Fractionation of Differentially Methylated Regions with the MethylMiner™ Methylated DNA Enrichment Kit and Deep Sequencing with the SOLiD™ System

Introduction

DNA methylation is an important epigenetic modification involved in the remodeling of chromatin structure and ultimately in the control of gene expression. Proper control of DNA methylation patterns is essential for normal embryonic development and tissue differentiation [1], X-chromosome inactivation [2], and gene imprinting [3]. Aberrant methylation has been linked to many diseases including cancers [4]. The modified base 5-methylcytosine constitutes about 1% of all DNA bases in mammalian genomes. Although the presence of 5-hydroxymethylcytosine in brain [5] and cytosine methylation at non-CpGs (namely

CpNpG in early embryogenesis) have been reported [6], mammalian DNA methylation is found almost exclusively in the symmetrical CpG dinucleotide. While DNA methylation in CpG-rich promoter regions correlates with transcriptional silencing, it is becoming clear that CpG methylation in regions of lower CpG density and distal to promoters correlate with gene expression [7,8]. Thus, it would be quite useful to effectively partition densely methylated regions from moderately methylated regions in order to better understand the physiological role of this DNA chemical modification.

To fully understand the role of DNA methylation in normal and disease states, it is important to first determine genomic methylation patterns. Methods of characterizing methylation are generally based upon one of three techniques: bisulfite conversion, digestion with methylation-sensitive restriction enzymes, and antibody or 5-methylcytosine binding protein–based purification of methylated DNA [9]. The advent of next-generation sequencing platforms enables ultra–high-throughput sequencing, mapping, and counting of short DNA reads (tags) and, in combination with any of the above methylation profiling strategies, could be used for comprehensive, genome-wide mapping of methylation sites.

Methylated cytosines constitute a small percentage of the bases in mammalian genomes. In the human genome, for example, only about 1% of the bases are 5-methylcytosine. It is in many cases neither cost effective nor necessary to sequence the entire genome with deep coverage in order to interrogate the genomic methylation patterns. Thus, sensitive, accurate, and versatile tools for the enrichment of methylated sequences from the genome are highly useful for genome-wide studies of methylation patterns.

Different methods for selective enrichment of methylated genomic DNA fragments are available. Those that are based upon antibodies are termed methylated DNA immunoprecipitation (MeDIP) [10]. Other methods utilize proteins that naturally recognize methylated cytosines in double-stranded DNA. In the case of the MethylMiner™ Methylated DNA Enrichment Kit, the capture medium is the methyl-CpG binding domain (MBD) of the human MBD2 protein coupled to superparamagnetic Dynabeads® M-280 Streptavidin via a biotin linker. This application note describes a genome-scale study on the patterns of DNA methylation in the MCF-7 breast cancer cell line, using MethylMiner™ Methylated DNA Enrichment Kit followed by sequencing on the SOLiD™ System.

Methods
Comparison of methylated DNA enrichment with the MBD-based MethylMiner™ kit to MeDIP identifies a number of workflow advantages to using the MethylMiner™ kit (Figure 1). First, with antibody-based enrichment, the fragmented DNA must be denatured and kept single-stranded
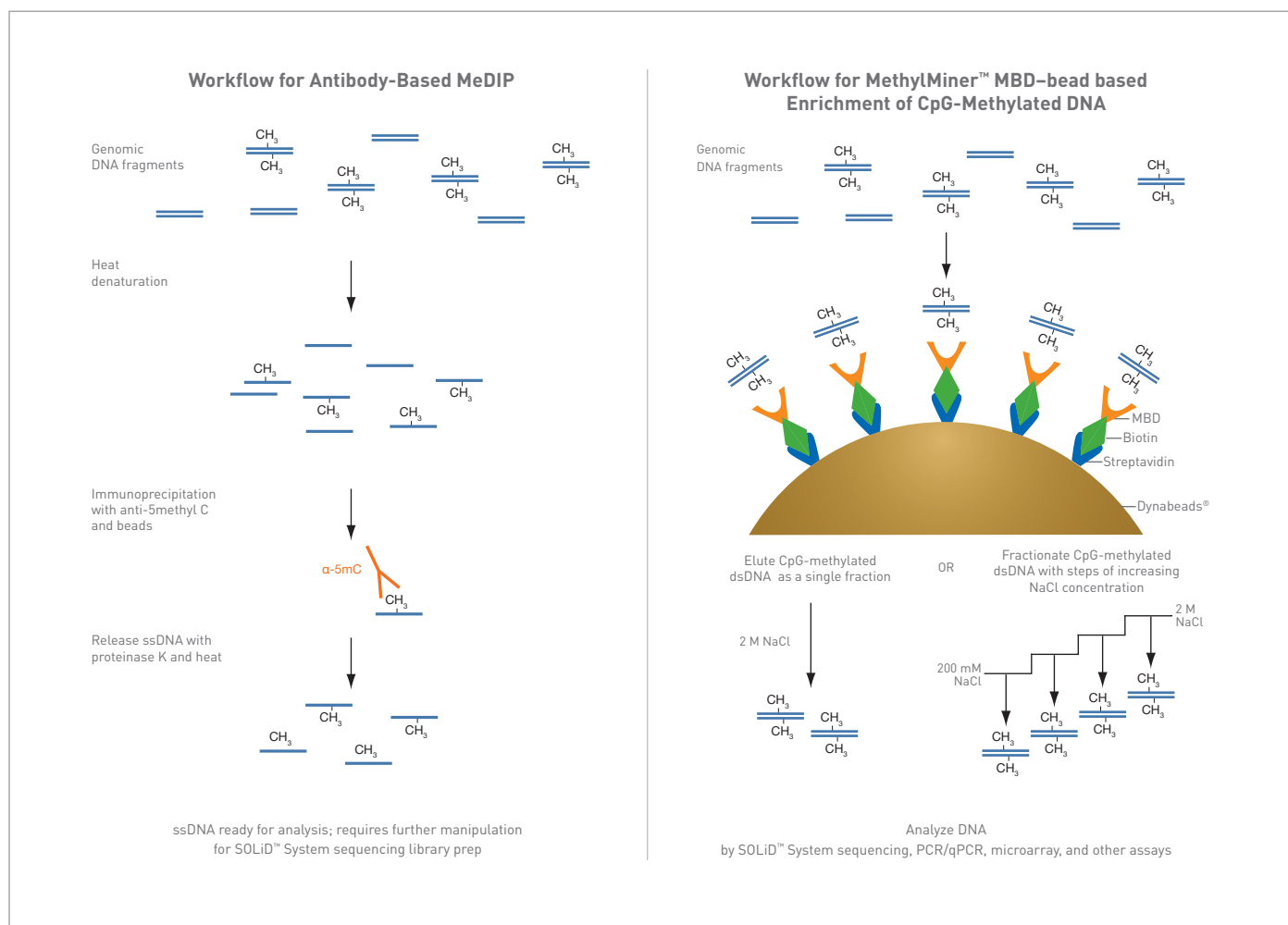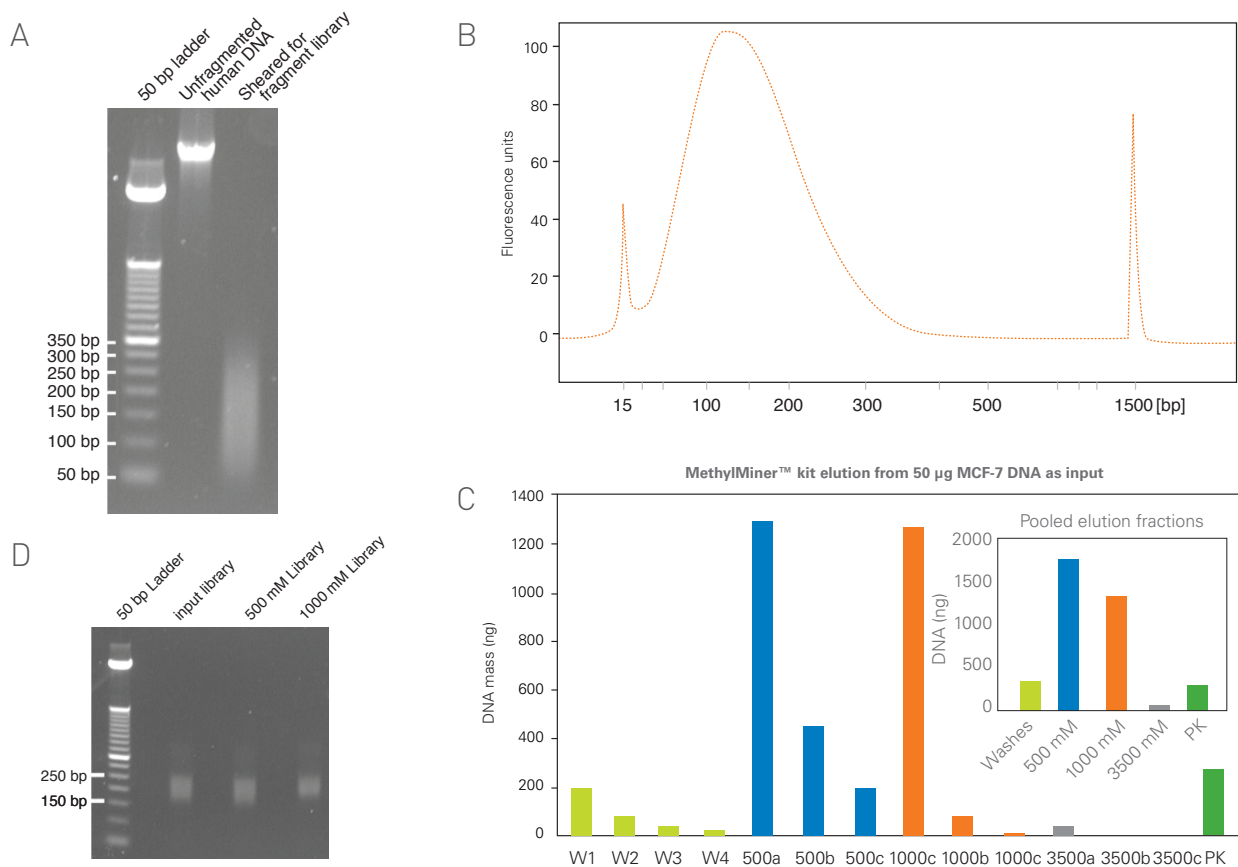

Figure 1. Comparison of Antibody-Based MeDIP and the MethylMiner™ Kit Workflows.

A

B

C MethylMiner™ kit elution from 50 µg MCF-7 DNA as input

Pooled elution fractions

D

Figure 2. SOLiD™ 3 Fragment Library Preparation After MethylMiner™ Kit Enrichment and Elution. Agarose gel image (A) and 2100 Bioanalyzer trace (B) of sheared human genomic DNA subjected to MethylMiner™ kit enrichment. Histograms (C) of DNA recovered during different stages of MethylMiner™ kit enrichment. W1–W4 are successive wash fractions. 500a–c are successive elutions with buffer containing 500 mM NaCl. Likewise,1000a–c and 3500a–c are successive elutions with buffer containing 1 M and 3.5 M NaCl, and PK refers to residual DNA recovered with a final proteinase K treatment. The inset shows the mass of DNA pooled from these iterations. Gel image of library DNA (D).

during the binding step for efficient capture. Denaturation of the DNA is necessary because all of the validated antibodies have been selected to bind a fully accessible methylated cytosine (5-mC) and cannot bind 5-mC occurring naturally in double-stranded DNA. This constraint can reduce binding efficiency as fragments can re-anneal during the binding reaction. To compensate for this, antibody-DNA binding is often performed for several hours to overnight at 4°C. After washing the beads, the DNA can only be recovered after digestion with proteinase K, usually with heat, which necessitates further cleanup, such as phenol:chloroform extraction followed by ethanol precipitation.

In contrast, the MethylMiner™ system is fast, convenient, and more versatile. DNA remains double-stranded during the binding step which takes only one hour at room temperature, or can be done overnight at 4°C. After incubation, the MethylMiner™ beads are washed and the captured methylated DNA eluted. The eluted double-stranded DNA fragments are ready for adaptor ligation (required by next-generation sequencing), which poses a significant advantage over working with single-stranded fragments. The system uses superparamagnetic Dynabeads® which provide high-quality and efficient capture kinetics, as well as rapid and convenient magnet-based bead capture for washing and changing solutions. Batch elution is easily achieved simply by mixing the beads with a high-salt (2 M NaCl) buffer [11]. Furthermore, since the affinity of MBD2 for methylated DNA can be modulated by ionic strength, fractionation of the captured DNA based on its degree of methylation can be achieved with graded changes in ionic strength. Varying degrees of CpG methylation density can influence gene regulation. Therefore, the ability to fractionate the genome according to the degree of methylation is important for functional studies. Finally, only ethanol precipitation is needed to obtain concentrated and purified double-stranded methylated DNA that is ready for further analysis, such as by PCR methods, microarray, or high throughput sequencing (Figure 1).

Results

Human DNA from the breast cancer–derived cell line MCF-7 was used to map the methylation status of CpG sites on a genome scale. Briefly, 50 µg of genomic DNA was sheared using a Covaris™ S2 non-contact sonicator (SOLiD™ 3 fragment library protocol) to generate random, short DNA

fragments with a median size of ~150 bp (Figures 2 A and B). Fragmented DNA was then subjected to MethylMiner™ methylated DNA enrichment according to the protocol supplied with the kit, and two methylated fractions (500 mM and 1 M NaCl eluates) or a single fraction (2 M NaCl eluate) were isolated [11]. Figure 2C depicts the typical elution profile of DNA from the MethylMiner™ beads following serial salt elution and multiple washes. Greater than 90% of the captured DNA is sequentially eluted with 500 mM and 1 M NaCl. Subsequent elution at very high NaCl concentration (3.5 M) followed by digestion with proteinase K showed that less than 10% of the captured DNA remains on the beads after elution with 1 M NaCl.

DNA fractions were then used to construct standard SOLiD™ fragment libraries according to the manufacturer's instructions (SOLiD™ Fragment Library Construction Kit, Applied Biosystems). Library DNA was size-selected either with E-Gel® SizeSelect™ 2% agarose gels or by gel-purification from 2% agarose E-Gel® EX gels (Figure 2D). The MethylMiner™ SOLiD™ libraries were sequenced at 50-base read length in a 4-well deposition chamber on a SOLiD™ 3 System

and the sequenced tags were mapped to the human reference genome (hg18) using the SOLiD™ secondary software analysis pipeline. Each quarter-slide chamber yielded ~20 million uniquely mappable sequencing reads.

To confirm that MethylMiner™ kit fractionation was indeed partitioning DNA fragments based on methylation density, the overall number of CpG dinucleotides within the fragments in each fractionation library was calculated after SOLiD™ System sequencing (Figure 3). There was a significantly higher density of CpG dinucleotides in the MethylMiner™ kit– enriched fractions compared to the unbound fraction. In addition, the 1 M eluate has higher density than the 500 mM eluate. Notably, 25% of the unenriched sequenced human genomic DNA fragments have no CpGs. In contrast, only 0.65% of the 500 mM-enriched sequences and 0.11% of the 1 M enriched sequences contain no CpGs. Despite the fact that the exact methylation status of the CpG sites is not interrogated directly, our results suggest that fragments containing two or more methylated CpGs per 150 bp are captured by the MethylMiner™ kit and are progressively enriched from 2- to 10-

fold with increasing degrees of methyl-CpG content. Overall, our data demonstrate that the MethylMiner™ kit can fractionate DNA based on CpG density (also see reference 12).

To validate the performance of MethylMiner™ kit fractionation, we labeled the MCF-7 DNA from enriched and unenriched fractions with Alexa Fluor® dyes (BioPrime® Total for FFPE Genomic Labeling System, Invitrogen) and hybridized them to both a human chr21/22 tiling array (Nimblegen) and a human CpG island/promoter focused array (Agilent) in parallel with DNA enriched by traditional antibody-based MeDIP [10]. Summary results for high-intensity microarray probe hits are shown in the Venn diagram in Figure 4.

There is overlap between the combined MethylMiner™ kit fractions (500 mM and 1 M) and MeDIP datasets, as shown in grey. However, there are more than 2400 unique strong hits from the combined MethylMiner™ kit fractions than from the MeDIP fraction. In contrast, only 147 array hits are unique to the MeDIP sample. In total, there are greater than five times more hits overall, and greater than 16 times as many unique hits using the MethylMiner™ kit as
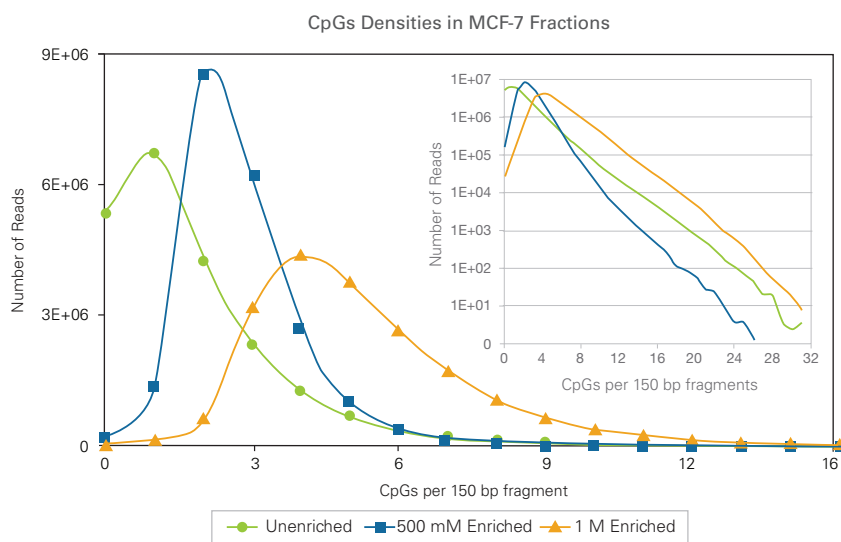

Figure 3. CpG dinucleotide density in MethylMiner™ kit–enriched SOLiD™ 3 fragment libraries. Number of CpG dinucleotides found within each 150 bp nuclear fragment sequenced from unenriched DNA and MethylMiner™ kit–enriched fractions (500 mM and 1 M) are shown. The number of CpGs per DNA fragment was determined by counting within a 150 bp region downstream of uniquely observed read starts. The inset graph shows the 10-fold depletion of CpG-rich fragments from the 500 mM fraction and concomitant 10-fold enrichment of these fragments in the 1M fraction, both relative to unenriched DNA.

Number of chr21 & chr22 array probes: 4240
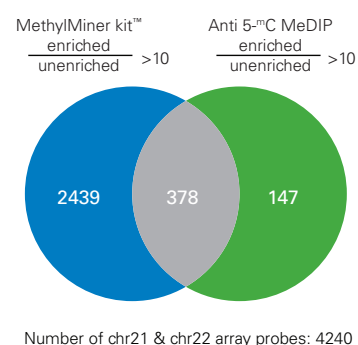
Figure 4. Strongly Positive Features on Microarrays: MethylMiner™ Kit vs. antibody MeDIP. The Venn diagram shows the strong positive features from chromosomes 21 and 22 (signal intensities >10-fold over background) detected on human CpG island/promoter focused arrays. The number of strong positive features for combined MethylMiner™ kit fractions (blue, pooled 500 mM and 1 M eluates) and antibody MeDIP (green) are shown. Overlap of strong features (grey) is also shown.
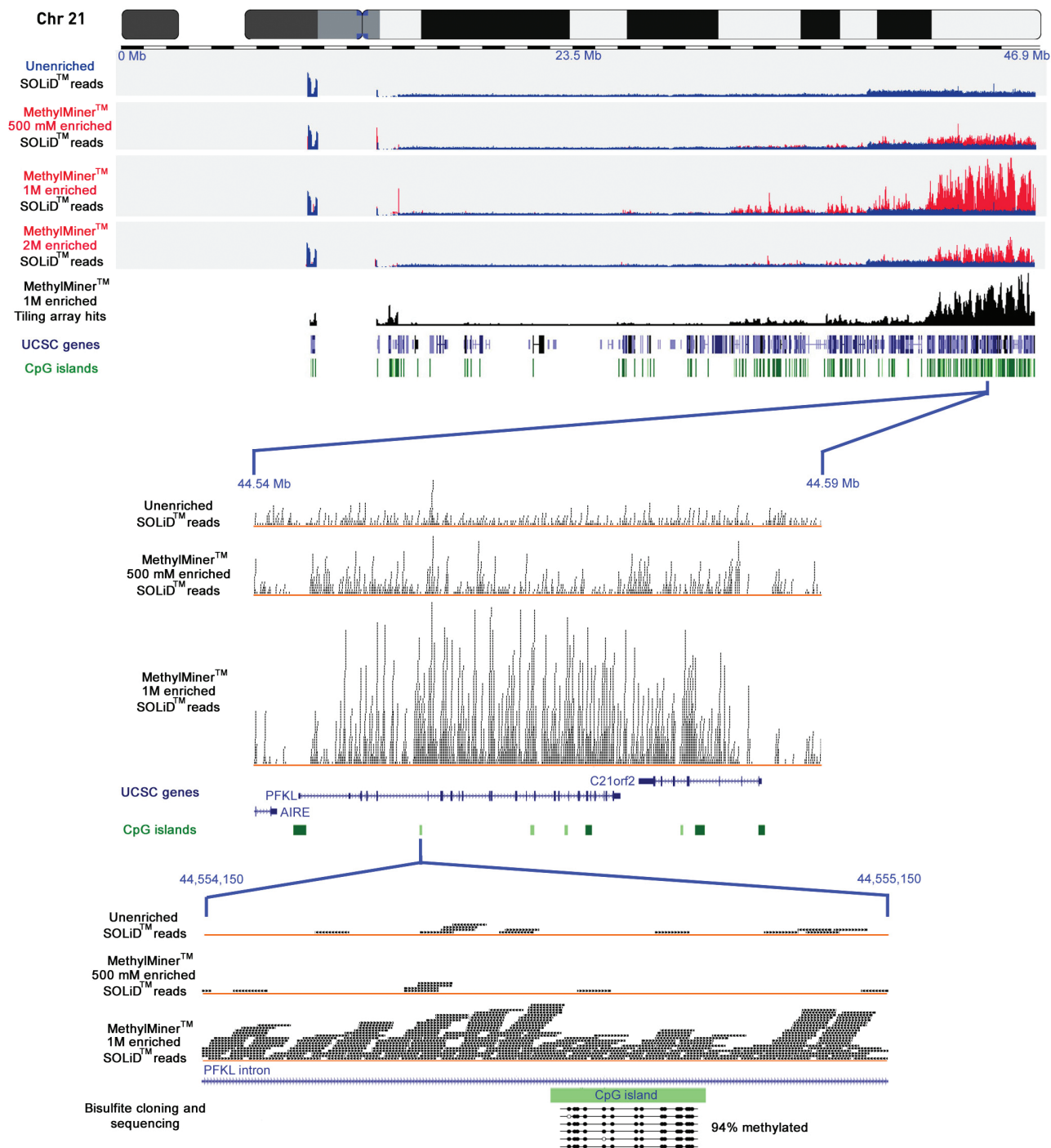
Figure 5. MethylMiner™ Enrichment—SOLiD™ System Sequencing Across Chromosome 21. Tracks illustrating sequencing reads from unenriched MCF-7 DNA, MethylMiner™ kit–enriched 500 mM and 1 M NaCl fractions, and full MethylMiner™ kit–enrichment with 2 M NaCl are shown. Sequencing read depth of coverage for the enriched fractions is shown in red relative to the more uniform low coverage obtained from unenriched DNA shown in blue. Positive features for the MethylMiner™ kit 1 M enriched fraction on a tiling array are also shown. Locations of annotated genes and CpG islands are indicated. Zoomed-in views of 50 kB and 1 kb regions are discussed in the text.
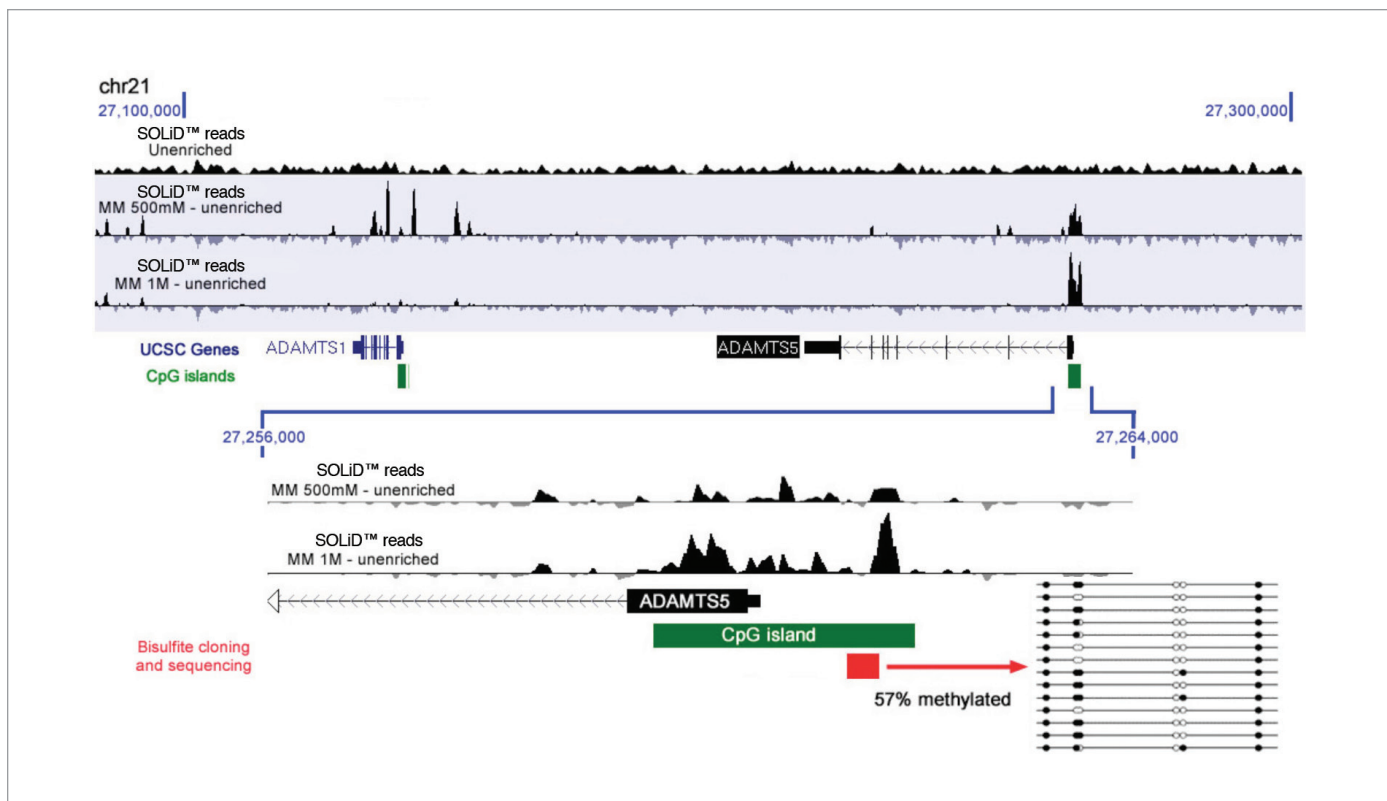
Figure 6. Annotated View of a 200 kb Region of Chromosome 21. Tracks of SOLiD™ System reads from unenriched and MethylMiner™ kit–enriched 500 mM and 1 M NaCl fractions. For MethylMiner™ kit–enriched data, the read distribution from an unenriched dataset of the same size has been subtracted to highlight regions of enrichment. Locations of annotated genes and CpG islands are also shown. Results of bisulfite sequencing of a small segment (red box and corresponding lollipop diagram) in the promoter CpG island of ADAMTS5 are also shown. Open circles: non-methylated CpGs; filled circles: methylated CpGs.

compared to MeDIP. Our results indicate that MethylMiner™ kit fractionation pulls down more sequences containing CpG-rich sites and sequences within promoters than the antibody-based approach.

In addition to "global" data demonstrating successful enrichment with MethylMiner™ capture, detailed information on enriched sequence locations and CpG densities can also be directly visualized and inspected with SOLiD™ System reads. Figure 5 depicts the uniquely mappable SOLiD™ System sequencing read locations and densities on a single chromosome (chr 21). Sequencing read distributions are depicted for unenriched DNA, successive MethylMiner™ kit elution fractions (500 mM and 1 M NaCl), and full MethylMiner™ kit elution with 2 M NaCl; further, tiling microarray data for the 1M NaCl fraction are shown.

For the MethylMiner™ kit–enriched uniquely mapped reads, the read coverage without

enrichment is superimposed on the enriched read coverage to illustrate the relative enrichment obtained for equivalent amounts of sequencing time and expense; each sequencing reaction comprised ~20 million uniquely mappable reads from one chamber of a 4-well slide. The annotated genes and CpG islands are also shown for reference. The unenriched track suggests that there has been amplification of chromosome 21 DNA in this venerable cancer cell line, and the enriched tracks indicate that the distal end of the long arm of the chromosome is densely methylated. This is not surprising since widespread methylation of large chromosomal segments has been seen in other cancer genome studies and since the vast majority of annotated genes lie within this region of chromosome 21 [13,14]. At this high level, good concordance can be seen between the data sets, and the profiles of the serially eluted 1 M fraction reads and the batch-eluted 2 M NaCl reads are remarkably similar, indicating good reproducibility of

sequence capture and significant sequence overlap between these two different elution schemes. The greater depth of coverage seen in the 1 M fraction attests to the power of fractionation to reduce complexity in the eluted sample. Methylated sequences that resist elution with 500 mM NaCl are better enriched in a subsequent 1 M elution fraction than in the more complex sample that is eluted as a whole with 2 M NaCl.

For a more detailed view, the analysis focused on a 50 kb region in chromosome 21 (middle of Figure 5), and at this level of resolution, a broad peak of enrichment that spans the neighboring genes PFKL and C21orf2 can be seen in the SOLiD™ System data from the 1 M elution fraction. Closer inspection of the 1 kb region that encompasses one of gene PFKL's intronic CpG islands clearly demonstrates the enrichment in this fraction (lower portion of Figure 5). The average depth of coverage is less than 1 in both the unenriched and

500 mM elution fractions; in contrast, coverage depth ranges from 5 to 15 in the 1 M elution fraction. Bisulfite cloning and sequencing of this CpG island verified that its CpGs are 94% methylated (Figure 5 lollipop diagram, bottom row; black circles indicate methylated CpGs and open circles non-methylated). Notably, this CpG island failed to be identified as methylated in our MeDIP-microarray experiments.

To further corroborate the MethylMiner™ SOLiD™ System data with the methylation status of the genomic DNA, we focused on a 200 kb region harboring the ADAMTS1 and ADAMTS5 genes on chromosome 21. The differential enrichment between the MethylMiner™ kit fractions is evident at this localized view. Sequences spanning the body and the upstream promoter region of the ADAMTS1 gene are exclusively enriched in the 500 mM fraction. With respect to sequences covering the CpG island in the promoter of the ADAMTS5 gene, however, there is preferential enrichment in the 1 M fraction (Figure 6).

We further carried out bisulfite cloning and sequencing across a portion of the CpG island in the ADAMTS5 promoter to determine the exact methylation pattern since this locus, like the PFKL CpG island shown in Figure 5, had also been identified in our microarray experiments as significantly enriched (>10-fold) with the MethylMiner™ kit but not captured by MeDIP. As shown by the lollipop diagram, the bisulfite-sequenced region shows an intermediate degree (57%) of CpG methylation. In all other loci tested (5 regions that were similarly detected on microarrays with MethylMiner™ kit but missed with MeDIP), bisulfite sequencing has confirmed the presence of >80% CpG methylation (data not shown). These results further indicate that MethylMiner™ kit enrichment permits the isolation of fractions of genomic DNA that harbor differing degrees of methylation. This capacity to capture moderately methylated regions, along with the access to SOLiD™ System sequencing data is a notable advantage over antibody-based methods.

Conclusions

When it comes to whole-genome methylation analysis, fractionation with the reagents in the MethylMiner™ Methylated DNA Enrichment Kit provides several significant advantages over traditional antibody-based enrichment. The DNA binding step is rapid, efficient, and more sensitive to low levels of CpG methylation with the biotinylated MBD protein in the MethylMiner™ kit. In addition, the incorporation of Dynabeads® superparamagnetic beads facilitates high-quality and effective capture kinetics. Furthermore, with MethylMiner™ kit enrichment, genomic DNA can be fractionated according to its methylation density. Finally, the capacity to capture and release double-stranded DNA greatly expedites library construction for SOLiD™ System sequencing.

In addition, multiplexing capabilities on the SOLiD™ System will permit the bar-coding of potentially hundreds of samples for simultaneous sequencing. Moreover, the use of the MethylMiner™ kit allows samples to be further divided into distinct fractions according to their methylation density. All of these features are important and relevant where sensitive and accurate screening of large numbers of samples is necessary, such as in clinical research and analysis. In summary, the MethylMiner™ kit combined with the SOLiD™ Sequencing System provides a seamless, easy workflow that delivers robust enrichment and deep coverage across focal areas for cost-effective sequencing of methylated sequences.

REFERENCES

1. Morgan HD, Santos F, Green, K, et al. (2005) Epigenetic reprogramming in mammals. Hum Mol Genet 14:R47–R58.

2. Reik W, Lewis A (2005) Co-evolution of X-chromosome inactivation and imprinting in mammals. Nat Rev Genet 6(5):403–410.

3. Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. *Nature* 447(7143):433–40.

4. Jones PA, Baylin SB (2007) The epigenetics of cancer. Cell 128(4):683–692.

5. Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324(5929):929–930.

6. Ramsahoye BH, Biniszkiewicz D, Lyko F, et al. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A 97(10):5237–5242.

7. Schmidl C, Klug M, Boeld TJ et al. (2009) Lineage specific DNA methylation in T cells correlates with histone methylation and enhancer activity. Genome Res 19(7):1165–1174.

8. Irizarry RA, Ladd-Acosta C, Wen B, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 41(2):178–186.

9. Zilberman D, Henikoff S (2007) Genome-wide analysis of DNA methylation patterns. Development 134(22):3959–3965.

10. Weber M, Davies JJ, Wittig D, et al. (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 37(8):853–862.

11. MethylMiner™ manual; follow links from bottom of the following page: www.invitrogen.com/methylminer

12. MethylMiner™ for Methylated DNA Enrichment webinar http://invitrogen.cnpg.com/Video/flatFiles/1080/index.aspx

13. Clark, S (2007) Action at a distance: epigenetic silencing of large chromosomal regions in carcinogenesis. Hum Molec Genet 16:R88–R95.

14. Gardiner K, Horisberger M, Kraus J, et al. (1990) Analysis of human chromosome 21: correlation of physical and cytogenetic maps; gene and CpG island distributions. EMBO J 9(1):25–34.

Headquarters
850 Lincoln Centre Drive | Foster City, CA 94404 USA
Phone 650.638.5800 | Toll Free 800.327.3002
www.appliedbiosystems.com

International Sales
For our office locations please call the division headquarters or refer to our Web site at
www.appliedbiosystems.com/about/offices.cfm