Sequencing of Methylated Human DNA Enriched through MBD-Affinity is a Cost-Effective Alternative to Whole Genome Bisulfite Conversion

Gavin D. Meredith, Miroslav Dudas, Elizabeth Levandowsky, Tamara Gilbert, Daniel Krissinger, George Marnellos, Vrunda Sheth, Christopher Adams, Clarence Lee, Timothy Harkins and Rob Bennett , Life Technologies Corporation, 5791 Van Allen Way, Carlsbad, CA, USA, 92008

ABSTRACT

Aberrant DNA methylation is characteristic of many cancers and other complex diseases and differences in methylation have been observed in a wide variety of genomic contexts; for example, both within "classic" promoter-associated CpG islands and also in distal, non-CpG island regions including CpG island shores [1, 2]. High-throughput sequencing of DNA fragment libraries from the well-characterized and epigenomically profiled IMR-90 fibroblast cell-line [3] proves that methyl-CpG binding domain (MBD) proteins bind methylated DNA fragments sensitively (binding fragments with as few as 2 methyl-CpGs per 150 bp) and selectively (~50-fold over non-methylated fragments). Furthermore, when MBD-captured DNA is sub-fractionated by step-wise salt elution, genomic DNA fragments separate based on their average methyl-CpG content. Peak analysis of the mapped reads permitted accurate identification of thousands of regions of methylation in different genomic locus classes. The lowest salt fraction (eluted with 350 mM NaCl) is depleted of CpG islands but otherwise samples the genome (introns, promoters, shores, repeats, and intergenic regions) evenly, intermediate salt fractions (eluted successively with 450 mM and 600 mM NaCl) show moderate degrees of enrichment (2- to 3-fold) for exons and CpG island shores, and the highest salt fraction (eluted with 2M NaCl) is greatly enriched (14-fold) for CpG islands and moderately enriched (4- to 5-fold) for shores and exons. Methylation at a large number of these positions was confirmed by whole-genome bisulfite sequencing on the SOLiD™ System along with comparison to the published methylome. Bisulfite-sequencing on the SOLiD™ Systemyields high-quality data on the exact positions of methylation as evidenced by high concordance (~94%) of our results with the published dataset; this high similarity between biological samples cultured and analyzed independently indicates a remarkably stable methylation pattern for this differentiated cell-line. MBD-affinity enrichment coupled with high-throughput sequencing enables efficient surveying of genomic DNA methylation and yields 5- to 10-fold improvement in throughput and cost.

INTRODUCTION

DNA methylation plays a critical role in gene regulation that influences normal organism development and many diseases including cancer. Profiling the DNA methylation patterns of higher organisms is challenging because methylation patterns vary between tissues and with developmental state, hence there are far more methylomes to be analyzed than genomes. Furthermore, in order to map methylation positions with high precision and accuracy, greater depth of sequencing is required than for normal genome sequencing [4]. Affinity-based enrichment of methylated DNA sequences prior to high-throughput sequencing, as with the SOLiD™ System, provides an avenue to pursue this kind of genome-wide information in a minimally biased and cost-efficient manner. The workflow described here using MethylMiner™ enrichment with stepwise salt gradient elution enables the partitioning of the genome into regions of low, moderate, and high density of methylation. This permits blind discovery of methylated regions and permits detection of differentially methylated regions (DMRs) between samples and across genomic feature subsets that harbor differing degrees of methylation density

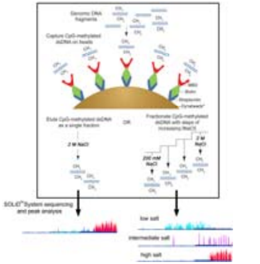
MATERIALS AND METHODS

Methylated DNA enrichment and SOLiD™ System sequencing.
Genomic DNA from cultured IMR-90 cells was purified with PureLink® columns. Purified HuRef genomic DNA was purchased from the Coriell Institute for Medical Research. Genomic DNA was fragmented to 50-400 bp (mean ~250 bp) with a Covaris® S2 System (Woburn, MA). Methyl-CpG binding-domain protein affinity capture was with MethylMiner™ Methylated DNA Enrichment kits (Invitrogen, Carlsbad, CA) following the manufacturer's protocol. For salt-gradient elution of MethylMiner™ DNA, successive fractions were obtained by elution using buffer containing the following NaCl concentrations: 200 mM, 350 mM, 450 mM, 600 mM, and finally 2 M NaCl. For HuRef DNA [5], successive elutions were done with buffer containing 450 mM, 600mM, and 2M NaCl. Each elution step consisted of 2–3 serial incubations of the MethylMiner™ beads at each salt concentration.

MATERIALS AND METHODS (continued)

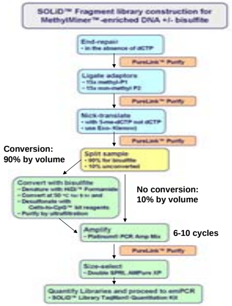
Whole-genome SOLiD™ bisulfite sequencing.
Genomic DNA from cultured IMR-90 cells was purified with PureLink® columns. Genomic DNA was fragmented to 50-400 bp (mean ~250 bp) with a Covaris® S2 System (Woburn, MA). The workflow is depicted in Figure 2. Five (5) micrograms of genomic DNA fragments was end-repaired using reagents from a SOLiD™ Fragment Library Construction kit with a dNTP-mix lacking dCTP, then PureLink® column purified. SOLiD™-BS-seq on methylation-enriched DNA is described in the Methods section of Ondov et al and is based on a workflow described by Bormann Chung et al [6, 7] and a protocol is available on-line [8].

Figure 1. MethylMiner™ Kit Fractionation of CpG-methylated DNA for SOLiD™ System Sequencing



Fragmented double-stranded CpG-methylated genomic DNA is directly and specifically captured on MethylMiner™ MBD-coated magnetic beads then eluted all-at-once with buffer containing 2M NaCl or separated into complementary fractions by step-wise elution with buffers containing progressively increasing concentrations of NaCl up to 2 M. Sequencing after single-step elution using 2M NaCl shows greatest enrichment for densely methylated regions of the sample (lower left). Elution using step-wise salt gradient buffers yields subsets of the methylome with differing degrees of methylation density; sparsely methylated fragments (light blue) elute with low salt, more densely methylated fragments (purple) elute with higher salt, and heavily methylated fragments (red) elute at maximal salt. Selective enrichment prior to sequencing permits clearer identification of these subclasses of methylated genomic DNA fragments.

Figure 2. Workflow for coupled preparation of unconverted and bisulfite-converted SOLiD™ fragment libraries



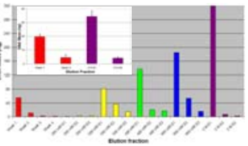
This basic workflow enables one to obtain matched normal and bisulfite-converted DNA libraries. In particular, these steps can be used for MethylMiner™ bead-enriched DNA fragments or applied to whole-genome DNA fragments. A document for this procedure, entitled "SOLiD™ Bisulfite-Converted Fragment Library Preparation Protocol" is available on-line from the Applied Biosystems and Invitrogen web-sites. Specifically, the protocol can be accessed at: http://tools.invitrogen.com/content/sfs/manuals/SOLiD_Bisulfite_Protocol_web.pdf

Figure 3. SOLiD™ System DNA Methylation Analysis Tool

A data analysis pipeline (available for download [9]) can be assembled using SOLiD™ System software (e.g. SOLiD™ BioScope™ software), free public tools (e.g., Bowtie [10], SAMtools, MACS [11]), and auxiliary scripts and programs (this is described more extensively in a recent Life Technologies Application Note entitled "Genome-wide methylation analysis at 150 bp resolution using the MethylMiner™ kit and SOLiD™ System sequencing). The pipeline can run on a Linux® desktop with multiple processors and sufficient RAM and disk space; minimum recommended requirements include 4–6 processors, 16GB RAM, and 2 TB storage. With Bowtie, SOLiD™ reads were mapped against the hg19 reference genome build. When using BioScope™ software, HuRef reads were mapped against the HuRef genome. The MACS algorithm is called for detection of peaks of methylated sequence enrichment.

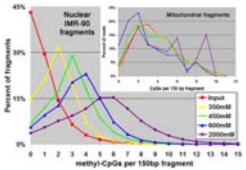
RESULTS

Figure 3. MethylMiner™ Elution profile from 25 µg of human cell-line IMR-90 genomic DNA (and 500 ng, inset)



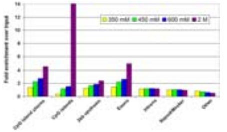
Mass of human DNA recovered using a step-wise salt gradient (350mM - 2M NaCl) elution. The majority of input DNA is not captured on the MethylMiner™ beads because only about 0.5-1% of all bases are 5-methyl-C in CpG dinucleotides. This fact, and the observation that diminishing amounts of DNA are recovered in successive wash and step-wise elutions (e.g., fractions 350a, 350b, and 350c in Fig. 2A-B), indicate that the methylated DNA–enrichment protocol worked properly. Importantly, as shown in the inset of Figure 2, MethylMiner™ bead enrichment from as little as 500 ng of fragmented genomic DNA can yield enough methylated DNA for SOLiD™ library construction.

Figure 4. SOLiD™ System DNA Methylation Analysis Tool



Counts of the number of methyl-CpG dinucleotides within 150 bp downstream of uniquely-mapped sequencing start-sites show that differing subsets of the genome are obtained with salt fractionation. Importantly, the number of sequenced fragments showing 0 CpG content drops with enrichment and the hypo-methylated human mitochondrial reads (inset) show no significant changes in their distributions; they are suppressed >50-fold with enrichment (not shown) indicating strong specificity of capture for CpG-methylated DNA.

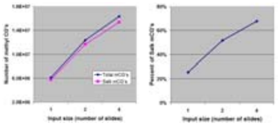
Figure 5. Enrichment trends for human genome features



Differential, specific enrichment for various, partially overlapping, annotated genomic features is obtained with salt-fractionation. Notably, CpG islands, shores, and exonic sequences increase in relative representation with increasing ionic strength. Single-fraction elution with high-salt (2M NaCl) yields an average behavior in this respect (not shown).

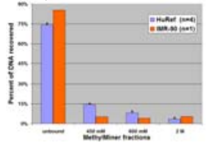
Table 1 and Figure 6. SOLiD™ Bisulfite sequencing of IMR-90 fragment library yields high concordance with published methylome

# slides	Total Reads	# Uniquely Mapped	Uniquely Mapped (%)	Total mCG's (this work)	Salk mCG's	Concordance with Salk mCG's	% Salk mCG's
1	495,265,636	151,321,772	30.6%	5,632,473	5,339,933	94.8%	23.6%
1	515,699,253	160,975,121	31.2%	6,096,313	5,770,102	94.6%	25.5%
2	1,010,964,889	312,296,893	30.9%	12,359,341	11,692,335	94.6%	51.6%
4	1,911,550,949	524,998,672	27.5%	16,354,550	15,357,420	93.9%	67.8%



The published methylome can be reproduced with 3-4 SOLiD™ System runs or about 4 billion raw 50 bp reads.

Figure 7. MethylMiner™ Elution profiles from 25 µg of human genomic DNA



Relative mass of human DNA recovered using a step-wise salt gradient (450mM - 2M NaCl) elution. The majority of input DNA is not captured on the MethylMiner™ beads because only about 0.5-1% of all bases are 5-methylC in CpG motifs. Typically 5-15% of input mass is recovered as the methylated subset.

Table 2. Paired-end mapping stats for HuRef libraries with and without ECC

Bisulfite converted

sample	F3	F5	non redundant pairings	reads per start point -F3	reads per start point -F5
HuRef_BS_SUP	53,577,466 (50.76%)	49,469,064 (46.87%)	27,397,712	1.15	1.12
HuRef_BS_SUP_ECC	56,305,625 (52.87%)	49,469,064 (46.87%)	28,646,054		
HuRef_BS_450mM	53,854,619 (48.95%)	46,488,311 (42.26%)	20,415,015	1.15	1.1
HuRef_BS_450mM_ECC	55,628,891 (50.14%)	46,488,311 (42.26%)	21,528,619		
HuRef_BS_600mM	48,413,158 (47.45%)	34,560,613 (33.87%)	13,143,685	1.15	1.11
HuRef_BS_600mM_ECC	51,790,121 (50.11%)	34,560,613 (33.87%)	13,922,181		
HuRef_BS_2M	36,350,685 (35.18%)	22,026,902 (21.32%)	5,272,821	1.14	1.19
HuRef_BS_2M_ECC	40,118,031 (38.54%)	22,026,902 (21.32%)	5,712,030		

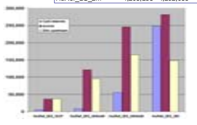
Non-converted

sample	F3	F5	non redundant pairings	reads per start point -F3	reads per start point -F5
HuRef_MM_SUP	81,428,617 (76.85%)	45,591,862 (42.97%)	45,692,214	1.23	1.17
HuRef_MM_SUP_ECC	85,086,709 (79.54%)	45,591,862 (42.97%)	47,268,892		
HuRef_MM_450mM	91,564,604 (83.59%)	48,130,909 (43.94%)	49,852,817	1.2	1.14
HuRef_MM_450mM_ECC	92,977,698 (84.13%)	48,130,909 (43.94%)	50,783,640		
HuRef_MM_600mM	85,394,256 (79.47%)	44,713,705 (41.61%)	40,183,242	1.29	1.2
HuRef_MM_600mM_ECC	88,131,539 (80.53%)	44,713,705 (41.61%)	41,132,528		
HuRef_MM_2M	81,505,487 (81.17%)	42,030,634 (41.86%)	34,437,218	1.19	1.13
HuRef_MM_2M_ECC	83,786,520 (81.65%)	42,030,634 (41.86%)	35,274,270		

Paired-end sequencing (75x35) shows very high rates of non-redundant pairings for unconverted MethylMiner™ bead-enriched libraries. ECC consistently permits higher yields of unique mapping too. For bisulfite-converted libraries, there is a reduction in pairing efficiency at higher ionic strengths, presumably due to the reduced sequence complexity both of the selected fragments and caused by bisulfite treatment.

Table 3 and Figure 8. Methyl-cytosines in HuRef detected using BioScope™ Software SNP-calls

sample	methylCs	me-CpGs	me-CpHs	overlap with Salk	CpG islands	exons	2kb upstream
HuRef_BS_SUP	1,640,358	1,635,245	5,113	1,477,719	5,797	36,377	36,832
HuRef_BS_450mM	5,232,402	5,230,498	1,904	4,840,232	8,211	121,714	95,167
HuRef_BS_600mM	6,483,756	6,481,321	2,435	5,928,973	55,222	245,710	164,585
HuRef_BS_2M	4,205,183	4,202,955	2,228	3,704,143	247,953	281,384	147,641



Since bisuflite-conversion can be thought of as the genome-wide chemical induction of non-methyl-C>T SNPs, the tools within BioScope™ software can be used to identify methylated Cs when an *in silico* fully-converted reference is used. An accounting of all methyl-Cs detected in the HuRef samples is given in the table and the distribution of methyl-Cs in CpG islands, exons and regions within 2 kb upstream of TSSs is graphed. Approximately 16 million methyl-Cs are detected, ~90% of which are found in the MethylMiner™ bead-enriched libraries each sequenced within a "quad" chamber. As seen for IMR-90 DNA, there is specific enrichment in regions corresponding to these genomic elements; however, a large number of methyl-Cs clearly reside elsewhere throughout the genome.

Figure 9 Example: Apparent Differential Methylation in the 5' region of the ISOC2 gene

Tracks indicate regions of significant MethylMiner™ enrichment

made with IGV browser [12]

CONCLUSIONS

A) Bisulfite-sequencing of human methylomes is tractable on the SOLiD™ 4 System; it requires 3-4 full runs per methylome. The results are highly concordant with published data.
B) MBD-based enrichment of methylated sequences with the MethylMiner™ kit is an efficient means to focus of SOLiD™ System sequencing on genomic feature subsets.
C) MBD-based enrichment permits sub-fractionation (by varying ionic strength) of the genome based on the density of methylated CpGs.
D) Methylated DNA enrichment can be coupled with bisulfite-sequencing permitting single-nucleotide resolution validation of specific positions of methylation and 4-fold reduction in sequencing cost.

REFERENCES

1. Irizarry RA, et al Nat Genet. 2009; 41:178-86
2. Schmidl C, et al Genome Res. 2009; 19:1165-74
3. Lister R, et al Nature 2009; 462:315-22
4. Zhang Y and Jeltsch A Genes 2010; 1:85-101
5. Levy S, et al PLoS Biology 2007; 5:2113-44
6. Borman Chung CA, et al PLoS One 2010; 5:e9320
7. Ondov BD, et al Bioinformatics 2010; 26:1901-2
8. SOLiD™ Bisulfite-Converted Fragment Library Prep Protocol
9. http://solidsoftwaretools.com/gf/project/dna_analysis/
10. Langmead B, et al Genome Biol. 2009; 10:R25
11. Zhang et al. Genome Biol. 2008; 9:R137
12. http://www.broadinstitute.org/igv

TRADEMARKS/LICENSING