Epigenetics 6:1, 34-44; January 2011; © 2011 Landes Bioscience

Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias

Shalima S. Nair,[1,†] Marcel W. Coolen,[1,3,†] Clare Stirzaker,[1,†] Jenny Z. Song,[1] Aaron L. Statham,[1] Dario Strbenac,[1] Mark D. Robinson[1,2] and Susan J. Clark[1,*]

[1]Epigenetics Laboratory; Cancer Research Program; Garvan Institute of Medical Research; Darlinghurst, NSW Australia; [2]Bioinformatics Division; Walter and Eliza Hall Institute of Medical Research; Melbourne, VIC Australia; [3]Department of Molecular Biology; Faculty of Science; Nijmegen Centre for Molecular Life Sciences; Radboud University Nijmegen; Nijmegen, The Netherlands

[†]These authors contributed equally to this work.

Key words: DNA methylation, CpG dinucleotides, epigenetics, genome wide, differentially methylated regions, cancer

Abbreviations: WGA, whole genome amplification; MAT, model-based analysis of tiling arrays; IGB, integrated genome browser; SAP, shrimp alkaline phosphatase; DMR, differentially methylated regions; MeDIP, methyl-DNA immunoprecipitation; MBD, methyl-CpG binding domain; 5MeC, 5'-methylcytosine

DNA methylation primarily occurs at CpG dinucleotides in mammals and is a common epigenetic mark that plays a critical role in the regulation of gene expression. Profiling DNA methylation patterns across the genome is vital to understand DNA methylation changes that occur during development and in disease phenotype. In this study, we compared two commonly used approaches to enrich for methylated DNA regions of the genome, namely methyl-DNA immunoprecipitation (MeDIP) that is based on enrichment with antibodies specific for 5′-methylcytosine (5MeC) and capture of methylated DNA using a methyl-CpG binding domain-based (MBD) protein to discover differentially methylated regions (DMRs) in cancer. The enriched methylated DNA fractions were interrogated on Affymetrix promoter tiling arrays and differentially methylated regions were identified. A detailed validation study of 42 regions was performed using Sequenom MassCLEAVE technique. This detailed analysis revealed that both enrichment techniques are sensitive for detecting DMRs and preferentially identified different CpG rich regions of the prostate cancer genome, with MeDIP commonly enriching for methylated regions with a low CpG density, while MBD capture favors regions of higher CpG density and identifies the greatest proportion of CpG islands. This is the first detailed validation report comparing different methylated DNA enrichment techniques for identifying regions of differential DNA methylation. Our study highlights the importance of understanding the nuances of the methods used for DNA genome-wide methylation analyses so that accurate interpretation of the biology is not overlooked.

Introduction

A prerequisite for understanding the function of DNA methylation is knowledge of its distribution in the genome. In mammals, 5'-methylcytosine (5MeC) accounts for ~1% of total DNA bases and therefore potentially affects 70–80% of all CpG dinucleotides in the genome.[1] However, DNA methylation patterns are dynamic in nature and vary during development and across the genome. The cycle of early embryonic demethylation followed by de novo methylation is critical in determining cell specific DNA methylation patterns. Possibly the most notable feature of mammalian DNA methylation patterns is the presence of CpG islands, that is, unmethylated GC-rich regions that possess high relative densities of CpG and are positioned at the 5' ends of many human genes.[2] There are approximately 29,000 CpG islands[3,4] in the human genome sequence and more than 60% of human genes are associated with CpG islands, of which the great majority are unmethylated at all stages of development and in all tissue types.[5]

*Correspondence to: Susan J. Clark; Email: s.clark@garvan.org.au
Submitted: 06/30/10; Accepted: 08/12/10
Previously published online: www.landesbioscience.com/journals/epigenetics/article/13313
DOI: 10.4161/epi.6.1.13313

A small but significant proportion of all CpG islands become methylated during development and developmentally programmed CpG-island methylation is involved in genomic imprinting and X chromosome inactivation.[6] Differential methylation that is functionally important can also occur at regions other than CpG islands, such as at CpG island shores[7] and potentially at CpG sites associated with tissue specific genes, miRNA genes and noncoding RNA expression. A significant fraction of all human CpG rich regions is prone to progressive methylation in certain tissues during aging (reviewed in ref. 8) or in abnormal cells such as cancers (reviewed in ref. 9) and permanent cell lines.[10,11] Alterations in DNA methylation can lead to aberrant gene expression and can play a critical role in disease states such as cancer, where promoter CpG island hypermethylation leads to inactivation of tumor suppressor genes and promoter hypomethylation leads to activation of critical cancer-associated genes.[12] Profiling DNA methylation across the genome is vital to understand the DNA methylation changes that are occurring in disease phenotypes and to identify genes where DNA methylation could provide potential molecular biomarkers that can be used for diagnosis or prognosis.

Several methods have been developed to map DNA methylation patterns for either limited regions or genome-wide studies. Most methods of characterizing methylation are based upon one of three techniques: chemical conversion with sodium bisulfite, digestion with methylation-sensitive restriction enzymes or affinity enrichment of methylated DNA (reviewed in ref. 13). These techniques were initially restricted to localized regions of the genome, but many have now been scaled to enable genome-wide methylation analysis, using microarray hybridization techniques and next generation sequencing technology. The two main approaches used to enrich for methylated regions of a genome are (1) methyl-DNA immunoprecipitation (MeDIP) which uses a monoclonal antibody specific for 5MeC[14] and (2) affinity capture with methyl-CpG binding domain-based (MBDCap) proteins.[15-20] MeDIP is based on the immunoprecipitation of single stranded molecules containing one or more methylated CpG sites (**Fig. 1A**). In contrast the MBD-based strategy captures double stranded methylated DNA fragments and different methylation densities can be analyzed depending on the salt fractionation employed;[15] lower salt fractions contain fragments with fewer methyl groups, while higher salt fractions contain more highly methylated DNA (**Fig. 1B and C**).

In this study we compared these two different approaches to identify differentially methylated genes in prostate normal (PrEC) and prostate cancer (LNCaP) cells using hybridization to promoter tiling arrays. We used the Sequenom MassCLEAVE technique[21] to validate the differentially methylated regions (DMRs) and examined the effect of CpG density and DNA methylation levels on the efficacy of the enrichment techniques used. We found that each technique was sensitive, but preferentially identified different CpG rich regions of the prostate cancer genome as significantly differentially methylated.

Results

MeDIP and MBDCap are commonly used in methylation studies but there has, to date, been little comparison of their ability to discover differentially methylated regions (DMRs) in cancer. Here, we compared enrichment of methylated DNA from the normal prostate cell line, PrEC and the prostate cancer cell line LNCaP using MeDIP and the MBDCap method.[15]

MBD-enrichment of methylated DNA by salt fractionation. For the MBDCap protocol, we compare the single high salt elution step (MBD-SF) with the captured sequences released in a step-wise elution series with increasing salt concentrations (0.2 M–2 M NaCl) (**Fig. 1A**). The specificity of enrichment, in each of the salt fractions, was initially assessed by assaying recovery of three control CpG island promoters of known methylation status. **Figure 1B** shows qPCR results of the elution of *GSTP1* (fully methylated), *cMet* (partial methylation) and *DDX18* (unmethylated), using either the single high salt elution (MBD-SF) or the step-wise elution series (Elution 1–6). All three gene promoters are unmethylated in PrEC cells and are eluted in the unbound fractions (**Fig. 1B**). In LNCaP cells, the *GSTP1* CpG island promoter DNA is fully methylated and is specifically eluted in the MBD-SF; in the elution series it is predominantly eluted in MBD-Elu5 (1 M NaCl) with negligible elution in lower salt elution steps. In contrast, the *cMet* promoter, which is heterogeneously methylated in LNCaP cells (**Fig. 1C**), is observed in both the unbound fraction and the MBD-SF fraction. In the elution series, the less densely methylated alleles are eluted in MBD-Elu2 (350 mM NaCl) and increasing amounts of *cMet* DNA are eluted with each step-wise elution with most of the DNA eluted at MBD-Elu5 (1 M NaCl). Finally, *DDX18*, which is unmethylated in both PrEC and LNCaP cells, is only present in the unbound fractions, thus demonstrating the specificity of the approach.

Identification of *differentially* methylated regions (DMRs) comparing MeDIP and MBDCap. To compare MeDIP and MBDCap methods in their ability to identify regions of DNA that were differentially methylated, we performed enrichment of methylated DNA from prostate cancer LNCaP and normal PrEC cells using both approaches. We compared DNA captured from Elu5 (MBD-Elu5) of the step-wise elution series and the single elution MBD-SF with MeDIP DNA using Affymetrix Human Promoter 1.0R arrays. The enriched methylated DNA for each method and corresponding input DNA were whole genome amplified, labeled and hybridized to the tiling arrays. Promoter array analysis was performed as described in the Materials and Methods and a mean t-statistic was calculated for each probe. An FDR analysis was performed on the probe-level tiling array data (see Materials and Methods) to control the expected percentage of false discoveries at 5% resulting in a smoothed t-statistic cut-off of >+4 or <-4, which was selected to identify regions of putative differential DNA methylation (**Table 1**).

For the comparison, we restricted our study to chromosome 7 since there are no CNVs (copy number variations) that could interfere with the DNA methylation data (unpublished data; Cancer Genome Project[22]). A total of 1,030 promoter regions on chromosome 7 were interrogated on the Affymetrix promoter tiling arrays of which 342 (sub-) regions were deemed as differentially methylated by one or more of the methylation enrichment methods (**Sup. Table S2**). Of these differentially

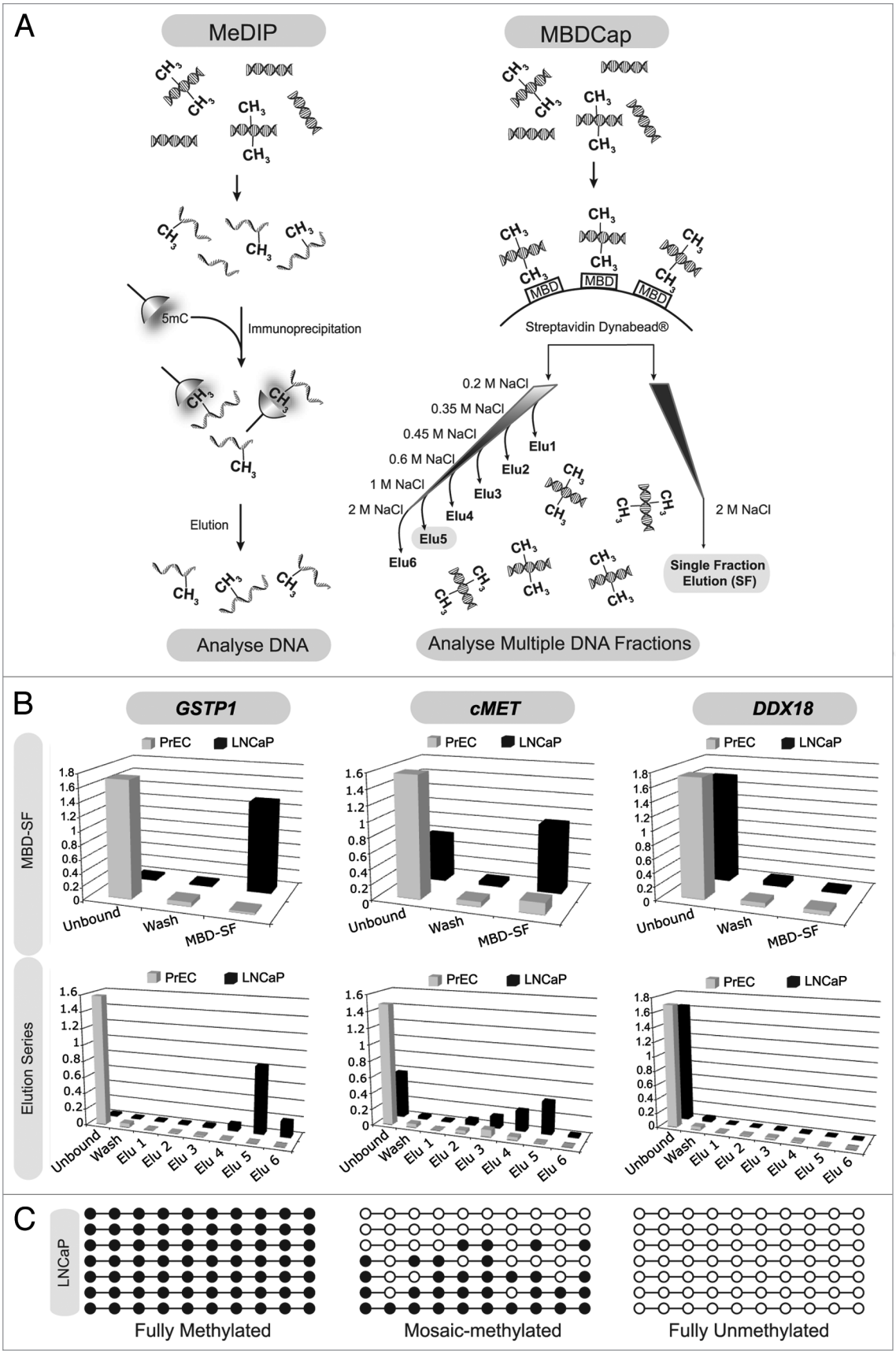

Figure 1. See figure legend on page 37.

Figure 1. See opposite page. Summary of MeDIP and the MBD-based approach for enrichment of methylated DNA. (A) Flow chart of MeDIP and MBD-based enrichment methods and the elution strategies for the MBD-approach. (B) qPCR validation of candidate genes enriched by the different elution strategies of the MBD-based approach. Elution of *GSTP1* (methylated), *cMet* (partially methylated) and *DDX18* (unmethylated) was quantitated by qPCR and normalized to input DNA. The unbound fraction, the first wash and the salt-eluted fractions were assayed as shown for LNCaP and PrEC. (C) Schematic representation of DNA methylation of the candidate genes in LNCaP cells by clonal bisulfite methylation sequencing. Black and white circles indicate methylated and unmethylated CpG sites respectively and each row represents an LNCaP clone. PrEC clones are unmethylated in all the candidate genes.

Table 1. Differentially methylated regions selected for validation by Sequenom analysis showing the corresponding t-stat score for each enrichment method, the difference in average methylation scores (LNCaP-PrEC) and the CpG density

DMR	Cooordinates	Size (bp)	Nearest gene	Assay with significant t-stat score (> +4 or < -4)	MeDIP t-stat score	MBD-Elu5 t-stat score	MBD-SF t-stat score	Coordinates validation assay	Methylation ratio (difference: LNCaP – PrEC)	CpG density
1	chr7:33909711-33910793	1082	BMPER	MBD-Elu5	3.27	**6.19**	3.89	chr7:33910570-33910788 (a)	0.18	19.5
								chr7:33910162-33910434 (b)	0.27	13.8
								chr7:33909922-33910188 (c)	0.11	9.5
2	chr7:89585495-89586037	542	---	MBD-Elu5	-0.60	**5.08**	0.75	chr7:89585509-89585795 (a)	0.47	7.5
								chr7:89585773-89585878 (b)	0.62	16.2
3	chr7:27137452-27137810	358	HOXA4	30/06/10	-0.28	**4.84**	-0.99	chr7:27137460-27137646	0.27	9.1
4	chr7:97198482-97199434	952	TAC1	MBD-Elu5	3.49	**5.30**	2.57	chr7:97198914-97199173	0.39	13.2
5	chr7:100420433-100435430	14997	MUC12	MBD-Elu5	0.81	**-15.32**	-1.60	chr7:100433446-100433693 (a)	-0.38	7.1
								chr7:100422958-100423226 (b)	-0.5	9.9
6	chr7:18033489-18034509	1020	PRPS1L1	MBD-Elu5	-3.15	**-5.63**	-0.46	chr7:18033834-18034089	-0.34	13.0
7	chr7:89586887-89587568	681	DPY19L2P4	MBD-Elu5 + MBD-SF	2.49	**7.36**	**5.11**	chr7:89586982-89587215	0.58	11.2
8	chr7:5534226-5535239	1013	ACTB	MBD-Elu5 + MBD-SF	2.97	**6.50**	**5.95**	chr7:5534481-5534766 (a)	0.11	16.6
								chr7:5534742-5535017 (b)	0.53	7.3
9	chr7:28965048-28965827	779	KIAA0644	MBD-Elu5 + MBD-SF	4.03	**5.84**	**4.65**	chr7:28965300-28965566	0.37	4.0
10	chr7:143587199-143588171	972	OR2A7	MBD-Elu5 + MBD-SF	0.53	**5.61**	**4.90**	chr7:143587872-143588058 (a)	NA	12.0
								chr7:143587316-143587581 (b)	NA	11.4
11	chr7:142433811-142434326	515	OR9A2	MBD-Elu5 + MBD-SF	-2.52	**-5.14**	**-5.81**	chr7:142434016-142434310	-0.41	2.5
12	chr7:100387405-100389557	2152	---	MBD-Elu5 + MBD-SF	-0.98	**-8.23**	**-9.28**	chr7:100387679-100387970	-0.52	7.9
13	chr7:100556609-100557540	931	SERPINE1	MBD-SF	3.88	2.64	**5.35**	chr7:100557030-100557256 (a)	0.26	8.1
								chr7:100557231-100557411 (b)	0.39	8.3
14	chr7:23591790-23592234	444	CLK2P	MBD-SF	-0.43	1.67	**4.23**	chr7:23592024-23592263	NA	11.8
15	chr7:55052872-55053563	691	EGFR	MBD-SF	2.30	1.91	**4.70**	chr7:55053280-55053540	0.79	5.4
16	chr7:43763603-43764420	817	BLVRA	MBD-SF	2.24	1.17	**4.62**	chr7:43764207-43764395	0.56	10.2
17	chr7:93390462-93390852	390	GNG11	MBD-SF	-0.64	-0.84	**-5.22**	chr7:93390504-93390765	-0.63	3.9
18	chr7:43124735-43125192	457	LOC100127950	MBD-SF	-0.85	-1.36	**-4.91**	chr7:43124648-43124933	-0.33	3.9
19	chr7:16471313-16472356	1043	SOSTDC1	MeDIP	**5.46**	0.05	3.22	chr7:16471990-16472268	0.49	6.2
20	chr7:16810995-16811635	640	AGR2	MeDIP	**5.10**	1.23	3.00	chr7:16811005-16811275	0.46	4.4
21	chr7:94376217-94376788	571	PPP1R9A	MeDIP	**4.68**	1.72	1.32	chr7:94376295-94376635	0.52	2.2
22	chr7:12694517-12695177	660	ARL4A	MeDIP	**4.42**	-0.60	0.86	chr7:12694468-12694747 (a)	0.28	1.3
								chr7:12694913-12695168 (b)	0.26	1.6
23	chr7:32074593-32075253	660	---	MeDIP	**-5.03**	0.07	-3.28	chr7:32074732-32075013 (a)	-0.3	5.0
								chr7:32074927-32075170 (b)	-0.18	5.8
24	chr7:15568393-15569676	1283	TMEM195	MeDIP	**-6.17**	-1.01	-3.41	chr7:15568884-15569233	-0.75	2.5
25	chr7:13996231-13997980	1749	ETV1	MeDIP + MBD-Elu5	**5.67**	**8.92**	3.19	chr7:13996580-13996841 (a)	0.27	9.4
								chr7:13997119-13997397 (b)	0.24	11.0
								chr7:13997457-13997609 (c)	0.2	15.5
26	chr7:28184310-28185059	749	JAZF1	MeDIP + MBD-Elu5	**5.19**	**6.68**	3.32	chr7:28184600-28184963	0.1	2.5
27	chr7:100300816-100301152	336	TRIP6	MeDIP + MBD-Elu5	**4.37**	**6.55**	1.84	chr7:100300861-100301025	0.62	6.7
28	chr7:26159833-26160264	431	NFE2L3	MeDIP + MBD-Elu5	**4.63**	**4.75**	1.72	chr7:26159982-26160251	0.26	4.3
29	chr7:92058451-92058987	536	FAM133B	MeDIP + MBD-Elu5	**-4.90**	**-4.90**	-3.01	chr7:92058557-92058895	-0.3	3.2
30	chr7:63531785-63532111	326	---	MeDIP + MBD-Elu5	**-4.54**	**-4.91**	-2.22	chr7:63531988-63532088	-0.54	1.6
31	chr7:148025462-148026216	754	CUL1	MeDIP + MBD-Elu5 + MBD-SF	**6.63**	**8.97**	**4.97**	chr7:148025779-148026007	0.77	8.0
32	chr7:136678272-136678958	686	PTN	MeDIP + MBD-Elu5 + MBD-SF	**5.68**	**8.36**	**5.21**	chr7:136678631-136678917	0.75	11.9
33	chr7:27184985-27185364	379	HOXA10	MeDIP + MBD-Elu5 + MBD-SF	**4.47**	**7.84**	**4.57**	chr7:27185230-27185362	0.42	8.0
34	chr7:79603330-79604083	753	GNAI1	MeDIP + MBD-Elu5 + MBD-SF	**5.47**	**7.11**	**5.01**	chr7:79603302-79603462	0.73	7.9
35	chr7:141234535-141235738	1203	---	MeDIP + MBD-Elu5 + MBD-SF	**-5.92**	**-7.78**	**-7.28**	chr7:141234793-141234984	-0.34	8.1
36	chr7:63206141-63206546	405	---	MeDIP + MBD-Elu5 + MBD-SF	**-4.54**	**-8.14**	**-6.81**	chr7:63206273-63206496	-0.46	3.4
37	chr7:18501735-18502428	693	HDAC9	MeDIP + MBD-SF	**4.78**	1.41	**5.06**	chr7:18501894-18502184	0.59	5.3
38	chr7:126820777-126821368	591	ZNF800	MeDIP + MBD-SF	**5.51**	2.68	**5.07**	chr7:126820899-126821150	0.67	6.8
39	chr7:92299045-92299765	720	CDK6	MeDIP + MBD-SF	**5.20**	3.15	**6.85**	chr7:92299521-92299760	0.51	2.6
40	chr7:143685636-143686467	831	OR2A9P	MeDIP + MBD-SF	**5.16**	-0.81	**4.84**	chr7:143520850-143521143	0.51	4.6
41	chr7:31347378-31347896	518	NEUROD6	MeDIP + MBD-SF	**-6.58**	-1.72	**-5.40**	chr7:31347560-31347773	-0.2	1.7
42	chr7:150179321-150179908	587	ABP1	MeDIP + MBD-SF	**-4.47**	-3.90	**-5.54**	chr7:150179590-150179860	-0.38	4.4

NA, not available—assays failed. (a), (b), (c),-independent Sequenom PCR within the same region.

methylated regions, 198 regions were hypermethylated in LNCaP (t-stat > +4) and 144 were hypomethylated in LNCaP (t-stat <-4) compared to PrEC. Only 26 regions were identified as differentially methylated by all three methylation enrichment methods tested (**Fig. 2A**), while 67, 101 and 57 regions were identified as differentially methylated exclusively by MeDIP, MBD-Elu5 and MBD-SF, respectively. Sixteen regions were called by both MeDIP and MBD-Elu5, 29 by MBD-Elu5 and MBD-SF and 46 regions by both MeDIP and MBD-SF (**Fig. 2A**). In **Figure 2B** we show a representative DMR example selected from each sector of the Venn diagram. We show the tiling array signal that was obtained with each enrichment method for LNCaP and PrEC and the t-stat score (LNCaP-PrEC) for that region. The calling of a DMR was not consistent between all methods of enrichment; for example, Chr7 (DMR19) was clearly detected by MeDIP (t-stat 5.46), but not detected by MBD-Elu5 (t-stat 0.045) and in contrast Chr7 (DMR2) was clearly detected by MBD-Elu5 (t-stat 5.07), but not detected by MeDIP (t-stat -0.60), whereas Chr7 (DMR37) was clearly detected by MeDIP (t-stat 4.77) and MBD-SF (t-stat 5.06) but not by MBD-Elu5 (t-stat 1.41). Notably Chr7 (DMR31) was detected as a methylated region by all three enrichment techniques (all t-stat scores above 4.0).

Validation of potential differentially methylated regions by sequenom analysis. To validate the different enrichment techniques and to investigate the advantages and limitations of each enrichment technique, we selected a total of 42 DMRs, six DMRs from each sector of the Venn diagram (**Table 1**), for detailed DNA methylation analysis. We used the Sequenom MassArray System and designed primers at positions in regions where the three methods showed the best t-stat score (**Fig. 2B**). **Figure 2C** shows examples of Sequenom methylation validation data for each representative DMR and all confirm hypermethylation in the LNCaP cells. The results of the Sequenom assay for all 42 DMR regions selected from each of the enrichment protocols are shown in **Table 1** (**Sup. Fig. S1**) together

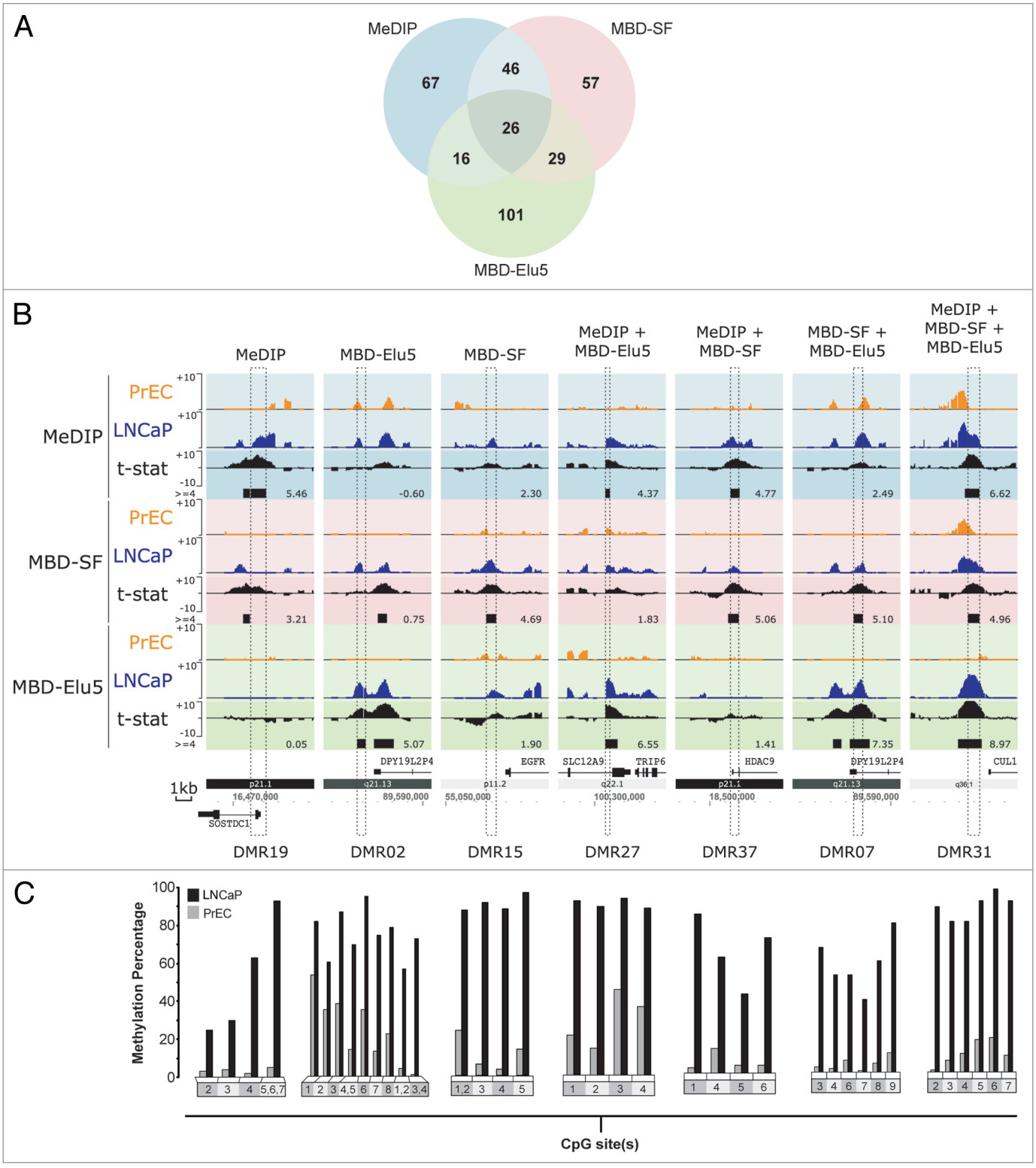

Figure 2. Differentially methylated regions (DMRs) identified from Affymetrix GeneChip Human Promoter tiling arrays analysis. (A) Venn diagram showing the number of regions that were identified as DMRs on chromosome 7 with a t-stat score of >+4 or <-4 by each of the three enrichment methods. Each group has been indicated in a different color. (B) Integrated Genome Browser (IGB) visualization showing the methylation status of representative regions in LNCaP and PrEC, that were identified as differentially methylated by MeDIP, MBD-SF and MBD-Elu5. The t-stat score for the difference (LNCaP-PrEC) is indicated, with black bars highlighting regions with t-stat score >/= +4. Dotted boxes indicate where the Sequenom primers were designed for validation. (C) Quantitative DNA methylation Sequenom MALDI-TOF analysis of the regions for which the IGB plot is shown. The bar plot indicates the average methylation ratio expressed as a 'methylation percentage' of individual CpG units in PrEC and LNCaP for the same representative regions.

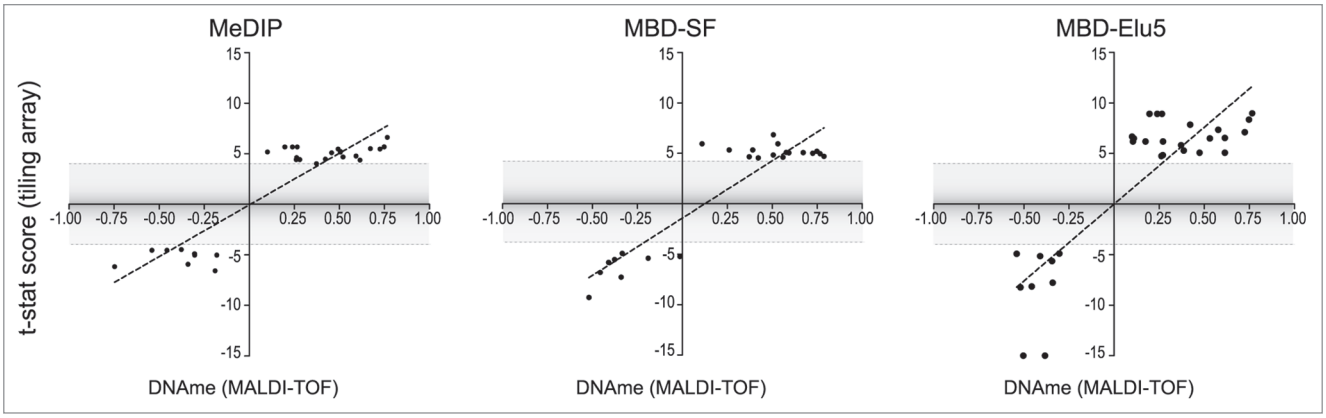

Figure 3. Correlation scatter plot indicating the correlation between tiling array t-statistic score (y-axis) and Sequenom methylation ratios (x-axis) obtained for the validated regions for each of the DNA methylation enrichment methods used. The correlation coefficients of 0.903, 0.916 and 0.865 were obtained for MeDIP, MBD-SF and MBD-Elu5, respectively.

with the corresponding t-stat score for all three methods. The Sequenom data revealed that 40/42 the regions that were identified as DMRs by having a t-stat score >+4 or <-4 by any of the enrichment methods were differentially methylated by greater than 10% in the cancer cell DNA (**Table 1**). Scatter plots of the scores obtained from the Sequenom methylation assays (LNCaP-PrEC) and the tiling array t-stat score (>+4 or <-4) of the corresponding DMRs for each enrichment method are shown in **Figure 3**. These plots illustrate that there are no false positives, with no data points present in the upper left or lower right quadrants, thus demonstrating that the assays are sensitive for detecting DMRs and that our false discovery control has been successful. The scatter plots show strong correlation between the t-stat score and actual DNA methylation difference of the DMRs that were validated for the selection enrichment method used.

CpG density analysis of differentially methylated regions. Most regions that we validated as DMRs had not been identified by all three enrichment techniques (**Fig. 2A**). This prompted us to investigate the corresponding CpG densities and DNA methylation density of the detected regions. The CpG densities for the DMR regions were calculated as described in Materials and Methods where a CpG density of 12 or greater represents a CpG island (**Fig. 4A**). We determined the CpG density of all the DMRs on chromosome 7 that had a t-stat score >+4 or <-4 from each enrichment method (**Fig. 4B**). For the hypermethylated regions (t-stat >+4) we observe that the regions uniquely called by the MeDIP procedure were generally of a low to intermediate CpG density score, similar to the CpG density range of the MBD-SF. In contrast, the DMRs picked by MBD-Elu5 have a broader range of CpG densities with regions of higher CpG density present. In the hypomethylated regions (i.e., t-stat <-4) we observed a similar trend with the overall CpG density being slightly lower for MeDIP and MBD-SF, while MDB-Elu5 has a broader CpG density range. Interestingly, the hypomethylated regions have significantly lower CpG density than the hypermethylated regions (two-tailed Mann Whitney test for MeDIP: $p < 0.0001$; MBD-SF: $p < 0.0001$; MBD-Elu5: $p = 0.0007$). These results suggest that most regions that are hypomethylated in LNCaP have a lower CpG density compared to those regions that are hypermethylated.

Next, we estimated the differential sensitivity of each enrichment approach in relation to the CpG density. In other words, how good is a certain technique at identifying a DMR within a given CpG density range? From the 42 validated regions and our FDR analysis, we can assume that the vast majority of the detected regions on chromosome 7 are truly differentially methylated. To this end, we used all 342 DMRs that were identified on chromosome 7 by one or more of the methylation enrichment methods and calculated the proportion of DMRs that were correctly identified as such by each method, in groups of CpG density (**Fig. 4C and D**). This revealed that MeDIP most efficiently identified DMRs in the low CpG density ranges, more efficiently than the MBD-approaches. At CpG densities 2.6–3.3, MeDIP and MBD-SF show similar sensitivity levels. All three enrichment techniques perform equally well at CpG densities of 4.1–5.2, but at CpG densities of greater than 5.2, MBD-Elu5 has the greatest sensitivity and clearly most efficiently detects CpG islands (which have a CpG density of above 12). Overall, the MeDIP and MBD-SF methods identified DMRs in a relatively broad range of CpG density, with MeDIP performing especially well at the regions of lowest CpG densities and MBD-Elu5 best at identifying regions with high CpG density calls.

DNA methylation differences of DMRs. We were interested to assess the magnitude of the DNA methylation differences that were required to identify a DMR and relate this to the particular enrichment and elution method used, as well as CpG density. From the Sequenom methylation data we determined the absolute methylation of each DMR in LNCaP and PrEC DNA. In **Figures 5A and B** we show DNA methylation differences for the validated regions that were hypermethylated in LNCaP relative to PrEC, while **Figures 5C and D** highlight the DNA methylation differences of hypomethylated DMRs. All three techniques identified DMRs with similar methylation differences (Kruskal-Wallis test $p > 0.05$), with an overall median

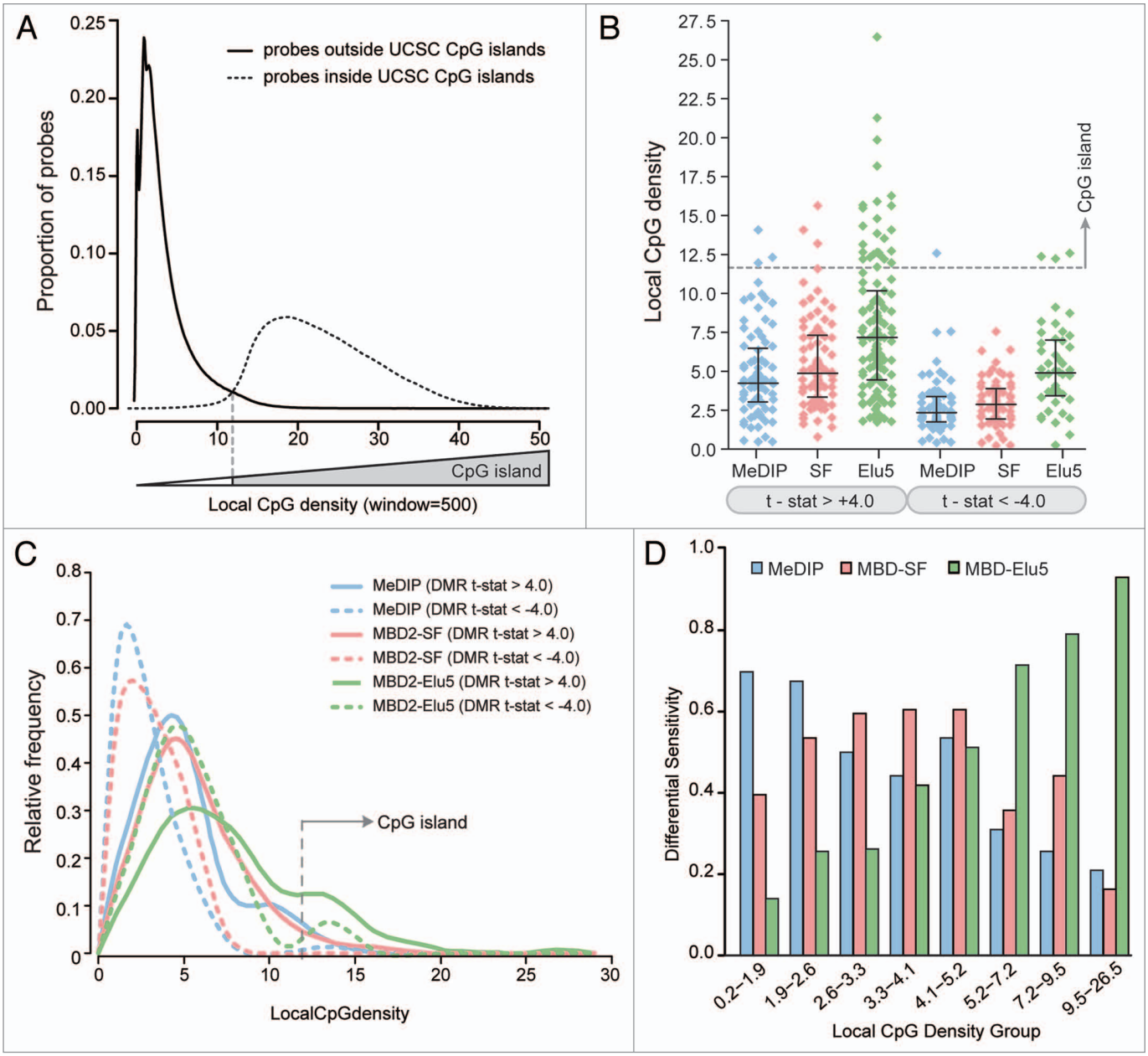

Figure 4. CpG density analysis of differentially methylated regions. (A) Plot showing the distribution of the local CpG density near probes that are either positioned inside or outside CpG islands. (B) CpG density of the DMRs called by each enrichment method. The y-axis shows the CpG density and the x-axis is grouped into two sections, one showing DMRs picked by the three methods that have a t-stat >+4 (hypermethylated regions) and other showing DMRs picked by the three methods that have a t-stat <-4 (hypomethylated regions) (SF = MBD-SF and Elu5 = MBD-Elu5). (C) Histogram of the distribution of CpG densities at DMRs identified by the three different methods. Hypermethylated regions (t-stat >4.0; solid lines) and hypomethylated regions (t-stat <-4.0; dashed lines) are shown as separate groups. (D) The differential sensitivity relative to CpG density of each enrichment approach. We show the proportion of differentially methylated regions that are detected in a given assay against any assay in fixed CpG density bins (equally sized groups). All the 342 DMRs regions were split into eight equally-sized groups. The differential sensitivity (y-axis) is between 0 and 1.

of 0.38 and a median absolute deviation of 0.13. DNA methylation differences down to ~20% can be readily and confidently identified in all the techniques using t-stat cut off 4.0 (**Fig. 5B and D**). We show that all enrichment techniques behave in a similar fashion with regard to the absolute methylation differences being detected, with CpG density having the greatest influence on the range of DMRs being identified by each technique (shown in **Fig. 4**).

Discussion

DNA methylation profiling across the genome is now possible with a variety of techniques.[13] Many of these techniques are being used to provide DNA methylation maps for normal cells and to delineate methylation changes that occur in disease phenotypes and to provide potential molecular biomarkers for diagnosis or disease prognosis.[20,23-26] Unlike other studies,[27,28] we

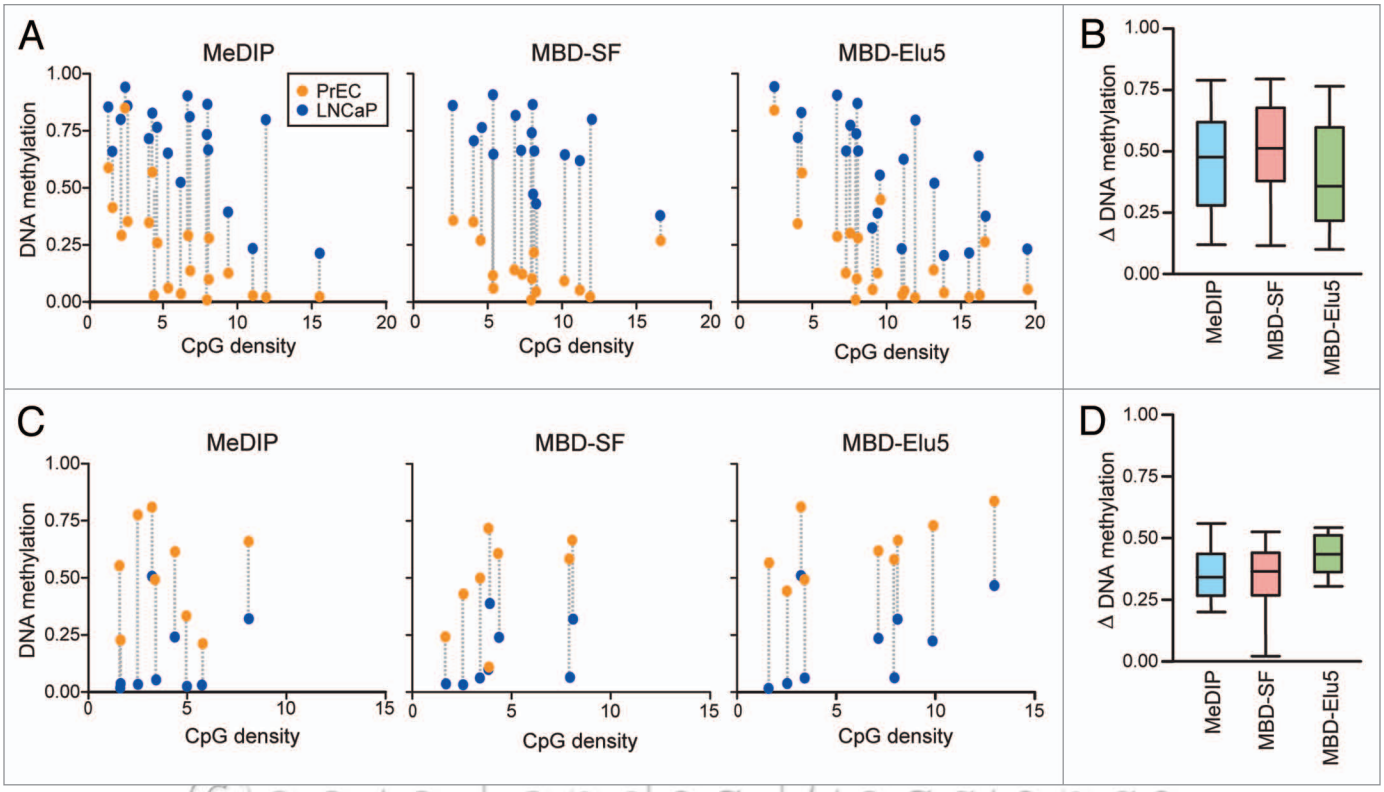

Figure 5. DNA methylation differences at DMRs identified by each technique. PrEC and LNCaP DNA methylation ratio values determined by Sequenom analysis are connected by a grey dashed line for each DMR and ordered by CpG density of the DMR (x-axis). (A) displays hypermethylated and (C) hypomethylated DMRs. (B and D) show box plots of DNA methylation differences of validated DMRs from each technique.

have compared two commonly used approaches for enrichment of methylated regions for genome-wide array and sequencing studies, namely MeDIP and MBDCap in their ability to identify differ-entially methylated regions between prostate normal and cancer cells. MeDIP uses an antibody to 5-methyl-cytosine, using single-stranded DNA, while the MBDCap approach uses the methyl-CpG binding domain of the MBD2 protein to capture methylated double-stranded DNA using different salt elutions. Following enrichment the methylated DNA was amplified and hybridized to promoter tiling arrays to compare their ability to identify differentially methylated regions. We used a stringent t-statistic cut-off to identify DMRs and performed an in-depth validation study of these DMRs using Sequenom MassCLEAVE analysis to evaluate the sequence and methylation properties identified with each approach.

We identified 342 regions as differentially methylated on chromosome 7 by one or more of the methylation enrichment techniques employed and validated the methylation status of 42 regions using Sequenom analysis that were detected by one, both or all three enrichment strategies. Using a stringent t-statistic cut-off, we found that 95% of the regions identified were differentially methylated, thus validating our false discovery rate control. However, as some regions were not detected by one or other of the enrichment techniques we determined the CpG density of the DMRs regions and found that MeDIP favors regions of low CpG density, MBD-SF covers a broad range of CpG densities, while MBD-Elu5 favors high CpG density regions. These findings suggest that comparisons of methylation datasets across different platforms may be difficult and may need to be informed of the relative sensitivities across the spectrum of CpG densities. The role of DNA methylation at regions of lower CpG density,[29] such as in gene bodies and CpG island shores,[7] has an important role in transcriptional regulation of non-coding RNAs and miRNAs, and we would suggest that MeDIP may be the method of choice in identifying methylation changes in regions of low CpG density. On the other hand, MBD-Elu5 would be the preferable method for interrogating methylation changes in CpG islands with a higher CpG density. We speculate that MeDIP is more sensitive at capturing regions of lower CpG density because of the initial DNA denaturing step prior to immunoprecipitation. That is, the CpG dense fragments either do not undergo complete denaturation to single stranded molecules and therefore do not bind effectively to the antibody or they have a propensity to re-anneal after denaturation. Alternatively, the antibody to 5MeC is more sensitive to fragments containing single methylated CpG sites.

Methylated DNA capture based on affinity to the methylated binding domain of MBD2 is based on a completely different premise and can be fine-tuned for interrogating regions with different DNA methylation densities. Since the affinity of MBD for methylated DNA is modulated by ionic strength, fractionation of the captured DNA with gradual changes in ionic

strength enables DNA to be isolated according to the degree of methylation. MBD-Elu5 of the fractionation series shows greater DNA methylation at high CpG densities compared to both MeDIP and MBD-SF, where a broad range of CpG densities and DNA methylation are detected. Other enrichment techniques, similar to the MBD2 affinity based enrichments, have now been developed, based on the multimerized MBD1 domains,[30] and MBD2b and MBD3L complexes (MIRA; reviewed in refs. 18 and 19). These assays are also based on salt elutions, and so are also likely to behave similarly to the MBD2-based capture that is described here.

Detection of DMRs after enrichment of methylated DNA can also be assessed by next-generation sequencing.[15,26,27] While microarrays have inherent drawbacks associated with amplification of CpG rich regions and probe hybridization of CpG-rich regions, it is clear that next-generation sequencing too has bias artifacts associated with GC content and some regions of the genome that are difficult to interrogate (unpublished data; reviewed in ref. 13). Next-generation sequencing is still expensive for many laboratories and both MeDIP and MBDCap coupled with promoter arrays are an excellent approach to exploring and comparing methylomes. In any enrichment-based platform, detecting absolute methylation levels is difficult and challenging,[27] as these results are biased by CpG density, GC density and amplification and array hybridization effects; instead, we show that detecting differential methylation levels is informative and less prone to artifacts or false positive results. Therefore, we recommend the use of these assays in a comparative manner (for example tumor versus normal or basal versus treated).

This study is the first detailed validation report comparing different methylated DNA enrichment techniques for analyzing differential DNA methylation. We show that both enrichment techniques are sensitive for detecting DMRs and preferentially identify different CpG rich regions of the prostate cancer genome, with MeDIP commonly enriching for methylated regions with a low CpG density, while MBD capture favors regions of higher CpG density and identifies the greatest proportion of CpG islands. Our study highlights the importance of understanding the nuances of the current enrichment methods used for DNA methylome studies so that accurate interpretation of the biology is not overlooked.

Materials and Methods

Cell lines and culture conditions. LNCaP prostate cancer cells were cultured as described previously.[31] Normal prostate epithelial cells PrEC (Cambrex Bio Science Cat. No. CC-2555) were cultured in Prostate Epithelial Growth Media (PrEGM Cambrex Bio Science Cat. No. CC-3166) according to the manufacturer's instructions.

Methylation profiling by MeDIP. The MeDIP assay was performed on 4 μg of sonicated genomic DNA (300–500 bp) in 1x IP buffer (10 mM sodium phosphate pH 7.0, 140 mM NaCl and 0.05% Triton X-100). Ten μg of anti-5-methylcytosine mouse monoclonal antibody (Calbiochem clone 162 33 D3 Cat No. NA81) was incubated overnight in 500 μl 1x IP buffer and the DNA/antibody complexes were collected with 80 μl Protein A/G PLUS agarose beads (Santa Cruz sc-2003).[14,27] The beads were washed three times with 1x IP buffer at 4°C and twice with 1 ml TE buffer at room temperature. The immune complexes were eluted with freshly prepared 1% SDS, 0.1 M $NaHCO_3$ and the DNA was purified by phenol/chloroform extraction, ethanol precipitation and resuspended in 30 μl H_2O. Input samples were processed in parallel.

Isolation of methylated DNA by MBDCap. The MethylMiner™ Methylated DNA Enrichment Kit (Invitrogen) was used to isolate methylated DNA. One μg of genomic DNA was sonicated to 100–500 bp. MBD-Biotin Protein (3.5 μg) was coupled to 10 μl of Dynabeads M-280 Streptavidin according to the manufacturer's instructions. The MBD-magnetic bead conjugates were washed three times and resuspended in 1 volume of 1x Bind/Wash buffer. The capture reaction was performed by the addition of 1 μg sonicated DNA to the MBD-magnetic beads on a rotating mixer for 1 h at room temperature. All capture reactions were done in duplicate. The beads were washed three times with 1x Bind/Wash buffer. The bound methylated DNA was eluted in one of two ways: (1) as a single fraction (MBD-SF) with a single High Salt Elution Buffer (2,000 mM NaCl) or (2) in a step-wise elution series using an increasing NaCl concentration of the elution buffer from 200 mM to 2,000 mM in a stepwise gradient [i.e.: Elution 1-(200 mM), Elution 2-(350 mM), Elution 3-(450 mM), Elution 4-(600 mM), Elution 5-(1,000 mM) and Elution 6-(2,000 mM)]. Elution 5 was denoted as MBD-Elu5. Each fraction was concentrated by ethanol precipitation using 1 μl glycogen (20 μg/μl), 1/10 volume of 3 M sodium acetate, pH 5.2 and 2 sample volumes of 100% ethanol and resuspended in 60 μl H_2O.

Whole genome amplification and promoter array analyses. Immunoprecipitated DNA and input DNA from MeDIP immunoprecipitations and MBDCap reactions, was amplified with GenomePlex Complete Whole Genome Amplification (WGA) Kit (Sigma Cat. No. WGA2) according to the manufacturer's instructions, as described previously.[32] Fifty ng of DNA was used in each amplification reaction. The reactions were cleaned up using cDNA cleanup columns (Affymetrix #900371) and 7.5 μg of amplified DNA was fragmented and labeled according to Affymetrix Chromatin Immunoprecipitation Assay Protocol (P/N 702238 Rev. 3). Affymetrix GeneChip Human Promoter 1.0R arrays (P/N. 900777) were hybridized using the GeneChip Hybridization wash and stain kit (P/N 900720). The array data for the MeDIP analysis has been submitted to Gene Expression Omnibus (GEO Series GSE19726). The array data for the MBDCap analysis is currently in submission to the Gene Expression Omnibus.

Promoter array statistical analysis. The array data was collected as biological duplicates and normalised using Model-based Analysis of Tiling-arrays (MAT).[33] For each probe (and each enrichment method), a moderated t-statistic for the $[IP_{LNCaP} - Input_{LNCaP}] - [IP_{PrEC} - Input_{PrEC}]$ contrast was computed using the R package limma.[34] At each probe, a trimmed mean of the t-statistics was also computed (window = 600 bp, 10% trim, using at least 10 probes). Next, the order of the probes along the genome

was permuted and a trimmed mean of t-statistic was calculated, which represented the null distribution. A false discovery rate (FDR) analysis was performed to find the expected percentage of false discoveries beyond a given cut off, for which the permuted data was used. The selected cut off was as +4 or -4, with an estimated FDR of 5%. Based on this cut off, regions of chromosome 7 were identified as regions of putative differential DNA methylation. Affymetrix promoter array signals and t-statistic scores were visualized using Integrated Genome Browser (IGB-Affymetrix).

Calculation of local CpG density. We use the definition of local CpG density given by Pellizola et al.[35] with a window of 500 bp. Briefly, the local CpG density is a weighted count of CpG sites in the genome upstream and downstream 500 bases from a given point of interest (e.g., microarray probe location). Weight decreases linearly from 1 at the center of the point of interest to 0 at 500 bases up or downstream. The score is a reflection of the number of CpG sites in close proximity to the point of interest and is a measure of GC content. The median CpG density score for a CpG island (as defined by UCSC) was calculated and a CpG density of 12 or greater represents a CpG island.

DNA extraction and bisulfite treatment. DNA was extracted from PrEC and LNCaP cell lines using the Puragene extraction kit (Gentra Systems). Bisulfite treatment was carried out on 2 μg of DNA as described previously.[36]

Quantitative massARRAY methylation analysis. Sequenom MassARRAY methylation analysis was performed as described previously.[21] Two μg of DNA extracted from PrEC and LNCaP cell lines were bisulfite treated using the standard bisulfite protocol.[36] As controls for the methylation analysis, whole genome amplified (WGA) DNA (0% methylated) and M.SssI treated DNA (100% methylated) were bisulfite treated in parallel. The primers were designed using the EpiDesigner^{BETA} software from Sequenom (see **Sup. Table S1** for sequences). Each reverse primer has a T7-promoter tag (5-CAG TAA TAC GAC TCA CTA TAG GGA GAA GGC T-3) and each forward primer has a 10-mer tag (5-AGG AAG AGA G-3). A total of 50 primer pairs were designed for the 42 regions at positions in regions where the three methods showed the best t-stat score (**Fig. 2B**). Upon testing these primers on bisulfite treated DNA, all the primers gave specific PCR products at a Tm of 60°C. In order to check for potential PCR bias towards methylated or non-methylated sequences, we used serological DNA (Millipore) as a 100% methylated control and whole genome amplified human blood DNA as a 0% methylated control. The PCRs were optimized and performed in triplicate using the conditions: 95°C for 2 min, 45 cycles of 95°C for 40 sec, 60°C for 1 min and 72°C for 1 min 30 sec and final extension at 72°C for 5 min. After PCR amplification, the triplicates were pooled and a shrimp alkaline phosphatase (SAP) treatment was performed using 5 μl of the PCR product as template. 2 μl of the SAP-treated PCR product was taken and subjected to in vitro transcription and RNaseA Cleavage for the T-cleavage reaction. The samples were purified by resin treatment and spotted on a 384-well SpectroCHIP by a MassARRAY Nanodispenser. This was followed by spectral acquisition on a MassARRAY Analyzer Compact matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. The results were then analyzed by the EpiTYPER software V 1.0, which gives quantitative methylation levels for individual CpG sites. The average methylation ratio was calculated by averaging the ratios obtained from each CpG site from both LNCaP and PrEC and calculating the difference between them. Methylation readings that had other signal overlaps and silent peaks were eliminated from the calculation.

Acknowledgements

We thank Kate Patterson for help with preparation of the figures and critical reviewing of the manuscript and Rebecca Hinshelwood for critical reviewing of the manuscript. We thank the Ramaciotti Centre, University of NSW (Sydney, Australia) for array hybridizations. This work is supported by National Health and Medical Research Council (NH&MRC) project (427614, 481347) and Fellowship (S.J.C.), Cancer Institute NSW grants (CINSW: S.J.C.; A.L.S.) and NBCF Program Grant (S.J.C.) and ACRF.

Authors' Contributions

S.J.C. initiated and supervised the study and, with M.W.C., C.S. and S.N., designed the experiments and wrote the paper; S.N., M.W.C., C.S. and J.Z.S. performed the experiments and S.N., A.L.S., D.S. and M.D.R. performed data analysis.

Note

Supplementary materials can be found at:
www.landesbioscience.com/supplement/NairEPI6-1-sup.pdf

References

1. Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res 1982; 10:2709-21.
2. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA 1992; 89:1827-31.
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001; 409:860-921.
4. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science 2001; 291:1304-51.
5. Antequera F, Bird A. Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci USA 1993; 90:11995-9.
6. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 2002; 16:6-21.
7. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet 2009; 41:1350-3.
8. Issa JP. CpG-island methylation in aging and cancer. Curr Top Microbiol Immunol 2000; 249:101-18.
9. Vucic EA, Brown CJ, Lam WL. Epigenetics of cancer progression. Pharmacogenomics 2008; 9:215-34.
10. Antequera F, Boyes J, Bird A. High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. Cell 1990; 62:503-14.
11. Jones PA, Wolkowicz MJ, Rideout WM, 3rd, Gonzales FA, Marziasz CM, Coetzee GA, et al. De novo methylation of the MyoD1 CpG island during the establishment of immortal cell lines. Proc Natl Acad Sci USA 1990; 87:6117-21.
12. Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, Herman JG. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. Hum Mol Genet 2001; 10:687-92.
13. Laird PW. Principles and challenges of genome-wide DNA methylation analysis. Nat Rev Genet 2010; 11:191-203.
14. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 2005; 37:853-62.
15. Serre D, Lee BH, Ting AH. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res 2010; 38:391-9.
16. Klose RJ, Sarraf SA, Schmiedeberg L, McDermott SM, Stancheva I, Bird AP. DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. Mol Cell 2005; 19:667-78.

17. Brinkman AB, Simmer F, Ma K, Kaan A, Zhu J, Stunnenberg HG. Whole-genome DNA methylation profiling using MethylCap-seq. Methods 2010; 52: 232-6.
18. Rauch T, Pfeifer GP. Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. Lab Invest 2005; 85:1172-80.
19. Rauch TA, Pfeifer GP. The MIRA method for DNA methylation analysis. Methods Mol Biol 2009; 507:65-75.
20. Gebhard C, Schwarzfischer L, Pham TH, Schilling E, Klug M, Andreesen R, et al. Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. Cancer Res 2006; 66:6118-28.
21. Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res 2007; 35:119.
22. [CGP] Cancer Genome Project. [updated 2007 Dec 9]. CGP LOH and Copy Number Analysis. [cited 2009 Dec 14]. Available from: www.sanger.ac.uk/research/projects/cancergenome.
23. Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, Segal E, et al. Evidence for an instructive mechanism of de novo methylation in cancer cells. Nat Genet 2006; 38:149-53.
24. Hayashi H, Nagae G, Tsutsumi S, Kaneshiro K, Kozaki T, Kaneda A, et al. High-resolution mapping of DNA methylation in human genome using oligonucleotide tiling array. Hum Genet 2007; 120:701-11.
25. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 2008; 454:766-70.
26. Ruike Y, Imanaka Y, Sato F, Shimizu K, Tsujimoto G. Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. BMC Genom 11:137.
27. Li N, Ye M, Li Y, Yan Z, Butcher LM, Sun J, et al. Whole genome DNA methylation analysis based on high throughput sequencing technology. Methods 2010; 52:203-1.
28. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell 2006; 126:1189-201.
29. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 2009; 41:178-86.
30. Jorgensen HF, Adie K, Chaubert P, Bird AP. Engineering a high-affinity methyl-CpG-binding protein. Nucleic Acids Res 2006; 34:96.
31. Song JZ, Stirzaker C, Harrison J, Melki JR, Clark SJ. Hypermethylation trigger of the glutathione-S-transferase gene (GSTP1) in prostate cancer cells. Oncogene 2002; 21:1048-61.
32. Coolen MW, Stirzaker C, Song JZ, Statham AL, Kassir Z, Moreno CS, et al. Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. Nat Cell Biol 2010; 12:235-46.
33. Johnson WE, Li W, Meyer CA, Gottardo R, Carroll JS, Brown M, et al. Model-based analysis of tiling-arrays for ChIP-chip. Proc Natl Acad Sci USA 2006; 103:12457-62.
34. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 2005; 21:2067-75.
35. Pelizzola M, Koga Y, Urban AE, Krauthammer M, Weissman S, Halaban R, et al. MEDME: an experimental and analytical methodology for the estimation of DNA methylation levels based on microarray derived MeDIP-enrichment. Genome Res 2008; 18:1652-9.
36. Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res 1994; 22:2990-7.