Sequencing of Methylated DNA Enriched with Methyl-CpG-binding domains

Gavin Meredith, Miroslav Dudas, José Bautista Saldaña, Daniel Krissinger, Eric Olivares, Adam Harris, Kevin Clancy, Ryan Lister‡, Joseph Ecker‡, Christopher Adams, and Rob Bennett

Life Technologies - Invitrogen and Applied Biosystems • 1610 Faraday Avenue • Carlsbad, California 92008 • USA

‡ Salk Institute for Biological Studies • 10010 N. Torrey Pines Rd. • La Jolla, California 92037 • USA

Introduction

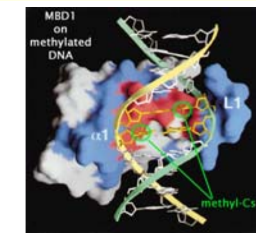
Methylation of cytosines in CpG dinucleotides in eukaryotic genomes is an epigenetic phenomenon that has wide-ranging implications in basic gene-regulation and medical research[1]. In particular, changes in DNA methylation, including genome-wide losses and hypermethylation of promoter regions, have been strongly correlated with the onset and progression of cancer. One common strategy for monitoring the methylation status of the genome and particular loci within it involves first fragmenting the DNA, then capturing the methylated fragments with an anti-5methyl cytosine antibody. This so-called MeDIP (methylated DNA immunoprecipitation) assay and variations (mCIP and MIRA) that use methyl-CpG binding domain (MBD-family) proteins have been used in conjunction with bisulfite sequencing to determine the distribution of methylated cytosines in human chromosomes and in the *Arabidopsis* genome[2-6].

Here we present data showing the features and use of a new MBD/bead-based system, the "MethylMiner™ Methylated DNA Enrichment Kit", upstream of SOLiD™ system sequencing of the *Arabidopsis thaliana* genome.

The need for new tools

Until recently, DNA methylation was thought to mainly serve to as a repressive mark that was relatively static and that its most variable role was in the regulation of a subset of promoters that possess regions of unusually high CpG content (so-called CpG islands). In this model, such promoters are either non-methylated and "open" or are heavily methylated and "repressed". However, a growing number of studies in the field are beginning to call this established paradigm into question. Thus tools to study regions of partial or dynamically variable methylation are required. MBD (methyl-CpG-binding) proteins offer the possibility to finely fractionate sample DNA based on its relative degree of methylation and can be easily used upstream of high-throughput sequencing.

Fig. 1 How MBDs bind DNA



Adapted from (Ohki et al. Cell, 105, 487–497, May 18, 2001)

MBD proteins share a common fold but bind double-stranded methylated DNA with different affinities[7]. The structures of MBD1 bound to methylated DNA (above) and MeCP2 have been determined; however, published data suggest that MBD2 binds with the highest affinity of all[8,9,10]. Since the binding involves broad contact between the protein and the major groove of the DNA, changes in pH or ionic strength modulate the binding interaction. It has also been shown that 1 M NaCl can almost entirely disrupt the binding at pH 7.4. The MBD used in the MethylMiner™ kit is a unique biotinylated recombinant version of the human MBD2 protein. The combination of this protein with streptavidin Dynabeads® creates a very efficient methylated DNA capture system.

Fig. 2 MethylMiner™ workflow for SOLiD™ system sequencing

Add fragmented dsDNA to beads

Bind — unbound

Methyl-CpG binding protein, Biotin, Streptavidin, Dynabead

Elute

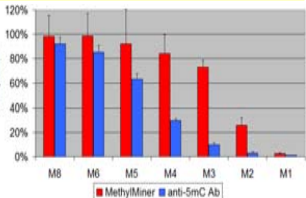
Directly or Fractionate

2M NaCl or Prot K — 500mM NaCl, 1M NaCl, Prot K

Deep-sequence

SOLiD™ System

Fig. 3 MethylMiner™ is more sensitive than α-5methyl C



The capture efficiency of the MethylMiner™ system was compared to that of the MeDIP assay. A series of synthetic 80 bp duplex DNAs carrying from 1 to 8 symmetrically methylated CpGs (10 pg in a background of 1 µg fragmented genomic DNA) was used. The MethylMiner™ kit (red bars) could capture ~25% of duplexes that carried only two methylated CpG motifs (M2). The MeDIP assay required the presence of 4 (M4) such sites for comparable performance.

Fig. 4 Sensitivity and Linearity

Amplifiable DNA from a methylated locus, in this case the PLAU promoter CpG island in MCF-7 cells, can be recovered from as little as 5 ng of input genomic DNA. For small-scale capture reactions, the MethylMiner kit shows linear recovery (inset) from between 25 ng and 1 µg of input. For sequencing, capture reactions can be scaled to at least 50 µg input.

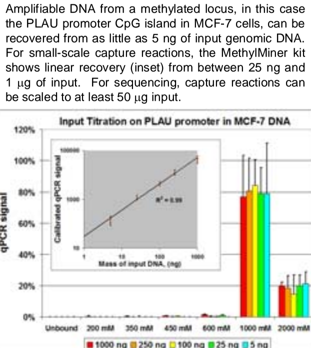
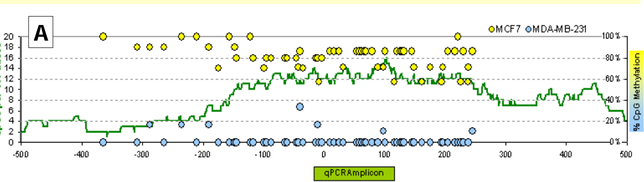


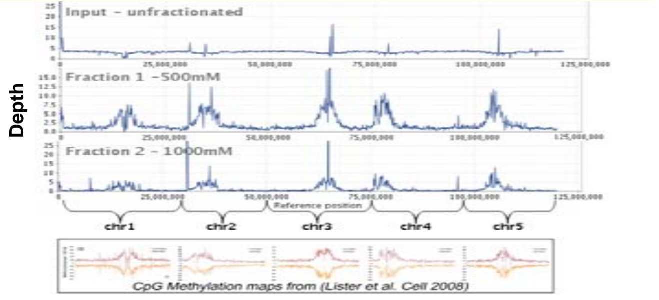
Fig. 5 Methyl-CpG Fractionation of the PLAU promoter CpG island



A) The CpG island at the promoter of the PLAU gene is illustrated by a green line indicating the number of CpGs within a 100 bp window. The increased density of CpGs occurs between positions -200 and +500 (relative to the transcription start-site). Yellow and blue circles are the frequencies of methylation for every CpG from -375 to +250 as determined by bisulfite sequencing in the two cell-lines. MCF-7 cells are hypermethylated at this locus whereas MDA-MB-231 cells are hypomethylated[11].

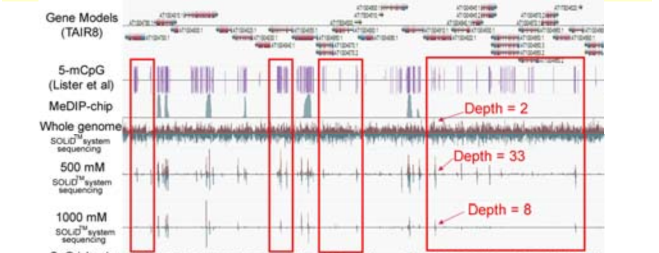
B) Recovery of DNA from the CpG island as assayed by qPCR after MethylMiner capture and salt fractionation. In this log scaling the difference between the cell-lines is obvious. For MCF-7, greater than 85% of the amplifiable DNA is eluted with 1M NaCl; in contrast, for MDA-MB-231, >95% is in the unbound fraction, but a slight peak is detectable in the 350 mM fraction. On a linear scale (inset) the contrast is more stark.

Fig. 6 *A thaliana* methylated fractions span the whole genome

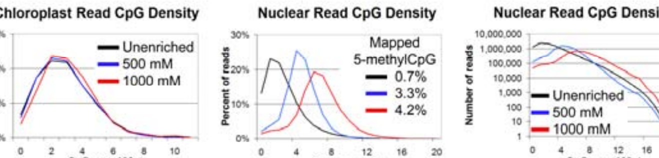


15 µg of ~50-500 bp fragments of *A thaliana* (Col0) genomic DNA were captured with the MethylMiner™ system. After washing, ~20% of the DNA was bound and eluted serially with 500mM NaCl, then 1M NaCl, to obtain >700 ng of DNA in each fraction. Fragment libraries were made from 250 ng of each, including input, and sequenced on a SOLiD™ (V2) platform in a quad-chambered flow-cell. In good concordance with a published dataset, the highly methylated pericentromeric regions were enriched in each fraction[11].

Fig. 7 SOLiD™ system seq shows MethylMiner™ fractions are enriched in methylated sequences missed by antibody MeDIP-chip



SOLiD™ system seq of MethylMiner™ fractions shows good correspondence to the BS-seq map of CpG methylation from Lister et al.[12] In many cases throughout the genome, this approach differentially enriches methylated sequences that an alternative antibody-based method, MeDIP-chip, misses (red boxes). AnnoJ browser view, ©2008 Julian Tonti-Filippini; http://www.annoj.org. Below, the chloroplast genome is not methylated and shows no enrichment for CpGs in the sequencing reads of the fractions; in contrast, the nuclear genome does. For nuclear genomic reads the average CpG density and degree of methylation increases from input to 500 mM to 1M and significant enrichment is obtained in the 1M fraction. The log-scale view shows the 2- to 4-fold enrichment of sequences with 4-8 CpGs per 100 bp in the 500 mM fraction and 2-to 10-fold enrichment of sequences with 5-20 CpGs per 100 bp in the 1M fraction.



References

1) Feinberg AP (2007) *Nature* **447**:433-40

2) Weber M, et al. (2005) *Nat Genet.* **37**:853-62

3) Down TA, et al. (2008) *Nat Biotech* **26**:779-784

4) Zhang X, et al. (2006) *Cell* **126**:1189-1201

5) Gebhard C, et al. (2006) *Nucleic Acids Res.* **34**:e82

6) Rauch T and Pfeifer GP (2005) *Lab Invest.* **85**:1172-80

7) Roloff TC, et al. BMC Genomics (2003) **4**:1

8) Ohki I, et al. (2001) *Cell* **105**:487-97

9) Fraga MF, et al. (2003) *Nuc. Acids Res.* **31**:1764-74

10) Ho KL, e at. (2008) *Molec. Cell* **29**:525-31

11) Guo Y, et al. (2002) *J Biol Chem.* **277**(44):41571-9

12) Lister R, et al. (2008) *Cell* **133**:1-14

invitrogen

www.invitrogen.com

For more information, please go to www.invitrogen.com/methylation then follow the links for "methylated DNA enrichment" and/or "kits & assays"

invitrogen™
www.invitrogen.com