Comprehensive high-throughput arrays for relative methylation (CHARM)

Rafael A. Irizarry, Christine Ladd-Acosta, Benilton Carvalho, et al.

*Genome Res.* published online Mar 3, 2008;
Access the most recent version at doi:10.1101/gr.7301508
Comprehensive high-throughput arrays for relative methylation (CHARM)

Rafael A. Irizarry,1,3 Christine Ladd-Acosta,2 Benilton Carvalho,1 Hao Wu,1 Sheri A. Brandenburg,2 Bo Wen,2 and Andrew P. Feinberg2,3

1Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, USA; 2Department of Medicine and Center for Epigenetics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

This study was originally conceived to test in a rigorous way the specificity of three major approaches to high-throughput array-based DNA methylation analysis: (I) MeDIP, or methylated DNA immunoprecipitation, an example of antibody-mediated methyl-specific fractionation; (II) HELP, or HpaII tiny fragment enrichment by ligation-mediated PCR, an example of differential amplification of methylated DNA; and (III) fractionation by McrBC, an enzyme that cuts most methylated DNA. These results were validated using 1466 Illumina methylation probes on the GoldenGate methylation assay and further resolved discrepancies among the methods through quantitative methylation pyrosequencing analysis. While all three methods provide useful information, there were significant limitations to each, specifically bias toward CpG islands in MeDIP, relatively incomplete coverage in HELP, and location imprecision in McrBC. However, we found that with an original array design strategy using tiling arrays and statistical procedures that average information from neighboring genomic locations, much improved specificity and sensitivity could be achieved, e.g., ∼100% sensitivity at 90% specificity with McrBC. We term this approach “comprehensive high-throughput arrays for relative methylation” (CHARM). While this approach was applied to McrBC analysis, the array design and computational algorithms are fractionation method-independent and make this a simple, general, relatively inexpensive tool suitable for genome-wide analysis, and in which individual samples can be assayed reliably at very high density, allowing locus-level genome-wide epigenetic discrimination of individuals, not just groups of samples. Furthermore, unlike the other approaches, CHARM is highly quantitative, a substantial advantage in application to the study of human disease.

[Supplemental material is available online at www.genome.org.]
Table 1. Current methods for DNA methylation analysis

<table>
<thead>
<tr>
<th>No. of sites interrogated</th>
<th>No. of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–10</td>
<td>10^2</td>
</tr>
<tr>
<td>10–100</td>
<td>10^3</td>
</tr>
<tr>
<td>100–1000</td>
<td>10^{3–4}</td>
</tr>
<tr>
<td>&gt;10^6</td>
<td>(Cost limiting)</td>
</tr>
</tbody>
</table>

chemical conversion of cytosine to uracil by sodium bisulfite or metabisulfite, followed by PCR (which incorporates T for U), and then DNA sequencing. While providing single-base resolution, the cost is the highest of all the commonly used methods, tens of thousands of dollars for a megabase of sequence data, itself comprising 40,384 CpG dinucleotides assayed (Eckhardt et al. 2006), and is therefore not currently suitable for whole-genome analysis on multiple samples. (2) A variety of methods that interrogate specific single-CpG dinucleotides or amplicons. These include MethylLight (Eads et al. 2000), COBRA (Xiong and Laird 1997), bisulfite pyrosequencing (Dupont et al. 2004), and the Illumina GoldenGate methylation assay (Bibikova et al. 2006). While sensitive, specific, and relatively inexpensive, none of these methods is suitable for analysis of the whole genome, which includes ~28 million CpG dinucleotides. (3) Microarray-based methods. These can interrogate much larger numbers of CpG than the other approaches, at extremely low unit cost, since the pricing is similar to other non-methylation-based array methods.

There are four major types of microarray-based methylation analysis. (1) Direct hybridization to CpG island arrays. This was one of the earliest methods; it was used to provide valuable data on tumor-type classification, for example (Gitan et al. 2002), and it still remains a useful discovery tool. However, its earliest developers have migrated away from this approach, since it requires presupposition about the potentially methylated sequences. (2) Methylated DNA immunoprecipitation (MeDIP), in which methylated DNA is fractionated using an antibody and then hybridized, with a differentially labeled total DNA control, to an oligonucleotide array (Weber et al. 2005). (3) Restriction enzyme digestion using methylcytosine-sensitive enzymes, followed by ligation-mediated PCR amplification of the targets. The paradigm of this method is the HELP (HpaII tiny fragment enrichment by ligation-mediated PCR) assay (Kuhlman et al. 2006). DNA is digested in parallel with MspI (resistant to DNA methylation), and then the HpaII and MspI products are amplified by ligation-mediated PCR and hybridized using separate fluorochromes to a customized array. As HpaII sites comprise 8% of CpG, that represents a fixed limit of sensitivity of the method. Alternatively, the restriction enzyme-digested DNA can be directly sequenced rather than hybridized to microarrays (Allinen et al. 2004), although one is still limited by the relatively small number of methylcytosine-sensitive restriction sites in the genome. (4) Restriction enzyme digestion of methylated DNA using MecRBC, without PCR, and differential hybridization to an array. DNA is digested with MecRBC, an enzyme with the unusual and desirable property of cutting methylated DNA promiscuously (recognition sequence R^mC(N)_{5,6}−_{10}R^mC), cleaving half of the methylated DNA in the genome and all methylated CpG islands (Sutherland et al. 1992). The enzyme is used on size-selected (1.5–4.0 kb) DNA to fractionate unmethylated (i.e., gel-purified high molecular weight) DNA after digestion, which is comparatively (two-color) hybridized with DNA similarly processed but not cut with the enzyme, on high density arrays. The original method was developed for Arabidopsis (Lippman et al. 2004), where it has value in eliminating the large fraction of methylated repetitive DNA in the plant genome. For mammalian genome application, a selection algorithm has been applied to obtain specific array probes thought to represent the state of a given methylation target (Ordway et al. 2006).

Although all of the microarray approaches are in common use, they have not been directly compared to each other, and our original goals were relatively modest: to directly compare methods using the same DNA samples and the same arrays. However, we found significant limitations generally to hybridization-based methylation analysis that could largely be overcome with novel statistical procedures and array design algorithms. As will be described in the second portion of the paper, a fractionation method-independent approach, termed CHARM (Comprehensive High-throughout Arrays for Relative Methylation), can detect DNA genome-wide methylation with ~100% sensitivity and 90% specificity.

Results

Overall design

Here we have designed a study to compare three array-based methylation detection technologies, MeDIP as an example of immunoprecipitation-based methods, MecRBC fractionation as an example of restriction enzyme fractionation, and HELP as an example of differential methylcytosine sensitive ligation-mediated PCR. As our test samples, two paired cell lines were used: HCT116, a highly methylated colorectal carcinoma line, and a DNA methyltransferase I and 3B double-knockout cell line (DKO), with comparatively low levels of methylation (Rhee et al. 2002). These data were also compared to direct bisulfite methylation analysis using the Illumina GoldenGate methylation assay (Bibikova et al. 2006) on 1466 CpG sites in 466 genes.

For all three assay types, design-specific arrays have already been designed, and we followed these designs, referred to here as canonical arrays. However, to enable direct comparison on the same arrays, samples were hybridized to NimbleGen’s Promoter 2 array and designed two tiling arrays (see Methods), which are referred to herein as common arrays. Note that in the case of MeDIP, one of the common arrays is the same as the canonical array, i.e., the Promoter 2 array. Because of the flexibility of de-
sign, the NimbleGen platform was used in all cases, which has also been used by the originators of these assay systems. In each case, a competitive hybridization approach was performed, in which samples were differentially labeled with Cy3 and Cy5 as described in the experimental protocols, specifically: (1) for MeDIP, methyl-enriched DNA with Cy5 and total DNA with Cy3; (2) for HELP, Hpall amplified with Cy5 andMspI with Cy3; and (3) for McrBC, methyl-depleted with Cy5, and total with Cy3. Note that McrBC dye-swaps were created as recommended by the original publication for mammalian DNA (O’Dwyer et al. 2006). However, we found that the benefit of dye-swaps does not merit their extra cost (see Supplemental Fig. 1), thus the comparisons shown here do not include them. The complete list of comparative experiments and arrays is provided in Table 2.

To decide among various strategies for measuring the same quantity, one looks to optimize sensitivity and specificity. Because specificity can be easily improved at the cost of sensitivity, and vice versa, one needs to assess both independently. We designed our experiments to assess sensitivity and specificity in the practical context of detecting methylated sites. To appropriately assess how experimental variability affects specificity, two technical replicates were performed for each method/sample-type pair (see Table 2). Measurements of methylation should be the same in both replicates, and deviation from equal values serves a measure of precision, which directly affects the specificity for measurements of methylation levels. The assessment of specificity was also facilitated by the use of the DKO samples. These provided many unmethylated sites useful for this assessment: Methods with low specificity will be more likely to call unmethylated sites as methylated. The HCT116 samples permitted a comprehensive assessment of sensitivity as many sites were methylated: Methods with high sensitivity will be more likely to call methylated sites as methylated (true positives). The Illumina GoldenGate assay was used as a reference against which all microarray methods were compared.

Quantification of methylation measurements

For each microarray hybridization, we used the raw feature intensities to form log ratios and denoted these with $M_i$ where $i$ represents a feature, and 1 and 2 represent the two replicate hybridizations. In principle, these values should all be 0, since $M_{1i}$ and $M_{2i}$ were measures of the same quantity. However, as expected, differences were observed due to natural variation in the sample preparation and array hybridization. These differences were studied using the canonical arrays for each method, because each method was likely optimized on their canonical arrays and we wanted to see each method at its optimal condition in addition to the common arrays.

We first assessed the precision of each method, by comparing $M$-values from replicate arrays, specifically studying the distribution of the differences between replicated $M$-values: $M_{1i} - M_{2i}$ where 1 represents a feature, and 1 and 2 represent the two replicate hybridizations. In principle, these values should all be 0, since $M_{1i}$ and $M_{2i}$ were measures of the same quantity. However, as expected, differences were observed due to natural variation in the sample preparation and array hybridization. These differences were studied using the canonical arrays for each method, because each method was likely optimized on their canonical arrays and we wanted to see each method at its optimal condition in addition to the common arrays.

The standard deviation (SD) of these differences, taken across probes, is a useful summary that relates directly to the range of $M$-values one should expect from samples with no difference in methylation status. For McrBC, the standard deviations (SDs) were 0.20 and 0.15 for the DKO and HCT116 samples, respectively (Table 3). For the HELP method, the SD was 0.27 for both samples. Finally, the MeDIP showed the worst precision with SDs of 0.55 and 0.60 for the DKO and HCT116 samples, respectively (Table 3). For the HELP method, the SD was 0.27 for both samples. Finally, the MeDIP showed the worst precision with SDs of 0.55 and 0.60 for the DKO and HCT116 samples, respectively (Table 3). These results were for the $M$-values obtained from the canonical arrays. A graphical assessment of precision is shown in Supplemental Figure 2.

### Table 2. Microarray characteristics

<table>
<thead>
<tr>
<th>Array name</th>
<th>Canonical method</th>
<th>Genomic regionsa</th>
<th>Probesb</th>
<th>Probe duplicatesc</th>
<th>Distanceđ</th>
<th>Sparsitye</th>
<th>HCT116 replicates</th>
<th>DKO replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ogha1</td>
<td>McrBC</td>
<td>21,143</td>
<td>1</td>
<td>3–4</td>
<td>50,000</td>
<td>∞</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>HELP_Promoter</td>
<td>HELP</td>
<td>25,625</td>
<td>14–15</td>
<td>1</td>
<td>27,000</td>
<td>50</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Promoter2</td>
<td>MeDIP</td>
<td>12,892</td>
<td>1–300</td>
<td>1</td>
<td>175,000</td>
<td>110</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Imprint_tiling</td>
<td>HELP</td>
<td>23</td>
<td>1148–122,572</td>
<td>1</td>
<td>2–43 million</td>
<td>47</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>CHARM</td>
<td>HELP</td>
<td>43,987</td>
<td>1–306</td>
<td>1</td>
<td>500–50,000</td>
<td>35</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

*aTotal number of genomic regions represented.

*bNumber of probes for each region.

*cTotal number of identical probes.

*dMean genomic distance, in base pairs, between genomic regions.

*eMean distance between probes in a region; infinite for Ogha1 which has only one probe per region.

*fTwo each of McrBC, HELP, and MeDIP.

*gOne each of McrBC, HELP, and MeDIP.

*hOne each of McrBC and MeDIP.
The standard deviation (SD), computed across probes, of the difference between methylation measurements of replicate arrays was used to quantify precision, with a lower number representing greater precision.

McrBC and HELP can discriminate DKO from HCT116

A global assessment of sensitivity was performed by comparing the distribution of the M-values from the HCT116 and DKO samples, i.e., a highly methylated and a highly unmethylated reference sample, respectively. Thus the expected M-values for DKO sample should mostly be centered at 0, and HCT116 should be shifted to a substantial number of positive values. Figure 1 demonstrates that the MedDIP method can barely distinguish between the two cell lines of differing methylation on a global scale, although at individual loci differences are clearly seen (discussed below). The McrBC and HELP arrays perform better at globally distinguishing the DKO from the HCT116 sample, with HELP to a somewhat greater degree.

Site-specific comparison of methods

The ability to distinguish sample methylation globally is not nearly as important as the ability to detect methylation at high genomic resolution. We therefore compared the performance of each method at the individual CpG level using the Illumina platform as reference standard, based on studies from us and others (Bibikova et al. 2006; Ladd-Acosta 2007), as well as data from this study shown in Supplemental Figure 3. For each method/array combination, each CpG assayed by the Illumina platform was matched to one M-value obtained from the microarray data. We now describe how this mapping was obtained for each method. For the McrBC method, we predicted the start and end of the resulting genomic segments after cutting at every ACG or GCG. These are referred to as the McrBC segments. For each probe on the Illumina platform representing an ACG or GCG we assigned two segments: those ending and starting on that CG. Next, for each probe on the microarray, we determined which McrBC segment contained it. Finally, the median M-value for all the microarray probes mapped to each Illumina probe was assigned as the microarray M-value. The McrBC canonical arrays used three to four replicate probes for 21,143 locations, as recommended by those authors (Ordway et al. 2006). Thus, at least four probes were used with each Illumina probe. A similar approach was used for the HELP method except the cleavage occurs at CCGG sites (Khulan et al. 2006). The HELP canonical arrays used 14–15 tiled probes in each of the HELP segments. The canonical arrays for MeDIP were the Promoter 2 arrays, which represent 12,892 promoter regions. We matched every CpG inside these promoter regions to the closest probe also in that region. We were then able to map CpGs represented by an Illumina probe, and included in one of the promoter regions, to one microarray probe M-value. Figure 2 demonstrates this comparison, using S87, 57, 51, and 1188 Illumina CpGs corresponding to specific CpGs on the MeDIP, McrBC, HELP, and CHARM arrays, respectively. The set of CpG covered by all platforms was too small to provide meaningful results, thus we based our comparisons on the different sets mapped by the different array types.

Sensitivity of HELP and MeDIP depends greatly on the CpG content

Figure 2 plots M-values from each of the microarray platforms against the corresponding M-values obtained from the Illumina platform. Values from the HCT116 and DKO samples were combined. For clarity, in Figure 2, data are shown from one HCT116 and one DKO array for each method. Results for all other arrays, i.e., the replicates, are similar and are shown in Supplemental Figure 4. Figure 2 stratifies points by CpG density. The observed-to-expected ratio for 500-bp regions was computed around each microarray probe shown in Figure 2 (ratios are denoted with color and with a small number inside each point). In this window we defined the expected number of CpGs as the proportion of Cs multiplied by the proportion of Gs. The observed-to-expected ratio is simply the proportion of CpGs divided by the expected proportion of CpGs. Notice that the traditional definition of a CpG island requires this ratio to be >0.6. The probes were stratified into two groups: low CpG density (ratio < 0.6) and high CpG density (ratio > 0.6). A regression line was fitted to each group (shown as red and blue lines for the low- and high-density groups, respectively). The correlation between Illumina M-values and microarray M-values is shown in Table 4. While McrBC showed similar sensitivity for both high- and low-density groups, HELP showed better sensitivity for the lower CpG density group than for the higher CpG density group.

Severe bias in current methods related to segment characteristics

For HCT116 samples, we stratified the M-values obtained from the McrBC and HELP canonical arrays by segment size to produce Figure 3A. Because in this sample one expects many methylated CpGs, many large M-values are expected independent of the segment size. However, the strata related to large and small fragments had substantially fewer large M-values than the middle-sized segments. Notice in particular that the HELP method had no sensitivity for CpGs associated with segments smaller than 300 bp. The McrBC method had no sensitivity for CpGs associated with segments larger than 1500 bp. Best results were observed for segments of sizes 200–600 and 700–1200 bp for McrBC and HELP, respectively. The segment sizes for MeDIP are unpredictable, thus, this method was not included in this figure.

We also assessed the effect of CpG density with this strati-
General limitations in single-CpG accuracy substantially improved by genome-weighted smoothing

Figure 2 also demonstrates that, at the individual CpG level, the agreement between microarray and Illumina reference measurements leaves much room for improvement. Notice that even for the best performing microarray-based method, McrBC, the variability seen in the microarray M-values suggests that none of the methods will be useful in practice if one uses individual probe level data or individual segment data. In particular, notice that a substantial number of the M-values for the CpGs called methylated by the reference standard (Illumina, right of the dashed vertical line) are in the same range as most of the M-values called unmethylated by the reference standard (Illumina, left of the dashed vertical line), i.e., between −0.75 and 0.75 on the Y-axis.

The fact that the methylation status of neighboring CpGs tends to be highly correlated (Eckhardt et al. 2006) motivated our introduction of a novel strategy for methylation analysis of genome-weighted smoothing: averaging probes within small contiguous genomic regions taking into account the biases illustrated in Figure 3. A novel aspect of our approach is that we combine information derived from the genome sequence with microarray data. By characterizing each of the segments induced by laboratory protocols, one can quantify the utility of the associated microarray data. This information is then used to adapt the averaging used in the smoothing step by assigning weights. Details on our novel smoothing strategy are provided in the Methods section.

Figure 4 demonstrates the advantage of genome-weighted smoothing. In this figure, M-values are plotted against location on the genome. The points are the M-values observed for each probe. The averaged M-values for probes in the same McrBC and HELP segments are shown with orange and green lines for McrBC and HELP, respectively. The results obtained using genome-weighted smoothing (described above) are shown with black curves. Note that for the McrBC and MeDIP methods, the range of the probe-level and segment M-values associated with unmethylated (Fig. 4A) and methylated (Fig. 4B) regions overlap; the results from smoothing do not. For example, for McrBC the segment M-values range from −0.75 to 0.5 and from −0.75 to 3 for the unmethylated and methylated regions, respectively. The values obtained from smoothing range from −0.2 to 0.25 and from 0.6 to 2.5 for the unmethylated and methylated regions, respectively. The averaging performed in the smoothing pr
dure greatly reduces noise, and the fact that the averaging is local, i.e., performed in small regions, permits us to preserve the ability to discriminate. Supplemental Figure 5 shows examples of various other regions illustrating the value of this approach.

The HELP method sometimes produced contradictory results at the same loci that were not apparent in the canonical design but were easy to see in the common array design (Fig. 4; Supplemental Fig. 5). This likely explained the lack of agreement with the reference method (Fig. 2). Because the HELP segments are small for the region shown in Figure 4, this result was expected, as Figure 3 demonstrates that the HELP method is not sensitive for small fragments. Supplemental Figure 5 shows several other examples.

CHARM, comprehensive high-throughput arrays for relative methylation

Based on the data described above, and in particular the importance of genome-weighted smoothing and array design, we have developed a novel platform for array-based DNA methylation analysis. The new method is independent of platform, and it combines the design of a novel array design and statistical procedures that perform genome-weighted averaging from neighboring genomic locations. More details are provided below.

The first component of our method is a new tiling array specifically designed to maximize the number of assayed CpGs.

Table 4. Correlation between microarray platforms and Illumina GoldenGate reference data (these are computed from the points shown in Fig. 2)

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall*</th>
<th>High ratiob</th>
<th>Low ratioa</th>
</tr>
</thead>
<tbody>
<tr>
<td>McrBC</td>
<td>0.63–0.68</td>
<td>0.55–0.64</td>
<td>0.61–0.69</td>
</tr>
<tr>
<td>HELP</td>
<td>0.48–0.50</td>
<td>0.32–0.36</td>
<td>0.57–0.60</td>
</tr>
<tr>
<td>MeDIP</td>
<td>0.21–0.38</td>
<td>0.30–0.43</td>
<td>0.03–0.29</td>
</tr>
<tr>
<td>CHARM</td>
<td>0.76</td>
<td>0.80</td>
<td>0.62</td>
</tr>
</tbody>
</table>

*aRange of correlations between microarray and Illumina $M$-values. The range is over all replicates. 
*bRange of correlations between microarray and Illumina $M$-values for probes within regions with observed-to-expected CpG ratios >0.6. 
*aRange of correlations between microarray and Illumina $M$-values for probes within regions with observed-to-expected CpG ratios <0.6.

Figure 3. DNA fragment-length–related biases. (A) $M$-values for the HCT116 sample are stratified by the DNA fragment size predicted by the McrBC (left panel) and HELP (right panel) enzyme digestions. (B) For all three methods, a 500-bp window was formed around each probe, the observed-to-expected ratio of CpG was calculated, and box-plots of the $M$-values are displayed by these ratios. Only probes related to fragments of sizes between 50 and 600 bp for McrBC, and between 600 and 1200 bp for HELP, are included.
For the reasons stated above, we did not want to restrict our attention to CpG islands. Instead, the number of CpGs assayed, for which we could reliably detect methylation status, were maximized. For example, because we rely on smoothing, isolated CpGs were not assayed. A careful analysis of different numbers of probes included in the smoothing demonstrated that at least 15 probe intensities were needed to obtain useful results (data not shown). The procedure for creating the array was as follows:

1) We identified all the CpGs in the genome. Any region of 300 bp with no CpGs was discarded.
2) We removed probes with multiple matches, including fuzzy matches as defined by NimbleGen (http://www.nimblegen.com/products/chip/index.html), to the genome.
3) Any region having a gap of \( \geq 300 \) bp between consecutive probes was divided into two new regions.
4) We discarded any region with fewer than 15 probes.
5) We tiled regions as possible, using 50-mers 35 bp apart. One can also prioritize for economy to limit to a single array by calculating the ratio of CpGs per probes in the region and by assigning higher priority to those with a higher ratio.

This array design would improve the detection strategy for any of the methods because it facilitates the smoothing strategy and assays many more CpGs. Probes associated with problematic segments (e.g., very small segments in the HELP assay) could be removed in the analysis stage. However, we selected McrBC for the application of this approach because of its superior sensitivity and specificity described earlier. Going forward, samples were also hybridized using the CHARM design as well as the MeDIP assay as well. We did not continue to use the HELP assay mainly because of its limited number of detectable sites (HpaII dependence).

To detect methylated regions in the CHARM method, the \( M \)-values were normalized, as described in the Methods section, and processed using genome-weighted smoothing, as described above. Figure 2D shows the smoothed \( M \)-values obtained from CHARM plotted against the reference \( M \)-values. Comparing Figure 2D with Figure 2, A–C, demonstrates how CHARM greatly improved the results obtained with the other methods.

Although it is potentially useful to treat methylation state as a continuous variable (Rakyan et al. 2004), the state of individual CpGs is strongly bimodal. Therefore, besides comparing quantitative results among methods, it is also important to also determine the ability to discriminate highly methylated from highly unmethylated sequences, a common question in molecular biology, e.g., in generating lists of candidate genes subject to epige-
netic regulation or alteration in disease. This binary classification also enabled us to construct receiver operating characteristic (ROC) curves (Fig. 5). ROC curves plot the sensitivity vs. (1 – specificity) for a binary classifier system (methylated or not) as discrimination thresholds (values of M) are varied. For this purpose, a genomic region was defined as “methylated” if all probes from the Illumina platform in the region were >90%. Similarly, unmethylated regions were defined as those with all probes <10%; 100 Illumina probe sets fulfilled these criteria. If the smoothed M-value within any of these regions was above a predetermined threshold, the region was considered methylated. Various thresholds were considered, and each defines a point in the ROC curve. The results greatly improved with CHARM. Notice that for a specificity of 90%, the McrBC sensitivity improved from 60% without CHARM to 100% with CHARM.

Finally, we note that the CHARM method, unlike MeDIP, HELP, or nonsmoothed McrBC, is highly quantitative, meaning that there was a linear relationship between methylation measured on the array and the reference methylation platform (Illumina), as shown clearly in Figure 2. The correlation coefficient comparing these two values was substantially better for CHARM compared to the other methods (Table 4), as was the ROC curve (Fig. 5).

Discussion

In summary, there are two major results of this work. First, we have shown that there are substantial limitations to all three commonly used approaches for array-based DNA methylation analysis. In the case of MeDIP, the assay is of relatively worse specificity, and the method is not sensitive, particularly outside of CpG islands. HELP, while accurately distinguishing markedly different cell types globally, does not cover many CpG dinucleotides because of the dependence on HpaII restriction sites and often shows lack of agreement with the reference method. Of the three approaches, McrBC performed the best, but as seen in the ROC curves, the sensitivity was only 60% at 90% specificity as previously practiced. Second, since neighboring CpGs have been shown to be closely correlated, we developed a genome-weighted smoothing algorithm to measure methylation from raw microarray data. Combining this novel approach with the most robust method for fractionating methylated DNA (McrBC), we designed custom arrays ideally suited for methylation detection, as defined in the Results section. This approach is termed “comprehensive high-throughput arrays for relative methylation” (CHARM). CHARM offers the possibility of relatively inexpensive genome-wide analysis with high precision and accuracy. On the NimbleGen HD2 arrays, 2.1 million features can be studied in this way.

The approach was data-driven, in that it used an independent assessment of 1466 CpG sites. Furthermore, the genome coverage on the array is genome sequence-driven, rather than based on arbitrary assumptions about the likely location of methylated sites (e.g., promoters) that might miss substantial numbers of regulatory sequences. Even with this unbiased, non-promoter-driven selection strategy, 87% of the Illumina-selected methylation cancer panel 1 genes are present on the HD2 array.

What were the likely inherent limitations of MeDIP and HELP shown by these experiments? The results obtained with the MeDIP method barely distinguished the HCT116 and DKO samples. A likely reason is that the immunoprecipitation (IP) step is not specific, i.e., unmethylated CpGs pass the filter of IP. This is consistent with the observation that detection was biased toward very high CpG content. Furthermore, note that the IP sample will be enriched with CpGs regardless of the number of segments that pass the filter. This is likely to result in cross-hybridization problems, e.g., probes with more CpGs might result in higher intensities only because of cross-hybridization with the high CpG content sample. In expression arrays it has been shown that background noise, such as cross-hybridization, greatly reduces sensitivity in cases were nominal amounts of target are low (Irizarry et al. 2006). If the same phenomenon is true in methylation arrays, then one would expect low sensitivity in regions with a small number of CpGs (low amount of target) as seen in the MeDIP arrays.

While HELP outperformed other methods in distinguishing the highly methylated HCT116 from the relatively unmethylated DKO globally, at the single-CpG level, the HELP method performed barely better than MeDIP. A possible explanation for this apparent contradiction is that the HELP method depends upon differences in ability of a fragment to be amplified, but the PCR step does not always amplify as expected. For example, in dense CpG regions, the smaller pieces, which are expected to amplify, might be too small for the PCR to work properly. Evidence that this phenomenon is occurring is the fact that the microarray data for HELP sometimes appears flipped in plots, such as in Figure 4: fragments were methylation-amplified opposite from the expected. It is important to note that the canonical design for HELP carefully selects regions where this phenomenon is unlikely to occur. But as mentioned, this greatly limits the coverage of the method. More sophisticated post-processing algorithms have been and likely will be further developed to correct for measurement discrepancies (Khulan et al. 2006). However, even with minimal post-processing as done here, one can obtain very good
Concordance with reference measurements, as seen in Figure 2, and more importantly good sensitivity and specificity at detecting methylation sites as seen in the ROC curves in Figure 5 and Supplemental Figure 6. CHARM also circumvents a major problem for genomic methylation analysis, namely, the importance of detecting changes in regions of relatively low CpG content. As shown here, MeDIP cannot detect methylation in these regions. The limitations for HELP in these regions is that it is more likely to miss these regions and that one cannot smooth the data to obtain high precision, since by definition one does not observe data from adjacent segments.

McrBC fractionation was originally applied to analysis of the plant genome (Martienssen et al. 2005) and subsequently used as a discovery tool for methylated CpG islands (Ordway et al. 2006), but it has not come into common use. The original whole-genome array design represented a few thousand segments with one probe each. In microarrays and other technologies that use hybridization, a large component of the variability seen across measures from different probes is due to sequence effects (Wu and Irizarry 2005). Averaging replicate probes does not reduce this variability as they have the same sequence. Therefore the precision is relatively low in data provided by a single probe replicated four times, as performed by Ordway et al. (2006). McrBC fractionation also suffers to some degree within CpG islands in the ability to discriminate highly methylated from highly unmethylated sequences, although it still outperforms MeDIP which works only within CpG islands (Supplemental Fig. 6).

Future work on CHARM includes the development of preprocessing algorithms that correct for sequence and segment effects. The resulting methods should improve the performance of CHARM within CpG islands. Finally, we note that while CHARM offers state-of-the-art, cost-effective methylation analysis, approximately one penny per measurement, second-generation sequencing will reduce the relative utility of arrays generally, particularly when the goal of a $1000 genome is met. Currently, complete genome coverage by second-generation sequencing, e.g., after bisulfite treatment, would cost hundreds of times this amount per sample. An alternative is fractionation of the DNA followed by sequencing, reducing the complexity of the target genome, as has been done for specific chromatin marks (Mikkelsen et al. 2007). These approaches hold considerable promise, but they also raise the interesting question of what to capture. The data shown here indicate that great care must be used in the capture strategy, and the CHARM assay is an important step toward that. In addition, the capture strategy also uses hybridization, so one must still deal with hybridization-related biases until complete whole-genome sequencing becomes inexpensive.

Methods

Cell culture and genomic DNA isolation

HCT116 cells (American Type Culture Collection) and DNMT1/RNMT38 (DKO) cells (Rhee et al. 2002) were cultured in McCoy’s 5A modified medium containing 10% fetal bovine serum and 1% penicillin/streptomycin. Genomic DNA was isolated from HCT116 and DKO cell lines and was prepared using the MasterPure DNA purification kit (EpiCentre) as specified by the manufacturer.

McrBC assay sample preparation

Genomic DNA (10 µg) was prepared, and McrBC digestion and gel fractionation were performed exactly as published (Ordway et al. 2006). As shown in Table 2, HCT116 and DKO samples prepared using McrBC were analyzed on the Oghal1 array (canonical), Promoter 2 array (common), Imprinting array (common), and the novel CHARM array.

HELP assay sample preparation

HCT116 and DKO samples were prepared as previously described by Khulan et al. (2006), in the laboratory of John Greally, to avoid any issues of technical handling (Albert Einstein College of Medicine, New York), using a total of 20 µg per sample (Khulan et al. 2006). The LM-PCR products were labeled with Cy3- or Cy5-conjugated oligonucleotide and random primers as previously described (Selzer et al. 2005). As shown in Table 2, HCT116 and DKO samples prepared using Hpal andMspI were analyzed on the HELP promoter array (canonical), Promoter 2 array (common), and Imprinting array (common).

MeDIP assay sample preparation

MeDIP assay was conducted according to published methods (Weber et al. 2005). As shown in Table 2, HCT116 and DKO samples prepared using MeDIP were analyzed on the Promoter 2 array (canonical), Imprinting array (common), and CHARM (common) array. As a positive control, MeDIP was validated using real-time PCR of Sat2 (Jiang et al. 2004).

Illumina GoldenGate assay sample preparation

Bisulfite conversion of 500 ng of genomic DNA was achieved through use of the EZ DNA Methylation-Gold kit (Zymo Research). All HCT116 and DKO samples were processed as described previously on the Illumina GoldenGate methylation cancer panel I (Bibikova et al. 2006). A β-value of 0–1.0 was reported signifying percent methylation, from 0% to 100%, respectively, for each CpG site. β-values were calculated by subtracting background using negative controls on the array and taking the ratio of the methylated signal intensity against the sum of both methylated and unmethylated signals.

Quantitative methylation analysis using pyrosequencing

One microgram of genomic DNA was bisulfite treated using the EpiTect kit (Qiagen) according to the manufacturer’s recommendations. Bisulfite treatment of genomic DNA results in unmethylated cytosine nucleotides being changed to thymidine while methylated cytosines remain unchanged. This difference is then detected as a C/T nucleotide polymorphism at each CpG site. CpG-unbiased primers were designed to PCR amplify 38, 16, and 14 CpG sites, respectively, in three genes, HLA-F, KCNK4, (previously known as SMARCA3), showing conflicting methylation across MeDIP, McrBC, and Illumina assays (Supplemental Table 1). Nested PCR was performed under standard conditions. Amplicons were analyzed on a PSQ HS 96 pyrosequencer as specified by the manufacturer (Biogate) and CpG sites quantified, from 0% to 100% methylation, using Pyro Q-CpG software.

Microarray design

Oghal1 is the canonical array for the McrBC method: 21,143 McrBC segments are represented by one probe each. Three to four replicate are used for each probe. The locations of these segments were chosen by the designer based on transcriptional start sites and CpG islands as described in their paper (Ordway et al. 2006). The HELP promoter array is the canonical array for the HELP method. The Promoter 2 array is one of NimbleGen’s off-
the-shelf products, with 12,892 promoter regions. The Imprint,
re-imprinted gene array represents 23 regions chosen by our group to study
imprinted genes. Region sizes ranged from 133,475 to 13,096,022
bp, and probes of size 50 bp were tiled at 47 bp from each other
with an occasional large jump to avoid repeat elements. Table 2
provides a summary of these arrays.

Data analysis

Normalization

As described above, the basic measurement used to quantify
methylation is the log-ratio of the intensities observed in the
treated and control channels. Existing statistical methods have
used the unadjusted log-ratios, assumed linear dye log-scale ef-
fect and removed these effects by simply subtracting the median,
and fitting and removing the effect within an ANOVA model
(Kerr and Churchill 2001). These methods fail to correct for the
strong nonlinear effects seen in M versus A plots. In expression
arrays, Loess normalization has been widely and successfully
used to solve this problem (Yang et al. 2002). The basic idea in
this procedure is to assume that for most probes are not
expressed and M = 0 and that A-dependent deviations are a
smooth function. The bias is estimated and removed using Loess
regression. This approach has been successful in expression ar-
rays. However, because in methylation experiments we expect
many sites to be methylated, one can no longer make this as-
sumption. If Loess normalization were used here, and we define
M = 0 as the average value of unmethylated sites, then one would
incorrectly force M = 0 for many of the probes associated with
methylation.

We therefore developed a method that does not require the
assumption that M = 0 for most probes. This method used ge-
nome sequence information and our knowledge of the fragment
selection protocol to select what we call pseudo-housekeeping
probes for which one can in fact assume M = 0. We then apply
the Loess normalization procedure developed for expression ar-
rays to the pseudo-housekeeping genes, obtain the correction
curves, and use this curve to correct M-values for all probes. An
additional advantage of this approach is that it provides a flexible
way to adapt existing techniques, developed for expression ar-
rays, to methylation data. Details of this normalization proce-
dure are described in Bolstad et al. (2003).

Genome–weighted smoothing

To obtain a smoothed M-value at any given genomic location, we
average all the M-values that were within a prespecified distance
from the location in question. A weighted average was used with
the weights determined from the results presented in Figure 3.
Notice that this part of the procedure is method-dependent. The
interval providing the values that are averaged is referred to as
the “smoothing window” and its length is referred to as the “win-
dow size.” Many smoothing algorithms exist, each one averaging
in a different way, e.g., assigning distance-dependent weights.
For the results presented in this paper, the running median al-
gorithm (Tukey 1977; Hardle and Steiger 1995) was used, with a
window size of 51 probes (~1500 bp), due to its simplicity and
robustness to outliers.

Acknowledgments

This work was supported by NIH grant HG005233. We thank
John Greally and Masako Suzuki for performing the HELP liga-
tion and amplification steps to ensure that we did not perform
this procedure incorrectly, and Jeff Jeddeloh for performing the
McrBC digestion for the same reason.

References

Allinzen, M., Beroukhim, R., Cal, L., Brennan, C., Lahti-Domenici, J.,
Molecular characterization of the tumor microenvironment in breast
data analysis: From disarray to consolidation and consensus. Nat.
High-throughput DNA methylation profiling using universal bead
Non-methylated CpG-rich islands at the human alpha-globin locus:
Implications for evolution of the alpha-globin pseudogene. EMBO J.
6: 999–1004.
comparison of normalization methods for high density oligonucleotide
Callinan, P.A. and Feinberg, A.P. 2006. The emerging science of
Eads, C.A., Danenberg, K.D., Kawakami, K., Saltz, L.B., Blake, C.,
high-throughput assay to measure DNA methylation. Nucleic Acids
Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M.,
methylation profiling of human chromosomes 6, 20 and 22. Nat.
Genet. 38: 1378–1385.
Esteller, M. 2006. Epigenetics provides a new generation of oncogenes
Methylation-specific oligonucleotide microarray: A new potential for
modification in constitutive heterochromatin versus unexpressed
Khulian, B., Thompson, R.F., Ye, K., Fazzari, M.J., Suzuki, M., Stasiak, E.,
isochromosom profiling of cytosine methylation: The
Ladd-Acosta, C.P.J., Sabuncy, S., Yolken, R., Webster, M., Dinke, T.,
81: 1304–1315.
Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N.,
2004. Role of transposable elements in heterochromatin and
Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos,
Genome-wide maps of chromatin state in pluripotent and
Ong, P.O., Miller, W., Lehoczky, J., Leung, C.T., Birren, B., Wheelan,
analysis of the mouse 1-megabase region orthologous to the human
Ordway, J.M., Bedell, J.A., Citek, R.W., Nunberg, A., Garrido, A,


Received October 12, 2007; accepted in revised form December 24, 2007.