

## Accepted Manuscript

A sensitive mass-spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples

Thuc Le, Kee-Pyo Kim, Guoping Fan, Kym F. Faull

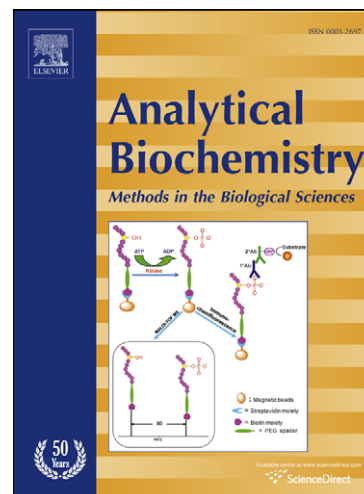
PII: S0003-2697(11)00040-6  
DOI: [10.1016/j.ab.2011.01.026](https://doi.org/10.1016/j.ab.2011.01.026)  
Reference: YABIO 10315

To appear in: *Analytical Biochemistry*

Received Date: 17 November 2010  
Revised Date: 13 January 2011  
Accepted Date: 19 January 2011

Please cite this article as: T. Le, K-P. Kim, G. Fan, K.F. Faull, A sensitive mass-spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples, *Analytical Biochemistry* (2011), doi: [10.1016/j.ab.2011.01.026](https://doi.org/10.1016/j.ab.2011.01.026)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



# A sensitive mass-spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples

Thuc Le<sup>1,2#</sup>, Kee-Pyo Kim<sup>1,3#</sup>, Guoping Fan<sup>1\*</sup> and Kym F. Faull<sup>2</sup>

<sup>1</sup>Department of Human Genetics, Broad Stem Cell Research Center, <sup>2</sup>Pasarow Mass Spectrometry Laboratory, Department of Psychiatry & Biobehavioral Sciences and the Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, UCLA, 695 Charles Young Drive, Los Angeles CA, USA, <sup>3</sup>Present address: Center of Cell and Gene Therapy, Baylor College of Medicine, Houston, TX 77030, USA

\*Correspondence should be addressed to: [gfan@mednet.ucla.edu](mailto:gfan@mednet.ucla.edu)

Telephone: (310) 267-0439 Fax: (310) 794-5446

# The first two authors contributed equally to this project.

**ABSTRACT**

The recent discovery of 5-hydroxymethyl-cytosine (5hmC) in embryonic stem cells and post-mitotic neurons has triggered the need for quantitative measurements of both 5-methyl-cytosine (5mC) and 5hmC in the same sample. We have developed a method using liquid chromatography electrospray ionization tandem mass spectrometry with multiple reaction monitoring (LC-ESI-MS/MS-MRM) to simultaneously measure levels of 5mC and 5hmC in digested genomic DNA. This method is fast, robust and accurate, and is more sensitive than the current 5hmC quantitation methods such as end-labeling with thin-layer chromatography and radio-labeling by glycosylation [1; 2]. Only 50 ng of digested genomic DNA is required to measure the presence of 0.1% 5hmC in DNA from mouse embryonic stem cells. Using this procedure we show that human induced pluripotent stem cells exhibit a dramatic increase in 5mC and 5hmC levels compared to parental fibroblast cells, suggesting a dynamic regulation of DNA methylation and hydroxymethylation during cellular reprogramming.

## INTRODUCTION

The pattern of methylated cytosine residues in DNA provides an inheritable epigenetic code that regulates gene expression during development. The covalent addition of a methyl group at the 5-position of cytosine primarily occurs in the CpG dinucleotide, and is catalyzed by a family of DNA methyltransferases (Dnmts) including maintenance Dnmt1 and *de novo* Dnmt3a and Dnmt3b. DNA methylation is involved in various biological processes such as genomic imprinting, silencing of retroviral transposons, X chromosome inactivation, and cellular differentiation. Mechanistically, promoter methylation can lead to transcriptional repression directly by inhibiting transcriptional binding, or indirectly by recruiting various proteins including methyl CpG binding proteins (MBDs), co-repressors and histone modification enzymes involved in chromatin remodeling [3; 4; 5; 6; 7; 8]. Importantly, many studies have shown that DNA methylation is a dynamic process in cellular proliferation and differentiation, and is tightly regulated in normal development. Aberrant DNA methylation patterns and mechanisms are deleterious to the developing central nervous system (CNS) [8; 9; 10; 11].

Recently there has been renewed interest in another, related, mammalian DNA modification, 5-hydroxymethyl-cytosine (5hmC). Significant levels of 5hmC are found in the developed murine central nervous system and in embryonic stem cells [1; 12; 13; 14]. *In vivo* addition of a hydroxyl group onto 5-methyl-cytosine (5mC) is catalyzed by 2-oxoglutarate oxygenase Tet1, Tet2, and Tet3 [1; 14]. There are also reports that 5hmC can be formed by other mechanisms beside Tet pathway, including UV irradiation

of 5mC in aerated aqueous solution [15] and DNA methyltransferase reaction of cytosine with formaldehyde [16]. To date, only the Tet pathway has been demonstrated to produce 5hmC in mammalian genomic DNA.

Speculation that 5hmC is involved in the DNA demethylation pathway comes from the two reported mechanisms of converting 5hmC into C. Bacterial DNA methyltransferases catalyze the removal of formaldehyde from 5hmC, thus converting 5hmC to C [16]. Another demethylation mechanism involves the photochemical hydration of 5hmC in basic solution [15]. However, these two possible DNA demethylation mechanisms have yet to be confirmed in mammalian models.

Some of the most commonly used methods for profiling and quantification of DNA methylation, such as bisulfite sequencing and methylation-sensitive enzyme-based assays, are unable to distinguish between 5hmC and 5mC [1; 17]. Several methods have been used to measure the 5hmC levels in the genome: these include end-labeling followed by thin layer chromatography [1], high performance liquid chromatography (HPLC) with UV detection [16], enzymatic radioactive glycosylation labeling [2], and single molecule, real-time sequencing [18]. The thin layer chromatography method has the advantage of being low cost and simple, but requires the availability of radioactive substrates and the accuracy is not comparable to other available methods. The specificity of UV detection relies heavily on the chromatographic separation to avoid co-elution of other components, including other DNA and RNA nucleotides that may be present in biological samples. The glycosylation method is based on enzymatic incorporation of radio-labeled glucose into genomic 5hmC, with quantification by radioactive counting. However a complete enzymatic reaction cannot be readily

assured and 5mC levels cannot be measured simultaneously. Independent measurement of 5hmC is possible with next generation sequencing, but the technology has yet to be perfected for accurate quantitation of many low abundant nucleotides including 5hmC.

Previous work demonstrated the precision, selectivity and sensitivity of liquid chromatography tandem mass spectrometry for measuring 5mC in biological samples, and as a diagnostic tool for cancer [19; 20; 21; 22; 23; 24; 25]. Using this technique all known DNA (excluding 5hmC) and RNA components, have been separated, distinguished and independently quantitated [24]. This approach allows DNA methylation to be measured both at the global [20; 21; 23; 24] and gene promoter regions [22]. However, none of the previous reports include 5hmC. We were prompted to develop a fast, sensitive and accurate method to measure both 5mC and 5hmC levels to support ongoing work on epigenetic control of stem cells and neural development. Here we report the use of liquid chromatography electrospray ionization tandem mass spectrometry with multiple reaction monitoring (LC-ESI-MS/MS-MRM) for the determination of genomic DNA methylation and hydroxymethylation. Separation of the deoxyribonucleosides is achieved within 6 minute using sub-two micron particle size reverse phase chromatography columns. In addition, mass-based detection discriminates between the three nucleoside bases of interest- 5hmC, 5mC and cytosine (C). The combination of LC and MS minimizes any possible cross-talk between the measurements of low abundant molecules (5hmC and 5mC) in the face of a chemically similar abundant species (C). Together, our data indicate that the MRM method provides unambiguous and independent quantification of 5hmC, 5mC, and C with high

reproducibility and low limits of detection of around 0.5 fmol per sample. This limit of detection can be equated to 50 ng of digested genomic DNA to measure 5hmC levels at the 0.1% level. Furthermore, the method is relatively fast, requiring less than 48 hours from extracting genomic DNA (few hours to a day), to digesting genomic DNA into nucleoside components (1-2 hours), and measuring the 5hmC and 5mC levels using the MRM method (6 minutes per sample).

ACCEPTED MANUSCRIPT

## MATERIALS AND METHODS

### Cell culture

mESCs of wild-type J1, *Dnmt1*<sup>-/-</sup> (cc), double knockout *Dnmt3a*<sup>-/-</sup>;*Dnmt3b*<sup>-/-</sup> (DKO), and triple knockout *Dnmt1*<sup>-/-</sup>;*Dnmt3a*<sup>-/-</sup>;*Dnmt3b*<sup>-/-</sup> (TKO), were maintained on gelatin-coated plates in mESC medium containing DMEM (Invitrogen), 10% fetal bovine serum (FBS, Gibco), 100  $\mu$ M 2-mercaptoethanol (Sigma), 0.1 mM nonessential amino acids (Gibco), 1 mM sodium pyruvate, 1X L-glutamine (Gibco), 1X penicillin/streptomycin (MP Biomedicals), and 100 units/mL of leukemia inhibitory factor (LIF).

Human BJ fibroblasts from neonatal foreskin were purchased from American Type Culture Collection (ATCC). These were maintained in medium containing DMEM, 1X penicillin/streptomycin, 1X glutamine, and 10% fetal bovine serum. Fibroblasts were passaged every 3 ~ 5 days using 0.05 % trypsin (Invitrogen).

Undifferentiated hESCs (HSF1) were maintained on a feeder layer of mitomycin C (Sigma)-treated mouse embryonic fibroblasts (MEF) in hESC medium containing DMEM/F12 (Invitrogen) supplemented with 20% Knockout Serum Replacement (KSR, Gibco), 1X glutamax, 1X non-essential amino acids, 0.11 mM  $\beta$ -mercaptoethanol (Sigma), 1X penicillin/streptomycin and 10 ng/ml of bFGF (PeproTech). hESCs were passaged every 5 ~ 7 days using Collagenase IV (Gibco) and Dispase (Gibco) at a final concentration of 1 mg/ml in hESC medium. All cells were cultured under a protocol approved by the Chancellor's Animal Research Committee (ARC) and Embryonic Stem Cell Research Oversight (ESCRO) Committee at UCLA.

### Derivation of human induced pluripotent stem cells (iPSCs)



The production of human iPSCs followed a published protocol [26] with slight modifications. Specifically, retroviruses containing OCT4, SOX2, KLF4 and c-MYC were produced in the Platinum-E (PLAT-E) retrovirus packaging cell line (Cell Biolabs). Viral supernatants were collected 48 hours and 72 hours after transfection and filtered through a 0.45  $\mu\text{m}$  PVDF filter (Millipore). The Slc7a1-expressing human BJ fibroblasts were plated at  $1.5 \times 10^5$  cells per well on a 6-well plate on day 1. On day 2, each retroviral supernatant was added into the fibroblasts in the presence of 4  $\mu\text{g}/\text{ml}$  of polybrene (Sigma). A second round of transduction was performed on day 3. Infection efficiency was monitored by fluorescence microscopy of cells transduced by retrovirus carrying GFP. On day 5, cells were trypsinized and re-plated in a density of  $1 \times 10^5$  per 10 cm plate on mitomycin (Sigma)-treated MEFs. On day 6, the medium was changed to hESC medium. iPS cell colonies were picked 3 weeks after infection. The picked colonies were cultured and passaged according to standard culturing protocols. All plasmids including pMXs-OCT4, pMXs-SOX2, pMXs-KLF4, pMXs-c-MYC and pLenti6/UbC/mSlc7a1 used for the derivation of iPSCs were purchased from Addgene.

#### **DNA extraction and hydrolysis**

hESC and iPSC colonies were harvested and passed through a cell strainer (BD falcon) to remove the feeder cells. Colonies were washed with PBS, treated with 500  $\mu\text{L}$  of DNA lysis buffer (100 mM Tris-HCl pH 8.0, 5 mM EDTA, 200 mM NaCl, 0.2% SDS), 5  $\mu\text{L}$  of proteinase K (100 mg/mL, Roche), and 5  $\mu\text{L}$  of RNase A (10 mg/mL, Roche), and incubated overnight at 37  $^{\circ}\text{C}$  in a shaking incubator. Genomic DNA was purified by a standard phenol/chloroform extraction followed by precipitation with two volumes of cold 100% ethanol. Subsequently the extracted genomic DNA was redissolved in TE buffer

(10 mM Tris-HCl and 1 mM EDTA, pH 8.0). Genomic DNA was quantified spectrophotometrically at 260 nm (Thermo Scientific NanoDrop).

DNA hydrolysis was performed by using DNA Degradase Plus (Zymo Research). Briefly, 1  $\mu$ g of genomic DNA was mixed with 2.5  $\mu$ l 10X DNA Degradase Reaction buffer, 1  $\mu$ l DNA Degradase Plus and water to make a total reaction volume of 25  $\mu$ l. The reaction mixture was incubated in 37 °C for more than an hour. Finally, the reaction was inactivated by adding 175  $\mu$ l of 0.1% formic acid to yield a final concentration of 5 ng of digested DNA/ $\mu$ l.

### **DNA standards**

Three 897bp DNA standards, each homogenous for either unmodified C, 5mC, or 5hmC, were purchased (Zymo, Irvine, CA), and used to generate a calibration curve. The standards were prepared by PCR using the appropriate nucleotides and were spin-column purified by the manufacturer to obtain 50 ng/ $\mu$ l solutions. By MRM criteria these standards were all more than 99.6% pure (Supplementary Figure 1).

### **Multiple Reaction Monitoring (MRM) Quantitation.**

DNA hydrolysis samples (10  $\mu$ l typically containing 50 ng of digested DNA) were injected onto a reverse phase UPLC column (Eclipse C18 2.1 x 50 mm, 1.8  $\mu$  particle size, Agilent) equilibrated and eluted (100  $\mu$ l/min) with water/methanol/formic acid (95/5/0.1, all by volume). The effluent from the column was directed to an electrospray ion source (Agilent Jet Stream) connected to a triple quadrupole mass spectrometer (Agilent 6460 QQQ) operating in the positive ion multiple reaction monitoring mode using previously optimized conditions, and the intensity of specific  $MH^+ \rightarrow$  fragment ion transitions were recorded (5mC m/z 242.1  $\rightarrow$  126.1, 5hC 258.1  $\rightarrow$  142.1 and dC m/z

228.1→112.1). The measured percentage of 5mC and 5hmC in each experimental sample was calculated from the MRM peak area divided by the combined peak areas for 5mC plus 5hmC plus C (total cytosine pool).

With each batch of experimental samples a series of standard samples was simultaneously prepared using the 897bp DNA standards, and run. The standard samples contained increasing amounts of 5mC and 5hmC in the presence of the same amount of C (0-10% for 5mC and 0-2% for 5hmC). Calibration curves were constructed for 5mC and 5hmC from the data obtained from the standard samples (measured 5mC or 5hmC peak area/total cytosine pool plotted against actual percentage of either 5mC or 5hmC in the samples). The measured percentage of 5mC and 5hmC in each experimental sample was then converted to actual percentage 5mC and 5hmC by interpolation from the calibration curves. This provided a correction for any differences that might exist in the molar MRM responses of the various nucleosides.

### **TET1 cloning, virus production and infection**

The Flag-tagged TET1 catalytic domain (NM\_030625) was amplified from hESCs cDNA using PCR with Hotstar Taq polymerase (Qiagen). Primers used for the amplification of TET1 are described in Supplementary Table 1. PCR products were purified and cloned into pCR4-TOPO plasmid using TOPO TA cloning kit (Invitrogen) following the manufacturer's protocol. TET1 sequences were verified by DNA sequencing. The TET1 catalytic domain was digested and ligated into BamHI and EcoRI sites of lentiviral plasmid, FUIGW (Addgene).

For lentivirus production, 293T cells were plated at  $8 \times 10^6$  cells per 10 cm plate and incubated at 37 °C overnight. Cells were transfected with 9  $\mu$ g of FUIGW-Flag-

TET1-GFP or FUIGW) along with 4.5  $\mu$ g of pMLDg/pRRE (Addgene), 1.8  $\mu$ g of pRSV-Rev (Addgene), and 2.7  $\mu$ g of pCMV-VSVG (Addgene) by Lipofectamine 2000 (Invitrogen), according to the manufacturer's protocol; 48 and 72 hours after transfection, viral supernatants were collected and filtered through a 0.45  $\mu$ m PVDF filter (Millipore). Fresh viral supernatants were infected into 293T cells in the presence of 4  $\mu$ g/ml of polybrene (Sigma) overnight. Infected cells were analyzed by immunostaining and extracted for LC-ESI-MS/MS-MRM.

### **Immunocytochemistry**

Antibodies used for immunostaining were: OCT4 (1:20, Santa Cruz), NANOG (1:100, Abcam), TRA1-60 (1:200, Chemicon), SSEA4 (1:200, Chemicon), SOX2 (1:200, Cell Signaling) DNMT3B (1:500, a gift from Dr. En Li). Human ES and iPS cells were plated on sterile coverglasses in 6-well plates and cultured for 24 ~ 48 hr. The medium was aspirated and cells were washed once with PBS and fixed with 4% paraformaldehyde/PBS for 30 min at room temperature. Cells were washed three times with 0.2 % Tween 20/PBS, then permeabilized with 0.2% Triton X-100/PBS for 30 min at room temperature and washed once with 0.2% Tween20/ PBS. Blocking was performed for 1 h at room temperature with 2% BSA/0.1% Tween 20/PBS. Primary antibodies diluted in blocking solution were incubated for 1 h at room temperature. Cells were washed three times with 0.2% Tween 20/PBS. Cy2- and Cy3- conjugated secondary antibodies diluted in blocking solution were incubated at room temperature for 30 min. Cells were washed three times with 0.2% Tween 20/PBS, stained with DAPI and mounted on glass slides (Fisher Scientific). Images were analyzed on a Nikon Eclipse 80i inverted microscope equipped with a CCD camera by using Spot Advance

imaging software (Diagnostics Instruments).

### **Southern blot analysis.**

Genomic DNA (5 µg) was digested with *Bst*BI (NEB) overnight at 37 °C, separated on a 1% agarose gel, and transferred to a Hybond-N<sup>+</sup> membrane (Amersham) in 10X SSC. The membrane was hybridized with P<sup>32</sup>-end-labeled oligo probes for *Sat 2* and *Sat 3* in QuikHyb solution (Stratagene) at 42 °C for 2 hours. The hybridized membrane was washed twice in 2X SSC/0.1 % SDS at room temperature and washed once in 0.1XSSC/0.1 % SDS at 60 °C. The membrane was exposed to a Kodak BioMax MS film. Oligo probes are described in Supplementary Table 1.

### **Bisulfite sequencing**

Genomic DNA (2 µg) was subjected to bisulfite conversion using EZ DNA Methylation Kit (Zymo research) following the manufacturer's protocol. Subsequently, PCR was carried out with HotStar Taq polymerase (Qiagen). Primers (OCT4 and NANOG) and PCR conditions are described in Supplementary Table 1. PCR products were purified by Wizard SV gel and PCR clean-up kit (Promega) and cloned into pCR4-TOPO plasmid using TOPO TA cloning kit (Invitrogen) following the manufacturer's protocol. Following transformation, 10 ~ 12 colonies were subjected to direct sequencing with the M13 reverse primer, followed by inoculations and minipreps.

### **RT-PCR**

Total RNA was isolated from cells using the RNeasy Mini kit (Qiagen) with QIAshredder (Qiagen) following the manufacturer's protocol. The small residual amounts of DNA were removed by the RNase-Free DNase Set kit (Qiagen). The concentration of RNA was quantified spectrophotometrically at 260 nm (Thermo

Scientific NanoDrop). Total RNA (1 µg) was reverse transcribed into complimentary DNA (cDNA). The reverse transcription (RT) was performed using the iScript cDNA synthesis Kit (Bio-Rad) according to the manufacturer's protocol. PCR was performed on a MyIQ Thermocycler (Biorad).

ACCEPTED MANUSCRIPT

## RESULTS

### Mass spectrometric characterization of 5hmC, 5mC, and C

An equal molar mixture of three commercial 897bp standard DNA fragments (Zymo, Irvine, CA), each homogenous for either unmodified C, 5mC, or 5hmC, was prepared and digested into the nucleoside components. This mixture produced ions during electrospray ionization corresponding to the protonated nucleosides deoxycytidine (dC), 5-methyl-2'-deoxycytidine (5mdC) and 5-hydroxymethyl-2'-deoxycytidine (5hmdC) at  $m/z$  228.1, 242.1 and 258.1, respectively. Collisionally induced dissociation (CID) of these protonated nucleosides produced a number of fragments, the most abundant of which correspond to the protonated bases liberated by cleavage of the glycosidic bond at  $m/z$  112.1 (C), 126.1 (5mC) and 142.1 (5hmC). Mass-based distinction between these nucleosides is therefore possible because the parent masses are unique as are the corresponding bases that result from glycosidic cleavage. The gas phase glycosidic cleavage of nucleosides is efficient, and the intensity of transitions of the protonated nucleosides to their corresponding bases can be used in the MRM mode for independent quantification:  $m/z$  228.1 $\rightarrow$ 112.1, 242.1 $\rightarrow$ 126.1 and 258.1 $\rightarrow$ 142.1 for dC to C, 5mdC to 5mC and 5hmdC to 5hmC, respectively.

When the equivalent of 50 ng of DNA was analyzed by LC-ESI-MS/MS-MRM, the sequentially eluting symmetrical peaks corresponding to dC to C, 5mdC to 5mC, and 5hmdC to 5hmC transitions revealed no detectable cross-talk (Figure 1A). Using the same commercial DNA fragments, the linearity of the response was tested by preparing

and analyzing samples with varying amount of 5mC and 5hmC in the presence of a constant amount of C containing DNA. Calibration curves constructed from this data set for both 5mC and 5hmC were linear (Figure 1B), and were used to calculate the percent DNA methylation and hydroxymethylation in experimental samples.

### Validation of the MRM method

The method was then used to measure the percentage of 5mC and 5hmC in some mouse embryonic stem cell (mESC) lines. The 5mC level of *Dnmt1*<sup>-/-</sup> mESC is about 25% of the wild-type 5mC level. Also, the double knockout, *Dnmt3a*<sup>-/-</sup> and *Dnmt3b*<sup>-/-</sup>, mESC at passage 35 (P35) has a 5mC level of about 16% of the wild-type (Figure 2). These results are consistent with previous studies that used nearest neighbor analysis and bisulfite next generation sequencing (BS-Seq) [27; 28].

A comparison of the 5mC and 5hmC levels in various mESC lines shows a strong correlation (Figure 2). This correlation is consistent with the biological conversion of 5mC to 5hmC by oxygenase TET enzyme [1; 14]. A higher 5mC level would favor more 5hmC conversion, and thus raise the global level of 5hmC.

To confirm another previous study, the FLAG-tagged TET1 catalytic domain was over-expressed in 293T cells (Figure 3A). Using the MRM method a drastic increase in 5hmC level was recorded accompanied by about 50% loss of 5mC level compare to control cells (Figure 3B). This observation was consistent with the previous study using 5mC antibody fluorescence immunocytochemistry that showed transfected 293T cells have 55% of the DNA methylation level found in control cells [1].

### Measuring 5hmC and 5mC in somatic and induced pluripotent stem cells



BJ fibroblasts were used to generate induced pluripotent stem cells (iPSCs) by retrovirally introducing *Oct4*, *Sox2*, *Klf4*, and *c-Myc* (Supplementary Figure 2). Two iPSC colonies (BJ iPS #7 and BJ iPS #8) were picked and expanded for further analysis. Both iPS lines showed a significant increase in 5mC after reprogramming from BJ fibroblasts (Figure 4A). This 5mC increase was accompanied by a significant increase in the 5hmC level.

Southern blot was performed on *Sat 2* and *Sat 3* repetitive sequences of BJ fibroblasts, BJ iPS #7 and BJ iPS #8, and showed an increase in DNA methylation at BstBI sites (TTCGAA) in the repetitive regions (Figure 4B). These results are consistent with the MRM result. However the promoter regions of both *Nanog* and *Oct4* underwent DNA demethylation (Figure 4C), suggesting that the 5mC level increase occurs on selective gene regions.

## DISCUSSION

We report the details of a fast and reliable method for measuring the relative levels of 5mC and 5hmC in small samples of digested DNA. Through the use of ultra performance liquid chromatography (UPLC) with sub-micron particle size packing, the analysis time is reduced to 6 minutes per sample. Using this method the limit of detection for these two nucleosides is around 0.5 fmol injected on-column. The linearity of the response is demonstrated across one order of magnitude which is more than sufficient for biological samples, and is probably much greater, and the levels of 5mC and 5hmC have been measured in ten different cell lines. Experience has shown that batches exceeding one hundred samples can be analyzed without any noticeable change or deterioration in chromatographic performance and MRM response. The durability of the UPLC columns used in this work is such that hundreds of samples have been analyzed on the same column, although as a precaution high organic washes every 20-30 samples are used to avoid any complications that could arise from the accumulation of materials not eluted during the isocratic analyses.

Both internal [23] and external standards [20; 21; 22; 24] have been used for quantitative measurements of DNA methylation. External standards that mimic the processing of biological samples have been used here. This has been done by preparing pre-mixed standard DNA samples, and then processing them through the entire work-up and digestion. The resulting standard curves reflect the unavoidable errors that arise during sample work-up such as ion suppression that might arise from components used in the reaction solutions. Consistent with the report of Song et al [25],

our results reveal no evidence that small variations in the completeness of DNA hydrolysis adversely affects the linearity of the observed responses.

In this report, the DNA methylation levels in iPSCs are not similar to ESCs and fibroblasts. Various findings have already indicated that there are epigenetic differences between normal ESCs and iPSCs, particularly in DNA methylation patterns [29; 30; 31; 32; 33]. Our preliminary results on the comparison between iPSCs and parental somatic cells show a significant number of genes undergo increased DNA methylation during re-programming (Shen et al., unpublished data). It has been reported by others that the epigenetic mechanism of DNA methylation is a limiting factor in the reprogramming process, and that the DNA methylation pattern may not truly emulate the pattern found in ESC [31; 34]. For example, treatment of DNA methyltransferase inhibitor, 5-aza-cytidine, facilitated the transition of partially reprogrammed cells to iPSC [34]. Interestingly, the level of 5hmC in iPSC from reprogrammed fibroblast reported here appears to be restored to the levels found in ESC.

In conclusion, we have established an accurate and robust assay for the simultaneous quantification of 5hmC and 5mC levels in biological samples. LC-ESI-MS/MS-MRM is acknowledged as a gold standard in quantitation methodology, and the method described here will have widespread applicability and is sufficiently flexible for expansion to include other rare nucleosides.

## FUNDING

This project is supported by CIRM RC1-0111 and NIH RO1 NS 051411 to GF.

## ACKNOWLEDGEMENT

The authors wish to thank Andy Gieschen from Agilent for his technical support with the Agilent 6460 mass spectrometer, Zymo Research for providing reagents and samples, and Dr. Julian Whitelegge for providing suggestions and encouragement for this study.

ACCEPTED MANUSCRIPT

## References

- [1]M. Tahiliani, K.P. Koh, Y. Shen, W.A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L.M. Iyer, D.R. Liu, L. Aravind, and A. Rao, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324 (2009) 930-5.
- [2]A. Szwagierczak, S. Bultmann, C.S. Schmidt, F. Spada, and H. Leonhardt, Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* (2010).
- [3]S.H. Cross, R.R. Meehan, X. Nan, and A. Bird, A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nat Genet* 16 (1997) 256-9.
- [4]X. Nan, F.J. Campoy, and A. Bird, MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* 88 (1997) 471-81.
- [5]P.L. Jones, G.J. Veenstra, P.A. Wade, D. Vermaak, S.U. Kass, N. Landsberger, J. Strouboulis, and A.P. Wolffe, Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 19 (1998) 187-91.
- [6]X. Nan, H.H. Ng, C.A. Johnson, C.D. Laherty, B.M. Turner, R.N. Eisenman, and A. Bird, Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393 (1998) 386-9.
- [7]K. Martinowich, D. Hattori, H. Wu, S. Fouse, F. He, Y. Hu, G. Fan, and Y.E. Sun, DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science* 302 (2003) 890-3.
- [8]G. Fan, K. Martinowich, M.H. Chin, F. He, S.D. Fouse, L. Hutnick, D. Hattori, W. Ge, Y. Shen, H. Wu, J. ten Hoeve, K. Shuai, and Y.E. Sun, DNA methylation controls the timing of astroglialogenesis through regulation of JAK-STAT signaling. *Development* 132 (2005) 3345-56.
- [9]G. Fan, C. Beard, R.Z. Chen, G. Csankovszki, Y. Sun, M. Siniaia, D. Biniszkiwicz, B. Bates, P.P. Lee, R. Kuhn, A. Trumpp, C. Poon, C.B. Wilson, and R. Jaenisch, DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. *J Neurosci* 21 (2001) 788-97.
- [10]L.K. Hutnick, P. Golshani, M. Namihira, Z. Xue, A. Matynia, X.W. Yang, A.J. Silva, F.E. Schweizer, and G. Fan, DNA hypomethylation restricted to the murine forebrain induces cortical degeneration and impairs postnatal neuronal maturation. *Hum Mol Genet* 18 (2009) 2875-88.
- [11]J. Feng, Y. Zhou, S.L. Campbell, T. Le, E. Li, J.D. Sweatt, A.J. Silva, and G. Fan, Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat Neurosci* 13 (2010) 423-30.
- [12]N.W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, and R. Yura, The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem J* 126 (1972) 781-90.
- [13]S. Kriaucionis, and N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324 (2009) 929-30.
- [14]S. Ito, A.C. D'Alessio, O.V. Taranova, K. Hong, L.C. Sowers, and Y. Zhang, Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* (2010).

- [15]E. Privat, and L.C. Sowers, Photochemical deamination and demethylation of 5-methylcytosine. *Chem Res Toxicol* 9 (1996) 745-50.
- [16]Z. Liutkeviciute, G. Lukinavicius, V. Masevicius, D. Daujotyte, and S. Klimasauskas, Cytosine-5-methyltransferases add aldehydes to DNA. *Nat Chem Biol* 5 (2009) 400-2.
- [17]Y. Huang, W.A. Pastor, Y. Shen, M. Tahiliani, D.R. Liu, and A. Rao, The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* 5 (2010) e8888.
- [18]B.A. Flusberg, D.R. Webster, J.H. Lee, K.J. Travers, E.C. Olivares, T.A. Clark, J. Korlach, and S.W. Turner, Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 7 (2010) 461-5.
- [19]D.G. Burke, K. Griffiths, Z. Kassir, and K. Emslie, Accurate measurement of DNA methylation that is traceable to the international system of units. *Anal Chem* 81 (2009) 7294-301.
- [20]R.M. Kok, D.E. Smith, R. Barto, A.M. Spijkerman, T. Teerlink, H.J. Gellekink, C. Jakobs, and Y.M. Smulders, Global DNA methylation measured by liquid chromatography-tandem mass spectrometry: analytical technique, reference values and determinants in healthy subjects. *Clin Chem Lab Med* 45 (2007) 903-11.
- [21]Z. Liu, S. Liu, Z. Xie, W. Blum, D. Perrotti, P. Paschka, R. Klisovic, J. Byrd, K.K. Chan, and G. Marcucci, Characterization of in vitro and in vivo hypomethylating effects of decitabine in acute myeloid leukemia by a rapid, specific and sensitive LC-MS/MS method. *Nucleic Acids Res* 35 (2007) e31.
- [22]Z. Liu, J. Wu, Z. Xie, S. Liu, P. Fan-Havard, T.H. Huang, C. Plass, G. Marcucci, and K.K. Chan, Quantification of regional DNA methylation by liquid chromatography/tandem mass spectrometry. *Anal Biochem* 391 (2009) 106-13.
- [23]E.P. Quinlivan, and J.F. Gregory, 3rd, DNA methylation determination by liquid chromatography-tandem mass spectrometry using novel biosynthetic [U-15N]deoxycytidine and [U-15N]methyldeoxycytidine internal standards. *Nucleic Acids Res* 36 (2008) e119.
- [24]L. Song, S.R. James, L. Kazim, and A.R. Karpf, Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry. *Anal Chem* 77 (2005) 504-10.
- [25]I. Yang, S.K. Kim, D.G. Burke, K. Griffiths, Z. Kassir, K.R. Emslie, Y. Gao, J. Wang, C.A. Foy, A.C. Pardos-Pardos, S. Ellison, P.J. Domann, S. Fujii, and S.R. Park, An international comparability study on quantification of total methyl cytosine content. *Anal Biochem* 384 (2009) 288-95.
- [26]K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, and S. Yamanaka, Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131 (2007) 861-72.
- [27]M. Jackson, A. Krassowska, N. Gilbert, T. Chevassut, L. Forrester, J. Ansell, and B. Ramsahoye, Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Mol Cell Biol* 24 (2004) 8862-71.
- [28]C. Popp, W. Dean, S. Feng, S.J. Cokus, S. Andrews, M. Pellegrini, S.E. Jacobsen, and W. Reik, Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463 (2010) 1101-5.
- [29]J. Deng, R. Shoemaker, B. Xie, A. Gore, E.M. LeProust, J. Antosiewicz-Bourget, D. Egli, N. Maherali, I.H. Park, J. Yu, G.Q. Daley, K. Eggan, K. Hochedlinger, J. Thomson, W.

- Wang, Y. Gao, and K. Zhang, Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotechnol* 27 (2009) 353-60.
- [30]A. Doi, I.H. Park, B. Wen, P. Murakami, M.J. Aryee, R. Irizarry, B. Herb, C. Ladd-Acosta, J. Rho, S. Loewer, J. Miller, T. Schlaeger, G.Q. Daley, and A.P. Feinberg, Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* 41 (2009) 1350-3.
- [31]K. Kim, A. Doi, B. Wen, K. Ng, R. Zhao, P. Cahan, J. Kim, M.J. Aryee, H. Ji, L.I. Ehrlich, A. Yabuuchi, A. Takeuchi, K.C. Cunniff, H. Hongguang, S. McKinney-Freeman, O. Naveiras, T.J. Yoon, R.A. Irizarry, N. Jung, J. Seita, J. Hanna, P. Murakami, R. Jaenisch, R. Weissleder, S.H. Orkin, I.L. Weissman, A.P. Feinberg, and G.Q. Daley, Epigenetic memory in induced pluripotent stem cells. *Nature* 467 (2010) 285-90.
- [32]M. Pick, Y. Stelzer, O. Bar-Nur, Y. Mayshar, A. Eden, and N. Benvenisty, Clone- and gene-specific aberrations of parental imprinting in human induced pluripotent stem cells. *Stem Cells* 27 (2009) 2686-90.
- [33]M. Stadtfeld, E. Apostolou, H. Akutsu, A. Fukuda, P. Follett, S. Natesan, T. Kono, T. Shioda, and K. Hochedlinger, Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* 465 (2010) 175-81.
- [34]T.S. Mikkelsen, J. Hanna, X. Zhang, M. Ku, M. Wernig, P. Schorderet, B.E. Bernstein, R. Jaenisch, E.S. Lander, and A. Meissner, Dissecting direct reprogramming through integrative genomic analysis. *Nature* 454 (2008) 49-55.

**Figure 1-** (A) LC-MS/MS-MRM chromatograms of nucleosides derived from an equal molar mixture of three commercial 948bp standard DNA fragments showing peaks corresponding to the response obtained from gas phase transitions of dC to C, 5hmdC to 5hmC, and 5mdC to 5mC. (B) Standard curves for 5mC and 5hmC. Percent DNA methylation and hydroxymethylation is plotted against the known ratios of methylated or hydroxymethylated DNA to the total pool of cytosine in the standard sample. cps, counts per second.

**Figure 2-** Percentage of 5hmC and 5mC in mESC DNA. 5mC and 5hmC contents are expressed as the percentage of 5mC or 5hmC in the total pool of cytosine. Data are the mean  $\pm$  s.d. from triplicate analyses.

**Figure 3-** Over-expression of human TET1 catalytic domain in 293T cells. (A) 293T cells over-expressing FLAG-catalytic domain of TET1 were co-stained for FLAG antibodies and DAPI. Scale bar, 100  $\mu$ m. (B) Percentage of 5hmC and 5mC in DNA from TET1 transfected cells, mock-transfected and un-transfected cells. Data are the mean  $\pm$  s.d. from triplicate analyses.

**Figure 4-** Reprogramming BJ fibroblast into iPSCs. (A) Percentage of 5mC and 5hmC in DNA from BJ fibroblast and two BJ iPS cell lines, #7 and #8. (B) Southern blot analysis of DNA methylation in BJ and two BJ iPS cell lines. DNA was digested with methyl-sensitive BstBI, separated on agarose gel, transferred to the membrane and hybridized to probes of the repetitive regions of *Sat 2* and *Sat 3*. Small DNA fragments of BJ fibroblast are indicative of DNA hypomethylation in *Sat 2* and *Sat 3* repetitive region. (C) Bisulfite sequencing of *Oct4* and *Nanog* promoter of BJ fibroblasts and two BJ iPS cell lines. Each row represents one clonal analysis and each box represents a



CpG site where the site number is indicated above. The methylation analysis is displayed according to the key. The overall percentage methylation of the gene promoter is indicated for each sample.

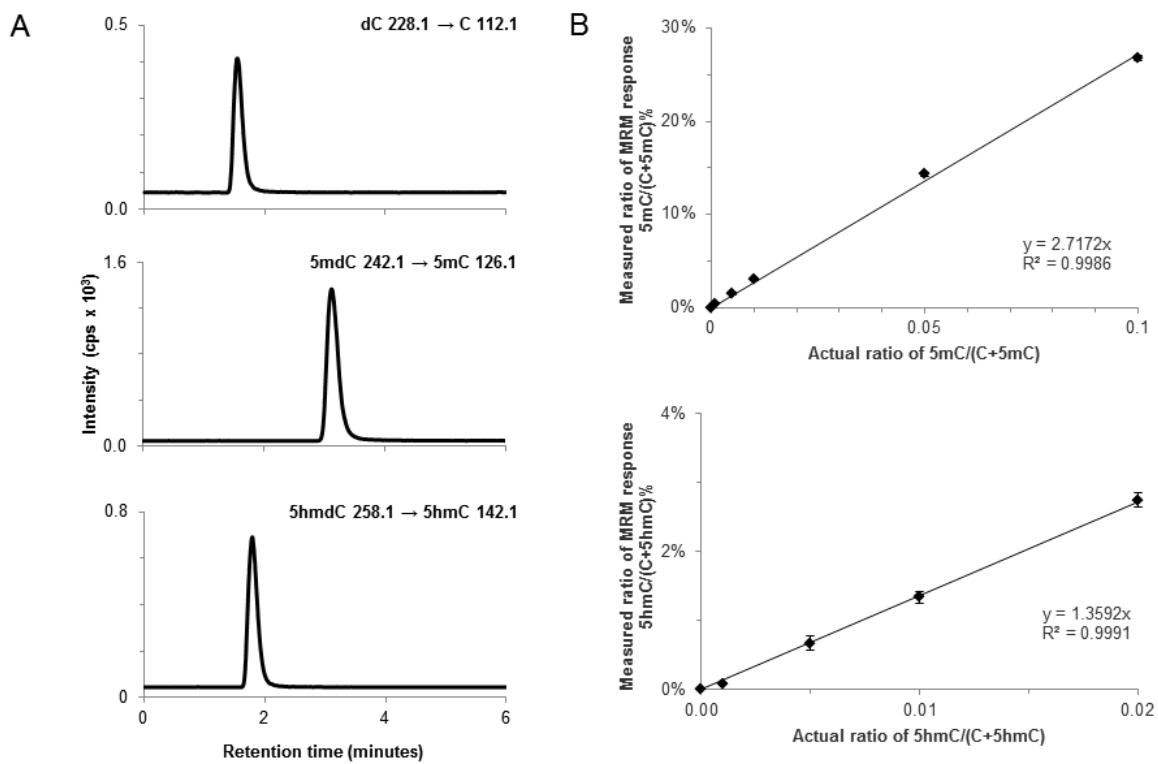
**Supplementary Figure 1-** Individual ion chromatograms of the three nucleosides to base transitions for all three DNA standards used to generate the calibration curve. Value beside the standards for each ion chromatograms is the percent purity of the indicated nucleoside of interest in the total pool of cytosine determined solely on the area of the peaks. Three replicates were done to calculate the purity of the nucleoside.

**Supplementary Figure 2-** Immunostaining characterization of the two BJ iPS cell lines using various ESC markers co-stained with DAPI. Scale bars are indicated in the section.

**Supplementary Table 1-** Primers and its sequence from various experiments.

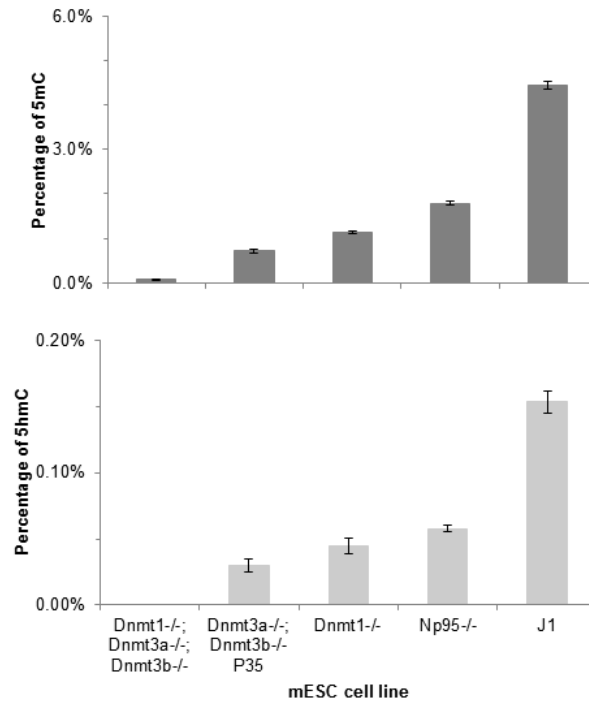
ACCEPTED MANUSCRIPT

Figure 1



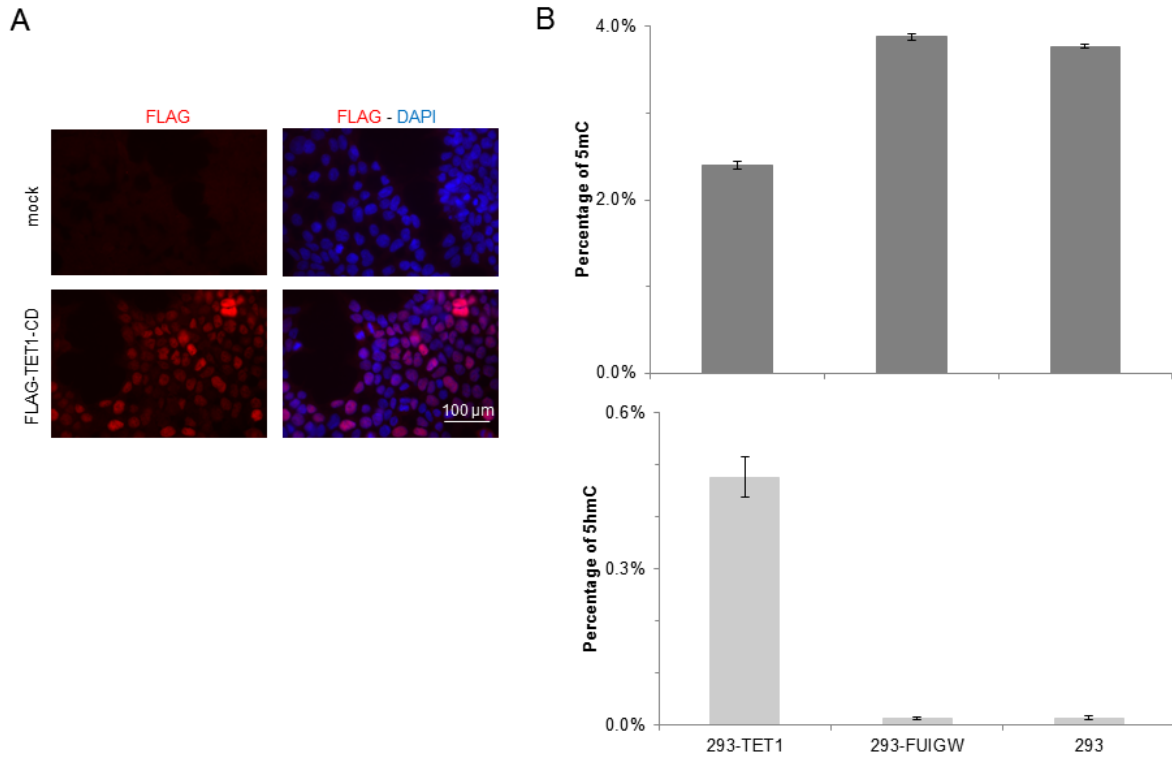
ACCEPTED

Figure 2



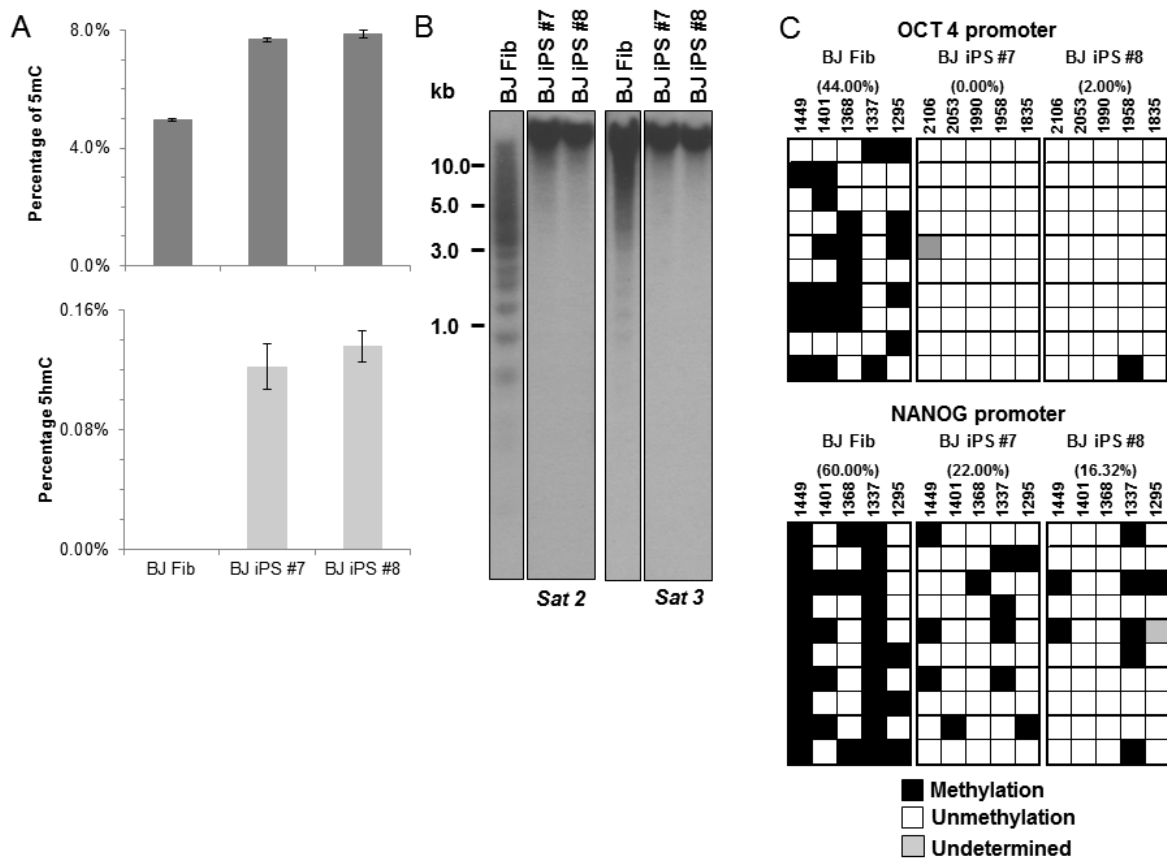
ACCEPTED

Figure 3



ACCEPTED

Figure 4



ACCEPTED