Genomics 98 (2011) 288–295

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

High density DNA methylation array with single CpG site resolution

Marina Bibikova *, Bret Barnes, Chan Tsan, Vincent Ho, Brandy Klotzle, Jennie M. Le, David Delano, Lu Zhang, Gary P. Schroth, Kevin L. Gunderson, Jian-Bing Fan, Richard Shen

Illumina, Inc. 9885 Towne Centre Drive, San Diego, CA 92121, USA

ARTICLE INFO

Article history:
Received 19 May 2011
Accepted 26 July 2011
Available online 2 August 2011

Keywords:
DNA methylation
Methylation profiling
Genome-wide
Epigenetics
Array
Infinium

ABSTRACT

We have developed a new generation of genome-wide DNA methylation BeadChip which allows high-throughput methylation profiling of the human genome. The new high density BeadChip can assay over 480K CpG sites and analyze twelve samples in parallel. The innovative content includes coverage of 99% of RefSeq genes with multiple probes per gene, 96% of CpG islands from the UCSC database, CpG island shores and additional content selected from whole-genome bisulfite sequencing data and input from DNA methylation experts. The well-characterized Infinium® Assay is used for analysis of CpG methylation using bisulfite-converted genomic DNA. We applied this technology to analyze DNA methylation in normal and tumor DNA samples and compared results with whole-genome bisulfite sequencing (WGBS) data obtained for the same samples. Highly comparable DNA methylation profiles were generated by the array and sequencing methods (average R^2 of 0.95). The ability to determine genome-wide methylation patterns will rapidly advance methylation research.

1. Introduction

DNA methylation is one of the most studied epigenetic modifications in human cells. Changes in DNA methylation patterns play a critical role in development, differentiation and diseases such as multiple sclerosis, diabetes, schizophrenia, aging, and multiple forms of cancer. Over the past decade, interest in DNA methylation has grown rapidly and expanded across new areas of research. Consequently, DNA methylation analysis methods have undergone dramatic changes. Many microarray and next-generation sequencing based technologies have emerged, and analyses that were previously restricted to specific loci in a limited number of genes can now be performed on a genome-wide scale [1–10]. Several recent reviews compared these approaches, and discussed the strengths and weaknesses associated with microarray and next-generation sequencing-based methods for DNA methylation profiling [11–14].

The increasing affordability and throughput of sequencing-based methylation analysis promises to revolutionize study designs in the coming years, but price and throughput still remain rate-limiting for many researchers, especially in the context of large sample size studies. Methylation analysis based on Illumina's Infinium technology was first introduced with the Infinium HumanMethylation27 BeadChip [8,15]. Infinium chemistry enables the reliable measurement of methylation status with single base resolution and without the requirement for a methylated DNA capture step, which bypasses the challenges associated with capture-dependent coverage bias and allows free access to most genomic target sites. Here we describe the development of a microarray that combines the benefits of Infinium chemistry with substantially expanded genome coverage to provide high quality, genome-wide content with target selection guided by researchers' needs rather than technical limitations. CpG site selection was defined by a set of content categories identified by a Consortium of epigenetics researchers. Each category was represented with either publicly-available data or experimentally-validated sites identified internally or contributed by members of the Consortium. An emphasis was placed on gene and CpG island regions, for which 99% and 96% coverage, respectively, were achieved. In addition, 12-sample per array format provides a throughput capacity for cost and time efficient analysis of large sample cohorts. The array data show strong reproducibility between replicates and high correlation with whole genome bisulfite sequencing data generated on the same samples. By providing a unique combination of high quality content, throughput and affordability, the Infinium HumanMethylation450 provides the research community with a powerful tool for assessing epigenetic changes across a wide range of study designs.

2. Results

2.1. Infinium methylation probe design

The key advantage of Infinium technology is that the assay complexity is limited only by the number of beads which are assembled on the slide section. The Infinium methylation array uses beads with long target-specific probes designed to interrogate individual CpG sites.

[*] Corresponding author. Fax: +1 858 202 4680.
E-mail address: mbibikova@illumina.com (M. Bibikova).

0888-7543/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2011.07.007

DNA methylation is measured using quantitative "genotyping" of bisulfite-converted genomic DNA. The previously developed HumanMethylation27 array [8] employed an Infinium I methylation-specific assay design consisting of two probes per CpG locus: one "unmethylated" and one "methylated" query probe (Fig. 1A). The 3′ terminus of the probe is designed to match either the protected cytosine (methylated design) or the thymine base resulting from bisulfite conversion and whole-genome amplification (unmethylated design). For target loci with flanking CpG sites, we assumed that methylation would be regionally correlated and resolved underlying CpG sites to be in phase with the 'methylated' (C) or 'unmethylated' (T) query sites. The co-methylation assumption is based on the study by Eckhardt et al. in which they bisulfite sequenced chromosomes 6, 20, and 22, and found over 90% of CpG sites within 50 bases had the same methylation status [16]. A recent investigation of correlation of methylation states between adjacent CpG sites conducted by Shoemaker et al.[17] also showed that in general methylation status at adjacent sites tends to be correlated, though suggested that the correlation may depend on the cell types or nearby polymorphic sites. Our probes have a span of 50 bases and within this distance methylation state is expected to be highly correlated.

To maximize the utilization of the new array's capacity, we tested Infinium II assay design which requires one probe per locus for CpG sites located in regions of low CpG density. The underlying CpG sites are represented by a "degenerate" R-base, allowing multiple combinations of oligos attached to the bead. The 3′ terminus of the probe complements the base directly upstream of the query site while a single base extension results in the addition of a labeled G or A base, complementary to either the 'methylated' C or 'unmethylated' T (Fig. 1B). We demonstrated that Infinium II probes can have up to three underlying CpG sites within the 50-mer probe sequence (i.e. 2^3 possible combinations overall) without compromising data quality. This feature enables the methylation status at a query site to be assessed independently of assumptions on the status of neighboring CpG sites.

2.2. Array content selection

We included 485,577 assays (482,421 CpG sites, 3091 non-CpG sites and 65 random SNPs) representing content categories selected with the guidance of a Consortium comprised of 22 methylation researchers representing 19 institutions worldwide. The Consortium identified a series of content categories including RefSeq genes (http://www.ncbi.nlm.nih.gov/RefSeq/), CpG islands, CpG island shores [18–20], Hidden Markov Model-defined CpG islands [21,22], FANTOM 4 promoters (http://fantom.gsc.riken.jp/4/) [23,24], MHC regions [25], informatically-identified enhancers [26–28] and others. The numbers of sites represented for each content category are listed in Table 1.

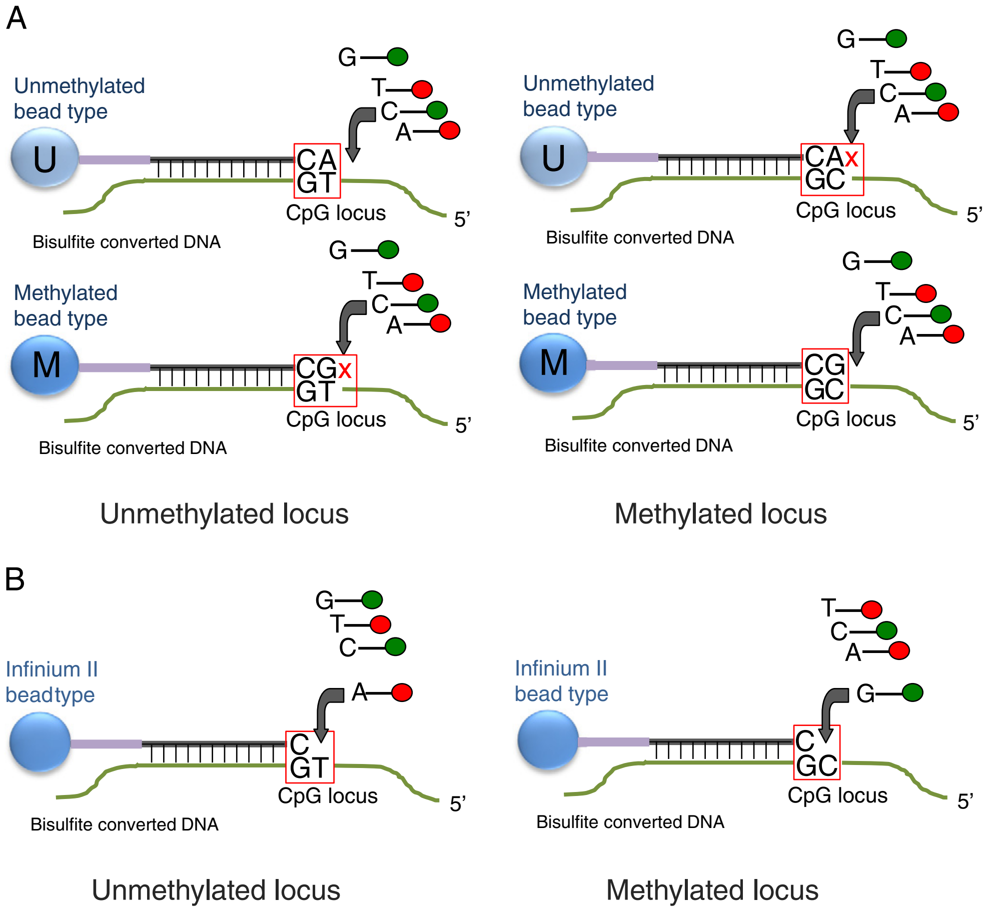

Fig. 1. Infinium Methylation Assay scheme. 1A. Infinium I assay. Two bead types correspond to each CpG locus: one bead type — to methylated (C), another bead type — to unmethylated (T) state of the CpG site. Probe design assumes same methylation status for adjacent CpG sites. Both bead types for the same CpG locus will incorporate the same type of labeled nucleotide, determined by the base preceding the interrogated "C" in the CpG locus, and therefore will be detected in the same color channel. 1B. Infinium II assay. One bead type corresponds to each CpG locus. Probe can contain up to 3 underlying CpG sites, with degenerate R base corresponding to C in the CpG position. Methylation state is detected by single-base extension. Each locus will be detected in two colors. In the current version of the Infinium II methylation assay design, labeled "A" is always incorporated at unmethylated query site ("T"), and "G" is incorporated at methylated query site ("C").

Table 1
HumanMethylation450 array content.

Feature type	Included on array
Total number of sites	485,577
RefSeq genes	21,231 (99%)
CpG islands	26,658 (96%)
CpG island shores (0–2 kb from CGI)	26,249 (92%)
CpG island shelves (2–4 kb from CGI)	24,018 (86%)
HMM islands[a]	62,600
FANTOM 4 promoters (High CpG content)[a]	9426
FANTOM 4 promoters (Low CpG content)[a]	2328
Differentially methylated regions (DMRs)[a]	16,232
Informatically-predicted enhancers[a]	80,538
DNAse hypersensitive sites	59,916
Ensemble regulatory features[a]	47,257
Loci in MHC region	12,334
HumanMethylation27 loci	25,978
Non-CpG loci	3091

[a] Features may contain multiple assay probes. One probe may belong to several content categories.

Per the Consortium's recommendations, the highest priority was placed on providing comprehensive coverage across the complete gene and CpG island regions. Toward this end, both gene and CpG island regions were subdivided according to UCSC classifications [29,30] (Fig. 2) and each subcategory was targeted individually (Table 2). Coverage of CpG island regions was further enhanced by including the 2 kb regions flanking CpG island shores (referred to here as “CpG island shelves”) (Table 3 and Fig. 2B) as well as Hidden Markov Model-defined CpG islands [21].

Also included are sites that were shown to be biologically significant/informative based on data that were generated internally or by the members of the Consortium. Other categories represented were non-CpG sites [9,31], DNase hypersensitive sites [32,33] and differentially methylated regions [18,34]. Detailed information on this content is available in the Infinium HumanMethylation450 User Guide and HumanMethylation450 manifest (www.illumina.com). A representative example of gene coverage by assay probes is shown on Fig. 2C.

2.3. Gene coverage

The array provides coverage of a total of 21,231 out of 21,474 UCSC RefGenes (NM and NR) (98.9%) with a global average of 17.2 probes per gene region (Table 2). Multiple transcripts of RefSeq genes are included, plus additional genes and transcripts not covered by the UCSC database (total of 29,246 transcripts). In order to achieve a comprehensive assessment of gene region methylation, probes covering gene regions were designed across multiple sub-regions. Promoter regions were divided into two, mutually exclusive bins of 200 bp and 1500 bp blocks upstream of the transcription start site (designated TSS200 and TSS1500, respectively). The 5′ and 3′ UTR, first exon and gene body were independently targeted as well (Fig. 2A). Details regarding the number of RefSeq genes and sub-regions, and the average number of CpG sites per gene locus represented on the array are given in Table 2.

2.4. CpG island coverage

CpG islands were defined based on UCSC annotation and as per the criteria previously described [35,36]. We employed a NCBI ‘strict’ definition for CpG islands (CGI) as DNA sequences (500 base windows; excluding most repetitive Alu-elements) with a GC base composition greater than 50% and a CpG observed/expected ratio of more than 0.6 [35,36]. As described by Takai and Jones [35], regions of DNA of greater than 500 bp with GC composition equal to or greater than 55% and observed CpG/expected CpG of 0.65 were more likely to be associated with the 5′ regions of genes. Using this definition, 60% of RefSeq genes contain one or more CGI and 40% contained no CGI. 26,658 CpG islands were covered overall with an average of 5.63 sites each. 28,249 “north” or upstream and 25,761 “south,” or downstream CpG island shores, immediately outside of the CpG islands, were targeted with averages of 2.93 and 2.81 sites, respectively. The 2 kb regions upstream and downstream of the CpG island shores, referred to here as “CpG island shelves,” were also targeted with global averages of 2.07 and 2.03 sites each (“north” and “south,” respectively) (Table 3 and Fig. 2B).

2.5. Methylation controls

To assess the overall functionality of the individual CpG assays on HumanMethylation450, we created three human genomic DNA methylation reference standards: unmethylated (U, 0%), hemi-methylated (H, 50%) and methylated (M, 100%) controls. These three reference standards were created by in vitro de-methylation (U) and subsequent in vitro methylation with SssI methylase (M) of standard Coriell genomic DNA. The hemi-methylated reference was a mixture of U and M in a 1:1 ratio. These three reference standards were run on the Infinium HumanMethylation450 methylation array, and the corresponding methylation beta-value (β = intensity of the Methylated allele (M) / (intensity of the Unmethylated allele (U) + intensity of the Methylated allele (M) + 100) was calculated for each of the >480K CpG sites. The distribution of beta-values is consistent with the three reference standards with the unmethylated (U) standard showing low beta-values, the hemi-methylated (H) standard showing intermediate beta-values, and the methylated (M) standard having high beta-values (Fig. 3). We noticed slightly different performance of Infinium I and Infinium II assays in terms of the beta-value distributions they produced. Infinium II assays demonstrate more pronounced off-axis behavior, resulting in an average upward shift in beta value of 0.02 for the unmethylated standard and an average downward shift of 0.08 for the methylated standard (Fig. 3). These differences do not significantly affect differential methylation detection; we can detect a delta beta of |0.2| with 99% confidence, a result similar to that for the HumanMethylation27 array in which all CpG sites were interrogated using Infinium I assays [8].

2.6. Assay reproducibility

To gage the technical performance of the assay, we assessed data reproducibility between technical replicates using lymphoblastoid cell lines NA17105 and NA17018, cancer cell line MCF7 and tumor and normal lung tissues (see Materials and methods section). The average correlation R^2 of beta-values for technical replicates was 0.992 (data not shown).

2.7. Correlation with HumanMethylation27 array

Over 94% of loci present on HumanMethylation27 array were included in the HumanMethylation450 array content. All loci which satisfied Infinium II design criteria were re-designed using one bead per locus. To confirm accurate methylation measurement across two platforms we compared the correlation between 450K and 27K arrays, showing an R^2 of >0.95. An example of beta value correlation for 25,978 loci in MCF7 cell line is shown in Fig. 4.

2.8. Correlation with whole-genome bisulfite sequencing data

We evaluated the correlation of methylation beta-values measured by the Infinium Methylation assay with results from whole genome bisulfite sequencing (WGBS) data generated on a HiSeq2000 (Illumina) using next-generation sequencing technology. Two comparisons were run, one with a normal lung tissue and the other with a lung tumor sample. WGBS data were filtered to include corresponding

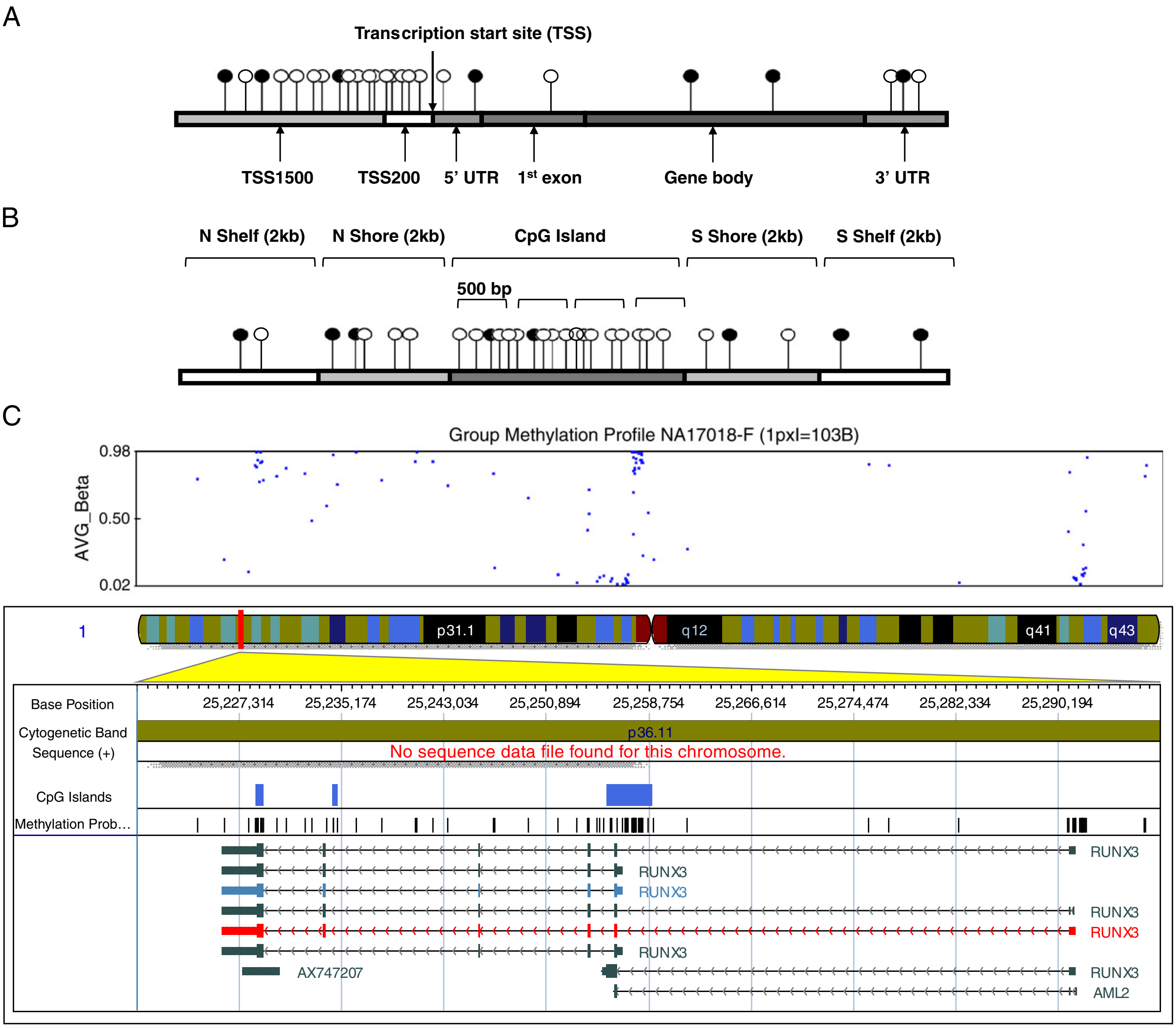

Fig. 2. Infinium Methylation probe selection. 2A. Coverage of NM and NR transcripts from UCSC database. Each transcript was divided into "functional regions" — TSS200 is the region from Transcription start site (TSS) to −200 nt upstream of TSS; TSS1500 covers −200 to −1500 nt upstream of TSS; 5′ UTR, 1st exon, gene body and 3″ UTR were also covered separately. 2B. Coverage of CpG islands and adjacent regions. CpG islands longer than 500 bp were divided into separate bins. The 2 kb regions immediately upstream and downstream of the CpG island boundaries, or "CpG island shores", and the 2 kb regions upstream and downstream of the CpG island shores, referred to here as "CpG island shelves," were also targeted separately. 2C. Coverage of the RUNX3 gene by HumanMethylation450 array probes. Blue dots in the "Group methylation Profile" window represent methylation beta values for CpG sites measured by the HumanMethylation450 array for NA17018 Coriell DNA sample. Individual assay probes are shown as black bars.

Table 2
Coverage of genes and transcripts from UCSC database.

Feature type	Genes mapped	Percent genes covered	Number of loci on array
NM_TSS200	15,957	84%	3.73
NM_TS1500	18,099	96%	4.31
NM_5′UTR	14,137	79%	4.68
NM_1stExon	15,580	82%	2.54
NM_3′UTR	13,071	72%	1.53
NM_GeneBody	17,117	97%	9.92
NR_TSS200	2140	71%	2.97
NR_TSS1500	2723	90%	3.84
NR_GeneBody	2382	79%	7.15

HumanMethylation450 loci covered with a minimum of 10 and maximum of 121 aligned reads, resulting in a total of 189,821 and 167,996 loci for comparison in the normal and tumor samples, respectively (Fig. 5A). The observed beta value correlations were 0.95 and 0.96 for

Table 3
Coverage of CpG islands from UCSC database.

Feature type	Features mapped	Percent features covered	Average number of loci on array
Island	26,658	96%	5.63
N_Shore	26,249	95%	2.93
S_Shore	25,761	93%	2.81
N_Shelf	23,965	86%	2.07
S_Shelf	24,018	87%	2.03

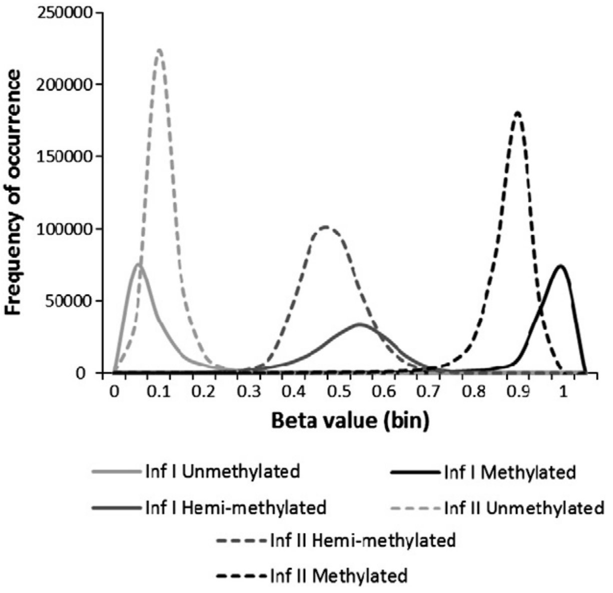

Fig. 3. Distribution of Methylation values for Infinium I and Infinium II loci. Unmethylated (U), Hemi-methylated (H), and Methylated (M) reference standards were created from Coriell genomic DNA sample as discussed in Methods. Note slightly different performance of Infinium I and Infinium II assays in regard to beta value distribution.

the normal and tumor samples, respectively. These results indicate that the beta values generated by the Infinium HumanMethylation450 array and whole genome bisulfite sequencing are consistent in reporting DNA methylation state across queried CpG loci (Fig. 5B).

3. Discussion

The body of literature focused on epigenetics research has rapidly increased over the last several years. This growth has fueled the need for new technologies and, in particular, the capability to run methylation analysis with high quality, genome-wide coverage on a platform that also offers high throughput capacity and cost-efficiency [11,13]. The Infinium HumanMethylation450 was designed with the guidance of a Consortium comprised of methylation researchers to meet these needs. The ability to quickly and affordably run genome-scale methylation analysis aligns with the requirements for large sample size studies such as The Cancer Genome Atlas project (TCGA; http://ocg.cancer.gov/programs/tcga.asp) and the International Cancer Genome Consortium initiatives (ICGC; http://www.icgc.org). The growing number of examples of reproducible associations identified through genome-wide association studies (GWAS) suggests that similar sample size ranges applied in genome-wide methylation screens could similarly lead to findings that might otherwise be missed. And while important questions pertaining to study design remain, the potential value of epigenome-wide association studies as well as the integration of genotype and methylation data across sample populations has already begun to be explored [13,37,38].

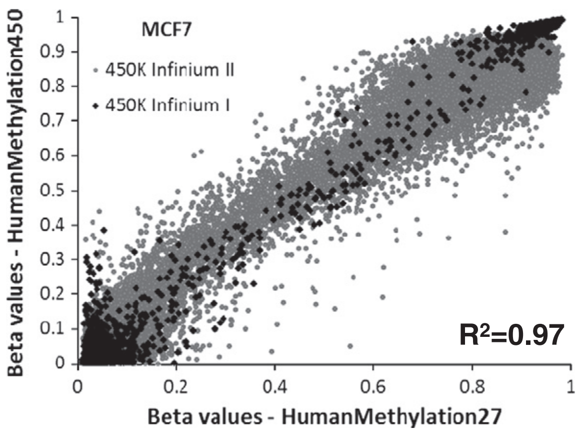

Fig. 4. Correlation between HumanMethylation450 and HumanMethylation27 arrays. The plot illustrates the correlation of beta values between HumanMethylation450 and HumanMethylation27 arrays across 25,978 different CpG sites in MCF7 cell line DNA sample. Over 90% of loci carried over from the HumanMethylation27 array were converted to Infinium II probe design for consistency with other probes on the 450K array. Good correlation ($R^2 = 0.97$) was observed between two array platforms.

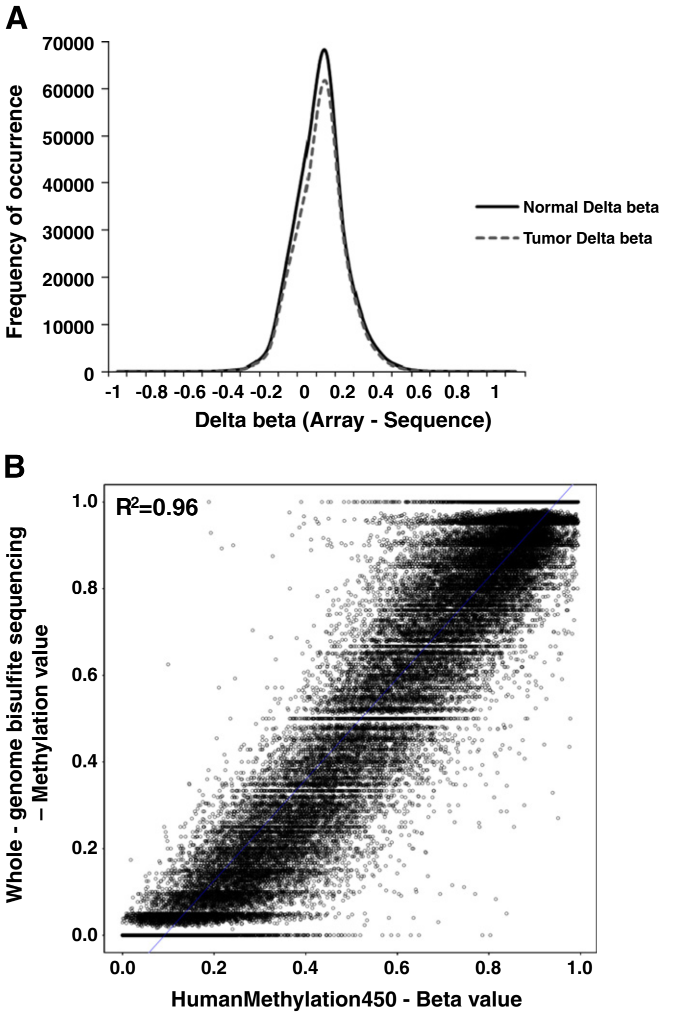

Fig. 5. Comparison between DNA methylation values generated by HumanMethylation450 array and Whole-genome bisulfite sequencing. 5A. Difference in methylation measurement between WGBS and HumanMethylation450 array. Comparison between methylation states (beta values) for 189,821 CpG loci in human normal lung sample and 167,996 CpG loci in human lung tumor sample measured on HumanMethylation450 array and by whole-genome bisulfite sequencing on Illumina HiSeq2000 for the same samples. Loci with 10–120× coverage in the sequencing data set and loci with detection p-value<0.01 were selected for comparison. Methylation values calculated for the sequencing data were subtracted from beta values generated by GenomeStudio for the HumanMethylation450 array. 88% of loci have delta beta<|0.2|, and 97% of loci have delta beta<|\0.3|. Correlation $R^2 = 0.96$. 5B. Correlation between WGBS and HumanMethylation450 array data. Scatter plot between methylation states (beta values) for a set of 48,809 CpG loci in a human lung cancer sample measured by HumanMethylation450 array and whole-genome bisulfite sequencing. Only the loci with 20–90 reads in the sequencing data set and loci with detection p-value<0.01 in the array data were selected for comparison.

The utility of Infinium HumanMethylation450 will be further extended by its applicability to FFPE samples, which was recently demonstrated using a new restoration kit (data not shown).

In summary, the HumanMethylation450 array should provide a powerful tool for investigators to fuel the continued, rapid evolution of epigenetic research [11,39] by offering simple and rapid genome-wide methylation analysis of hundreds of thousands of CpG sites across large numbers of samples. While additional sites of interest will continue to be identified, this array was designed to provide an efficient, robust and affordable discovery solution targeting core content of common interest within the epigenetics research community.

4. Materials and methods

4.1. Array design

Probe performance assessment experiments were run to determine optimal probe design parameters for both Infinium I and II designs. Assay probes were selected with the goal of providing the most complete coverage possible across the content categories identified by the Consortium. Among these content categories, gene regions and CpG islands were given top priority. Each target region was allocated a maximum loci count which was inversely related to its level of priority (e.g. gene promoter regions and CpG islands were allotted a higher number of loci than other target regions). Large regions such as large CpG islands were subdivided into separate sub-regions to ensure even coverage. After each round of target selection and probe design, empirical analytical testing removed poorly-performing probes. Subsequent rounds of selection were then run until the pool size was exhausted. This approach ensured strong probe performance as well as an optimal balance between coverage density in the highest priority regions and breadth of coverage across remaining targeted regions.

4.2. DNA samples

DNA samples NA17105 and NA17018 were purchased from the Coriell Institute for Medical Research (NJ, USA). DNA from normal and tumor lung tissues and MCF7 cell line were purchased from BioChain Institute (Hayward, CA).

4.3. Bisulfite conversion of genomic DNA

DNA samples for Infinium Methylation assay were bisulfite converted using EZ DNA methylation kit (Cat. #D5001) from Zymo Research (CA, USA). 500 ng of gDNA was denatured by addition of Zymo M-Dilution buffer (contains NaOH) and incubated for 15 min at 37 °C. CT-conversion reagent (bisulfite-containing) was added to the denatured DNA and incubated for 16 h at 50 °C in a thermocycler and denatured every 60 min by heating to 95 °C for 30 s. DNA samples for the whole-genome bisulfite sequencing were bisulfite converted using EpiTect Bisulfite conversion kit (Cat. #59104) from QIAGEN (Valencia, CA) following manufacturer's recommendations with modifications [40].

4.4. Methylation reference samples preparation

Methylation reference standards for assessment of the Infinium probes quality were prepared as described previously [8]. 50 ng of Coriell gDNA NA18105 was amplified with the REPLI-g Mini Kit (QIAGEN Cat. #150025) following manufacturer's recommendations. We used male genomic DNA in order to assess quality of the Y-chromosomal loci. The WGA amplified DNA was subjected to Mung bean nuclease treatment to remove single-stranded DNA. The resultant unmethylated DNA (U) was treated with SssI methylase, which globally methylates all double-stranded CpG sites, to create a nearly completely methylated reference standard (M). The hemi-methylated reference (H) was created by mixing U and M samples in a 1:1 stoichiometric ratio.

4.5. Infinium methylation assay

The assay was carried out as described previously [8]. In brief, 4 μl of bisulfite-converted DNA (~150 ng) was used in the whole-genome amplification (WGA) reaction. After amplification, the DNA was fragmented enzymatically, precipitated and re-suspended in hybridization buffer. All subsequent steps were performed following the standard Infinium protocol (User Guide part #15019519 A). Fragmented DNA was dispensed onto the HumanMethylation450 BeadChips, and hybridization performed in hybridization oven for 20 h. After hybridization, the array was processed through a primer extension and an immunohistochemistry staining protocol to allow detection of a single-base extension reaction [41–43]. Finally, BeadChips were coated and then imaged on an Illumina iScan.

Methylation level of each CpG locus was calculated in GenomeStudio® Methylation module as methylation beta-value (β = intensity of the Methylated allele (M)/(intensity of the Unmethylated allele (U) + intensity of the Methylated allele (M) + 100).

4.6. Whole-genome bisulfite sequencing

For the whole-genome DNA methylation analysis at single nucleotide resolution, 2–5 μg of lung normal and lung tumor genomic DNA was fragmented using Covaris shearing. The fragmented DNA was end-polished, and a single 'A' nucleotide was added to the 3′ ends of the blunt fragments. The fragments were ligated with Illumina methylated forked adaptors, and 200–400 bp fragments were selected by gel electrophoresis and purified using a QIAquick Gel Extraction Kit (QIAGEN). Purified DNA fragments were treated with bisulfite using the EpiTect Bisulfite Kit (QIAGEN) for approximately 14 h to ensure maximal conversion rate [40]. The bisulfite-treated DNA was enriched by 4 cycles of PCR with Pfu Turbo Cx DNA polymerase (Stratagene Products, Agilent, La Jolla, CA). The libraries were sequenced on Illumina HiSeq2000 sequencing instrument according to standard Illumina cluster generation and sequencing protocols with a 2 × 75 bp read length.

4.7. Data analysis

Infinium methylation data was processed with Methylation Module of GenomeStudio software using HumanMethylation450 manifest v1.1. Whole genome bisulfite sequencing data was analyzed using pipeline developed at the Salk Institute [9]. Briefly, raw sequencing data was processed using Illumina pipeline and FastQ output data was generated and aligned to the human genome (hg19) using Bowtie alignment algorithm. Methylation status for each aligned site was calculated at a minimum of 10× coverage per site.

All CpG sites with a p-value less than 0.01 on the 450k array were mapped to WGBS data on the same strand and coordinate. For each of these mapped sites a beta value was calculated by taking the number of methylated tags (C) and dividing by the sum of methylated (C) and unmethylated (T) tag counts. WGBS sites with less than 10× and greater than 120× coverage were removed. Scatter plots and r-squared statistics were then calculated by comparing array vs. sequencing beta values for the remaining matching sites.

Acknowledgments

We sincerely thank all Consortium members who contributed to the development of the HumanMethylation450 array: Devin Absher (HudsonAlpha Institute for Biotechnology), Hiroyuki Aburatani (University of Tokyo), Stephen Baylin (the Johns Hopkins University School of Medicine), Stephan Beck (University College London), Benjamin Berman (University of Southern California), Andrew P. Feinberg (the

Johns Hopkins University School of Medicine), François Fuks (Free University of Brussels), Michel Georges (University of Liège), Ivo Gut (Centre National de Génotypage), Thomas Hudson (Ontario Institute for Cancer Research), Peter W. Laird (University of Southern California), Louise Laurent (The Scripps Research Institute), Jeanne Loring (The Scripps Research Institute), Alexander Meissner (Harvard University/Broad Institute), Vardhman Rakyan (Queen Mary University of London), Brian Reed (Fred Hutchinson Cancer Research Center), Bing Ren (University of California San Diego), Tim Spector (King's College London), John Stamatoyannopoulos (University of Washington), Chia-Lin Wei (Genome Institute of Singapore), and Martin Widschwendter (University College London).

We would like to thank Juying Yan and Carmen Xiao for technical assistance with generation of the WGBS data, and Ryan Lister and Joe Ecker (Salk Institute) for help with initial WGBS data analysis.

References

[1] A. Schumacher, P. Kapranov, Z. Kaminsky, J. Flanagan, A. Assadzadeh, P. Yau, C. Virtanen, N. Winegarden, J. Cheng, T. Gingeras, A. Petronis, Microarray-based DNA methylation profiling: technology and applications, Nucleic Acids Res. 34 (2006) 528–542.

[2] M. Weber, J.J. Davies, D. Wittig, E.J. Oakeley, M. Haase, W.L. Lam, D. Schubeler, Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells, Nat. Genet. 37 (2005) 853–862.

[3] J.M. Ordway, J.A. Bedell, R.W. Citek, A. Nunberg, A. Garrido, R. Kendall, J.R. Stevens, D. Cao, R.W. Doerge, Y. Korshunova, H. Holemon, J.D. McPherson, N. Lakey, J. Leon, R.A. Martienssen, J.A. Jeddeloh, Comprehensive DNA methylation profiling in a human cancer genome identifies novel epigenetic targets, Carcinogenesis 27 (2006) 2409–2423.

[4] A. Meissner, A. Gnirke, G.W. Bell, B. Ramsahoye, E.S. Lander, R. Jaenisch, Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis, Nucleic Acids Res. 33 (2005) 5868–5877.

[5] T. Rauch, H. Li, X. Wu, G.P. Pfeifer, MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells, Cancer Res. 66 (2006) 7939–7947.

[6] S.Q. Kuang, W.G. Tong, H. Yang, W. Lin, M.K. Lee, Z.H. Fang, Y. Wei, J. Jelinek, J.P. Issa, G. Garcia-Manero, Genome-wide identification of aberrantly methylated promoter associated CpG islands in acute lymphocytic leukemia, Leukemia 22 (2008) 1529–1538.

[7] N. Omura, C.P. Li, A. Li, S.M. Hong, K. Walter, A. Jimeno, M. Hidalgo, M. Goggins, Genome-wide profiling of methylated promoters in pancreatic adenocarcinoma, Cancer Biol. Ther. 7 (2008) 1146–1156.

[8] M. Bibikova, J. Le, B. Barnes, S. Saedinia-Melnyk, L. Zhou, R. Shen, K.L. Gunderson, Genome-wide DNA methylation profiling using Infinium® assay, Epigenomics 1 (2009) 177–200.

[9] R. Lister, M. Pelizzola, R.H. Dowen, R.D. Hawkins, G. Hon, J. Tonti-Filippini, J.R. Nery, L. Lee, Z. Ye, Q.M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A.H. Millar, J.A. Thomson, B. Ren, J.R. Ecker, Human DNA methylomes at base resolution show widespread epigenomic differences, Nature 462 (2009) 315–322.

[10] Y. Ruike, Y. Imanaka, F. Sato, K. Shimizu, G. Tsujimoto, Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing, BMC Genomics 11 (2010) 137.

[11] S. Beck, Taking the measure of the methylome, Nat. Biotechnol. 28 (2010) 1026–1028.

[12] Y.W. Huang, T.H. Huang, L.S. Wang, Profiling DNA methylomes from microarray to genome-scale sequencing, Technol. Cancer Res. Treat. 9 (2010) 139–147.

[13] P.W. Laird, Principles and challenges of genome-wide DNA methylation analysis, Nat. Rev. Genet. 11 (2010) 191–203.

[14] M. Bibikova, J.B. Fan, Genome-wide DNA methylation profiling, Wiley Interdiscip. Rev. Syst. Biol. Med. 2 (2010) 210–223.

[15] C. Bock, J. Walter, M. Paulsen, T. Lengauer, CpG island mapping by epigenome prediction, PLoS Comput. Biol. 3 (2007) e110.

[16] F. Eckhardt, J. Lewin, R. Cortese, V.K. Rakyan, J. Attwood, M. Burger, J. Burton, T.V. Cox, R. Davies, T.A. Down, C. Haefliger, R. Horton, K. Howe, D.K. Jackson, J. Kunde, C. Koenig, J. Liddle, D. Niblett, T. Otto, R. Pettett, S. Seemann, C. Thompson, T. West, J. Rogers, A. Olek, K. Berlin, S. Beck, DNA methylation profiling of human chromosomes 6, 20 and 22, Nat. Genet. 38 (2006) 1378–1385.

[17] R. Shoemaker, J. Deng, W. Wang, K. Zhang, Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome, Genome Res. 20 (2010) 883–889.

[18] A. Doi, I.H. Park, B. Wen, P. Murakami, M.J. Aryee, R. Irizarry, B. Herb, C. Ladd-Acosta, J. Rho, S. Loewer, J. Miller, T. Schlaeger, G.Q. Daley, A.P. Feinberg, Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts, Nat. Genet. 41 (2009) 1350–1353.

[19] A.F. Rubin, P. Green, Mutation patterns in cancer genomes, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 21766–21770.

[20] R.A. Irizarry, C. Ladd-Acosta, B. Wen, Z. Wu, C. Montano, P. Onyango, H. Cui, K. Gabo, M. Rongione, M. Webster, H. Ji, J.B. Potash, S. Sabunciyan, A.P. Feinberg, The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores, Nat. Genet. 41 (2009) 178–186.

[21] R.A. Irizarry, H. Wu, A.P. Feinberg, A species-generalized probabilistic model-based definition of CpG islands, Mamm. Genome 20 (2009) 674–680.

[22] F. Hsieh, S.C. Chen, K. Pollard, A nearly exhaustive search for CpG islands on whole chromosomes, Int. J. Biostat. 5 (2009) 1.

[23] J. Severin, A.M. Waterhouse, H. Kawaji, T. Lassmann, E. van Nimwegen, P.J. Balwierz, M.J. de Hoon, D.A. Hume, P. Carninci, Y. Hayashizaki, H. Suzuki, C.O. Daub, A.R. Forrest, FANTOM4 EdgeExpressDB: an integrated database of promoters, genes, microRNAs, expression dynamics and regulatory interactions, Genome Biol. 10 (2009) R39.

[24] FANTOM4, Functional Annotation of the Mammalian Genome, RIKEN Omics Science Center, Yokohama City, 2009.

[25] E.M. Tomazou, V.K. Rakyan, G. Lefebvre, R. Andrews, P. Ellis, D.K. Jackson, C. Langford, M.D. Francis, L. Backdahl, M. Miretti, P. Coggill, D. Ottaviani, D. Sheer, A. Murrell, S. Beck, Generation of a genomic tiling array of the human major histocompatibility complex (MHC) and its application for DNA methylation analysis, BMC Med. Genomics 1 (2008) 19.

[26] N.D. Heintzman, R.K. Stuart, G. Hon, Y. Fu, C.W. Ching, R.D. Hawkins, L.O. Barrera, S. Van Calcar, C. Qu, K.A. Ching, W. Wang, Z. Weng, R.D. Green, G.E. Crawford, B. Ren, Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome, Nat. Genet. 39 (2007) 311–318.

[27] E. Birney, J.A. Stamatoyannopoulos, A. Dutta, R. Guigo, T.R. Gingeras, E.H. Margulies, Z. Weng, M. Snyder, E.T. Dermitzakis, R.E. Thurman, M.S. Kuehn, C.M. Taylor, S. Neph, C.M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J.A. Greenbaum, R.M. Andrews, P. Flicek, P.J. Boyle, H. Cao, N.P. Carter, G.K. Clelland, S. Davis, N. Day, P. Dhami, S.C. Dillon, M.O. Dorschner, H. Fiegler, P.G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K.D. James, B.E. Johnson, E.M. Johnson, T.T. Frum, E.R. Rosenzweig, N. Karnani, K. Lee, G.C. Lefebvre, P.A. Navas, F. Neri, S.C. Parker, P.J. Sabo, R. Sandstrom, A. Shafer, D. Vetrie, M. Weaver, S. Wilcox, M. Yu, F.S. Collins, J. Dekker, J.D. Lieb, T.D. Tullius, G.E. Crawford, S. Sunyaev, W.S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A. Sandelin, I.L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H.A. Hirsch, E.A. Sekinger, J. Lagarde, J.F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermuller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korbel, O. Emanuelsson, J.S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M.C. Dickson, D.J. Thomas, M.T. Weirauch, J. Gilbert, et al., Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, Nature 447 (2007) 799–816.

[28] N.D. Heintzman, G.C. Hon, R.D. Hawkins, P. Kheradpour, A. Stark, L.F. Harp, Z. Ye, L.K. Lee, R.K. Stuart, C.W. Ching, K.A. Ching, J.E. Antosiewicz-Bourget, H. Liu, X. Zhang, R.D. Green, V.V. Lobanenkov, R. Stewart, J.A. Thomson, G.E. Crawford, M. Kellis, B. Ren, Histone modifications at human enhancers reflect global cell-type-specific gene expression, Nature 459 (2009) 108–112.

[29] P.A. Fujita, B. Rhead, A.S. Zweig, A.S. Hinrichs, D. Karolchik, M.S. Cline, M. Goldman, G.P. Barber, H. Clawson, A. Coelho, M. Diekhans, T.R. Dreszer, B.M. Giardine, R.A. Harte, J. Hillman-Jackson, F. Hsu, V. Kirkup, R.M. Kuhn, K. Learned, C.H. Li, L.R. Meyer, A. Pohl, B.J. Raney, K.R. Rosenbloom, K.E. Smith, D. Haussler, W.J. Kent, The UCSC genome browser database: update 2011, Nucleic Acids Res. 39 (2011) D876–D882.

[30] B. Rhead, D. Karolchik, R.M. Kuhn, A.S. Hinrichs, A.S. Zweig, P.A. Fujita, M. Diekhans, K.E. Smith, K.R. Rosenbloom, B.J. Raney, A. Pohl, M. Pheasant, L.R. Meyer, K. Learned, F. Hsu, J. Hillman-Jackson, R.A. Harte, B. Giardine, T.R. Dreszer, H. Clawson, G.P. Barber, D. Haussler, W.J. Kent, The UCSC genome browser database: update 2010, Nucleic Acids Res. 38 (2010) D613–D619.

[31] L. Laurent, E. Wong, G. Li, T. Huynh, A. Tsirigos, C.T. Ong, H.M. Low, K.W. Kin Sung, I. Rigoutsos, J. Loring, C.L. Wei, Dynamic changes in the human methylome during differentiation, Genome Res. 20 (2010) 320–331.

[32] H. Lian, W.A. Thompson, R. Thurman, J.A. Stamatoyannopoulos, W.S. Noble, C.E. Lawrence, Automated mapping of large-scale chromatin structure in ENCODE, Bioinformatics 24 (2008) 1911–1916.

[33] H. Xi, H.P. Shulha, J.M. Lin, T.R. Vales, Y. Fu, D.M. Bodine, R.D. McKay, J.G. Chenoweth, P.J. Tesar, T.S. Furey, B. Ren, Z. Weng, G.E. Crawford, Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome, PLoS Genet. 3 (2007) e136.

[34] V.K. Rakyan, T.A. Down, N.P. Thorne, P. Flicek, E. Kulesha, S. Graf, E.M. Tomazou, L. Backdahl, N. Johnson, M. Herberth, K.L. Howe, D.K. Jackson, M.M. Miretti, H. Fiegler, J.C. Marioni, E. Birney, T.J. Hubbard, N.P. Carter, S. Tavare, S. Beck, An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs), Genome Res. 18 (2008) 1518–1529.

[35] D. Takai, P.A. Jones, Comprehensive analysis of CpG islands in human chromosomes 21 and 22, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 3740–3745.

[36] M. Gardiner-Garden, M. Frommer, CpG islands in vertebrate genomes, J. Mol. Biol. 196 (1987) 261–282.

[37] J.T. Bell, A.A. Pai, J.K. Pickrell, D.J. Gaffney, R. Pique-Regi, J.F. Degner, Y. Gilad, J.K. Pritchard, DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines, Genome Biol. 12 (2011) R10.

[38] V.K. Rakyan, T.A. Down, D.J. Balding, S. Beck, Epigenome-wide association studies for common human diseases, Nat. Rev. Genet. 12 (2011) 529–541.

[39] J. Sandoval, H.A. Heyn, S. Moran, J. Serra-Musach, M.A. Pujana, M. Bibikova, M. Esteller, Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome, Epigenetics 6 (2011) 692–702.
[40] F. Boellmann, L. Zhang, H.J. Clewell, G.P. Schroth, E.M. Kenyon, M.E. Andersen, R.S. Thomas, Genome-wide analysis of DNA methylation and gene expression changes in the mouse lung following subchronic arsenate exposure, Toxicol. Sci. 117 (2010) 404–417.
[41] K.L. Gunderson, F.J. Steemers, H. Ren, P. Ng, L. Zhou, C. Tsan, W. Chang, D. Bullis, J. Musmacker, C. King, L.L. Lebruska, D. Barker, A. Oliphant, K.M. Kuhn, R. Shen, Whole-genome genotyping, Methods Enzymol. 410 (2006) 359–376.
[42] F.J. Steemers, K.L. Gunderson, Whole genome genotyping technologies on the BeadArray platform, Biotechnol. J. 2 (2007) 41–49.
[43] K.L. Gunderson, Whole-genome genotyping on bead arrays, Methods Mol. Biol. 529 (2009) 197–213.