International Journal of Biochemistry and Biotechnology ISSN: 2167-0404 Vol. 2 (6), pp. 449-456, June, 2013. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Full Length Research Paper

High quality bisulfite sequencing using nanogram amounts of genomic DNA

Jihua Sun[1] Honglong Wu[1] Guanyu Ji[1] Boxin Wu Shujing Yan Wen Gao Rena Lam Wenwei Zhang Xiuqing Zhang*

Beijing Genomics Institute, Shenzhen 518000, China
[1]These authors contributed equally to this work

Accepted 5 January, 2013

Multiple methods have been developed to decipher the genome-scale DNA methylation (Methylome). Bisulfite sequencing (BS-seq), which combines sodium bisulfite conversion with high throughput sequencing, can measure DNA methylation at single-nucleotide resolutions. However, it normally needs large amounts of genomic DNA (5-10μg) to start with, thus creating an obstacle for it to be widely used. Here we optimized the normal BS-seq method for generating high quality bisulfite sequencing on whole genomes using nanogram DNA (MBS-seq). Systematic comparison the whole genome methylation study based on minute amount of genomic DNA through next generation sequencing technology.

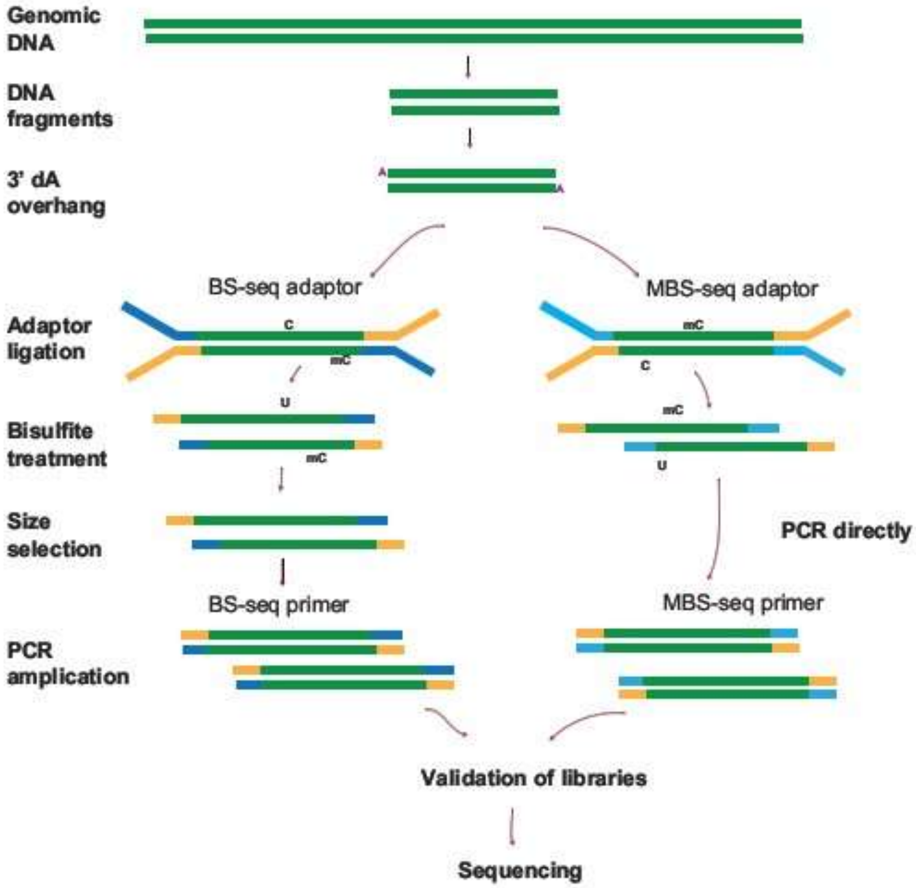
Key words: DNA Methylation, Whole Genome Bisulfite Sequencing, Nanogram Amounts of Genomic DNA.

INTRODUCTION

As one of the most intensely studied epigenetic mechanism, DNA Methylation plays a significant role in regulating gene expression. In mammalian somatic tissue, DNA methylation occurs almost exclusively (99.98%) at CpG dinucleotides. In plants, it occurs in CG, CHG- and CHH-contexts (Bernstein et. al., 2007.Bird, 1986). However, In human embryonic stem cells, it has been shown that 25% of all 5mC occurs in the CHG and CHH contexts (Lister et. al.,2009). 5-hydroxymethylcytosine (5hmC), be called sixth base, also been reported widely (Jin et. al., 2010). More studies began to focusing on the balance between hydroxymethylation and methylation in the genome and its molecular mechanism (Ficz et. al., 2011.Jin et. al., 2011). But the importance of its exact function in the genome remains largely unknown (Ku et. al., 2011). Multiple methods have been developed to detect methylation sites from a certain regions to whole genome. However, the method needs large amounts of DNA (5-10μg) which is the biggest challenge for BS-seq

*Corresponding author. E-mail: zhangxq@genomics.org.cn

Figure 1. General pipeline of normal BS–seq (left) and MBS-seq (right) library construction protocol including all key steps.

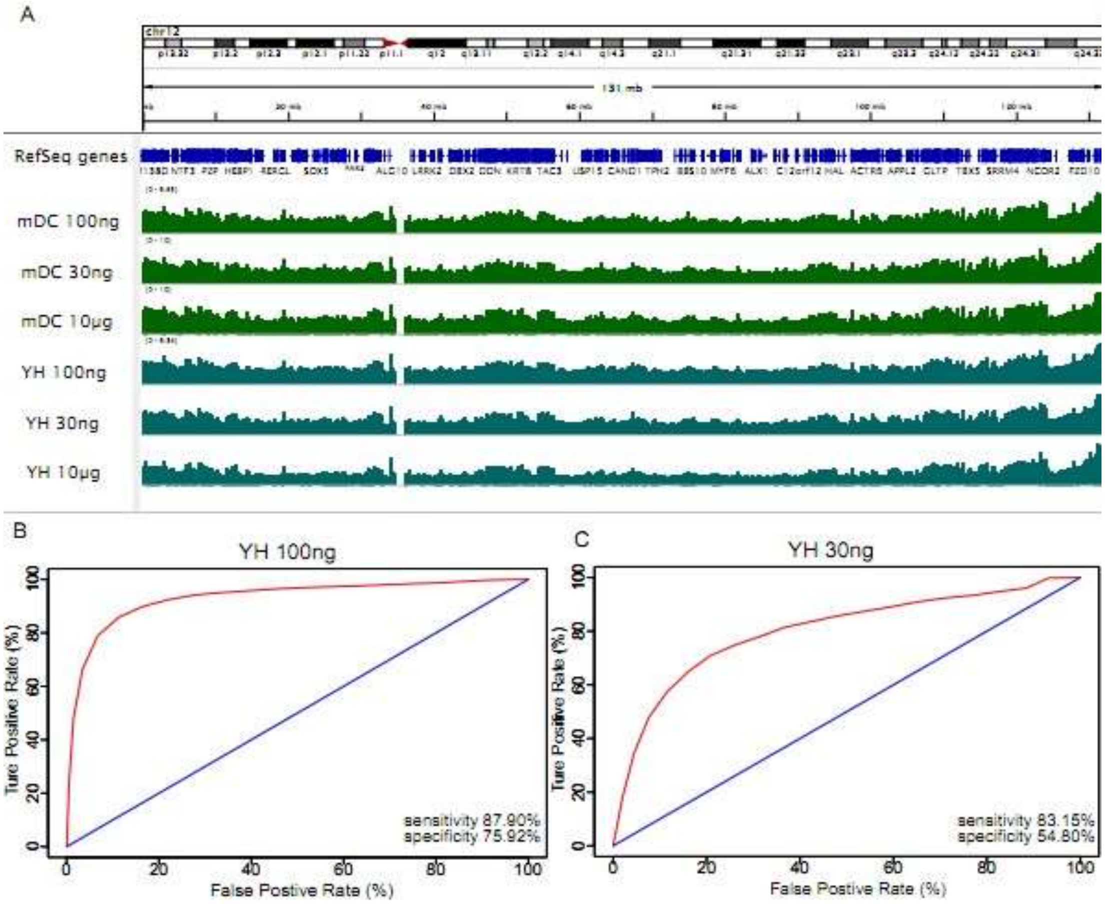


(Bock et. al., 2010. Harris et. al., 2010.Laird, 2010). Many studies have tried to use minute genomics DNA to construct the libraries. Most of them choose capturing methods, which Include, Reduced Representation Bisulfate Sequencing (RRBS) (Meissner et. al., 2005), it enriches CG-rich parts of the genome, and the similar strategy be used in DNA immunoprecipitation (MeDIP-seq) (Weber et. al., 2005). Most of those methods indeed reduce the starting amount of DNA to a very low level (Laird, 2010). But the main disadvantage of those methods is that it is not a whole genome scale mapping. There are also approach which have used Transposase-based method to fragment the genome DNA to construct minute bisulfite libraries (Adey and Shendure,2012) , but as its needs to use the transposase modification, which may affect the randomness of the data, at the same time the efficiency of transposase fragmenting DNA also needs to be further optimized. In order to provide a simple and repeatability minute methylation sequencing method, here we report the MBS-seq (Minute DNA Bisulfite Sequencing) method, which adapted the normal BS-seq and can generate whole genome bisulfite sequencing libraries by using a minimum of 30ng genomic DNA.

MATERIAL AND METHODS

In this study, two kinds of genomic DNA, Peripheral Blood Mononuclear Cells (PBMCs) and a human Immunocyte cell line (human macrophages, name mDC) were used as comparison to illustrate the MBS method, both of them have the standard methylome data (from normal BS-seq).

Figure 2. The distribution of Methylation level on the whole genome. We chose top 3,000 genes of normal BS-seq library with higher methylation as ‘positive’ and generate the operating characteristic curve (ROC) of YH 100ng and YH 30ng library data. (A) The distribution of methylation level (methylated reads / methylated reads + unmethylated reads) of cytosin on chromosome 12. The step of each window is 500bp. (B) ROC of YH 100ng data. (C) ROC of YH 10 ng data.



Peripheral blood was obtained from the same individual from the YanHuang (YH) project (Wang et. al., 2008). YH is an Asian male whose whole genome has been resequenced previously by using genetic material from PBMCs (Li et. al., 2010). The mononuclear cells were separated through Ficoll-Paque (GE Healthcare) gradient centrifugation. The total DNA was prepared by Proteinase K/Phenol extraction.

The human immunocyte cells were prepared from peripheral blood monocytes and separated by Ficoll of healthy donors followed as previously described (Krause et. al., 1996). To generate macrophages, 1×10^6 monocytes/ml were seeded in RPMI 1640 medium (HyClone) supplemented with 2% human pooled AB-group serum (Cambrex IEP GmbH, Wiesbaden, Germany) and cultured on teflon foils. The DNA of macrophages cell lines were prepared by following the methods described in Smith et al (Smith et. al., 2009)

The methylome data of PBMCs was generated as part of the YH DNA methylome project (Li et. al., 2010). It is used as a standard reference for YH MBS-seq data in this study. The methylome of the immunocyte line (in press) is used as the standard reference to compare the data of mDC MBS-seq library. The methylome data of YH and mDC was used as the standard reference to compare with MBS-seq data, except the subject of CpG coverage and CpG bias analysis.

As the study involves using human biological samples, the study was first reviewed and approved by the Institutional Review Board at the Beijing Genomics Institute, and conducted after receiving said approval.

Reagents and equipments used for normal BS-seq and

Table 1: Reagents used in experiments.

Reagents	Catalog No.	Supplier
10×polynucleotide kinase buffer	B904	Enzymatics
dNTP solution set	N201L	Enzymatics
T4 DNA polymerase	P708L	Enzymatics
Klenow enzyme	P706L	Enzymatics
T4 polynucleotide kinase	Y904L	Enzymatics
10× blue buffer	B011	Enzymatics
1 mM dATP		Enzymatics
Klenow (3’-5’ exo-)	P701-LC-L	Enzymatics
2×Rapid ligation buffer	B101	Enzymatics
T4 DNA ligase (rapid)	L603-HC-L	Enzymatics
PE adaptor oligo mix		Takara/IDT/Illumina oligo kit
Index PE adaptor oligo mix		Takara/IDT/Illumina oligo kit
PCR primer PE 1.0		Takara/IDT/Illumina oligo kit
PCR primer PE 2.0		Takara/IDT/Illumina oligo kit
Index PCR primer PE		Takara/IDT/Illumina oligo kit
50 bp ladder marker	MD108-01	TIANGEN
DL2000 marker	MD114-02	TIANGEN
λ-HindⅢ marker	D3403A	TaKaRa
50bp DNA ladder	N3236L	NEB
EZ DNA Methylation-Gold kitTM	D5005	ZYMO
Unmethylated lambda DNA	D1521	Promega
JumpStartTM Taq DNA polymerase	D9307	Sigma
Quant-iTTM dsDNA HS Assay kit	Q32851	Invitrogen
QIAquick Gel Extraction kit	28706	Qiagen

Table 2. Equipments used in experiments

Equipments	Catalog No.	Supplier
PCR thermal cycler	Veriti thermal cycler	ABI
Agilent 2100	2100	Bioanalyzer Agilent
NanoDrop 1000	Spectrophotometer	Thermo Fisher Scientific
Covaris sonication system	S-2	Covaris
Thermomixer	Thermomixer comfort	Eppendorf
Centrifuge	5417R	Eppendorf
Qubit Invitrogen		Invitrogen

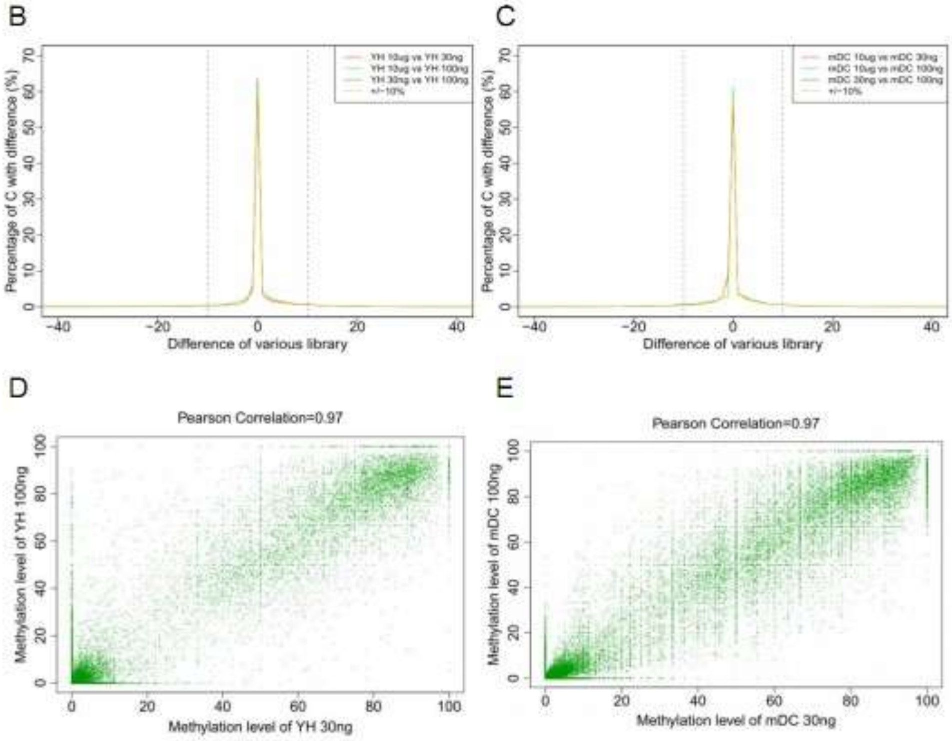
Table 3. Sanitation parameters to fragment the genomic DNA into 100- 400 bp.

		Normal BS-seq	MBS-seq
Treatment 1	Duty/cycle (%)	10	5
	Intensity	5	5
	Cycle/burst	200	200
	Time (s)	60	60
Treatment 2	Time (s)	0	0
Treatment 3	Time (s)	0	0
Treatment 4	Time (s)	0	0
Cycle		10	6

Figure 3. Validation of the methylation rate of MBS-seq libraries. (A) Proportion of methylation level's difference less than 10%. (B-C): Distribution of methylation level's difference of YH and mDC libraries. (D-E): Pearson correlation between MBS-seq dataset of YH and mDC

A

	Difference lower than -10%	Difference in (-10%~10%)	Difference larger than 10%
YH 10 μg vs YH 100ng	4.02	94.33	1.65
YH 10 μg vs YH 30ng	4.42	94.19	1.39
YH 30ng vs YH 100ng	0.72	97.54	1.73
mDC 10 μg vs mDC 100ng	1.24	96.53	2.23
mDC 10 μg vs mDC 30ng	1.47	96.24	2.29
mDC 30ng vs mDC 100ng	1.06	97.98	0.96



MBS-seq library preparation in this study are listed in Table 1-2.

For normal BS-seq, 10μg genomic DNA is fragmented by using the Covaris S2 sonication system; the parameters listed in Table 3. Following fragmentation, libraries were constructed by conducting the Illumina Paired-End protocol as described in Li et al (Li et. al.,2010) (Figure 1).

For MBS-seq library, it started with 30ng and 100ng of genomic DNA of YH and mDC respectively. First, the DNA is fragmented by Covaris S2 at the same parameters of normal BS-seq with shorter time (Table 4). After fragmentation, DNA end repair, and <A> base addition, after that the amount of DNA is quantified. Then based on the amount of DNA, the corresponding amount of MBS-seq adaptor is added (sequences information summarized in Table 5). After adaptor ligation, 200ng fragmented unmethylated λ DNA is added into the ligated DNA to undergo bisulfite treatment (EZ DNA Methylation-Gold kit, ZYMO) together. PCR amplification is done directly after bisulfite conversion.

The PCR was carried out in a final reaction volume of 50μL consisting of 20 μL purified DNA, 4 μL 2.5 mM dNTP, 5 μL 10X buffer, 0.5 μL JumpStart™ Taq DNA polymerase, 2μL PCR primers and 18.5 μL MltraPureTM Water, and the following thermal cycling program: 94℃

Table 4. The amount of DNA after <A> base addition and the corresponding amounts of MBS-seq adaptor added.

	30 ng	100 ng
Amounts of DNA after <A> base addition	5.2 ng	20.4 ng
MBS-seq adaptor (0.4μM)	1.5μl	6μl

Table 5. Sequences information of adaptor and PCR primers.

Normal BS-seq adaptor	R	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
	F	P-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Minute MBS-seq adaptor	R	TACACTCTTTCCCTACACGACGCTCTTCCGATCT
	F	P-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
Normal BS-seq primer	F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
	R	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
Minute BS-seq primer	F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
	R	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Table 6: Data production details of normal BS-seq and MBS-seq libraries.

Sample	YH 10μg	YH 100ng	YH 30ng	mDC 10μg	mDC 100ng	mDC 30ng
Raw Data (Gbp)	103.7	2.52	2.31	10.36	2.06	1.35
Reads Number (M)	1,587	57.32	46.34	138.15	42.09	27.63
Uniq Mapped reads (M)	1,035	36.16	30.94	95.83	28.18	18.99
Uniq Mapped bases (M)	64.38	1.59	1.5	7.19	1.38	0.93
Mapped rate (%)	62.08	63.08	66.77	69.37	66.97	68.72
Conversion Rate (%)	99.71	99.12	99.44	99.36	99.3	99.19
Duplication rate (%)	*5.76	2.3	2.67	4.66	1.69	1.83

30s, 11cycles for 100ng or 13cycles for 30ng, 60℃ 30s, 72℃ 30s then prolong with 1 min at 72℃. PCR products were purified and 160bp-300bp range was excised from 2% TAE agarose gel to get the library (Figure 1).

RESULTS

High-throughput Pair-end sequencing was done by Illumina's GAII sequencer. To determine the reproducibility and reliability of this method, two independent experiments using YH and mDC sample in a scatter analysis were compared. The sequence data were sorted according to the different DNA quantity as 10μg, 100ng, 30ng (Table 6). The raw data is processed by Illumina's base-calling software, i.e.Pipeline. Subsequently all reads were aligned to the reference genome hg18 with SOAP Aligner (Li et. al.,2008) and methyl cytosines were identified according to published strategy by Li et al (Li et. al.,2010).

The concept of methylation level was calculated by 100*reads/total reads of certain cytosine types was co-

Table 7. Details of 50M data of normal BS-seq libraries.

Sample	YH 10μg	mDC 10μg
Conversion rate	99.49	99.36
Reads Number(M)	45.12	49.92
Raw Data(Gbp)	1.99	2.2
Uniq Mapped reads(M)	32.56	35.05
Uniq Mapped bases(M)	1.39	1.54
Mapped rate(%)	72.17	70.22
Duplication rate(%)	5.76	1.83

vered. Considering not all of unmethylated cytosine would be converted into uracil after bisulfite treatment. Thus the conversion rate is very important for the bisulfite sequencing. In this study we use the cytidine conversion rate on non-CpG as the conversion rate, and in all libraries, the conversation rate is larger than 99% (see Table 6-7). Moreover, high quality raw data with a high mapped rate and low duplication rate (less than 10%) were produced (see Table 6).

To assess the MBS-seq method on subject of CpG coverage and C, G bias form PCR amplification, we randomly selected 50M reads form normal BS-seq (see Table 7) to compare the corresponding amount data of MBS-seq. The coverage of CpG on each chromosome reveal that reads distributed uniformly on genome in each sample, as we anticipated that chromosome X and Y had lower coverage due to less abundance than euchromosome.(see Supplementary Figure 1). Furthermore, to make sure there were no bias between C or G in MBS-seq compare with normal libray BS-seq, we also checked the reads distribution with different CpG $_{\text{observe/expect}}$ (CpG$_{\text{o/e}}$) (Gardiner-Garden and Frommer, 1987). There was no obvious bias between the BS-seq and MBS-seq libraries (see Supplementary Figure 2), which also illustrated the MBS-seq library has no bias on GC content.

We detected the MBS-seq methylation rate repeatability of 37,576 genes from hg18 (from UCSC, 20110418). The visual analysis of the methylation level distribution suggested MBS-seq dataset has a good concordance to normal BS-seq dataset (Figure 2A). Considering the normal BS-seq data as a golden standard, we chose top 3000 genes with high methylaiton levels in normal BS-seq dataset defined as ‘positive’, estimated that the MBS-seq dataset had a sensitivity of ~80% at a specificity of ~50%, and true positive and false positive rate for MBS-seq dataset were calculated at varying cutoff values for the reveiver operating curve (ROC) analysis (see Figure 2 B-C and Supplementary Figure 3). This provides evidence that the methyaltion level of MBS-seq had a well reproducibility compare to normal BS-seq libraries.

Then, we compared the methylation rate of CG with depth more than 4X between normal BS-seq and MBS-seq to see the concordance between these different DNA quantities. The cytosine’s methylation level in the MBS-seq dataset was consistent with normal BS-seq (Figure 3). Even though there were a few biases which have different level between MBS-seq and normal BS-seq, most of the differences were less than 10% (Figure 3 A-C).

There is also high correlation of methylation level of single CpG site between different dataset. The correlation was 0.97 within the following sets: YH 10μg and YH 100ng, YH 10μg and YH 30ng, YH 100ng and YH30ng (see Figure 3D and Supplementary Figure 4A-B). A high correlation also exists in mDC sample: 0.94 between mDC 10μg and mDC 100ng; 0.96 between mDC 10μg and mDC 30ng; 0.97 between mDC 100ng and mDC 30ng (see Figure 3E and Supplementary Figure 4C-D). It is predicts the randomness of sequencing error mainly contribute to the small difference in varied dataset.

DISCUSSION

The use of large quantity of starting DNA (5-10ug) resulted the study of BS-seq cannot be widely conducted as it struggles to get large amounts DNA to be sequenced. In MBS-seq method we optimize the normal BS-seq adapter (Table 5**)** to improve the amplification efficiency, and we did PCR directly after adapter ligation, in order to avoid the DNA lost in gel cutting step. Adding corresponding adapter to ligation system is very important as there is no gel cutting before PCR amplification (Table 4).

Beyond our method there are also methods which use WGA (Whole Genome Amplification) in bisulfite sequencing. It uses bisulfite modification DNA as template to do WGA amplification, in order to get enough DNA to construct libraries. The huge defect of WGA bisulfate sequencing is that the bisulfite modification DNA is not suitable for genome scale amplification. Because the templet is more complicated after the bisulfite modification than normal DNA (no methylated C change into U, and degradation seriously). On the other hand, the high bias in WGA reaction also affects the accuratce of methylated level detection.

To characterization of DNA methylation of a single genome by bisulfite sequencing currently requires around 8 lanes of illumina platform. Follow the new MBS-seq methods we can generate library from 30ng genomic DNA to sequence. That means we need around 100ng genomic DNA to do a Methylome mapping research. But it is also different for some research to gather 30-100ng genomic DNA. In this way, to absolutely resolve the minute DNA quantity for the epigenetic research, we also need a long way to go.

In conclusion, in this study we report an optimized normal BS-seq protocol for generating the high quality bisulfite sequencing on whole genome with nanogram amounts of genomic DNA (MBS-seq). MBS-seq libraries have been successfully constructed by using two different sources of genomic DNA, ranging from 30ng to 100ng (YH and mDC). Validation of this new method by comparing the MBS-seq libraries (30ng and 100ng) to normal BS-seq libraries (10μg) by associating data quality, the methylation distribution, and the correlation of methylation level. The high correlation between normal and minute libraries proves that the new MBS-seq allows unambiguous mapping of the whole genome DNA methylation from 30ng genomic DNA. The new method will be widely accepted and can work very well on clinical samples with nanogram DNA.

ACKNOWLEDGMENTS

We would like to thank the faculty and staff from BGI (Beijing Genomics Institute) in Shenzhen. Their names are not included in the author list but they contributed to the large-scale sequencing of this work. We would also like to thank Dr. Qian Zhang for supplying the mDC cell line. This work was supported by the Chinese 863 program (2012AA02A201) and Guangdong Key Laboratory of Human Disease Genomics and Shenzhen Key Laboratory of (CXB20093110066A].

Competing interests

The authors declare that they have no competing interests.

REFERENCES

Adey Shendure J (2012). Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. Genome Res. 22: 1139-1143.

Bernstein BE, Meissner A, Lander ES (2007). The mammalian epigenome. Cell. 128: 669-681.

Bird AP (1986). CpG-rich islands and the function of DNA methylation. Nature. 321: 209-213.

Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Gnirke A, Stunnenberg HG, Meissner A (2010). Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol. 28: 1106-1114.

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature. 473: 398-402.

Gardiner-Garden M, Frommer M (1987). CpG islands in vertebrate genomes. J Mol Biol. 196: 261-282.

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y, Olshen A, Ballinger T, Zhou X, Forsberg KJ, Gu J, Echipare L, O'Geen H, Lister R, Pelizzola M, Xi Y, Epstein CB, Bernstein BE, Hawkins RD, Ren B, Chung WY, Gu H, Bock C, Gnirke A, Zhang MQ, Haussler D, Ecker JR, Li W, Farnham PJ, Waterland RA, Meissner A, Marra MA, Hirst M, Milosavljevic A, Costello JF (2010). Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol. 28: 1097-1105.

Jin SG, Kadam S, Pfeifer GP (2010). Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Research. 38: e125-e125.

Jin SG, Wu X, Li AX, Pfeifer GP (2011). Genomic mapping of 5-hydroxymethylcytosine in the human brain. Nucleic Acids Res. 39: 5015-5024.

Krause SW, Rehli M, Kreutz M, Schwarzfischer L, Paulauskis JD, Andreesen R (1996). Differential screening identifies genetic markers of monocyte to macrophage maturation. J Leukoc Biol. 60: 540-545.

Ku CS, Naidoo N, Wu M, Soong R (2011). Studying the epigenome using next generation sequencing. J Med Genet.

Laird PW (2010). Principles and challenges of genome-wide DNA methylation analysis. Nature Reviews Genetics. 11: 191.

Laird PW (2010). Principles and challenges of genomewide DNA methylation analysis. Nat Rev Genet. 11: 191-203.

Li N, Ye M, Li Y, Yan Z, Butcher LM, Sun J, Han X, Chen Q, Zhang X, Wang J (2010). Whole genome DNA methylation analysis based on high throughput sequencing technology. Methods. 52: 203-212.

Li R, Li Y, Kristiansen K, Wang J (2008). SOAP: short oligonucleotide alignment program. Bioinformatics. 24: 713-714.

Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, Zheng H, Yu J, Wu H, Sun J, Zhang H, Chen Q, Luo R, Chen M, He Y, Jin X, Zhang Q, Yu C, Zhou G, Huang Y, Cao H, Zhou X, Guo S, Hu X, Li X, Kristiansen K, Bolund L, Xu J, Wang W, Yang H, Wang J, Li R, Beck S, Zhang X (2010). The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol. 8: e1000533.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 462: 315-322.

Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 33: 5868-5877.

Smith ZD, Gu H, Bock C, Gnirke A, Meissner A (2009). High-throughput bisulfite sequencing in mammalian genomes. Methods. 48: 226-232.

Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, Ma L, Li G, Zhang G, Yang B, Yu C, Liang F, Li W, Li S, Ni P, Ruan J, Li Q, Zhu H, Liu D, Lu Z, Li N, Guo G, Ye J, Fang L, Hao Q, Chen Q, Liang Y, Su Y, San A, Ping C, Yang S, Chen F, Li L, Zhou K, Ren Y, Yang L, Gao Y, Yang G, Li Z, Feng X, Kristiansen K, Wong GK, Nielsen R, Durbin R, Bolund L, Zhang X, Yang H (2008). The diploid genome sequence of an Asian individual. Nature. 456: 60-65.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet. 37: 853-862.