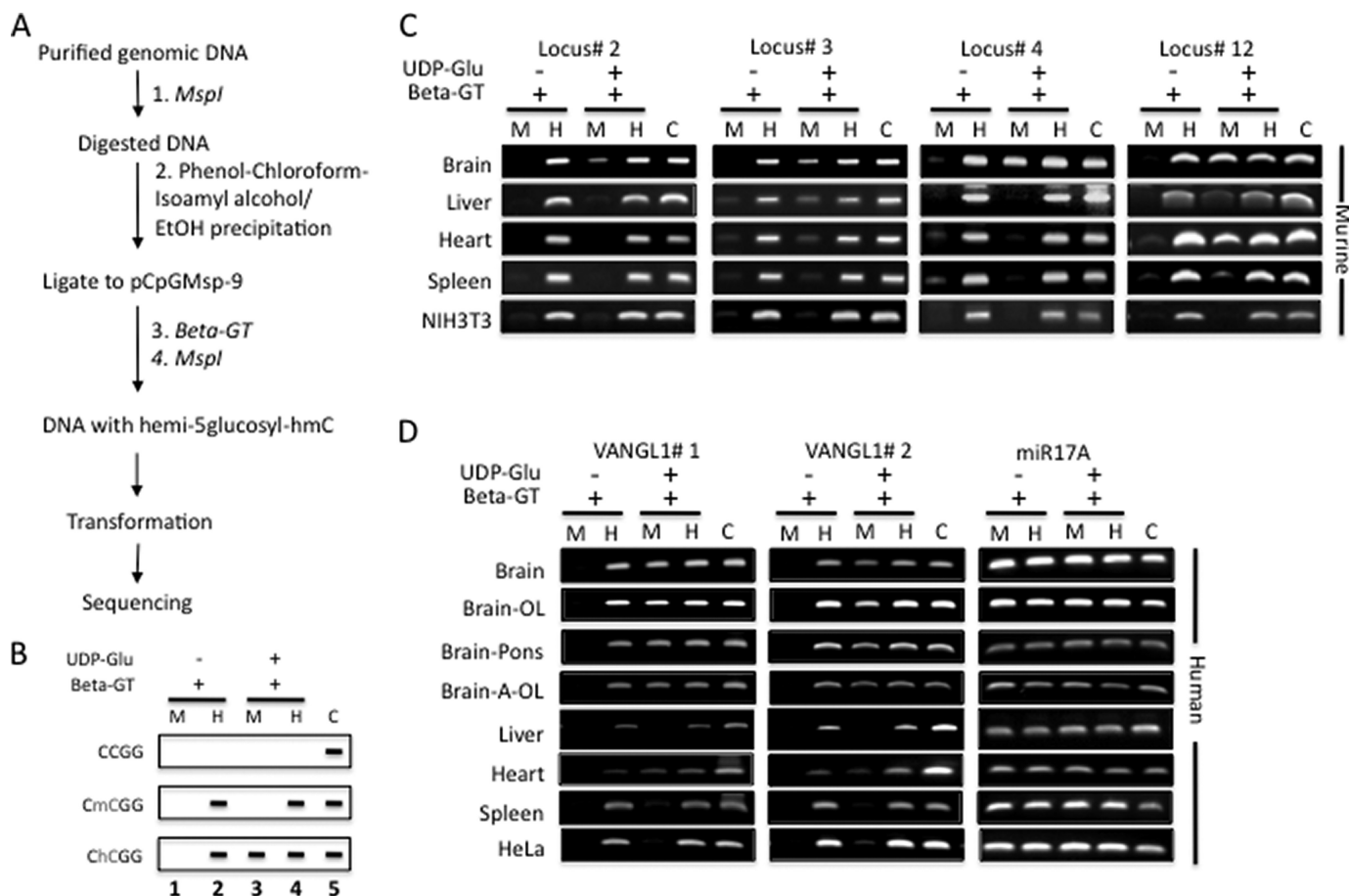
digestion of 5-hydroxymethylcytosine DNA. Enzymatic titration with varying amounts of MspI was performed by digesting a test DNA containing a symmetrical 5-hmC at the internal CpG site and cross-checking the digestion efficiency by qPCR with primers flanking the CCGG (MspI or HpaII recognition) sequence. The test DNA was 100 nucleotides long with a central MspI or HpaII recognition sequence (Fig. 2A). Indeed, 100 units of MspI was sufficient to fully digest 200 ng of 5-hydroxymethylcytosine containing DNA as no undigested DNA was observed at that specific enzyme to hydroxymethylated DNA ratio (supplemental Fig. 1, A and B). Despite using considerable amounts of enzyme, we often observed a small amount of background (1–5%) by qPCR analysis. In a similar experiment 50 units of HpaII were determined to be sufficient for complete digestion of unmethylated CCGG sequences but not 5-mC, 5-hmC, or 5-ghmC modifications at the internal cytosine (data not shown). Thus, all subsequent enzymatic digestions were carried out with 100 or 50 units of MspI or HpaII, respectively.

To determine the robustness of MspI and HpaII digestion for measuring 5-hmC and 5-mC at the internal CG of cognate CCGG sites, we mixed the test duplex DNAs at different ratios to represent various amounts of 5-hmC, 5-mC, and C (Fig. 2A). All mixes were divided into two aliquots, one being subjected to  $\beta$ -glucosylation reaction and the other being the corresponding control reactions. The control and experimental samples were then aliquoted and digested with either MspI or HpaII along with a control set of reactions without any restriction enzymes. After 4 h of restriction digestion, the digested DNAs were analyzed by qPCR using a pair of flanking primers to determine the percentages of various modified species. Indeed, the observed amounts of 5-hmC (normalized copy number from MspI +  $\beta$ -GT sample) and total 5-mC (normalized copy number from

HpaII +  $\beta$ -GT sample) for 6 DNA mixes (5-hmC:5-mC:C ratios 5:15:80, 10:50:40, 20:50:30, 33:33:33, 40:40:20, 60:10:30) very closely matched the expected values with significant Pearson correlation coefficients (5-hmC  $p$  value = 0.004, 5-mC  $p$  value = 0.006), supporting the validity and robustness of the assay system (Fig. 2, B and C).

*Identification and Tissue-specific Distribution of 5-Hydroxymethylcytosine in the Mammalian Genome*—Based on the above observations, we utilized the differential sensitivity of MspI to 5-hmC and 5-ghmC to isolate and identify 5-hmC containing loci from both mouse and human genomic DNA. We designed a scheme to clone 5-hmC containing sequences (Fig. 3A) using mouse E14 embryonic stem cell (19) and normal human brain DNAs. In brief, the DNA was digested with MspI, purified away from the enzyme, and ligated into a MspI digested vector with a single CCGG cloning site (pCpGMsp-9). After ligation of the DNA and vector, the ligation mix was subjected to a saturating amount of  $\beta$ -GT and supplemented with cofactor UDP-Glc to ensure that all the 5-hmC was converted to 5-ghmC. After glucosylation, the ligated DNA was again subjected to MspI digestion. This is to make certain that ligated plasmids contain 5-ghmC at both of the ligation sites, as any ligated products with unmethylated or methylated cytosines would be digested and linearized. This library was transformed into *E. coli*, promoting the selective degradation and elimination of the linear products. Transformed cells were zeocin-resistant and were expected to have inserts containing genomic DNA flanked by two MspI sites.

Sequencing of these clones demonstrated they did indeed have CCGG flanking sites and were perfect matches to the relevant genomes, either mouse (supplemental Table 1) or human (supplemental Table 2). Although most inserts were small enough to be fully sequenced from one end (~800 base pairs)



**FIGURE 3. Cloning, sequence identification, and tissue-specific distribution of 5-hydroxymethylcytosine.** *A*, a scheme for cloning and identification of 5-hmC containing DNA fragments using *MspI* restriction enzyme and  $\beta$ -GT (*Beta-GT*) is shown. *B*, a scheme for identification of 5-hmC containing CCGG locus using *MspI* (*M*) and *HpaII* (*H*) isoschizomers using glucosylation reaction is shown along with control (*C*). *C*, locus-specific end point PCR to interrogate and detect 5-hmC at CCGG sites in mouse genomic DNA is shown. The loci are discovered based on cloning scheme as shown in *A* and in [supplemental Table 1](#). Mouse brain, liver, heart, and spleen along with NIH3T3 cultured cell DNAs were interrogated for the 4 loci ([supplemental Table 3](#): loci 2 and 3 are 5' and 3' *MspI* sites of chromosome 10, respectively, bp 34,574,152; locus 4 is the 5' *MspI* site of chromosome 12, bp 17,432,255; locus 12 is the 5' *MspI* site of *Lpr1* intron, bp 2,372,508) as shown. *D*, shown is the locus-specific end point PCR to interrogate and detect 5-hmC at CCGG sites in human genomic DNA. The loci are discovered based on the cloning scheme as shown in *A* and in [supplemental Table 2](#). Human brain (pons, occipital lobe (*OL*)), liver, heart, and spleen along with HeLa-cultured cell DNAs were interrogated for both of the *VANGL1* loci as shown. The control DNA interrogated fragment without CCGG sequence is *miR17A*.

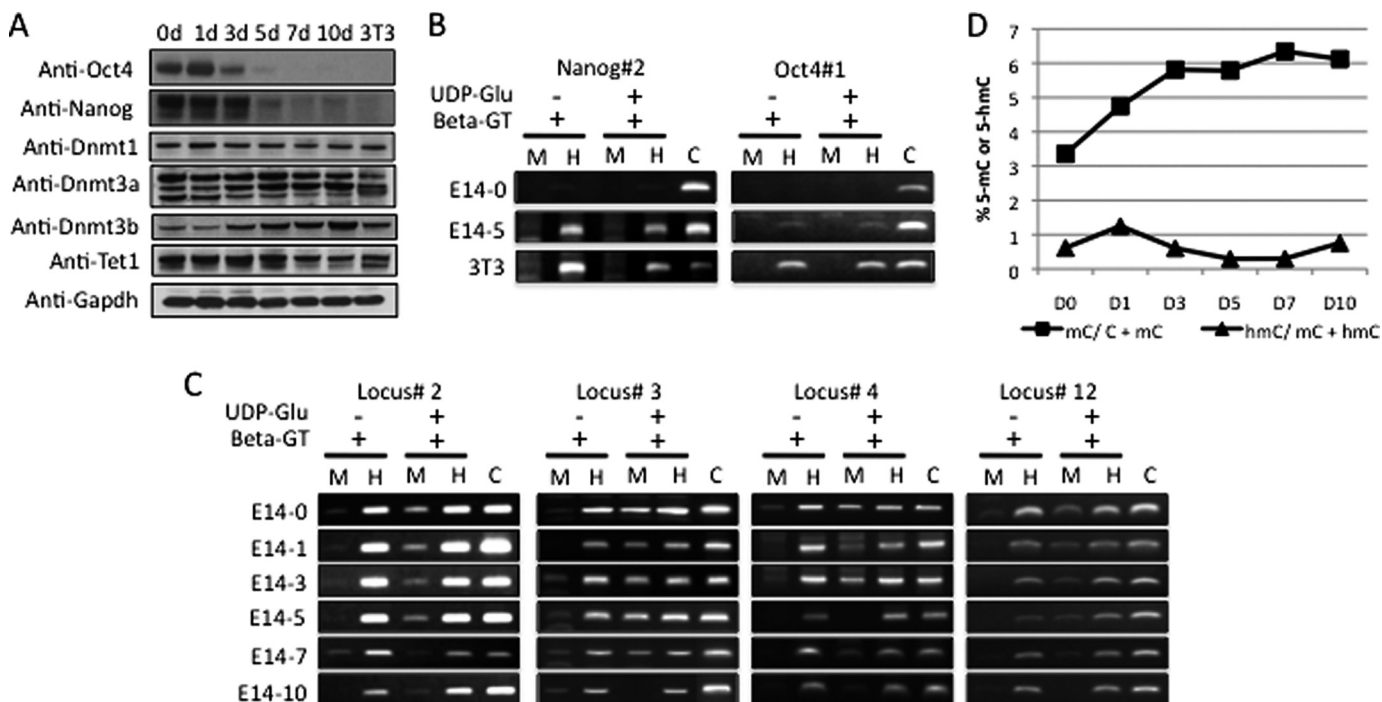
and displayed two clear CCGG sites, there were several inserts that were too long to be fully sequenced and thus only one CCGG site was located. In addition, there were several plasmids containing multiple CCGG sites in which inserts aligning to various sequences in the genome were identified, suggesting that multiple fragments were ligated and cloned into single plasmids. The identified DNA sequences with putative 5-hmC matched to various genomic locations, including repetitive DNA elements, intergenic regions of the DNA, and within genes ([supplemental Tables 1 and 2](#)), suggesting a broad 5-hmC distribution in the genome.

To validate that the selected sequences truly represent 5-hmC containing DNA, we designed end point PCR assays for a subset of the identified regions. We first determined the optimal *MspI* concentration to genomic DNA for complete digestion of hydroxymethylated loci. A digestion of 1.5  $\mu$ g of mouse brain genomic DNA with 100 units of *MspI* fully digested the DNA based on qPCR analysis of a selected locus that was identified in the screen ([supplemental Table 1](#)). Gradual loss of qPCR signal was observed as the *MspI* concentration increased but remained the same between 50–100 units of the *MspI* ([sup-](#)

[plemental Fig. 1C](#)). Similar to our previous experiment ([supplemental Fig. 1B](#)), we observed a very low background signal with qPCR analysis of genomic DNA digested with 100 units of *MspI*. This is likely due to a combination of incomplete digestion and high sensitivity of qPCR. Therefore, in subsequent experiments, 100 units of *MspI* were used per 1.5  $\mu$ g of genomic DNA, and background signal observed in the non-glucosylated DNA digested with *MspI* was subtracted from the matched samples.

For the analysis, genomic DNAs were glucosylated by  $\beta$ -GT followed by digestion with either *MspI* or *HpaII*. Control DNAs, not treated with  $\beta$ -GT, were similarly digested. Specific CCGG sites were then probed using flanking primer sets and end point PCR. Although the unmethylated DNA was expected to yield no PCR products for all digested DNA samples, depending on the type of methylation (5-mC or 5-hmC), one would expect to see either three or four PCR products, including the control PCR product from undigested samples (Fig. 3B). Internal CpG methylation (5-mC) would not yield products for *MspI*-digested DNA irrespective of glucosylation reaction (Fig. 3B, *the first versus the third lane*). However, in the presence of

## Tissue-specific Distribution of 5-Hydroxymethylcytosine



**FIGURE 4. Analysis of 5-mC and 5-hmC during embryonic stem cell differentiation to embryoid body.** A, LIF withdrawal leading to embryonic stem cell marker Oct4 and Nanog repression is shown. Shown is a Western blot of ES cell extract after LIF withdrawal along with NIH3T3 (3T3), as shown on the top, with the indicated antibodies on the left. Gapdh is the loading control. B, Oct4 and Nanog gene are methylated upon embryoid body formation. Day post-LIF withdrawal is on the left along with NIH3T3. UDP-Glc and  $\beta$ -GT are shown on the top along with restriction enzyme MspI (M) and HpaII (H) cleavage along with control (C). C, 5-hmC detection during embryoid body formation is shown. Day post-LIF withdrawal is shown on the left. D, global methylation analysis using LC-MS during embryoid body formation is shown. Each analysis point was performed in duplicate. Deoxycytosine (C), 5-methyldeoxycytosine (5-mC or mC) and 5-hydroxymethyldeoxycytosine (5-hmC or hmC) are shown.

5-hmC the glucosylation reaction will result in glucosylated DNA that would yield an additional band in lane 3 (middle panel, Fig. 3B).

As proof of the above principle, we glucosylated and digested mouse brain, liver, heart, and spleen DNA along with cell culture DNA from the mouse NIH3T3 cell line and subjected them to the protocol described in Fig. 3B. The control and glucosylated DNA samples were digested and subjected to end point PCR using four sets of PCR primers interrogating randomly chosen mouse-specific loci identified in our screen (supplemental Tables 1 and 3 and Fig. 3C). Mouse brain DNA was consistently hydroxymethylated at all of the CCGG loci examined, whereas the liver, spleen, and heart DNAs displayed variable amounts of 5-hmC in the different CCGG sites. The cultured NIH3T3 cells did not display detectable amounts of 5-hmC in any of the CCGG loci examined (Fig. 3B). Similarly, we used human DNA from whole brain as well as different regions of the brain (pons and occipital lobe (OL)), including an occipital lobe sample from an Alzheimer patient's brain (Brain-A-OL), and compared them with human heart, liver, spleen, and cultured HeLa DNA for 5-hmC content at the VANGL1 CCGG loci that we identified previously (supplemental Table 2). All of the human brain tissue DNAs displayed high levels of 5-hmC at CCGG loci of the VANGL1 gene. Similar to what we observed in mouse DNAs, human spleen and liver DNA did not appear to be hydroxymethylated, whereas heart did display some hydroxymethylation (Fig. 3C). Interestingly, abundant PCR products were detected with HpaII digestion at all of the loci examined, suggesting that they are highly methylated and

that 5-hmC is only present in a portion of the methylated alleles. As a loading control, we also examined the miR17A gene that does not have CCGG sites and found equal amounts of PCR products with all samples. These results confirm that MspI and HpaII digestion of glucosylated DNA can be used to determine the tissue-specific presence of 5-hmC and 5-mC within CCGG loci in mammalian genomic DNA.

**Dynamic Changes in 5-Hydroxymethylcytosine during Embryoid Body Formation**—It was estimated that about one tenth of the methylated cytosine of the mouse ES genome is 5-hmC and that upon differentiation, global 5-hmC decreases by ~40% (14). To investigate the dynamics of 5-hmC at specific loci, we differentiated E14 ES cells to embryoid bodies via withdrawing LIF. After LIF withdrawal, totipotent ES markers Oct4 and Nanog were down-regulated as expected (Fig. 4A). Correlating with gene repression, both Nanog and Oct4 acquired 5-mC to similar levels as that observed in NIH3T3 cells, as determined by end point PCR (Fig. 4B). Interestingly, neither Oct4 nor Nanog displayed any hydroxymethylation at the CCGG loci examined. We next measured 5-hmC at the four loci described earlier in the undifferentiated ES cells and embryoid bodies at various time points by end point PCR. Indeed, all four loci (2, 3, 4, and 12) appeared to be losing 5-hmC during embryoid body formation, as treatment of the DNA with  $\beta$ -GT did not protect against MspI digestion (the third lane in each panel, Fig. 4C).

We further inquired if the locus-specific changes in 5-hmC and 5-mC had any correlation with global level methyl cytosine dynamics. For quantification of methylated bases, we next hydrolyzed genomic DNA and performed LC-MS analysis to

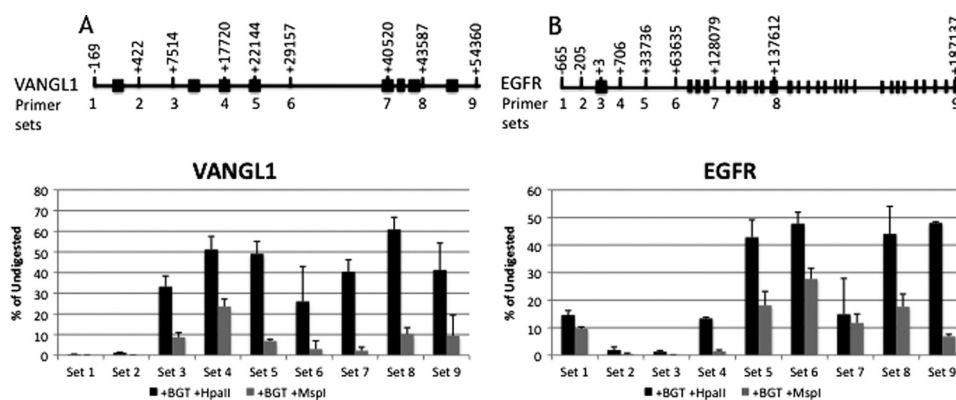


FIGURE 5. **Methylation and hydroxymethylation analysis across gene body.** A, *VANGL1*. B, *EGFR*. The genes are depicted with exons shown as *black rectangles*. The interrogated CCGG site numbers are indicated on the *top* along with the primer sets on the *bottom*. Diagrams are not drawn to scale. Total methylated cytosines (5-mC plus 5-hmC) obtained from  $\beta$ -GT-treated (BGT) HpaII-digested DNA and  $\beta$ -GT-treated MspI-digested DNA show the percentage of 5-hmC (*bottom panels*) at each specific site.

determine the amounts of genomic C, 5-mC, and 5-hmC. We observed a gradual increase in the total 5-mC in conjunction with a rapid increase followed by a slow and gradual decrease in 5-hmC levels in the genome during embryoid body differentiation (Fig. 4D).

The dynamic changes we observed in the global levels of 5-mC and 5-hmC represent potential developmental regulation mechanisms where one or more enzymatic components of DNA methylation and hydroxymethylation may be involved. To investigate whether the expression levels of DNA methyltransferases or TET1 changed during these processes, we next subjected E14 whole cell extracts to Western blotting and probed with murine anti-Dnmt antibodies (Dnmt1, Dnmt3a, and Dnmt3b) and anti-Tet1 antibody (supplemental Fig. 2). Dnmt1 expression did not change during differentiation (Fig. 4A). However, expression of three different isoforms of Dnmt3a varied, and Dnmt3b expression increased upon differentiation, suggesting that the increase in global 5-mC we observed may be mediated via the *de novo* methylation activity of the Dnmt3 enzymes (Fig. 4A). Tet1 decreased gradually upon embryoid body differentiation, which closely matched the steady reduction in 5-hmC (Fig. 4A). These results support the hypothesis that global methylation and hydroxymethylation are dynamically regulated during embryoid differentiation.

**Gene Body Methylation and Hydroxymethylation in *VANGL1* and *EGFR* Genes**—Gene body methylation is well documented in the mammalian epigenome and commonly occurs within genes that are highly expressed (20). Furthermore, several recent publications have revealed the presence of hydroxymethylation in the transcribed regions of genes using hydroxymethyl-DNA immunoprecipitation (21–23).

From our screen of human brain DNA, we identified a number of 5-hmC loci that were in gene bodies, including *VANGL1* and *EGFR* (supplemental Table 2). Therefore, we employed the MspI and HpaII isoschizomers and  $\beta$ -GT to examine both 5-mC and 5-hmC levels across *VANGL1* and *EGFR* genes in human brain DNA. Quantitative PCR on glucosylated DNA digested by HpaII yielded a total percentage of methylated cytosine (5-mC + 5-hmC), and the same DNA digested with MspI yielded 5-hmC percentages as compared with undigested DNA. Both *VANGL1* and *EGFR* were interrogated at nine different

CCGG sites covering the entire genes (Fig. 5, A and B, *upper panel*). Indeed, 5-hmC levels correlated with 5-mC levels over the entire genes in normal human brain DNA. For both of the genes, the transcription start sites displayed neither 5-mC nor 5-hmC and approximately a third of the methylation at other loci was 5-hmC (Fig. 5, A and B, *lower panels*, supplemental Figs. 3 and 4, *HNBI sample*). We also performed similar analysis on *VANGL1* and *EGFR* from spleen, liver, heart, and HeLa DNA to quantitate 5-hmC and 5-mC. Several loci displayed a low percentage of 5-hmC for the *VANGL1* gene in heart and spleen DNAs and *EGFR* (supplemental Fig. 3). Analysis of HeLa DNA did not reveal the presence of 5-hmC, although most sites were methylated to some extent in both the *VANGL1* and *EGFR* genes. These results suggest that 5-hmC may be involved in the regulation of gene expression in a tissue-specific manner.

To confirm that the signal we observed in the  $\beta$ -GT-treated MspI-digested samples was not due to processing, we performed a proof of principal experiment normalizing by two different primer sets amplifying a region with no MspI/HpaII sites (supplemental Fig. 5). We normalized qPCR results for *EGFR* primer sets 1 and 2 from a normal human brain sample by each of the “no CCGG” containing primer sets and found very similar results (supplemental Fig. 5). Indeed, the trend of 5-mC and 5-hmC profiling was consistent between all three normalizations (supplemental Fig. 5). For example, the 5-hmC values with normalization by the undigested sample alone or with *No CCGG 1* and *No CCGG 2* were 0.3, 0.38 and 0.36, respectively.

Based on our results, the fraction of 5-hmC and 5-mC at a certain CpG site can be estimated. For example, one could infer that in human normal brain DNA, about 50% of the C is modified, and about 50% of the modified Cs are 5-hmC with primer set 4 of *VANGL1*. Similarly, human heart DNA, which showed a consistently higher proportion of 5-hmC than liver or spleen, displayed 18% 5-mC with *VANGL1* primer set 6, and most of this appeared to be 5-hmC.

## DISCUSSION

Modified DNA bases are widespread in living organisms including mammals. The most common DNA modification in mammals is 5-mC, which is believed to be the precursor of

## Tissue-specific Distribution of 5-Hydroxymethylcytosine

5-hmC. During the early seventies, formic acid hydrolysis followed by chromatographic analysis was used to detect 5-hmC in murine brain and liver DNA. Using this method, 5-hmC was estimated to comprise about 15% of the total cytosine residues (24). More recently, use of thin layer chromatography and high pressure liquid chromatography coupled to mass spectrometry has reliably identified and quantified 5-hmC to be 0.6% of the total nucleotides, corresponding to about a quarter of modified cytosine residues in Purkinje cells (13). From our studies and previous reports, it appears that there are substantial amounts of 5-hmC in mammalian genomes and that these modified bases may have potential physiological significance. For example, 5-hmC may play a role in the fetal development of heart, lung, and brain (13, 14), and chromosomal translocation of MLL-TET1 and MLL-TET2 as well as mutation of TET2 may be involved in carcinogenesis, particularly leukemia (25, 26).

However, sequence-specific 5-hmC detection and quantification methods are still lacking. Most commonly used techniques for DNA methylation detection and mapping, including sodium bisulfite sequencing and its other derivative approaches (27), cannot distinguish between 5-mC and 5-hmC. In a recent study, it was suggested that bisulfite treatment could convert 5-hmC in DNA to cytosine-5-methylsulfonate, which would interfere with PCR amplification via polymerase stalling (28). Another study by Jin *et al.* (16) demonstrated that bisulfite-treated DNA containing 5-hmC can be amplified efficiently, and similar to 5-mC, 5-hmC does not undergo conversion to a deaminated cytosine ring that would be read as T base after bisulfite conversion and PCR. Furthermore, the affinity matrices using methyl binding domains may not be useful in identifying 5-hmC either, due to their poor affinity toward this modification (15). Although 5-hmC antibody-based methods may be a viable alternative to detect and map 5-hmC, they pose specific challenges to single base resolution or a sequence where 5-mC and 5-hmC are in close proximity.

Because most of the loci, depending on tissue specificity or mixed populations of cells, may contain C, 5-mC, and 5-hmC, our method of glucosylation of DNA followed by MspI/HpaII digestion will aid in detection and quantification of 5-hmC. The major perceived drawback of this method is that it can only interrogate 5-hmC in CCGG context. However, several other recently developed methods also utilize MspI/HpaII enzymes for genome wide methylation analysis. For example, reduced representation bisulfite sequencing utilizes MspI digestion in conjunction with bisulfite treatment to sequence the majority of CpG islands in the human genome (29). The reduced representation bisulfite sequencing method showed deep coverage of gene promoters and selective sampling of all other type of genomic regions while detecting epigenetic alterations (29).

Furthermore, use of the MspI isoschizomer HpaII in the methyl-sensitive cut-counting (MSCC) method generated non-targeted genome-scale data for ~1.4 million unique HpaII sites (CCGG) in the DNA of B-lymphocytes (~2.3 million total number of HpaII sites) and confirmed that gene-body methylation in highly expressed genes is a consistent phenomenon throughout the human genome. The authors observed that HpaII sites have a distribution similar to the distribution of all CpG dinucleotides (30), making them a good target for rela-

tively unbiased genome-scale profiling. For example, 7.5% of all CG sites are within CpG islands as compared with 11.8% all HpaII sites. The frequency of all CG sites and all HpaII sites within 1 kb of the transcription start site, inside genes or within a repetitive DNA, were observed to be similar, 2.3 *versus* 2.8%, 43.3 *versus* 45.5%, 51.5 *versus* 52.6%, respectively. Although our technique interrogates ~10% of the total CG sites, it can offer a potential for genome wide 5-hmC profiling by using MspI digestion of either glucosylated or control DNA followed by high throughput sequencing.

Another potential issue of this assay is background due to incomplete enzymatic digestion in combination with the extreme sensitivity of qPCR analysis. Indeed, we observed a background between 1 and 5% for most of the qPCR products despite a high concentration of enzyme being used for cleavage. The background products were dependent on various commercial batches of DNA (in our experiments all human DNAs) as well as primer sets used for amplification. Normalization by a locus containing no HpaII/MspI sites does not alter the results significantly. Therefore, we have subtracted the background values from all the qPCR analysis data and then normalized the copy numbers to undigested matched DNA samples. Although we clearly show that our assay can be used to identify and quantify both 5-hmC and 5-mC at specific cytosine residues, these caveats must be considered for any type of downstream analysis.

Our current results support previous studies indicating that ES and brain cell genomic DNAs contain considerable amounts of 5-hmC (13, 14). In addition, this report identified novel loci containing 5-hmC in mouse ES cells and in multiple regions of human brain DNA, including genes, intergenic regions, and repetitive elements. Further analysis of these loci revealed that 5-hmC patterns shift during embryoid body formation and although brain tissue DNA contains significant amounts of 5-hmC as expected, other tissue DNAs are hydroxymethylated at various loci. These results suggest that 5-hmC, like 5-mC, may play a role in determining differentiation status and tissue-specific gene expression.

In a previous study, Tahiliani *et al.* (14) reported that Tet1 mRNA expression and genomic 5-hmC are both decreased at 5 days after LIF removal. Here, we confirmed that Tet1 protein expression indeed decreases after LIF withdrawal. Interestingly, we also observed a decrease in genomic 5-hmC at day 5, but this was preceded by a sharp increase in 5-hmC at day 1 of embryoid body differentiation. Tet1 expression does not appear to increase at early time points after LIF withdrawal. Therefore, it is possible that other members of the Tet family enzymes (Tet2 and Tet3) may have altered expression upon commencement of differentiation, as they have recently been shown to possess oxygenase activity *in vivo* (31).

Finally, every locus we examined through end point or qPCR that was identified in our screen displayed high levels of total methylation, even in samples that did not appear to be hydroxymethylated. This suggests that these loci are normally highly methylated and then in certain tissues this methylation can be converted to 5-hmC and thus accounts for only a fraction of the total methylation. In agreement with several recently published articles (21–23), we found that 5-hmC coincided



with gene body methylation. *VANGL1* and *EGFR* both have highly regulated expression in brain tissues and are important for normal brain development and function (32–34). It is currently unknown how gene body methylation may affect expression, and the identification of hydroxymethylation in the transcribed regions reveals yet another layer of possible epigenetic regulation of these genes. Future studies will be required to determine the exact function of 5-hmC in various tissues and within different regions of the mammalian genome.

*Acknowledgments*—We thank Jack Benner for nucleotide analysis of the genomic DNA and Thomas C. Evans and William Jack for reading the manuscript. We thank Drs. Donald G. Comb, Richard J. Roberts, James V. Ellard, and New England Biolabs, Inc. for supporting the basic research.

## REFERENCES

- Bestor, T. H., and Verdine, G. L. (1994) *Curr. Opin. Cell Biol.* **6**, 380–389
- Ooi, S. K., O'Donnell, A. H., and Bestor, T. H. (2009) *J Cell Sci.* **122**, 2787–2791
- Lister, R., Pelizzola, M., Downen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., and Ecker, J. R. (2009) *Nature* **462**, 315–322
- Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M., and Jacobsen, S. E. (2008) *Nature* **452**, 215–219
- Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., and Lander, E. S. (2008) *Nature* **454**, 766–770
- Gebhard, C., Benner, C., Ehrlich, M., Schwarzfischer, L., Schilling, E., Klug, M., Dietmaier, W., Thiede, C., Holler, E., Andreesen, R., and Rehli, M. (2010) *Cancer Res.* **70**, 1398–1407
- Dickson, J., Gowher, H., Strogantsev, R., Gaszner, M., Hair, A., Felsenfeld, G., and West, A. G. (2010) *PLoS Genet.* **6**, e1000804
- Ho, K. L., McNaie, I. W., Schmiedeberg, L., Klose, R. J., Bird, A. P., and Walkinshaw, M. D. (2008) *Mol. Cell* **29**, 525–531
- Dhasarathy, A., and Wade, P. A. (2008) *Mutat. Res.* **647**, 39–43
- Jones, P. A., and Takai, D. (2001) *Science* **293**, 1068–1070
- Dahlmann, H. A., Vaidyanathan, V. G., and Sturla, S. J. (2009) *Biochemistry* **48**, 9347–9359
- Goosen, N. (2010) *DNA Repair (Amst)* **9**, 593–596
- Kriaucionis, S., and Heintz, N. (2009) *Science* **324**, 929–930
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) *Science* **324**, 930–935
- Valinluck, V., Tsai, H. H., Rogstad, D. K., Burdzy, A., Bird, A., and Sowers, L. C. (2004) *Nucleic Acids Res.* **32**, 4100–4108
- Jin, S. G., Kadam, S., and Pfeifer, G. P. (2010) *Nucleic Acids Res.* **38**, e125
- Waalwijk, C., and Flavell, R. A. (1978) *Nucleic Acids Res.* **5**, 3231–3236
- Moréra, S., Imberty, A., Aschke-Sonnenborn, U., Rüger, W., and Freemont, P. S. (1999) *J. Mol. Biol.* **292**, 717–730
- Wakayama, T., Rodriguez, I., Perry, A. C., Yanagimachi, R., and Mombaerts, P. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14984–14989
- Feng, S., Cokus, S. J., Zhang, X., Chen, P. Y., Bostick, M., Goll, M. G., Hetzel, J., Jain, J., Strauss, S. H., Halpern, M. E., Ukomadu, C., Sadler, K. C., Pradhan, S., Pellegrini, M., and Jacobsen, S. E. (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8689–8694
- Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A., Rappaport, J., and Helin, K. (2011) *Nature* **473**, 343–348
- Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. A., Marques, C. J., Andrews, S., and Reik, W. (2011) *Nature* **473**, 398–402
- Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y. E., and Zhang, Y. (2011) *Genes Dev.* **25**, 679–684
- Penn, N. W., Suwalski, R., O'Riley, C., Bojanowski, K., and Yura, R. (1972) *Biochem J.* **126**, 781–790
- Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., and Hayashi, Y. (2002) *Cancer Res.* **62**, 4075–4080
- 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., Liu, X. S., Aravind, L., Agarwal, S., Maciejewski, J. P., and Rao, A. (2010) *Nature* **468**, 839–843
- Xiong, Z., and Laird, P. W. (1997) *Nucleic Acids Res.* **25**, 2532–2534
- Huang, Y., Pastor, W. A., Shen, Y., Tahiliani, M., Liu, D. R., and Rao, A. (2010) *PLoS One* **5**, e8888
- Gu, H., Bock, C., Mikkelsen, T. S., Jäger, N., Smith, Z. D., Tomazou, E., Gnirke, A., Lander, E. S., and Meissner, A. (2010) *Nat. Methods* **7**, 133–136
- Ball, M. P., Li, J. B., Gao, Y., Lee, J. H., LeProust, E. M., Park, I. H., Xie, B., Daley, G. Q., and Church, G. M. (2009) *Nat. Biotechnol.* **27**, 361–368
- Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., and Zhang, Y. (2010) *Nature* **466**, 1129–1133
- Torban, E., Patenaude, A. M., Leclerc, S., Rakowiecki, S., Gauthier, S., Andelfinger, G., Epstein, D. J., and Gros, P. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3449–3454
- Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C. L., Haase, J., Janes, J., Huss, J. W., 3rd, and Su, A. I. (2009) *Genome Biology* **10**, R130
- Huse, J. T., and Holland, E. C. (2010) *Nat Rev Cancer* **10**, 319–331