

# Genome-wide mapping of DNA methylation: a quantitative technology comparison

Christoph Bock<sup>1-4,6</sup>, Eleni M. Tomazou<sup>1-3,6</sup>, Arjen Brinkman<sup>5</sup>, Fabian Müller<sup>1-4</sup>, Femke Simmer<sup>5</sup>, Hongcang Gu<sup>1</sup>, Natalie Jäger<sup>1-3</sup>, Andreas Gnirke<sup>1</sup>, Hendrik G. Stunnenberg<sup>5</sup> and Alexander Meissner<sup>1-3</sup>

1 Broad Institute, Cambridge, Massachusetts 02142

2 Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138

3 Harvard Stem Cell Institute, Cambridge, Massachusetts 02138

4 Max Planck Institute for Informatics, Saarbrücken, Germany

5 Radboud University Department of Molecular Biology, Nijmegen Center for Molecular Life Sciences, Nijmegen, The Netherlands

6 equal contributions

Correspondence should be addressed to:

C.B. [cbock@broadinstitute.org](mailto:cbock@broadinstitute.org) or A.M. [alexander\\_meissner@harvard.edu](mailto:alexander_meissner@harvard.edu)

## Keywords:

Epigenome profiling, epigenetics, sequencing, differentially methylated regions, molecular diagnostics, biomarker discovery, cancer

## Abstract

DNA methylation is a key component of mammalian gene regulation and the most classical example of an epigenetic mark. DNA methylation patterns are mitotically heritable and stable over time, but they undergo considerable changes in response to cell differentiation, diseases and environmental influences. Several methods have been developed for DNA methylation profiling on a genomic scale. Here, we benchmark four of these methods on two sample pairs, comparing their accuracy and power to detect DNA methylation differences. The results show that all evaluated methods (MeDIP-seq: methylated DNA immunoprecipitation, MethylCap-seq: methylated DNA capture by affinity purification, RRBS: reduced representation bisulfite sequencing, and the Infinium HumanMethylation27 assay) produce accurate DNA methylation data. However, these methods differ in their ability to detect differentially methylated regions between pairs of samples. We highlight strengths and weaknesses of the four methods and give practical recommendations for the design of epigenomic case-control studies.

## Introduction

Twenty-five years of research on cancer epigenetics have firmly established the prevalence of aberrant DNA methylation in cancers<sup>1-8</sup>. Moreover, recent studies have investigated the role of DNA methylation for neural and autoimmune diseases, its correlation with physiological conditions and its response to environmental influences<sup>9-11</sup>. Comprehensive mapping of DNA methylation in relevant clinical cohorts is likely to identify new disease genes and potential drug targets, help establish the relevance of epigenetic alterations in diseases other than cancer, and provide a rich source for biomarker development<sup>12</sup>. In a biotechnology context, DNA methylation profiling could also facilitate quality control of cultured cells, exploiting the fact that cell states and differentiation potential of stem cells are reflected in their DNA methylation patterns<sup>13</sup>.

Several methods have been developed to enable DNA methylation profiling on a genomic scale. Most of these methods combine DNA analysis by microarrays or high-throughput sequencing with one of four ways of translating DNA methylation patterns into DNA sequence information or library enrichment: (i) Methylated

DNA immunoprecipitation (MeDIP) uses an antibody that is specific for 5-methyl-cytosine to retrieve methylated fragments from sonicated DNA<sup>14,15</sup>. (ii) Methylated DNA capture by affinity purification (MethylCap) employs a methyl-binding domain protein to obtain DNA fractions with similar methylation levels<sup>16-18</sup>. (iii) Bisulfite-based methods utilize a chemical reaction that selectively converts unmethylated (but not methylated) cytosines into uracils, thus introducing methylation-specific single-nucleotide polymorphisms into the DNA sequence<sup>13,19,20</sup>. (iv) Methylation-specific digestion uses prokaryotic restriction enzymes to fractionate DNA in a methylation-specific way<sup>21-23</sup>.

The diversity of methods and absence of an uncontested commercial market leader raise questions about each method's strengths and weaknesses – questions that researchers have to answer for themselves when selecting the most appropriate technology for any given project. The goal of this study was to perform a comprehensive benchmarking of four popular methods, with a special emphasis on their practical utility for biomedical research and biomarker development. We selected MeDIP-seq<sup>14</sup>, MethylCap-seq<sup>16</sup>, RRBS<sup>24</sup> and the Infinium HumanMethylation27 assay<sup>19</sup> for inclusion in this comparison, based on the following considerations: (i) All four methods are relatively easy to set up because detailed protocols have been published and / or commercial kits are available. (ii) We chose RRBS rather than genome-wide bisulfite sequencing because its per-sample cost are comparable to the other methods and realistic for large sample sizes. (iii) The Infinium HumanMethylation27 assay was included because of its wide use and easy integration with existing genotyping pipelines; it is the only microarray-based method in our comparison. (iv) Methods that utilize tiling microarrays have been useful in the past<sup>25-27</sup>, but were excluded from the current analysis because sequencing methods nowadays provide substantially higher genomic coverage at competitive cost. (v) Methylation-specific digestion was excluded because no algorithm exists that could accurately infer quantitative DNA methylation data from digested read frequencies<sup>21</sup>. An outline of the experimental and analytical procedure of this technology comparison is shown in Figure 1.

## Results

### *DNA methylation mapping using MeDIP, MethylCap, RRBS and Infinium*

Genome-wide DNA methylation mapping is most commonly used as a discovery tool, in order to identify differentially methylated regions (DMRs) as candidates for further research. Typical examples are cancer-specific DMRs, which play an increasing role as biomarkers for cancer diagnosis and therapy optimization<sup>28,29</sup>. To emulate the case-control approach that is widely used for epigenetic biomarker development<sup>12</sup>, our study focuses on sample pairs that we statistically compare with each other. Specifically, we selected two human embryonic stem (ES) cell lines that were derived from genetically unrelated embryos<sup>30</sup>, and a matched pair of colon tumor and adjacent normal colon tissue obtained from the same donor. We applied each of the four methods (MeDIP, MethylCap, RRBS, Infinium) to all four samples (HUES6 ES cells, HUES8 ES cells, colon tumor, matched normal colon tissue), generating a total of 16 genome-scale DNA methylation maps. All data were processed with a standardized bioinformatic pipeline, and the technical data quality turned out to be similarly high across all samples and methods (Table 1).

When plotting the DNA methylation data as genome browser tracks we found excellent visual agreement between all four methods (Figure 2 and <http://meth-benchmark.computational-epigenetics.org/>). MeDIP and MethylCap gave rise to peaks of methylated DNA that were similar in shape, size and location, indicating that MeDIP's monoclonal antibody and MethylCap's methyl-binding domain enrich for similar DNA fragments. However, MeDIP exhibited higher baseline levels and lower peak heights than MethylCap. This reduced dynamic range is already apparent from Figure 2 (note the different scale of the y-axis) and becomes more obvious when plotting MeDIP and MethylCap tracks along an entire chromosome (Supplementary Figure 1). We could also quantitatively confirm this visual observation by plotting the mean read frequency among enriched and depleted fractions of the genome (Supplementary Figure 2). The visual agreement between RRBS and Infinium was as high as between MeDIP and MethylCap, with the limitation that Infinium covers two orders of magnitude fewer CpGs than RRBS. Finally, the bisulfite-based methods (RRBS, Infinium) generally confirm the results of the

enrichment-based methods (MeDIP, MethylCap), although there are deviations in repeat-rich as well as in CpG-poor genomic regions (Supplementary Figure 3).

#### *Accuracy of DNA methylation mapping*

For a more quantitative assessment of measurement accuracy, we compared the results of the three sequencing-based methods (MeDIP, MethylCap, RRBS) with a common reference, the Infinium HumanMethylation27 assay. The Infinium assay was used as reference because its accuracy has previously been validated by comparisons with several other methods. For example, DNA methylation measurements using the Infinium assay gave rise to Pearson correlations around 0.9 when compared to MethyLight and to the GoldenGate assay<sup>19,31</sup>.

RRBS and Infinium data can be compared directly and without normalization, because both methods measure absolute DNA methylation levels. When averaging the RRBS measurements within a 200 basepair sequence window around each Infinium probe, we observed excellent agreement between the two methods (Pearson's  $r = 0.92$ ; Figure 3C). Comparison with MeDIP and MethylCap is less straightforward because both methods measure the relative enrichment of methylated DNA rather than absolute DNA methylation levels. When we correlated the number of sequencing reads per 1-kilobase region with the DNA methylation measurements of the Infinium assay, the Pearson correlation did not exceed 0.6 (Supplementary Figure 3A and B). High density of repetitive DNA was identified as the major source of spurious read enrichment in regions with low absolute DNA methylation levels, and low CpG density gave rise to low read numbers in regions with high levels of DNA methylation (Supplementary Figure 3C and D). The confounding effect of DNA sequence is also visible in Figure 2: Low read counts can indicate either the relative absence of CpGs (example: region 1 in Figure 2) or the absence of DNA methylation in the presence of CpGs (region 2), and strong peaks can occur in genomic regions that are incompletely methylated if the CpG density is sufficiently high to give rise to substantial read enrichment (region 3).

It has previously been reported that statistical correction for CpG density can improve the quantification of DNA methylation levels based on MeDIP data<sup>14,32</sup>. We therefore constructed a linear regression model that corrects for the confounding effect of DNA sequence (see Methods section for details), and we observe substantially improved results (Figure 3A and 3B). The correlation between the statistically corrected read counts and the DNA methylation measurements of the Infinium assay amounted to 0.84 for MeDIP and to 0.88 for MethylCap. Although these correlations were close to the value of 0.92 observed for RRBS, some inaccuracies remained. In particular, there was a tendency of MeDIP and MethylCap to over-estimate DNA methylation levels in regions that exhibit almost no DNA methylation. Furthermore, it became evident from our data that the single-basepair resolution of bisulfite-based methods translates into more precise DNA methylation maps: When testing various window sizes around each Infinium probe (data not shown), we found that for RRBS the best compromise between accuracy and coverage was a 200 basepair window, while a lower resolution of one kilobase per window was optimal for MeDIP and MethylCap.

#### *Genomic coverage of DNA methylation mapping*

The single-basepair resolution of the two bisulfite-based methods comes at the cost of reduced genomic coverage compared to the two enrichment-based methods: RRBS reads are clustered in approximately 1% of the genome<sup>13,33</sup> and Infinium focuses on the methylation status of slightly less than 15,000 gene promoters, while MeDIP and MethylCap reads should in theory be able to identify methylated genomic regions located anywhere in the genome. To assess the empirical genomic coverage of each method, we calculated the number of reads (MeDIP, MethylCap) or CpG methylation measurements (RRBS) for each of the following genomic regions: (i) CpG islands, (ii) gene promoters, and (iii) a 1-kilobase tiling of the genome. The results are shown in Figure 4, and coverage details for a total of 13 types of genomic regions are available online (<http://meth-benchmark.computational-epigenetics.org/>).

As expected, MeDIP and MethylCap provide broad coverage of the genome, whereas RRBS and Infinium are more restricted to CpG islands and promoter regions. However, the practically relevant differences in ge-

nomic coverage are lower than Figure 4 may suggest. This is because a minimum number of reads are required in at least one sample to reliably detect differential methylation in a given sample pair. We illustrate this point by two statistical power calculations, which were performed with G\*Power 3<sup>34</sup>. Assume that a genomic region is covered by five MeDIP or MethylCap reads in one sample. Then it has to contain at least 20 reads in the second sample to be detected as hypermethylated (assuming a statistical power of 80% and a p-value of 5% without multiple-testing correction). Similarly, RRBS would detect a DNA methylation increase from 30% to 70% only when at least 25 measurements are available in each sample (again assuming a statistical power of 80% and a p-value of 5% without multiple-testing correction).

#### *Identification of differentially methylated regions with MeDIP, MethylCap and RRBS*

Genome-wide DNA methylation mapping is most commonly used for detecting DNA methylation differences between samples, for example between diseased and healthy tissue or between genetically modified and control cells. To assess how well MeDIP, MethylCap and RRBS perform on this task, we developed a bioinformatic method that identifies statistically significant DMRs from multiple types of sequencing data (the Infinium assay requires a different approach and is discussed in a separate section below). For a pre-defined set of genomic regions we count the numbers of sequenced reads (for MeDIP and MethylCap), or alternatively the numbers of methylated vs. unmethylated CpGs (for RRBS), and we test for statistically significant differences between two samples using Fisher's exact test. When applied to a complete tiling of the human genome, this method performs genome-wide DMR detection. Alternatively, it can be targeted to specific region types such as CpG islands, gene promoters or putative enhancers, which can lead to more sensitive detection of small difference because the multiple-testing burden is reduced compared to genome-wide DMR detection. We pursued both the unbiased and the annotation-guided approach in parallel, focusing our comparison on three types of genomic regions: (i) CpG islands, (ii) gene promoters, and (iii) a 1-kilobase tiling of the genome (Figure 5, Supplementary Figures 4 to 8). In addition, DMR data for a total of 13 types of genomic regions are available online (<http://meth-benchmark.computational-epigenetics.org/>).

Overall, we observed high correlation for each of the two sample pairs, but also outliers suggesting the presence of DMRs. Based on the RRBS data we obtained Pearson correlations around 0.9 for all three region types, both between the two ES cell lines (HUES6 and HUES8) and between the colon tumor and matched normal colon tissue. For MethylCap and MeDIP, the correlations were somewhat lower and ranged from 0.75 to 0.92 (Figure 5, Supplementary Figures 4 to 8). Using the DMR detection algorithm (see Methods for details), we identified several hundred to several thousand DMRs in each of the two sample pairs. Reassuringly, there was substantial overlap between all three methods. For the two human ES cell lines, 277 out of 44,440 CpG islands were detected as differentially methylated by each of the three methods (Figure 5D), and similar comparisons for all sample pairs and region types (Supplementary Figures 4 to 8) confirmed that the pairwise overlap between all methods was always statistically significant ( $p < 0.01$ , Fisher's exact test). In total, we observed that up to 1,000 CpG islands, 405 promoter regions or 1,924 of the 1-kilobase tiling regions (i.e. less than 0.1% of the genome) were detected as differentially methylated by at least two methods. Note however that it is not possible to combine these values into a single sum of DMRs because many CpG islands overlap with promoter regions, and every CpG island and promoter region overlaps with at least one tiling region. Nor does the number of differentially methylated tiling regions provide an accurate estimate of the "true" number of DMRs because a sizable number of DMRs are not statistically significant anymore when split into 1-kilobase regions. Despite these conceptual difficulties, our data clearly indicate that – on average – MethylCap identifies more DMRs than RRBS, while MeDIP identifies the lowest number of DMRs. This order was observed not only based on the total number of DMRs per method, but is also maintained when focusing only on those DMRs that were detected by at least two methods.

#### *Validation of method-specific differentially methylated regions*

In order to pinpoint potential problems of MeDIP, MethylCap or RRBS, we manually inspected a large number of regions that were identified as significant DMRs by only one method (<http://meth-benchmark.computational->

[epigenetics.org/](http://epigenetics.org/)). We observed that many method-specific DMRs were supported by consistent but insignificant trends among the other methods, confirming that MeDIP, MethylCap and RRBS differ in their power to detect DMRs at the current sequencing depth. We investigate this point in more detail in a separate section below, by measuring the effect of sequencing depth on the performance of each method. However, we also detected some evidence of technical artifacts: MeDIP and MethylCap identified DMRs in highly repetitive regions that were often not supported by RRBS. It is likely that the high copy number of certain repeat sequences amplifies small sample-specific differences in the efficiency of methylated DNA enrichment, which may give rise to spurious DMRs. In contrast, RRBS seems more robust toward such fluctuations because it measures DNA methylation based on the DNA sequence of the reads in a given region, rather than based on their read frequency. We also assessed whether copy-number variation is a major confounding factor for DMR discovery. This does not seem to be case: The vast majority of DMRs were shorter than 10kb (Supplementary Figure 9), while it is not uncommon for cancer-specific as well as germline-transmitted copy-number variations to extend for much longer distances<sup>35,36</sup>.

As an additional validation, we selected eight method-specific DMRs based on the ES cell comparison, and we investigated DNA methylation patterns in the two ES cell lines by clonal bisulfite sequencing (Table 2). These genomic regions were hand-picked such that one method clearly identified them as DMRs while the two other methods did not show a trend in either direction. Note that this pre-selection makes the validation substantially harder than confirming randomly selected DMRs, because method-specific DMRs tend to be weaker than DMRs that are detected by several methods. As an additional complication, some of the selected DMRs are highly repetitive or overlap with known copy-number variations. Sequencing an average of 11 clones per sample and region, we were able to confirm three out of three MethylCap-specific DMRs and two out of two RRBS-specific DMRs. In contrast, two MeDIP-specific DMRs could not be confirmed, and for the third region the agreement was marginal (Table 2, Supplementary File 1).

To assess the practical relevance of the method-specific differences, we asked whether biologically interesting hits were missed by any of the three methods. For this analysis we focused on the colon samples because of the large number of genes with a known or suspected role in colon cancer. Our results show that several interesting DMRs are detected by all methods, including tumor-specific hypermethylation in the promoters of GATA2<sup>25</sup> and GATA5<sup>37</sup>. However, a significant number of interesting DMRs were missed by MeDIP, while MethylCap and RRBS both detect them. To give a few examples, this is the case for tumor-specific hypermethylation in the promoter regions of SOX17<sup>38</sup>, POU2AF1<sup>39</sup> and SEPT9<sup>29</sup>. Somewhat more rarely, we also observed interesting DMRs being missed by MethylCap or RRBS. For example, MethylCap overlooked tumor-specific hypermethylation at the promoter of SFRP1<sup>40</sup>, and RRBS missed tumor-specific hypermethylation at the promoter of DKK2<sup>41</sup>.

#### *The effect of sequencing depth on the performance of MeDIP, MethylCap and RRBS*

The three sequencing-based methods use DNA sequencing as a way of counting DNA fragments, in order to determine the percentage of methylation-enriched reads that align to specific regions (MeDIP, MethylCap) or to calculate the ratio of methylated and unmethylated cytosines at single CpGs (RRBS). Conceptually, sequencing can be thought of as random sampling from a large pool of DNA fragments. It is therefore expected that the performance of these methods increases when sequencing more DNA fragments, until it levels off as the sequencing depth approaches saturation. To quantify this effect, we repeated the accuracy analysis (Figure 3) and the DMR detection (Figure 5) on randomly sampled subsets of sequencing reads. First, we benchmarked each method against the Infinium data, assessing their ability to quantify DNA methylation levels based on reduced read numbers (Supplementary Figure 10). The results show that all three methods give rise to accurate DNA methylation measurements based on as few as 20% of the total read coverage, and almost no improvement was observed between 50% and 100% sequencing depth. While these data suggest that relatively low sequencing depths are often sufficient for obtaining accurate DNA methylation levels, this cannot be generalized to the entire genome: Infinium probes tend to be located in CpG-rich genomic regions, which are also preferentially covered by Me-

DIP, MethylCap and RRBS measurements (Figure 4), such that saturation is reached earlier in the vicinity of Infinium probes than in CpG-poor genomic regions.

Second, we tested how many DMRs were still detected among the two sample pairs when the number of sequencing reads in each of the samples was reduced (Supplementary Figure 11). For MeDIP, the number of detected DMRs dropped to less than half when the sequencing depth was reduced to 50%, and there was little indication that the number of MeDIP DMRs approaches saturation at the total sequencing depth. For MethylCap and RRBS, the reduction in the number of DMR was less dramatic, and there is a trend toward saturation visible for both methods (Supplementary Figure 11). Overall, the saturation analysis reinforced a conceptual difference between RRBS on the one hand and MeDIP and MethylCap on the other hand: In RRBS, all sequencing is focused on a well-defined, CpG-rich “reduced representation” of the genome. In contrast, MeDIP and MethylCap reads are widely distributed over the genome (albeit with a significant tendency toward high coverage in CpG-rich regions), and deep sequencing increasingly uncovers weak DMRs located in CpG-poor genomic regions.

#### *DNA methylation mapping of repetitive DNA*

DNA methylation differences in repetitive regions have frequently been ignored by genome-wide studies, due to technical difficulties such as ambiguous read alignment (for sequencing) and cross-hybridization (for microarrays). This is unfortunate given that loss of DNA methylation in repetitive DNA was the first epigenetic alteration shown to play a role in cancer<sup>3</sup> and has been an area of active research ever since<sup>42</sup>. In the current study, we explored two complementary approaches to test for repeat-associated DNA methylation differences. First, we included repetitive regions alongside non-repetitive regions in the DMR detection described above (Figure 5, Supplementary Figures 4 to 8), rather than discarding the repetitive cases. This way, it was possible to identify repeat-associated DMRs in a similar way as non-repetitive DMRs, and we could validate several such cases by clonal bisulfite sequencing (Table 2). However, the focus on specific genomic regions makes it difficult to detect global trends that affect certain repeat classes independent of their exact location in the genome. We therefore developed a second approach, which was motivated by the common origin of many repetitive regions from a small number of retrotransposons<sup>43</sup>. The basic concept of this second approach is to align sequencing reads to prototypic sequences (e.g., of Alu and L1 elements), in order to obtain DNA methylation measures per *repeat class* rather than per *repeat instance*.

To that end, we obtained a manually curated list of 1,267 prototypic repeat sequences that spans the spectrum of repetitive DNA present in the human genome<sup>44</sup>, and we aligned the sequencing reads of all three methods to this collection of repeat sequences. Approximately 20% of all MeDIP, MethylCap and RRBS reads could be aligned with high confidence, enabling us to estimate the global DNA methylation levels for 553 prototypic repeat sequences. The results of the three methods were in excellent agreement with each other (Supplementary File 2) and detected substantial differences in the DNA methylation levels of different repeat classes: Among Alu, SVA and satellite repeat sequences we observed consistently high levels of DNA methylation, while most LINE, LTR and DNA repeat sequences exhibited low levels of DNA methylation in the four samples that we investigated. However, we found that the repeat sequences with the highest copy-number throughout the genome were highly methylated for all repeat classes.

When we compared the DNA methylation levels in the two sample pairs (Supplementary File 3), we observed widespread but relatively moderate hypomethylation in the colon tumor relative to matched normal colon tissue. The most common targets were Alu, SVA and satellite repeat sequences, which is consistent with previous reports about cancer-specific hypomethylation<sup>42</sup>. An interesting difference was also identified between the two ES cell lines on the one hand and the two colon samples on the other hand: the only human-specific LINE repeat sequence in our collection (L1HS\_5end) exhibited high levels of DNA methylation in the two colon samples, but was largely unmethylated and even marked by histone H3K4 trimethylation in the two ES cell lines (Supplementary File 2). These data suggest that young retrotransposons find ways to evade silencing by DNA methylation in pluripotent cells, which may contribute to their ability to maintain activity in spite of an elaborate epigenetic genome defense<sup>45</sup>.

### *Utility of the Infinium HumanMethylation27 assay for DMR discovery*

The previous sections focused largely on MeDIP, MethylCap and RRBS, while the Infinium data were only used to validate the other methods' accuracy. However, many current studies employ the Infinium HumanMethylation27 assay as a DMR discovery tool on its own right, which prompted us to test its utility for that purpose. The Infinium assay measures the DNA methylation level of about 27,000 CpGs, which is substantially lower than the approximately 2 million CpGs of RRBS and the even larger number of CpGs that are covered by MeDIP and MethylCap. However, DNA methylation levels are highly correlated among neighboring CpGs<sup>20,46</sup>, suggesting that careful selection of “indicator CpGs” may enable the detection of a disproportionate number of DMRs based on relatively few CpG measurements.

*To assess the practical utility of the Infinium HumanMethylation27 assay for DMR discovery, we initially performed statistical testing in much the same way as for Figure 5.* However, most CpG islands were covered by only two Infinium probes, which resulted in low statistical power to detect significant differences. Specifically, paired-samples t-tests identified just three significant DMRs among the ES cell lines and two DMRs between the colon tumor and matched normal colon tissue (data not shown). Thus, we reformulated our question and asked how many true DMRs exhibited suggestive (albeit insignificant) DNA methylation differences in the Infinium data. As an approximation of true DMRs, we focused on those CpG islands that were detected by at least two sequencing-based methods. Between the two ES cell lines a total of 1,000 consensus DMRs were identified (corresponding to the sum of all center fields in Figure 5), of which 251 were covered by at least one Infinium probe. Similarly, we identified 463 consensus DMRs between the colon tumor and matched normal colon tissue, of which 177 were covered by at least one Infinium probe. In most cases, the directionality of the difference was consistent between the consensus DMRs and the Infinium data (Supplementary Figure 10). But when we imposed a minimum threshold of 20 percentage points DNA methylation difference in the same way as for RRBS, the number of Infinium-detected DMRs dropped to 162 (ES-cell comparison) and 95 (colon cancer comparison). In other words, the Infinium assay found approximately 20% of the consensus DMRs that we identified by the sequencing-based methods.

## **Discussion**

Over the last decade, DNA methylation mapping played an important role in establishing the prevalence of altered DNA methylation in cancer<sup>25,47-53</sup>, and it is likely that the analysis of cancer samples will continue to drive innovation in epigenome mapping. Simultaneously, researchers have started to systematically study the role of DNA methylation in a wide range of non-neoplastic diseases (e.g., <http://www.nih.gov/news/health/sep2009/od-16.htm>). It is indeed a good time to explore epigenetic alterations that may contribute to human diseases: Genome-wide association studies have been completed for all common diseases, and their results point to a major role of non-genetic factors in the etiology of most diseases<sup>54</sup>. Furthermore, it has been suggested that epigenetic events could provide a tractable link between the genome and the environment, with the epigenome emerging as a biochemical record of relevant life events<sup>55,56</sup>. Systematic investigation of these topics requires powerful and cost-efficient methods for identifying DNA methylation differences between samples.

The goal of this study was to evaluate current methods for DNA methylation mapping and to compare their performance in a practical application scenario. To mimic a typical disease-centered case-control study, we worked with primary patient material (colon samples) and used lower amounts of input DNA than in most previous studies (MeDIP: 300ng, MethylCap: 1µg, RRBS: 50ng, Infinium: 1µg). Furthermore, we focused on cell types that are known to exhibit relatively moderate DNA methylation differences<sup>25,57</sup>, in contrast to the massive DNA methylation alterations that are frequently observed in cultured somatic cells<sup>13</sup> and cancer cell lines<sup>52</sup>. Finally, all four methods included in the current study are widely available and not excessively costly, such that there are few obstacles to using this technology comparison as a blue print for individual lab efforts, including epigenomic case-control studies.

Overall, the data confirmed that all four methods provide accurate DNA methylation measurements and can be used to detect DMRs in clinical samples. In terms of accuracy, the bisulfite-based methods (RRBS, Infinium)

performed slightly better than the enrichment-based methods and did not require any statistical correction of CpG bias. The genomic coverage was moderately higher for MethylCap than for MeDIP, RRBS coverage was by design focused on CpG-rich regions, and the Infinium assay covered a relatively small number of preselected genomic regions. Despite the striking differences in genomic coverage, a substantial fraction of DMRs detected by MeDIP or MethylCap were also identified by RRBS (and vice versa). This somewhat counter-intuitive observation can be explained by the role of region-specific read coverage for the ability to identify statistically significant DMRs: If a genomic region is CpG-poor and thus rarely sequenced by MeDIP or MethylCap, both methods have low statistical power to detect differential DNA methylation. In contrast, CpG-rich genomic regions tend to be more amenable to DMR detection by MeDIP and MethylCap and are also frequently covered by RRBS measurements. Finally, we observed that MethylCap was able to detect roughly twice as many DMRs as MeDIP at comparable sequencing depths, RRBS detected more DMRs than MeDIP but fewer DMRs than MethylCap, and the Infinium assay detected only 20% of the consensus DMRs identified by the sequencing-based methods.

Our study reinforces the importance of sequencing depth as a key parameter for DMR discovery independent of which method is being used. To allow for a fair and practically relevant comparison, we sequenced approximately 30 to 40 million reads for each sample and method. However, it became evident that deeper sequencing would identify further DMRs, especially for MeDIP and MethylCap (Supplementary Figure 11). For disease-centered studies it is therefore necessary to make an informed decision about how to distribute the available resources between sequencing few samples more deeply and sequencing more samples less deeply. Such a decision can be guided by statistical power calculations when some prior knowledge exists about the characteristics of expected DMRs (e.g., magnitude of difference, location in CpG-rich vs. CpG-poor genomic regions), or they can be dictated by practical considerations such as the number of available samples. MeDIP, MethylCap and RRBS as performed in this study seem to provide a practically useful compromise between breadth and depth of sequencing. In contrast, whole-genome bisulfite sequencing<sup>58</sup> provides comprehensive genomic coverage at the cost of having to sequence over a billion reads per sample. On the other end of the spectrum, RRBS can be combined with barcoded sequencing to quickly screen a large number of samples with just one million reads per sample.

A side note of our study is the generation and analysis of a comprehensive map of DNA methylation in repetitive DNA. Genome-wide studies tend to ignore repetitive regions due to technical difficulties, and the few studies that focused specifically on mapping DNA methylation in repetitive regions did so at relatively low coverage<sup>59-61</sup>. The current dataset seems well-suited to analyze DNA methylation in repetitive regions because the joint results obtained by three different experimental methods helped us to control for technical artifacts that have often obstructed the analysis of repetitive DNA. We observed that repeat sequences are most highly methylated if they are CpG-rich and highly prevalent in the human genome (Supplementary File 2). In contrast, the DNA methylation levels varied widely among repeat sequences that are either CpG-poor or infrequent in the genome. These results lend support to the hypothesis that DNA methylation provides a mechanism for keeping active retrotransposons in check<sup>62</sup>. They argue for a highly specific mechanism of repeat repression, which targets DNA methylation mostly to those repeat sequences that threaten genome integrity, while many “benign” repeat sequences may remain unmethylated.

In summary, we benchmarked four methods for genome-scale DNA methylation profiling in terms of their accuracy and power to detect DNA methylation differences. These results will facilitate the selection of suitable methods for studying the role of DNA methylation in human diseases.

## Methods

### *Sample origin and cell culture*

Human ES cells were cultured in knockout serum replacement (KOSR) medium according to established protocols<sup>30</sup> and genomic DNA was extracted as described previously<sup>63</sup>. DNA for the colon tumor and matched normal colon tissue was purchased from BioChain (lot number A704198). Both samples originate from the same donor, an 81-year-old male patient diagnosed with moderately differentiated adenocarcinoma.



#### *Methyl-DNA Immunoprecipitation (MeDIP)*

MeDIP<sup>14</sup> was performed using the EZ DNA methylation kit (Zymo Research). A total of 300ng DNA per sample was sonicated using Bioruptor (Diagenode) with 8 intervals of 10min (30s on, 30s off), resulting in an average fragment size of 150 basepairs. Sonicated DNA was end-repaired and ligated with sequencing adapters as described previously<sup>14,64</sup>. Gel-based selection for fragment sizes between 100 and 200 basepairs was followed by methylated DNA immunoprecipitation according to the manufacturer's protocol. A total of 1µg of monoclonal antibody against 5-methyl-cytosine (included in the EZ DNA methylation kit) was used for immunoprecipitation. The immunoprecipitated DNA was PCR-amplified and the specificity of the enrichment was confirmed by qPCR for selected loci as described previously<sup>65</sup>. Sequencing on the Illumina Genome Analyzer II was performed according to the manufacturer's standard protocol.

#### *Methylated-DNA capture (MethylCap)*

MethylCap<sup>16</sup> was performed in a robotized procedure using a SX-8G / IP-Star (Diagenode). 2µg of His6-GST-MBD (Diagenode) was combined with 1µg of sonicated DNA in 200µl of binding buffer (BB, 20mM Tris-HCl pH 8.5, 0.1% Triton X-100) containing 200mM NaCl. This solution was incubated at 4°C for 2 hours. Magnetic GST-beads were prepared by washing 35µl of a well-mixed MagneGST glutathione particle suspension (Promega) with 200µl of binding buffer plus 200mM NaCl at 4°C. Washing was repeated once and the supernatant was removed. The GST-MBD-DNA solution was added to the washed and collected beads, and this suspension was rotated for another hour at 4°C. After removal of the supernatant (this is the flow-through) the beads-GST-MBD-DNA complexes were eluted by washing. 200µl of binding buffer with different concentrations of NaCl was added and the suspension was rotated for 10min at 4°C. Beads were captured using a magnet, and the supernatant was collected. The elution procedure consisted of 1x 300mM (wash), 2x 400mM (wash), 1x 500mM ("low" eluate), 1x 600mM ("medium" eluate), 1x 800mM NaCl ("high" eluate). The collected eluates were purified using QIAquick PCR purification spin columns (Qiagen), eluted with 100µl elution buffer and prepared for sequencing on the Illumina Genome Analyzer II as described previously<sup>16</sup>. A single lane was sequenced for the low, medium and high eluates, respectively. These lanes are shown separately in Figure 2, and we tested whether the accuracy relative to the Infinium assay could be improved by taking this additional information into account. However, a linear model that was based on the separate read counts of the three lanes did not outperform a model that was based on the sum of the three lanes, which is why we used only pooled data for the analyses described in this paper.

#### *Reduced representation bisulfite sequencing (RRBS)*

RRBS<sup>24</sup> was performed according to a previously published protocol<sup>63</sup> with some optimizations for clinical samples and low amounts of input DNA<sup>24</sup>. The main steps were: (i) A total of 50ng (ES cells) or 1µg (colon samples) genomic DNA was digested by 5U to 20U of MspI (New England Biolabs, NEB) for up to 16h. (ii) End-repair and adenylation of digested DNA were performed in a 20µl reaction consisting of 10U of Klenow fragments (3'→5' exo-, NEB), 2µl premixed nucleotide triphosphates (1mM dGTP, 10mM dATP, 1mM 5' methylated dCTP). The reaction was incubated at 30°C for 30min followed by 37°C for additional 30min. (iii) Preannealed 5-methylcytosine-containing Illumina adapters were ligated with adenylated DNA fragments in a 20µl reaction containing of 1µl concentrated T4 ligase (NEB), 1-2µl of 15µM adapters at 16°C for 16 to 20 hours. (iv) Gel-based selection for fragments with insertion sizes of 40 to 120 basepairs and 120 to 220 basepairs was performed as described previously<sup>24</sup>. (v) Bisulfite treatment with the EpiTect Bisulfite Kit (Qiagen) was conducted following the protocol designated for DNA isolated from formalin-fixed and paraffin-embedded tissues. Two rounds of conversion were performed in order to maximize bisulfite conversion rates. The final bisulfite-converted DNA was eluted with 2x 20µl pre-heated (65°C) EB buffer. (vi) To determine the minimum number of PCR cycles for final library enrichment, analytical (10µl) PCR reactions containing 0.5µl of bisulfite-treated DNA, 0.2µM each of Illumina PCR primers LPX1.1 and 2.1 and 0.5U PfuTurbo Cx Hotstart DNA polymerase

(Stratagene) were set up. The thermocycler conditions were: 5min at 95°C, varied cycle numbers (10-20) of 20s at 95°C, 30s at 65°C, 30s at 72°C, followed by 7min at 72°C. PCR products were visualized by running on a 4-20% polyacrylamide Criterion TBE Gel (Bio-Rad) and stained by SYBR Green. The final libraries were generated by 8 of 25µl PCR reaction with each one containing 2-3µl of bisulfite-converted template, 1.25U PfuTurbo Cx Hotstart polymerase and 0.2µM each of Illumina LPX1.1 as well as 2.1 PCR primers. The libraries were PCR amplified and sequenced on the Illumina Genome Analyzer II as described previously<sup>24</sup>.

#### *Microarray-based epigenotyping (Infinium)*

Infinium<sup>19</sup> analysis was performed by the Genetic Analysis Platform at the Broad Institute. A total of 1µg of genomic DNA per sample was bisulfite-treated according to the manufacturer's protocol and hybridized onto Infinium HumanMethylation27 bead arrays (Illumina). We previously observed almost perfect agreement between technical replicates (Pearson's  $r > 0.98$ ), which is why only a single hybridization was performed for each sample.

#### *Data preparation and quality control*

All sequencing data were processed using Illumina's standard pipeline for image processing and base calling. MeDIP and MethylCap reads were aligned against the human genome using Maq<sup>66</sup>, and the aligned reads were extended to the mean fragment lengths obtained during sonication. From each group of duplicate reads (i.e. reads aligned to the exact same start position on the same chromosome) all but one read were discarded, in order to minimize the impact of PCR bias on downstream analysis. For RRBS, the reads were aligned using a custom bisulfite alignment algorithm<sup>13</sup>, and DNA methylation status was assessed by comparing the aligned reads to the reference genome as described previously<sup>24</sup>. Infinium HumanMethylation27 data were processed with Illumina's BeadStudio 3.2 software, using the default background subtraction method for normalization. UCSC Genome Browser tracks were constructed by custom scripts implemented in the Python programming language (<http://www.python.org/>).

#### *Quantification of absolute DNA methylation levels*

We used linear regression models to estimate the absolute DNA methylation levels from the MeDIP and MethylCap read counts. Based on a number of different feature selection experiments, we found that the following combination of variables was robustly predictive of DNA methylation levels: (i) the square root of the total number of MeDIP or MethylCap reads within the given region, (ii) the square root of the total number of whole-cell extract (WCE) reads within the region (based on a cross-tissue WCE track that we routinely use for ChIP-seq data normalization), (iii) the logit of the CpG frequency within the region, (iv) the relative GC content of the region, (v) the ratio of Cs relative to CpGs, and (vi) the relative repeat content of the region as determined by RepeatMasker (<http://www.repeatmasker.org>). For both MeDIP and MethylCap, we observed that the read frequencies were strongly positively associated with the absolute methylation level according to Infinium data, while the repeat content was moderately positively associated. In contrast, the logit of the CpG frequency was highly negatively associated with DNA methylation, and all other variables as well as the model's intercept exhibited a moderately negative association. For model fitting and performance evaluation, the current dataset was split into equally sized training and test sets. All model fitting was performed using the R statistics package (<http://www.r-project.org/>).

#### *Identification of differentially methylated regions*

In our experience, classical peak detection<sup>67,68</sup> is not well-suited for DMR identification because of the high number of spurious hits encountered when borderline peaks are detected in one sample but not in the other (C. Bock, unpublished observation). Instead, we used a statistical test to compare two samples directly with each other. For a given region with RRBS data, we count the number of methylated vs. unmethylated CpGs in both samples and perform Fisher's exact test to obtain a p-value that is indicative of the likelihood of the region being a DMR. Similarly, for MeDIP and MethylCap we count the numbers of reads that align inside the region for

both samples and use Fisher's exact test to contrast these values with the total numbers of reads that align elsewhere in the genome. And for the Infinium assay we use a paired-samples t-test to compare the two samples'  $\beta$ -values of all Infinium probes inside the region. These tests are performed on a large number of genomic regions in parallel (e.g., on all CpG islands), and the p-values are corrected for multiple testing using the q-value method<sup>69</sup>. Genomic regions with a q-value of less than 0.1 are flagged as hypermethylated or hypomethylated (depending on the directionality of the difference), but only if the absolute DNA methylation difference exceeds 20% (for RRBS and Infinium) or if there is at least a twofold difference in the read number (for MeDIP and MethylCap). These thresholds were chosen by their practical utility in a number of comparisons between different cell types and have no further justification. We also mark genomic regions with insufficient sequencing coverage, but do not exclude them from DMR analysis. For MeDIP and MethylCap we require at least ten reads per 10 million total reads for the sample with higher read coverage, and for RRBS we require a minimum of five CpGs with at least five reads each in both samples.

This statistical approach to DMR identification requires us to define sets of genomic regions on which the analysis is being performed. We pursued a two-way strategy to maximize the chances of finding interesting DMRs. On the one hand, we focused specifically on CpG islands and gene promoters, which are prime candidates for epigenetic regulation. This approach provides increased statistical power for regions with well-known functional roles because the relatively low number of CpG islands and gene promoters reduces the burden of multiple-testing correction compared to the genome-wide case. On the other hand, we used a 1-kilobase tiling of the genome to detect DMRs that are located outside of any candidate regions. And to cast an even wider net, we collected a comprehensive set of 13 types of genomic regions, which includes not only CpG islands and gene promoters, but also CpG island shores<sup>25</sup>, enhancers<sup>70</sup>, evolutionary conserved regions and other types of genomic regions. DMR data for all of these region sets were calculated using a set of Python and R scripts and are available online (<http://meth-benchmark.computational-epigenetics.org/>).

#### *Experimental validation*

Based on the CpG islands that were detected as differentially methylated between the two ES cell lines (Figure 5), we manually selected eight method-specific DMRs for experimental validation. To that end, those CpG islands that were identified as statistically significant DMRs by one method (but not by the other two methods) were visually inspected in the UCSC Genome Browser, and regions were selected for validation only if the data fully supported their classification as method-specific DMRs. In particular, regions were not selected if a second method already picked up a suggestive but insignificant trend in the same direction as the first method, or when the data of the first method already suggested that the DMR was a false-positive hit (e.g., because of contradictory trends in the vicinity of the DMR). Experimental validation was performed by clonal bisulfite sequencing following established protocols<sup>71</sup>. Primers were designed using MethPrimer<sup>72</sup> such that the amplicon overlapped with those CpGs that exhibited the highest levels of differential methylation according to our original data. To prepare for bisulfite sequencing, 1 $\mu$ g of DNA was bisulfite-converted using the EpiTect kit (Qiagen); 50ng of bisulfite-converted DNA was PCR-amplified (see Supplementary File 1 for primer sequences); and purified amplicons were cloned using the TOPO TA cloning kit (Invitrogen). For each region an average of 11 clones were randomly chosen for sequencing. All sequencing data were processed using the BiQ Analyzer software<sup>73</sup>, and the results are summarized in Supplementary File 1.

#### *Analysis of repetitive DNA*

Repeat sequences were obtained from database version 14.07 of RepBase Update<sup>44</sup>, which is publicly available online (<http://www.girinst.org/server/RepBase/index.php>). From a total of 11,670 prototypic repeat sequences we selected those 1,267 that were annotated either to human or to its ancestors in the taxonomic tree, and we combined these prototypic repeat sequences into a pseudo-genome file. Maq version 0.7.1<sup>66</sup> was used to align MeDIP, MethylCap, RRBS, CHIP-seq (H3K4me3) and whole-cell extract (WCE) sequencing reads against this pseudo-genome. For RRBS, both the reads and the reference genome were bisulfite-converted in silico prior to

the alignment. The epigenetic status of each prototypic repeat sequence was quantified as follows: (i) For MeDIP, MethylCap and ChIP-seq we calculated the odds ratios relative to the WCE data. (ii) For RRBS we computed the number of methylated CpGs, total number of CpG measurements and percentage of DNA methylation based on the comparison of the aligned reads with the prototypic repeat sequence.

We discarded rare repeats with WCE coverage below 100 aligned reads or RRBS coverage below 25 CpG measurements, resulting in 553 prototypic repeat sequences that were used for further analysis. Among these were 97 LINE class sequences (92 of them from the L1 family), 51 SINEs (48 of them from the Alu family), 6 SVAs, 62 DNA repeats, 15 satellite repeats, 315 LTRs, 1 low-complexity repeat and 6 RNA repeats (Supplementary File 2). To quantify differential methylation between a pair of MeDIP and MethylCap samples, we calculated the pairwise odds ratio of the read coverage for each prototypic repeat sequence, while the absolute DNA methylation difference was used in the case of RRBS (Supplementary File 3). The significance of the difference was assessed using Fisher's exact test in the same way as for the non-repetitive genome (described above).

## Acknowledgements

We thank A. Crenshaw and M. Parkin (Broad Institute, Cambridge, MA) for assistance with the Infinium assay and K. Halachev (Max Planck Institute for Informatics, Saarbrücken, Germany) for the provision of genome annotation files. C. Bock is supported by a Feodor Lynen Fellowship from the Alexander von Humboldt Foundation. A. B. Brinkman is supported by the Dutch Cancer Foundation (KWF, grant KUN 2008-4130). A. Meissner is supported by the Massachusetts Life Science Center and the Pew Charitable Trusts. The described work was in part funded by the Pew Charitable Trusts, the NIH Roadmap Initiative on Epigenomics (U01ES017155), and the European Union's CANCERDIP project (HEALTH-F2-2007-200620)

## Competing interests

The authors declare that no competing financial interests exist.

## References

1. Baylin, S. B. and Ohm, J. E., *Nat. Rev. Cancer* **6**, 107 (2006).
2. Esteller, M., *N. Engl. J. Med.* **358**, 1148 (2008).
3. Feinberg, A. P. and Tycko, B., *Nat. Rev. Cancer* **4**, 143 (2004).
4. Feinberg, A. P., Ohlsson, R., and Henikoff, S., *Nat. Rev. Genet.* **7**, 21 (2006).
5. Issa, J. P., *Nat. Rev. Cancer* **4**, 988 (2004).
6. Jones, P. A. and Laird, P. W., *Nat. Genet.* **21**, 163 (1999).
7. Jones, P. A. and Baylin, S. B., *Cell* **128**, 683 (2007).
8. Laird, P. W., *Nat. Rev. Cancer* **3**, 253 (2003).
9. Richardson, B., *Nat. Clin. Pract. Rheumatol.* **3**, 521 (2007).
10. Tobi, E. W. et al., *Hum. Mol. Genet.* **18**, 4046 (2009).
11. Urdinguio, R. G., Sanchez-Mut, J. V., and Esteller, M., *Lancet Neurol.* **8**, 1056 (2009).
12. Bock, C., *Epigenomics* **1**, 99 (2009).
13. Meissner, A. et al., *Nature* **454**, 766 (2008).
14. Down, T. A. et al., *Nat. Biotechnol.* **26**, 779 (2008).
15. Weber, M. et al., *Nat. Genet.* **37**, 853 (2005).
16. Brinkman, A. B. et al., *Methods* (2010).
17. Rauch, T. and Pfeifer, G. P., *Lab. Invest.* **85**, 1172 (2005).
18. Serre, D., Lee, B. H., and Ting, A. H., *Nucleic Acids Res.* **38**, 391 (2009).
19. Bibikova, M. et al., *Epigenomics* **1**, 177 (2009).
20. Eckhardt, F. et al., *Nat. Genet.* **38**, 1378 (2006).
21. Brunner, A. L. et al., *Genome Res.* **19**, 1044 (2009).
22. Irizarry, R. A. et al., *Genome Res.* **18**, 780 (2008).

23. Oda, M. et al., *Nucleic Acids Res.* (2009).
24. Gu, H. et al., *Nat. Methods* **7**, 133 (2010).
25. Irizarry, R. A. et al., *Nat. Genet.* **41**, 178 (2009).
26. Rauch, T. et al., *Proc. Natl. Acad. Sci. U. S. A.* **104**, 5527 (2007).
27. Weber, M. et al., *Nat. Genet.* **39**, 457 (2007).
28. Hegi, M. E. et al., *N. Engl. J. Med.* **352**, 997 (2005).
29. Lofton-Day, C. et al., *Clin. Chem.* **54**, 414 (2008).
30. Cowan, C. A. et al., *N. Engl. J. Med.* **350**, 1353 (2004).
31. Weisenberger, D. J., Berg, D. V. D., Pan, F., Berman, B. P., and Laird, P. W., (2009).
32. Pelizzola, M. et al., *Genome Res.* **18**, 1652 (2008).
33. Meissner, A. et al., *Nucleic Acids Res.* **33**, 5868 (2005).
34. Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A., *Behav. Res. Methods* **39**, 175 (2007).
35. Beroukhim, R. et al., *Nature* **463**, 899 (2010).
36. Redon, R. et al., *Nature* **444**, 444 (2006).
37. Hellebrekers, D. M. et al., *Clin. Cancer Res.* **15**, 3990 (2009).
38. Zhang, W. et al., *Cancer Res.* **68**, 2764 (2008).
39. Tenesa, A. et al., *Nat. Genet.* **40**, 631 (2008).
40. Caldwell, G. M. et al., *Cancer Res.* **64**, 883 (2004).
41. Hirata, H. et al., *Clin. Cancer Res.* **15**, 5678 (2009).
42. Ehrlich, M., *Epigenomics* **1**, 239 (2009).
43. Lander, E. S. et al., *Nature* **409**, 860 (2001).
44. Jurka, J., *Trends Genet.* **16**, 418 (2000).
45. Bestor, T. H. and Tycko, B., *Nat. Genet.* **12**, 363 (1996).
46. Bock, C., Walter, J., Paulsen, M., and Lengauer, T., *Nucleic Acids Res.* **36**, e55 (2008).
47. Esteller, M., Corn, P. G., Baylin, S. B., and Herman, J. G., *Cancer Res.* **61**, 3225 (2001).
48. Hu, M. et al., *Nat. Genet.* **37**, 899 (2005).
49. Keshet, I. et al., *Nat. Genet.* **38**, 149 (2006).
50. Noushmehr, H. et al., *Cancer Cell* **17**, 510 (2010).
51. Shames, D. S. et al., *PLoS Med.* **3**, e486 (2006).
52. Smiraglia, D. J. et al., *Hum. Mol. Genet.* **10**, 1413 (2001).
53. Weisenberger, D. J. et al., *Nat. Genet.* **38**, 787 (2006).
54. Manolio, T. A. et al., *Nature* **461**, 747 (2009).
55. Foley, D. L. et al., *Am. J. Epidemiol.* **169**, 389 (2009).
56. Heijmans, B. T., Tobi, E. W., Lumey, L. H., and Slagboom, P. E., *Epigenetics* **4** (2009).
57. Doi, A. et al., *Nat. Genet.* (2009).
58. Lister, R. et al., *Nature* **462**, 315 (2009).
59. Horard, B. et al., *Epigenetics* **4**, 339 (2009).
60. Rodriguez, J. et al., *Nucleic Acids Res.* **36**, 770 (2008).
61. Weisenberger, D. J. et al., *Nucleic Acids Res.* **33**, 6823 (2005).
62. Yoder, J. A., Walsh, C. P., and Bestor, T. H., *Trends Genet.* **13**, 335 (1997).
63. Smith, Z. D., Gu, H., Bock, C., Gnirke, A., and Meissner, A., *Methods* **48**, 226 (2009).
64. Mikkelsen, T. S. et al., *Nature* **448**, 553 (2007).
65. Rakyán, V. K. et al., *Genome Res.* **18**, 1518 (2008).
66. Li, H., Ruan, J., and Durbin, R., *Genome Res.* **18**, 1851 (2008).
67. Bock, C. and Lengauer, T., *Bioinformatics* **24**, 1 (2008).
68. Park, P. J., *Nat. Rev. Genet.* **10**, 669 (2009).
69. Storey, J. D. and Tibshirani, R., *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440 (2003).
70. Heintzman, N. D. et al., *Nature* **459**, 108 (2009).
71. Hajkova, P. et al., *Methods Mol. Biol.* **200**, 143 (2002).
72. Li, L. C. and Dahiya, R., *Bioinformatics* **18**, 1427 (2002).
73. Bock, C. et al., *Bioinformatics* **21**, 4067 (2005).
74. Bock, C., Walter, J., Paulsen, M., and Lengauer, T., *PLoS Comput. Biol.* **3**, e110 (2007).
75. Pruitt, K. D., Tatusova, T., and Maglott, D. R., *Nucleic Acids Res.* **35**, D61 (2007).

76. Flicek, P. et al., *Nucleic Acids Res.* **36**, D707 (2008).

## Tables

Table 1. Summary of DNA methylation mapping experiments

Run No.	Method	Sample name	#lanes	#reads (total)	#reads (aligned)	Alignment rate	#reads (unique)	#reads (duplicates)	Unique read rate
1	MeDIP	HUES6 ES cell line	2	37,086,239	22,798,831	61.5%	12,849,623	9,949,208	56.4%
2	MeDIP	HUES8 ES cell line	2	36,078,308	24,266,670	67.3%	12,287,174	11,979,496	50.6%
3	MeDIP	Primary colon tumor	2	33,453,797	18,582,183	55.5%	7,006,484	11,575,699	37.7%
4	MeDIP	Matched normal colon tissue	2	37,789,936	21,793,567	57.7%	10,360,103	11,433,464	47.5%

Run No.	Method	Sample name	#lanes	#reads (total)	#reads (aligned)	Alignment rate	#reads (unique)	#reads (duplicates)	Unique read rate
5	MethylCap	HUES6 ES cell line	3	38,436,495	23,401,511	60.9%	21,712,433	1,689,078	92.8%
6	MethylCap	HUES8 ES cell line	3	38,735,596	21,670,301	55.9%	19,585,988	2,084,313	90.4%
7	MethylCap	Primary colon tumor	3	37,718,830	23,206,054	61.5%	21,600,129	1,605,925	93.1%
8	MethylCap	Matched normal colon tissue	3	38,330,519	22,724,002	59.3%	21,290,282	1,433,720	93.7%

Run No.	Method	Sample name	#lanes	#reads (total)	#reads (aligned)	Alignment rate	#CpGs (total)	#CpGs (unique)	Mean CpG coverage
9	RRBS	HUES6 ES cell line	2	30,004,147	12,150,905	40.5%	22,181,147	2,181,128	10.2x
10	RRBS	HUES8 ES cell line	2	28,395,040	12,670,034	44.6%	29,704,332	2,185,751	13.6x
11	RRBS	Primary colon tumor	4	40,015,958	9,545,423	23.9%	16,891,325	1,297,296	13.0x
12	RRBS	Matched normal colon tissue	4	32,072,287	6,214,732	19.4%	10,190,227	1,134,963	9.0x

Run No.	Method	Sample name	#arrays	#CpGs (total)	#CpGs (valid)	#CpGs (unique)	Valid probe rate
13	Infinium	HUES6 ES cell line	1	27,578	27,192	27,192	98.6%
14	Infinium	HUES8 ES cell line	1	27,578	27,090	27,090	98.2%
15	Infinium	Primary colon tumor	1	27,578	27,561	27,561	99.9%
16	Infinium	Matched normal colon tissue	1	27,578	27,478	27,478	99.6%

Technical notes: (i) Between two and four lanes [#lanes] were sequenced on the Illumina Genome Analyzer II, which yielded approximately 30 to 40 million reads [#reads (total)] per sample and method. These reads were aligned to the human genome [# reads (aligned)], and the number of unique [#reads (unique)] as well as duplicate reads [#reads (duplicates)] was calculated by counting the first read that aligns to a specific genomic position as unique and all further occurrences of the same genomic position as duplicates (genomic positions were defined by the combination of chromosome, read start position and strand). (ii) Three lanes were sequenced for samples 5-8, one lane per eluate (high, medium and low, as described in the Methods section). (iii) Samples 11 and 12 were part of a sequencing optimization run that resulted in lower sequencing yield and reduced alignment rates. Four lanes were sequenced to reach the target of 30 to 40 million reads per sample and method. (iv) All other samples were sequenced on two lanes each. There were no failed runs that had to be redone, and all sequencing data that were generated for the current project are summarized in this table. (v) The yield of a single lane of Illumina Genome Analyzer II sequencing strongly increased during the course of this study. As of June 2010, we observe total read numbers per lane that average around 40 million for MeDIP and MethylCap, and which are close to 30 million for RRBS. Current alignment rates range from 60% to 80% for typical runs with all three methods.

Table 2: Validation of method-specific DMRs for MeDIP, MethylCap and RRBS

DMR location	Description	Experimental validation	MeDIP	MethylCap	RRBS
MeDIP-specific DMR chr10:88,149,016-88,149,732	Intergenic CpG island ~30kb upstream of GRID1, partial overlap with degenerate L1 element	HUES6: 38/56 (68%) methylated CpGs HUES8: 26/44 (59%) methylated CpGs → <b>insignificant</b> (p=0.41)	hypermethylated (q=1.1E-04)	insignificant (q=0.59)	insignificant (q=0.43)
MeDIP-specific DMR chr16:31,142,904-31,143,799	CpG island overlapping with the terminal exon of TRIM72	HUES6: 342/362 (94%) methylated CpGs HUES8: 466/523 (89%) methylated CpGs → marginally <b>hypermeth.</b> (p=0.0051)	hypermethylated (q= 1.2E-05)	insignificant (q=0.73)	insufficient coverage
MeDIP-specific DMR chr1:211,290,079-211,290,896	CpG island overlapping with the putative promoter region of RPS6KC1	HUES6: 53/60 (88%) methylated CpGs HUES8: 45/50 (90%) methylated CpGs → <b>insignificant</b> (p=1.0)	hypermethylated (q= 3.0E-06)	insignificant (q=0.97)	insignificant (q=0.29)
MethylCap-specific DMR chr20:29,526,646-29,527,380	CpG island overlapping with the putative promoter region of REM1	HUES6: 5/72 (7%) methylated CpGs HUES8: 78/84 (93%) methylated CpGs → <b>hypomethylated</b> (p= 1.4E-30)	insufficient coverage	hypomethylated (q= 1.8E-09)	insufficient coverage
MethylCap-specific DMR chr2:151,825,938-151,826,902	CpG island overlapping with the putative promoter region of RBM43 and a known copy-number variation	HUES6: 161/208 (77%) methylated CpGs HUES8: 9/104 (9%) methylated CpGs → <b>hypermethylated</b> (p= 3.3E-33)	insignificant (q=0.18)	hypermethylated (q= 7.3E-09)	insufficient coverage
MethylCap-specific DMR chr13:44,348,934-44,349,700	Intergenic CpG island ~60kb upstream of NUFIP1, partial overlap with degenerate Alu element	HUES6: 80/88 (91%) methylated CpGs HUES8: 41/79 (52%) methylated CpGs → <b>hypermethylated</b> (p= 1.2E-08)	insignificant (q=0.40)	hypermethylated (q= 8.3-07)	insufficient coverage
RRBS-specific DMR chr3:186,889,821-186,890,200	CpG island overlapping with an internal exon and intron of IGF2BP2	HUES6: 5/90 (6%) methylated CpGs HUES8: 88/90 (98%) methylated CpGs → <b>hypomethylated</b> (p= 4.3E-42)	insufficient coverage	insignificant (q=0.18)	hypomethylated (q= 3.5E-40)
RRBS-specific DMR chr3:32,609,320-32,609,612	Intergenic CpG island ~20kb upstream of DYNC1L1	HUES6: 41/121 (34%) methylated CpGs HUES8: 130/143 (91%) methylated CpGs → <b>hypomethylated</b> (p= 3.5E-23)	insufficient coverage	insignificant (q=0.52)	hypomethylated (q= 2.9E-26)

Eight method-specific DMRs were selected based on the DNA methylation maps of MeDIP, MethylCap and RRBS and experimentally validated using bisulfite sequencing. The validation candidates were manually chosen from the list of CpG islands (Figure 5), such that each region is identified as DMR by only one method, while the other two methods do not detect any significant (or suggestive) difference in DNA methylation between the two human ES cell lines (HUES6 and HUES8). The validation was performed by clonal bisulfite sequencing with an average of 11 clones per region in each of the two human ES cell lines. The p-values in column 3 were calculated from the clonal bisulfite sequencing data using Fisher's exact test, based on the DNA methylation levels of individual CpGs. The q-values in column 4 to 6 were derived from the DNA methylation maps as described in the Methods section. Columns are highlighted in green if the validation confirmed a method's original assessment, in red if the validation did not confirm the assessment, and in grey if no assessment was possible because of insufficient coverage. All genomic coordinates refer to the amplicon on which the validation was performed, which was chosen such that it overlaps with the most differentially methylated region within the selected CpG island. The coordinates are given relative to the hg18/NCBI36 assembly of the human genome. A detailed documentation of the validation experiments is available online (Supplementary File 1).



## Figures

### Figure 1. Outline of the DNA methylation technology comparison

Four methods for DNA methylation mapping were compared on two pairs of samples. The resulting 16 DNA methylation maps were bioinformatically analyzed and benchmarked against each other. In addition, clonal bisulfite sequencing was performed on selected genomic regions to validate DNA methylation differences that were detected exclusively by one method.

### Figure 2. Comparison of DNA methylation maps obtained with MeDIP, MethylCap, RRBS and Infinium

DNA methylation maps were generated using MeDIP (first two tracks, in green), MethylCap (three tracks in blue, grey and red), RRBS (stacked blue tracks) and Infinium (single black track with percentage values) and converted into UCSC Genome Browser tracks. The screenshot shows the HOXA cluster in a human ES cell line (HUES6). Each track represents data from a single sequencing lane (MeDIP, MethylCap, RRBS) or microarray hybridization (Infinium). MeDIP and MethylCap data are visually similar to ChIP-seq data, with peaks in regions that exhibit high density of the target molecule (5-methyl-cytosine) and troughs in regions with low density of methylated cytosines. The height of the peaks represents the number of reads in each genomic interval, for each track normalized to the same genome-wide read count (note the twofold compressed scaling of the MethylCap tracks relative to the MeDIP tracks, which is indicative of higher dynamic range for MethylCap compared to MeDIP). RRBS gives rise to clusters of CpGs with absolute DNA methylation measurements, separated by regions that are not covered due to the reduced-representation property of the RRBS protocol. Each data point corresponds to the methylation level at a single CpG, and dark blue points indicate higher methylation levels than light blue points. Infinium data is represented in a similar way as the RRBS data, and the methylation levels at single CpGs are shown as percentage values. The three grey columns highlight regions that illustrate specific aspects of the enrichment methods: (1) A promoter region that is CpG-poor and therefore not detectable by MeDIP or MethylCap – independent of its DNA methylation level; (2) a promoter region that contains many CpGs but low levels of DNA methylation, which also results in the absence of MeDIP and MethylCap peaks; and (3) a CpG island that exhibits a strong enrichment peak for both MeDIP and MethylCap although the RRBS data indicates that it is only partially methylated. For reference, the CpG density is indicated by stacked points (black) at the bottom of the diagram, and CpG islands (red) as well as known genes (blue) are listed as described previously<sup>74,75</sup>. All DNA methylation maps are available as genome browser tracks (<http://meth-benchmark.computational-epigenetics.org/>).

### Figure 3. Quantification of DNA methylation with MeDIP, MethylCap and RRBS

Absolute DNA methylation levels were calculated from the data obtained by MeDIP (panel A), MethylCap (panel B) and RRBS (panel C), respectively, and compared to DNA methylation levels determined by the Infinium assay. For MeDIP and MethylCap, sequencing reads were counted in 1-kilobase regions surrounding each CpG that is interrogated by the Infinium assay, and a regression model was used to infer absolute DNA methylation levels. Scatterplots and correlation coefficients were calculated on a test set that was not used for model fitting or feature selection. For RRBS, the DNA methylation level was determined as the percentage of methylated CpGs within 200 basepairs surrounding each CpG that is interrogated by the Infinium assay. Data shown are for the HUES6 human ES cell line, and regions that did not have sufficient sequencing coverage were excluded.

### Figure 4. Genomic coverage of MeDIP, MethylCap, RRBS and Infinium

Genomic coverage was quantified by the number of DNA methylation measurements that overlap with CpG islands (top row), gene promoters (center row) and a 1-kilobase tiling of the genome (bottom row). For MeDIP and MethylCap, the number of measurements is equal to the number of unique sequencing reads that fall inside each region. For RRBS, it refers to the number of valid DNA methylation measurements at CpGs within each region (one RRBS sequencing read typically yields one measurement, but can also give rise to more than one measurement if it contains several CpGs). For Infinium, the number of measurements is equal to the number of CpGs within each region that are present on the HumanMethylation27 microarray. CpG islands were calculated using CgiHunter (<http://cgihunter.bioinf.mpi-inf.mpg.de/>), requiring a minimum CpG observed vs. expected ratio of 0.6, a minimum GC content of 0.5 and a minimum length of 700 basepairs<sup>74</sup>. Promoter regions were calculated based on Ensembl gene annotations<sup>76</sup>, such that the region starts one kilobase upstream of the annotated transcription start site (TSS) and extends to one kilobase downstream of the TSS. The genomic tiling was obtained by sliding a 1-kilobase window through the genome such that each tile starts at the position where the previous tile ends. No repeat-masking was performed for any of the three types of genomic regions. Data are shown for the HUES6 human ES cell line.

### Figure 5. Detection of differentially methylated regions with MeDIP, MethylCap and RRBS

Average DNA methylation measurements were calculated for each CpG island and compared between two human ES cell lines (HUES6 and HUES8). Total read frequencies are shown for MeDIP (panel A) and MethylCap (panel B), and mean DNA methylation levels are shown for RRBS (panel C). Regions with insufficient sequencing coverage were excluded. The Venn diagram (panel D) displays the total number and mutual overlap of differentially methylated CpG islands that each method identified as statistically significant. CpG islands were defined as described in the legend of Figure 4.

**DNA for two pairs of samples**

- Two human ES cell lines derived from unrelated embryos
- A colon tumor and matched normal colon tissue from the same patient

**MeDIP**

1. Sonication of DNA
2. Library preparation
3. Denaturation and enrichment with antibody for 5-methyl-cytosine
4. Library amplification

**MethylCap**

1. Sonication of DNA
2. Enrichment with methyl-binding domain protein
3. Washing and elution
4. Library preparation and amplification

**RRBS**

1. Digestion with MspI
2. Library preparation
3. Gel-based size selection
4. Bisulfite treatment
5. Library amplification

**Infinium**

1. DNA preparation
2. Bisulfite conversion
3. Hybridization onto Illumina bead arrays (Infinium Human-Methylation27)
4. Data normalization using the Illumina BeadStudio software

**Validation**

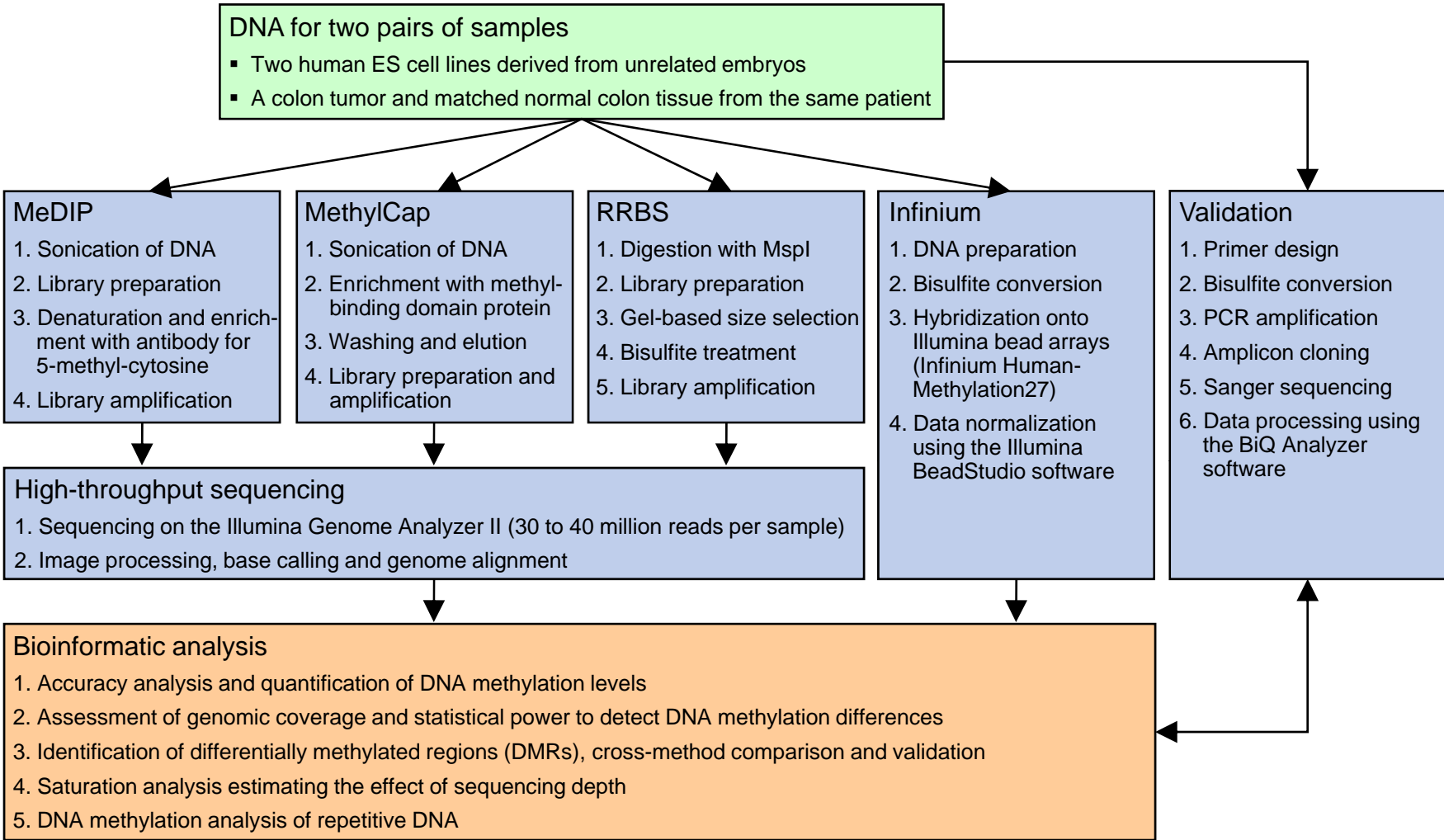
1. Primer design
2. Bisulfite conversion
3. PCR amplification
4. Amplicon cloning
5. Sanger sequencing
6. Data processing using the BiQ Analyzer software

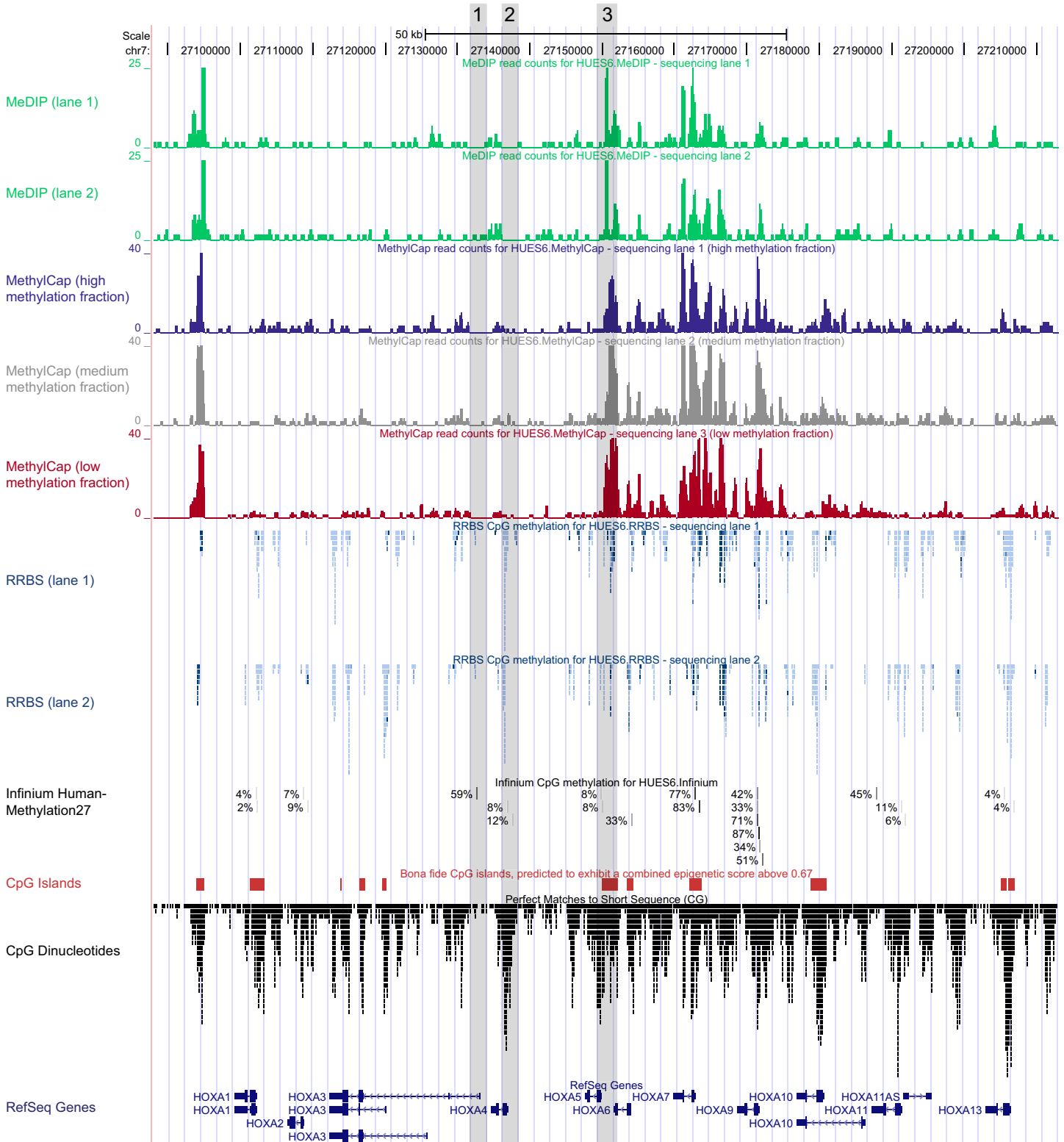
**High-throughput sequencing**

1. Sequencing on the Illumina Genome Analyzer II (30 to 40 million reads per sample)
2. Image processing, base calling and genome alignment

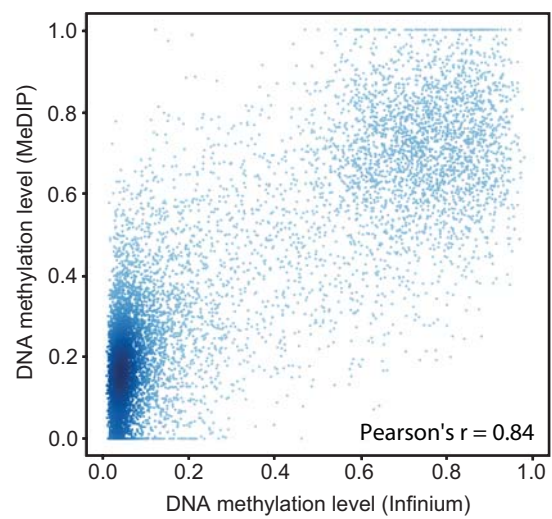
**Bioinformatic analysis**

1. Accuracy analysis and quantification of DNA methylation levels
2. Assessment of genomic coverage and statistical power to detect DNA methylation differences
3. Identification of differentially methylated regions (DMRs), cross-method comparison and validation
4. Saturation analysis estimating the effect of sequencing depth
5. DNA methylation analysis of repetitive DNA

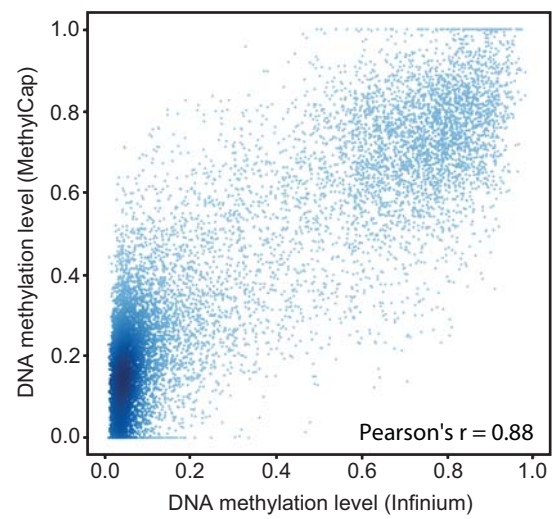




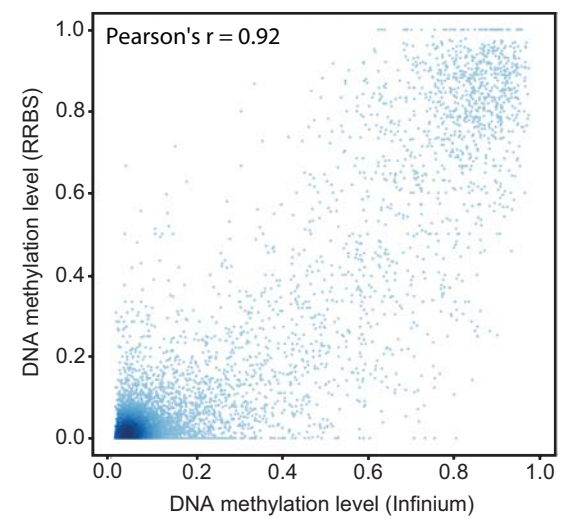
A. MeDIP (corrected for sequence bias)



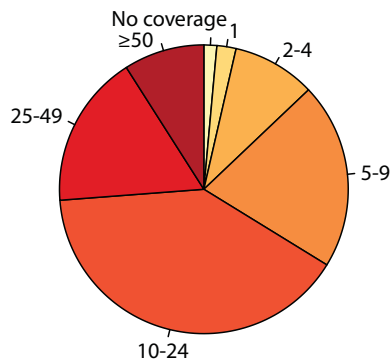
B. MethylCap (corrected for sequence bias)



C. RRBS (no correction required)



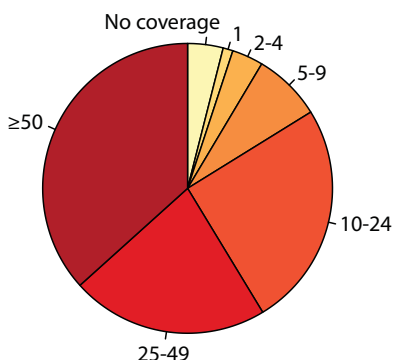
### MeDIP



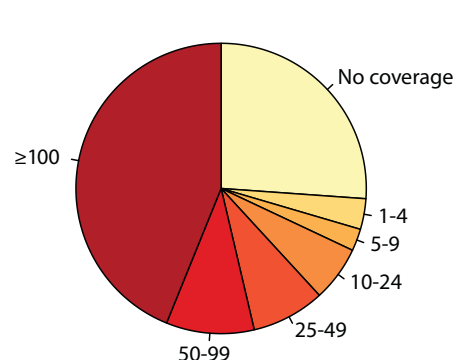
CpG Islands  
(length  $\geq 700$  bp)

44,440 regions  
genome-wide

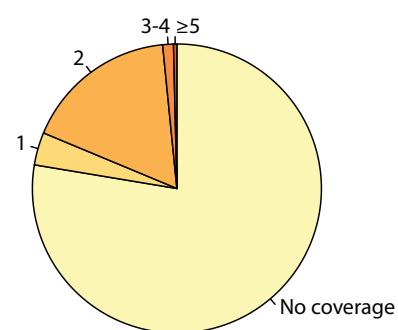
### MethylCap



### RRBS

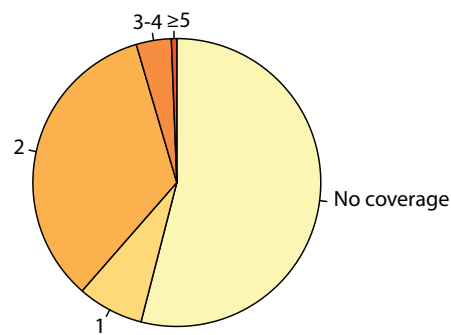
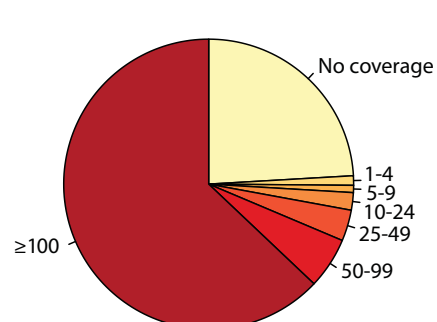
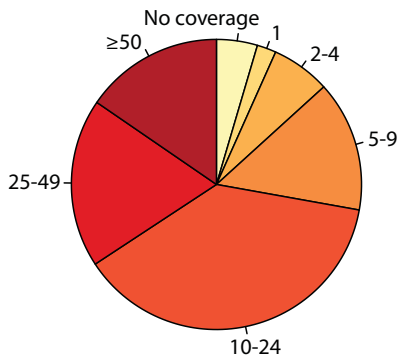
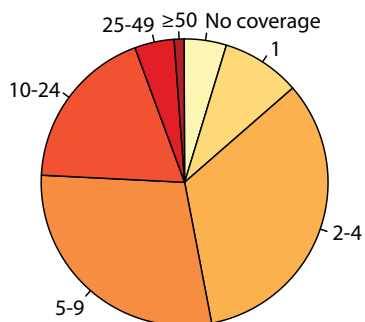


### Infinium



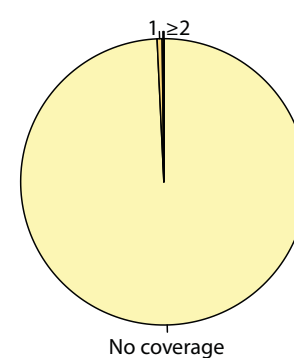
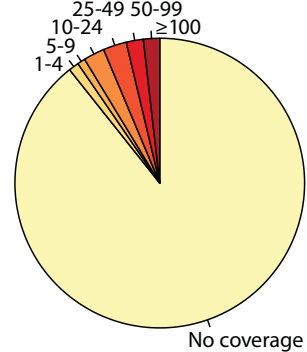
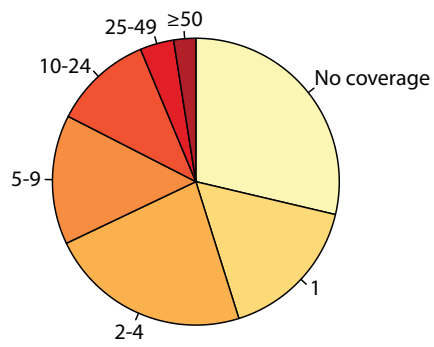
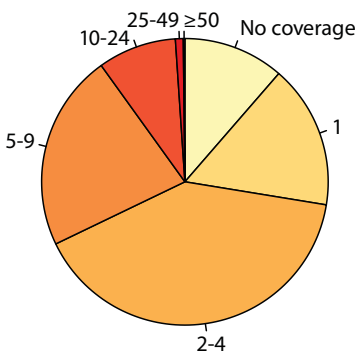
Promoter regions  
(2kb centered on TSS)

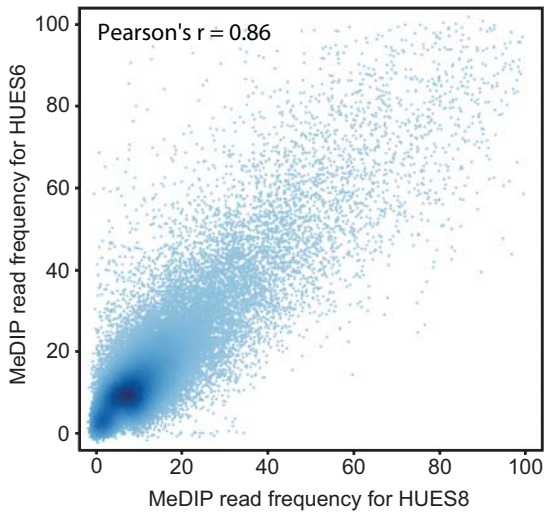
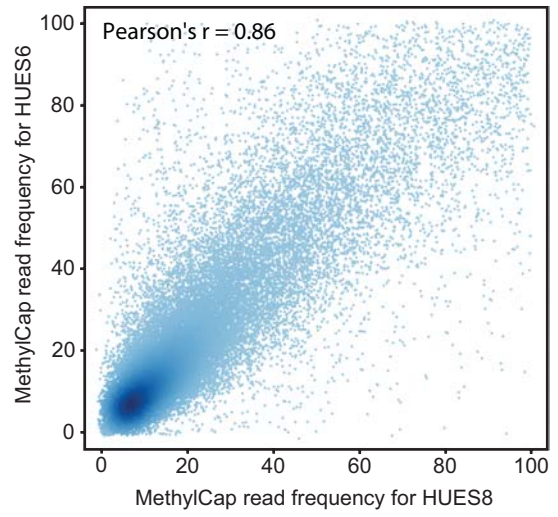
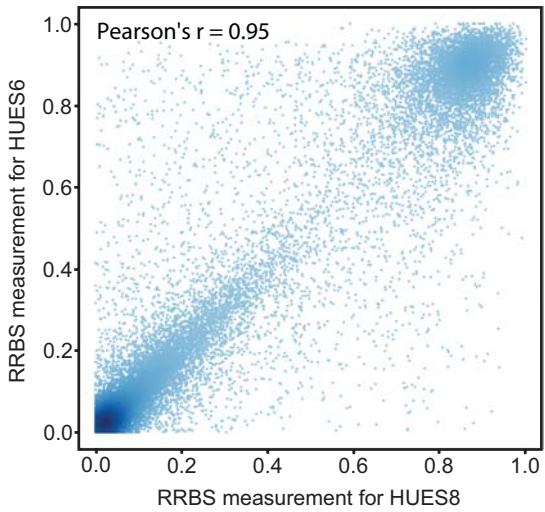
23,690 regions  
genome-wide



Whole genome  
(1kb sliding window)

2,858,143 regions  
genome-wide



**A****B****C****D**