A novel sequencing method for genome-wide profiling of 5-hydroxymethylcytosine with single-base resolution

ZYMO RESEARCH
The Beauty of Science is to Make Things Simple

Darany Tan, TzuHung Chung, Xueguang Sun & Xi-Yu Jia
Zymo Research Corporation, Irvine, CA

Abstract

5-hydroxymethylcytosine (5hmC) is an epigenetic mark abundant in embryo stem cells and brain tissues. The exact biological functions of 5hmC are still under close investigation although several lines of evidence have indicated it could be involved in active DNA demethylation. Meanwhile, extensive studies have been carried out to determine its genomic distribution. A number of approaches have been developed using either affinity based enrichment, such as hMeDIP, that rely on antibody and other specific binding proteins to target 5hmC, or modified bisulfite sequencing, namely oxidative bisulfite sequencing (OxBS) and TET assisted bisulfite sequencing (TAB-sequencing). However, all those methods have limitations which hamper their application. For example, affinity based methods lack single base resolution while modified bisulfite sequencing methods require efficient chemical or enzymatic oxidation which cannot be easily achieved or guaranteed. As an alternative, we have developed a novel genome-wide sequencing method that utilizes an enzyme based modification approach coupled with bisulfite-sequencing for detecting 5hmC. This methodology allows quantification of 5hmC levels with single CpG resolution and can also be employed for locus-specific assays. Using this method, we were able to map and quantify 5hmC sites at the genomic scale for several different biological samples. This novel method can determine the exact location and abundance of 5hmC, which will facilitate our understanding of 5hmC in regulating gene expression in different biological contexts.

Introduction

The importance of DNA methylation in the epigenetic regulation of genes have been well established, but the biological functions of the “sixth base”, 5hmC, have been less clear. Here, we introduce a novel method, Mirror Bisulfite Sequencing, for detecting 5hmC with single CpG resolution. This method utilizes highly efficient, specific enzymes, β-glucosyltransferase (βGT) and M.SssI methylase, coupled with bisulfite conversion to determine the hydroxymethylation status of a single CpG site. We used this method for genome-wide 5hmC profiling of the human brain.

Methodology

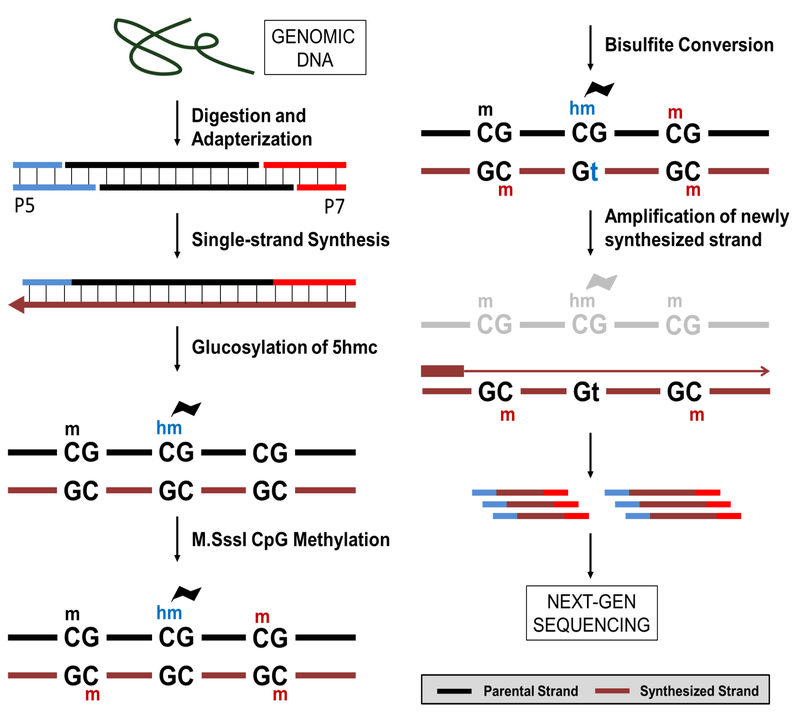


Figure 1. Overview of Mirror Bisulfite-Sequencing. This method detects 5-hydroxymethylcytosine in the CpG context by synthesizing a single strand to mirror the parental strand. Glucosylation of 5hmC on the parental strand would block M.SssI methylation of CpGs. The DNA is then subjected to bisulfite conversion, and the synthesized strand is selectively amplified. Synthesized CpGs complementary to parental 5hmC sites would be sequenced as thymine whereas CpGs mirroring non-5hmCs would be methylated, protected from conversion, and sequenced as a cytosine.

Detection of 5hmC sites

The presence of 5hmCs can be detected by using a single strand synthesis to mirror the parental strand. After enzymatic reactions and bisulfite conversion, the complementing CpG on the synthesized strand will be read as a thymine. Those mirroring non-hydroxymethylated CpG sites would remain as a cytosine.

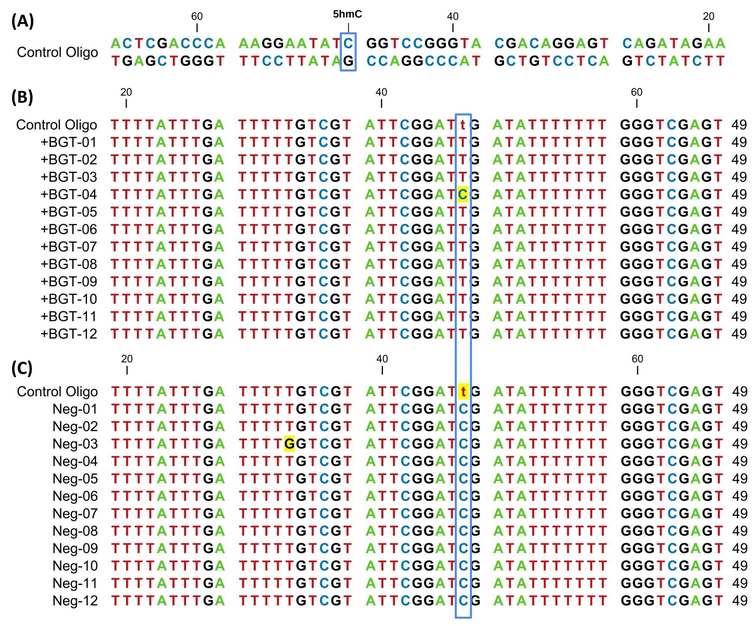


Figure 2. Detection of a single hydroxymethylated CpG site. (a) An oligo with a single 5hmC site was treated using the described method. (b) Glucosyl-5hmC blocked methylation of the complementary CpG site, so the cytosine was converted in 11 of 12 sequencing reads. (c) The negative βGT control had complete CpG methylation, so the complementary CpG site remained as a cytosine.

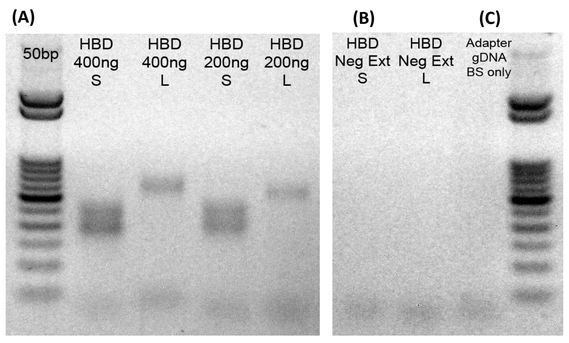


Figure 3. Genomic libraries from human brain DNA. (a) Libraries can be prepared from 200ng DNA. Controls are included to detect background levels due to parental strand amplification. (b) One control omits the strand extension step and (c) the second control shows that the parental strand cannot be amplified after bisulfite conversion. S = 150-250bp fragments L = 250-350bp fragments

Genome-wide 5hmC Profiling

	Total Reads	Mapped Reads	Bisulfite Efficiency	Coverage	# of CpGs
+βGT 1	24,824,751	15,448,763	99.21%	11X	6,404,742
+βGT 2	23,279,812	14,271,016	99.19%	10X	6,179,758
-βGT Control	31,600,025	18,637,460	99.21%	13X	6,573,224

Table 1. Brief overview of sequencing results and quality of libraries. Libraries were prepared in parallel from human brain DNA. All samples were sequenced using Illumina HiSeq 2000 platform.

Chrom	Position	Negative: CpG Methylation	Negative: # of reads	BGT 1: CpG Methylation	BGT 1: # of Reads	BGT 2: CpG Methylation	BGT 2: # of Reads	Avg 5hmC Level
chr15	75470593	1	20	0.2727	11	0.3571	14	0.7567
chr14	100216578	1	28	0.2632	19	0.2353	17	0.7508
chr14	99891229	1	14	0.3571	14	0.2143	14	0.7143
chr6	41650681	1	38	0.2222	18	0.3500	20	0.7112
chr19	14545419	1	11	0.3636	11	0.2353	17	0.7005
chr10	80971605	1	14	0.3000	10	0.3333	18	0.6833
chr3	177077229	1	27	0.4000	15	0.4000	10	0.6778
chr14	105619457	1	14	0.3333	15	0.2500	12	0.6627
chr6	43251531	1	39	0.6667	18	0.2353	17	0.6327
chr10	114900905	1	10	0.4375	16	0.3000	10	0.6313
chr6	41650673	1	38	0.6111	18	0.4000	20	0.6237
chr17	55963183	1	10	0.4286	14	0.3333	12	0.6190
chr19	970807	1	15	0.3333	12	0.4286	14	0.6190

Table 2. CpG sites with high hydroxymethylation levels in the human brain. Two +BGT libraries were prepared in parallel and sequenced. The negative control has 100% methylation, indicating the CpG site has little background contributing to the hydroxymethylation levels detected in the samples.

Mapping of 5hmCs and SNPs

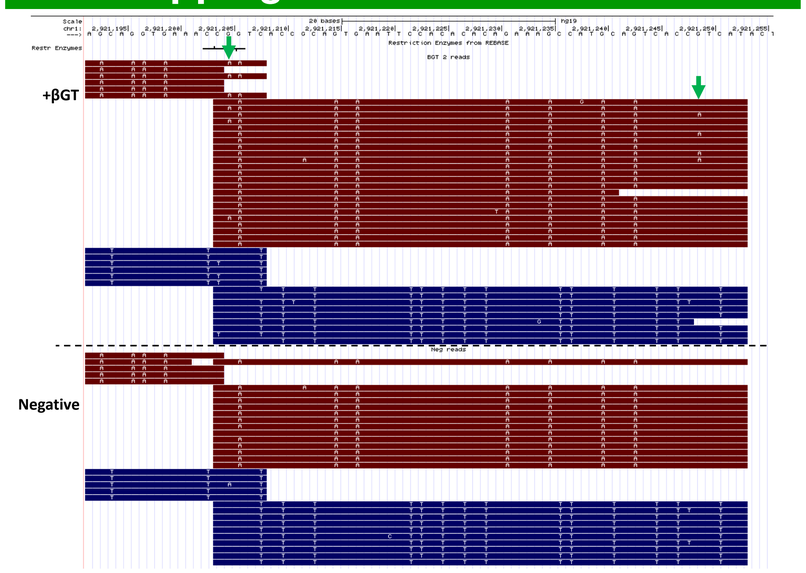


Figure 4. 5hmC mapping on the hg19 UCSC genome browser for chr1:2,921,203. The presence of 5hmC at CpG sites are indicated by the presence of a thymine on the + strand (blue) or adenine on the – strand (red). Sequencing of the negative control shows there is little background contributing to the quantitation of 5hmC levels.

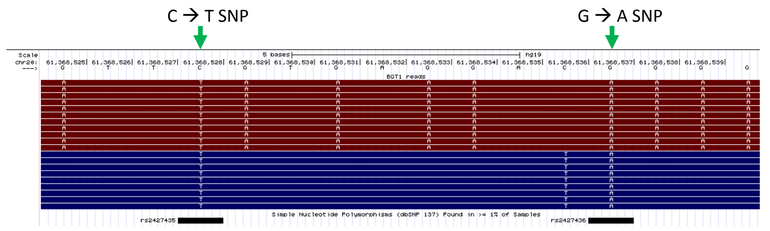
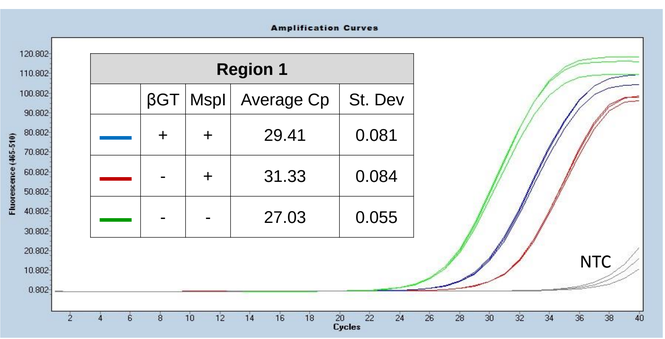


Figure 5. Sequencing results can be analyzed to identify SNPs. The presence of a SNP can result in false positives, but it can be filtered and verified by available SNP data or PCR analysis.

Loci Validation

Region	Chrom	Position	Msp I Site?	BGT1: CpG Methylation	BGT1: # of reads	BGT2: CpG Methylation	BGT2: # of reads	Avg 5hmC Level
1	chr1	914570	Y	0.8750	16	0.8462	13	0.1394
	chr1	914599	Y	0.7931	29	0.8667	30	0.1134
2	chr1	2921203	Y	0.8667	15	0.7333	15	0.2000



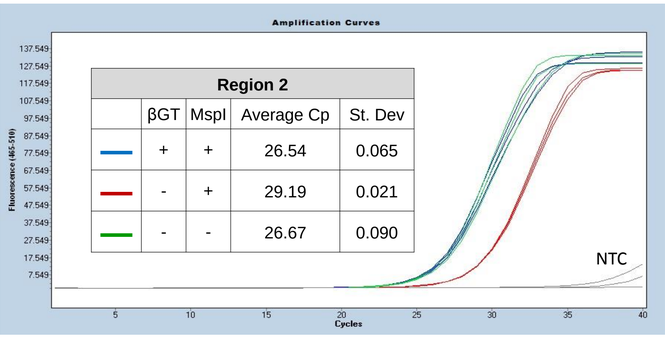


Figure 6. Loci validation of regions showing clustered 5hmC sites using qPCR analysis. Genomic DNA was treated with and without βGT and subjected to MspI digestion. The presence of a glucosyl-5hmC blocks cleavage, and the loci can be amplified. Differences in Cp values between the glucosylated and unglucosylated samples indicates the presence of a 5hmC. The untreated DNA establishes the level of complete hydroxymethylation at the site.

Conclusions

Our Mirror Bisulfite-Sequencing method offers quantitative resolution of hydroxymethylated CpG sites across the genome with the use of highly efficient, commercially available enzymes and without oxidative chemical reactions. Libraries can be constructed using as low as 200ng of genomic DNA. In addition, sequencing results can be analyzed to identify single-nucleotide polymorphisms. Mirror Bisulfite-Sequencing can also be adapted for whole-genome sequencing as well as locus-specific analysis.