



Illumina HumanMethylation BeadChip for Genome-Wide DNA Methylation Profiling: Advantages and Limitations

115

Kazuhiko Nakabayashi

Contents

Introduction	2204
Basic Principles Underlying the Infinium Methylation Assay	2205
Manufacture and Decoding of BeadChip Arrays	2207
Experimental Procedures	2208
HumanMethylation27, 450, and EPIC BeadChip Arrays	2208
Data Analysis	2209
Step 1: Filtering Problematic Probes	2210
Step 2: Within-Array Normalization	2213
Step 3: Between-Array Normalization	2213
Conclusions	2214
Dictionary of Terms	2214
Key Facts of HumanMethylation BeadChip	2215
Summary Points	2215
References	2216

Abstract

HumanMethylation BeadChip is an array platform for highly multiplexed measurement of DNA methylation at individual CpG locus in the human genome based on Illumina's bead technology. It measures the DNA methylation level of individual CpG site by quantitative genotyping of C/T polymorphisms generated in bisulfite-converted and amplified genomic DNA. The current version, HumanMethylationEPIC, measures the DNA methylation level of >850,000 CpG sites, while the previous versions, HumanMethylation450 (HM450) and

K. Nakabayashi (✉)

Division of Developmental Genomics, Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, Tokyo, Japan

e-mail: nakabaya-k@ncchd.go.jp

© Springer Nature Switzerland AG 2019

V. B. Patel, V. R. Preedy (eds.), *Handbook of Nutrition, Diet, and Epigenetics*,
https://doi.org/10.1007/978-3-319-55530-0_89

2203

HumanMethylation27 (HM27), measured that of >480,000 and >27,000 CpG sites, respectively. HumanMethylation BeadChip requires only 4 days to produce methylome profiles of human samples using 250–500 ng of genomic DNA as a starting material. Because of its time and cost efficiency, high sample output, and overall quantitative accuracy and reproducibility, HM450 has become the most widely used means of large-scale methylation profiling of human samples in recent years. However, it is important to consider potential confounders originating in the technical limitations of HumanMethylation BeadChip such as cross-reactive probes, SNP-affected probes, within-array bias (Infinium I and II bias), and between-array bias (batch effects) especially when subtle methylation differences need to be detected by statistical tests between large numbers of cases and controls. Many integrated analysis packages have been developed by the epigenetics research community as computational solutions for technical and biological confounders associated with HumanMethylation BeadChip data. Considering the substantial increase of the coverage of regulatory regions along with the advantages inherited from HM450, EPIC is expected to maintain its popularity as a platform for epigenome-wide association studies for the foreseeable future.

Keywords

DNA methylation · Bisulfite conversion · Methylome · Infinium assay · Array · Epigenome-wide association study · Genomic imprinting · Data normalization · Batch effects · SNP genotyping

List of Abbreviations

5-mC	5-methylcytosine
CpG	Cytosine-guanine
ddNTP	Dideoxynucleotide triphosphate
DNP	Dinitrophenol
EWAS	Epigenome-wide association study
NGS	Next-generation sequencing
nt	Nucleotide
SNP	Single-nucleotide polymorphism
UTR	Untranslated region

Introduction

DNA methylation plays a critical role in the regulation of gene expression during development and differentiation. Global hypomethylation and regional hypermethylation at specific gene promoters are observed in many cancers (Kelly et al. 2010). Aberrant DNA methylation has also been described in many other diseases such as neurological, autoimmune, and metabolic diseases (Wen et al. 2016; Zhang and Zhang 2015; de Mello et al. 2014). In mammals, DNA methylation primarily occurs at the carbon 5 position of the cytosine of the CpG dinucleotide and produces

5-mC. Whole-genome and reduced representation bisulfite sequencing has up to now been the most common method of genome-wide DNA methylation profiling (Stirzaker et al. 2014). These NGS-based methods work for any species if genomic sequence information of the species is available. Illumina HumanMethylation BeadChip is an alternative method of DNA methylation profiling of human samples. While this array-based method does not interrogate DNA methylation as comprehensively as whole-genome bisulfite sequencing, it has two key advantages over the NGS-based methods: cost-effectiveness and high sample throughput. Such features have made HumanMethylation450 BeadChip and its successor HumanMethylationEPIC the first choice for large-scale epigenome-wide association studies in which hundreds or thousands of human samples are enrolled. This chapter outlines the basic principles and the experimental and analytical procedures of the HumanMethylation BeadChip and covers strengths and limitations of this array-based platform that are widely recognized by the epigenetics research community.

Basic Principles Underlying the Infinium Methylation Assay

Illumina's bead technology (Gunderson et al. 2004) enables highly multiplexed measurement of DNA methylation at individual CpG locus in the human genome. HumanMethylation BeadChip measures DNA methylation levels at the cytosine of CpG sites using quantitative genotyping of C/T polymorphisms in bisulfite-converted genomic DNA followed by whole-genome amplification: unmethylated cytosines are converted to uracils (and amplified as thymines), and methylated cytosines are protected and detected as cytosines (Hayatsu 2008).

HumanMethylation BeadChip uses two assay designs, Infinium I and Infinium II (Fig. 1). Infinium assays consist of the following three major steps: hybridization of the bisulfite-converted and amplified DNA fragments to 50-mer oligonucleotides attached to each bead, single-base extension at the 3' terminus of the oligonucleotide using fluorescently labeled ddNTPs, and the measurement of fluorescent intensity for each bead. The Infinium I assay uses two bead types per CpG locus: one for measuring the methylated state and the other for the unmethylated state. The 3' terminus of 50-mer probes is designed to match the interrogated site of either cytosine or thymine. A single-base extension occurs at the base immediately adjacent to the interrogated CpG site. Probes for Infinium I assays are designed on the assumption that methylation is regionally correlated within a 50 base interval. This "comethylation" assumption adopted by Illumina is based on the results of a large-scale bisulfite sequencing study showing that >90% of CpG sites within 50 bases had the same methylation status (Eckhardt et al. 2006) and of another study demonstrating that the methylation status of adjacent CpG sites tends to be correlated (Shoemaker et al. 2010). The Infinium II assay requires only one bead type per CpG locus (Fig. 1). The 3' terminus of 50-mer probes is cytosine complementary to the guanine of the query CpG site. A single-base extension of labeled G or A base occurs, depending on the complementarity to either the methylated C or unmethylated T. Illumina determined that Infinium II probes can contain up to 3

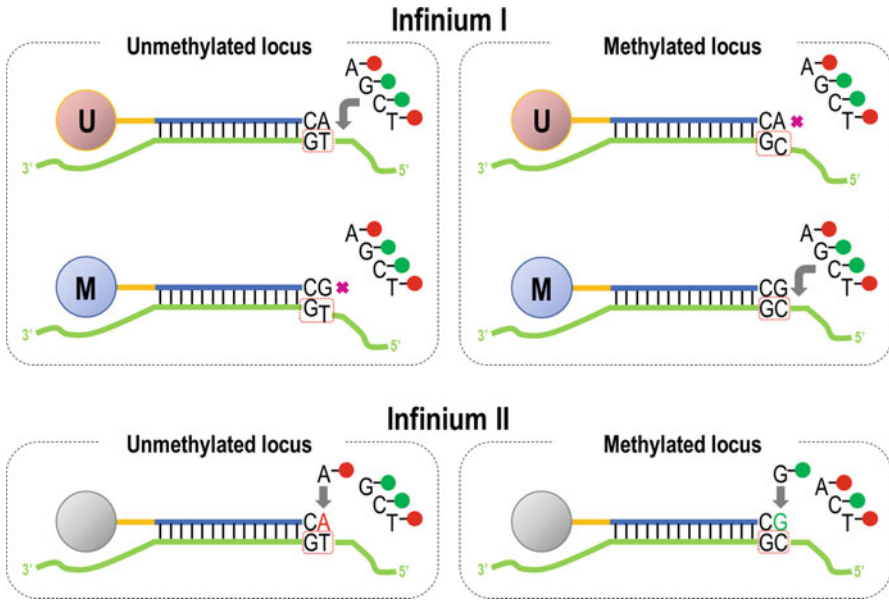


Fig. 1 Infinium assay designs. A figure in an Illumina's technical note (http://www.illumina.com/documents/products/technotes/technote_hm450_data_analysis_optimization.pdf) was reproduced with slight modifications. Bisulfite-converted DNA and interrogated CpG sites are indicated by green lines and the boxes outlined in red dots, respectively. DNP and biotin conjugated to ddNTPs are shown as red and green circles. One of the four types of ddNTPs (DNP-labeled ddATP/ddTTP and biotin-labeled ddCTP/ddGTP) is incorporated at the 3' end of the probe oligonucleotides when the last two bases of the probe oligonucleotides are complementary to the interrogated CpG site in bisulfite-converted DNA. In the Infinium I assay design, single-base extension occurs at the base adjacent to the interrogated CpG site. The same type of dideoxynucleotide is incorporated in both the unmethylated and methylated loci. However, in the Infinium II assay design, the single-base extension occurs at the C/T nucleotides of the interrogated CpG site. Either DNP-labeled ddATP or biotin-labeled ddGTP is incorporated depending on the methylation state of the interrogated CpG site. Incorporated ddNTPs are immunofluorescently detected using Cy5-labeled anti-DNP and Cy3-labeled streptavidin followed by signal amplification. Therefore, while the same color channel (red or green) is used to measure the fluorescent intensities for the unmethylated and methylated status of the interrogated CpG site in the Infinium I assay design, in the Infinium assay II design, two color channels are used, red for the unmethylated status and green for the methylated status of the interrogated CpG site. Dye bias between two color channels can be adjusted using signal intensities of built-in control probes

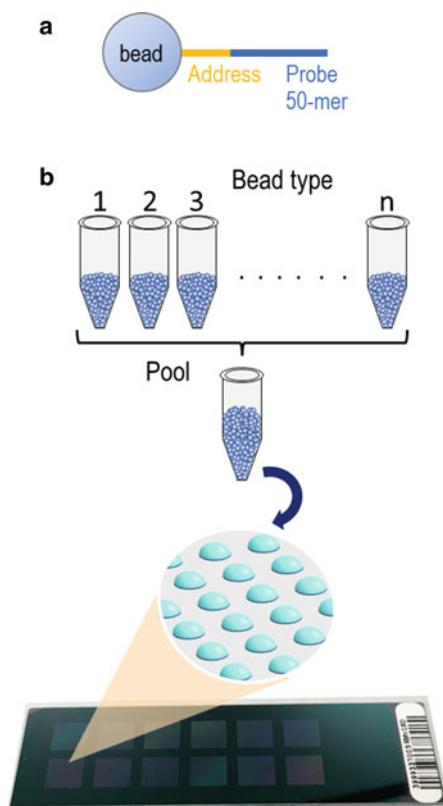
CpG sites within a 50-mer probe sequence and that the underlying CpG sites may be represented by “degenerate” R-bases (A or G base). This feature makes it possible to assess the methylation status of a query site independently of assumptions on the status of neighboring CpG sites and to increase the number of CpG sites interrogated. While the Infinium II design was applied whenever possible, the Infinium I design was required to measure the DNA methylation levels of CpG sites within CpG-rich regions (i.e., CpG islands).

Manufacture and Decoding of BeadChip Arrays

BeadChip technology is based on 3 micron silica beads that self-assemble in microwells with a uniform spacing of approximately 5.7 microns on planar silica slides (Fan et al. 2005). Each bead in the Infinium assay holds hundreds of thousands of copies of a specific oligonucleotide comprising an address code and a 50-mer probe sequence (Fig. 2). All types of beads are pooled and randomly assembled in the microwells by Van der Waals forces and hydrostatic interactions with the walls of the well on the array slide. The arrays are subsequently subjected to the decoding procedure, which determines the positions of all types of beads in each array (Fan et al. 2005). Sequential hybridizations of fluorescently labeled decoder oligonucleotides (with four distinguishable labels) to the address code of the beads impart a color signature specifying a bead type to each of the beads on the array. The information of the bead locations for each array is provided as a DMAP file. In the case of HumanMethylation BeadChip, a median of 14 beads is randomly distributed on the array. The presence of multiple beads of each type provides quantitative accuracy through multiple measurements with statistical processing. The random distribution

Fig. 2 Manufacture of BeadChip arrays. (a)

Illumina's BeadChip technology uses 3 micron silica beads attached to hundreds of thousands of copies of a specific oligonucleotide comprising an address code and a 50-mer probe sequence (depicted in orange and blue, respectively). (b) In the case of HumanMethylation450 BeadChip, over 620,000 types of beads are generated, pooled, and self-assembled onto the array. On average, 14 copies of each bead type are present on an array. Each array is subjected to the decoding procedure, which determines the positions of each type of beads for each array (Gunderson et al. 2004)



of the beads increases assay robustness by minimizing the chance of any local failure affecting the overall result (Gunderson et al. 2004). In the randomly assembled arrays, the number of each type of bead is a random variable with a Poisson distribution. When the number of a bead is less than three (i.e., 0, 1, or 2), the methylation beta value is not calculated for the corresponding CpG site by the GenomeStudio software (Illumina). In the case of HumanMethylation450 BeadChip, several hundred sites (out of >480,000 CpG sites) per sample inevitably lack the methylation beta value due to the bead number threshold.

Experimental Procedures

Experimental procedures using HumanMethylation BeadChip include sodium bisulfite conversion of genomic DNA (day 1), whole-genome amplification (day 2) followed by DNA fragmentation, hybridization of bisulfite-converted and amplified DNA to BeadChip (day 3), and washing, staining, and scanning (day 4) (Carless 2015). The staining procedure includes a single-base extension using labeled ddNTPs (DNP-labeled ddATP and ddTTP and biotin-labeled ddCTP and ddGTP), fluorescent detection of DNP and biotin with Cy5-labeled anti-DNP and Cy3-labeled streptavidin, and signal amplification. The stained BeadChip is scanned using iScan or HiScan SQ system (Illumina) to measure the intensities of methylated and unmethylated signals for each of the target CpG sites. In this scanning procedure, decoded data (DMAP) files containing bead-type information for each position on the BeadChip are required to obtain raw data files (IDAT files). The GenomeStudio software (methylation module) calculates the methylation level of individual CpG sites from IDAT files as beta value ranging from 0 (completely unmethylated) to 1 (completely methylated).

HumanMethylation27, 450, and EPIC BeadChip Arrays

HumanMethylation27 (HM27) BeadChip, the first version of HumanMethylation BeadChip, was released in 2008 and contained Infinium I probes for 27,578 CpG sites located within the promoter regions of 14,475 consensus coding sequences (CCDS) genes and well-known cancer genes (Bibikova et al. 2009). HumanMethylation450 (HM450) BeadChip was released in 2011. HM450 was designed to include 485,577 assays (482,421 CpG sites, 3091 non-CpG sites, and 65 random SNPs) containing 25,978 (94%) of CpG sites present on HM27. The new content covered a more diverse set of genomic features: three CpG island (CGI)-related categories, CGI, shore (0–2 kb from CGI), and shelf (2–4 kb from CGI), and six gene feature categories, namely, TSS200 (from transcription start site (TSS) to –200 nt upstream of TSS), TSS1500 (–200 to –1500 nt upstream of TSS), 5'UTR, first exon, gene body, and 3'UTR, which were introduced to probe annotation. HM450 provides coverage of a total of 21,231 out of 21,474 UCSC RefGenes (99%) and 26,658 (96%) CGIs (Bibikova et al. 2011). This improved coverage of

HM450 along with the advantages of the HumanMethylation BeadChip platform, time and cost-effectiveness, and high sample throughput resulted in the use of HM450 in many studies of various types. Examples of the HM450 studies include those of genomic imprinting (Court et al. 2014), X inactivation (Cotton et al. 2015), myoblast differentiation (Miyata et al. 2015), aging (Florath et al. 2014), and many types of cancers. Examples of EWAS, in which the DNA methylation profiles of over a thousand subjects were obtained by HM450, are those for rheumatoid arthritis (Liu et al. 2013), metabolic traits (Dick et al. 2014, Petersen et al. 2014), cardiovascular diseases (Rask-Andersen et al. 2016), and diabetes (Florath et al. 2016). Other examples of large-scale projects that used HM450 are the ARIES study, a population-based methylome project that profiled 1000 mother and child pairs (Relton et al. 2015), and the Cancer Genome Atlas Consortium (TCGA, <http://cancergenome.nih.gov/>), which profiled about 8000 samples from over 200 cancer types.

DNA methylation at distal regulatory elements, particularly enhancers and transcription factor binding sites, has been shown to be dynamically regulated among different tissues and cell types (Ziller et al. 2013; Gu et al. 2016). The latest version, HumanMethylationEPIC BeadChip, was released in 2015. EPIC covers >850,000 CpG sites including >90% of HM450 content and 413,743 additional CpG sites. The newly added CpG sites include 35,000 CpG sites located at potential enhancers identified by the FANTOM5 project (<http://fantom.gsc.riken.jp/5/>) and the ENCODE project (<http://www.encodeproject.org/>). Pidsley et al. (2016) reported that EPIC probes cover 58% of FANTOM5 enhancers and 7% distal and 27% proximal ENCODE regulatory elements and that a single EPIC probe does not always represent the methylation level of distal enhancer elements, which tend to show variable methylation levels across a region. However, because of the substantial increase of coverage of regulatory regions along with the advantages it has inherited from HM450, EPIC is expected to maintain its popularity as an EWAS platform. The major features of HM27, HM450, and EPIC BeadChip arrays are summarized in Table 1. The genomic locations of the CpG probes contained in HM27, HM450, and EPIC BeadChip arrays are shown for a 36 kb genomic interval including the *TAL1* gene (Fig. 3) to exemplify the difference in probe distribution and density among the three types of HumanMethylation BeadChip arrays.

Data Analysis

This section describes the chief data processing procedures generally recommended for an HM450/EPIC dataset: probe filtering, within-array normalization, and between-array normalization (batch effect correction). These procedures, if correctly conducted, minimize variance and improve statistical power for detecting small DNA methylation changes. Several bioconductor or R packages that perform all or most of the procedures are available for free (Table 2) including methylumi, minfi, waterRmelon, ChAMP, and RnBeads, which have been reviewed by Morris and Beck (2015). JIllumina is an open-source Java library recently released for the handling and processing of HM450/EPIC raw data (Almeida et al. 2016) (Table 2)

Table 1 Comparison of probe content and the number of use of HM27, HM450 and EPIC

	EPIC (2015)	HM450 (2011)	HM27 (2009)
Target C	867,867	485,512	27,578
Target CpG	864,935	482,421	27,578
Target CpG_Infinium I	143,293	135,476	27,578
Target CpG_Infinium II	721,642	346,945	0
Non-CpG target	2,932	3,091	0
SNP probes	59	65	32
Control probes	638	850	112
Sample number/array	8	12	12
Genomic DNA required	250 ng	500 ng	1000 ng
GEO platform ID	GPL21145	GPL13534 GPL16304	GPL8490
Samples registered in GEO (as of march 2017)	70	56,543 2,268	18,708

GEO: Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>)

Probe annotation files are available at GEO

and forms the backbone of DiMmer (Almeida et al. 2017), a graphical user interface software that interactively guides EWAS data analysis procedures. Although Illumina's GenomeStudio is useful for analyzing a small number of samples, its functions are limited compared to the software packages listed above.

Step 1: Filtering Problematic Probes

Removal of probes with high detection p -values (probes that have failed to hybridize) is generally recommended. The detection p -value is defined as "1-R/N," where R is the relative rank of the signal intensity relative to the negative controls and N is the number of negative controls (N = 614 in the case of HM450). Illumina recommends the removal of probes with a detection p -value >0.05. In addition, removing probes with missing beta values, probes containing common SNPs, and cross-reactive probes should also be considered. The Infinium methylation assay is based on the quantitative genotyping of C/T polymorphisms generated at CpG sites after bisulfite treatment. When C/T SNP exists at the target CpG sites, the genomic TG allele behaves as unmethylated CpG. Therefore, probes with target CpG sites overlapping a common SNP (minor allele frequency > 0.05) should be treated with caution in studies enrolling multiple human subjects. HumanMethylation BeadChip relies on the hybridization of bisulfite-treated DNA to 50-mer oligos attached to the beads. Bisulfite conversion of the majority of Cs to T generates "three-letter genome sequences," increasing the probability of probe cross-reactivity, i.e., hybridization of a probe to loci other than its primary target. Depending on the criteria used, 8–25% of the HM450 probes were classified as cross-reactive (Price et al. 2013; Chen et al. 2013). Therefore, when differentially methylated probes are extracted by the comparison of HumanMethylation data between groups of interest without prior

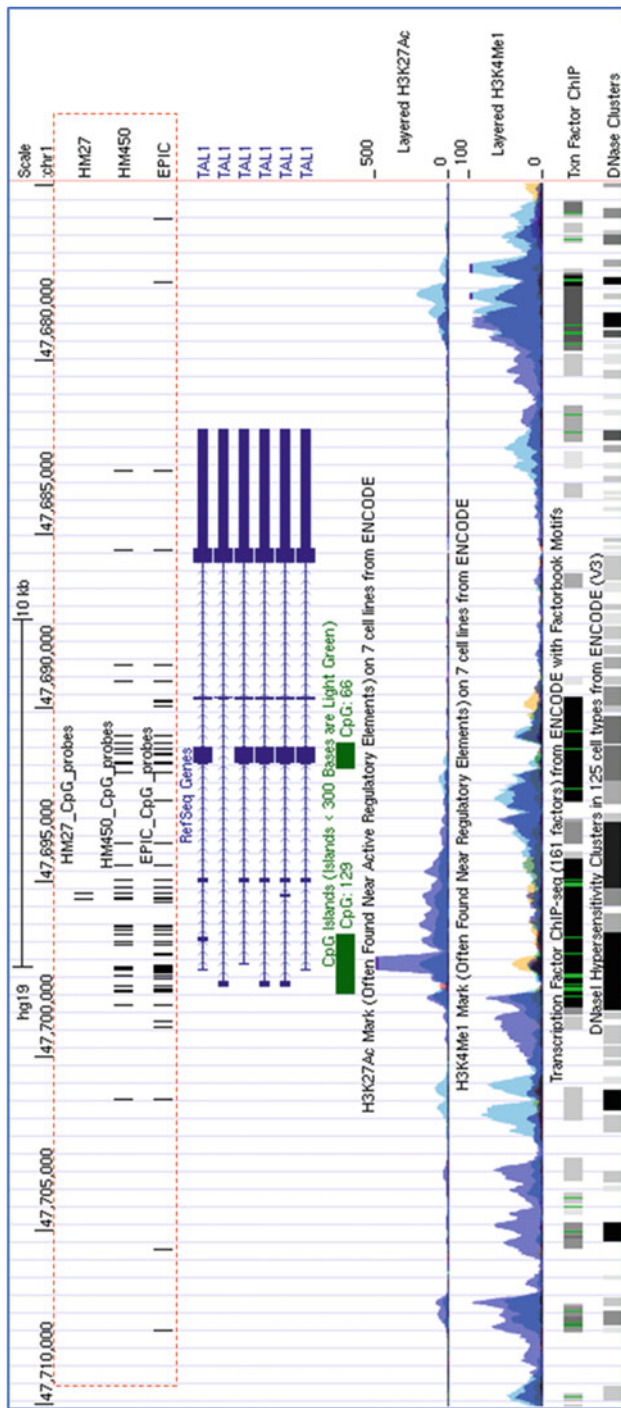


Fig. 3 Distribution of CpG probes included in three types of HumanMethylation BeadChip arrays within a 36 kb genomic interval. The genomic interval of chr1:47,674,001–47,710,000 (hg19) including the *TALI* gene is shown using UCSC Genome Browser (<http://genome.ucsc.edu/>) (Kent et al. 2002). The locations of the CpG probes contained in HM27, HM450, and EPIC BeadChip arrays within this interval were shown as vertical black lines by uploading the genomic positions of the probes in the BED (Browser Extensible Data) format using the custom track tool of the Browser. The distribution of the CpG probes in this interval exemplifies the wider coverage of enhancer regions, which are positive for H3K4Me1 signals, by EPIC probes than by HM450 probes

Table 2 Features of data analysis packages for HumanMethylation BeadChip data

	GenomeStudio (Illumina)	Mimfi	WaterMelon	ChAMP	RnBeads	Illumina
Import raw data (IDAT) files	Yes	Yes	Yes	Yes	Yes	Yes
Probe filtering	Yes (1 and 2 only)	Yes	Yes	Yes	Yes	Yes
1. Detection p-value						
2. Bead count						
3. SNP-containing probes						
4. Cross-reactive probes						
Within-array normalization:	Yes (1 only)	Yes	Yes	Yes	Yes	Yes
1. Background correction						
2. Type II bias adjustment						
Between-array normalization:	No	Yes	No	Yes	Yes	No
Batch effect correction						
Cell composition correction	No	Yes	No	Yes	Yes	Yes
Differentially methylated probes	Yes	Yes	No	Yes	Yes	Yes
Differentially methylated regions	No	Yes	No	Yes	Yes	Yes
Copy number analysis	No	No	No	Yes	No	No

exclusion of SNP-containing and cross-reactive probes, it is important to consider the possibility that such problematic probes may have been falsely identified as differentially methylated.

Step 2: Within-Array Normalization

Removing variations of a technical, rather than a biological, origin is crucial. Within-array normalization is performed for background correction, dye-bias adjustment, and type II bias adjustment (Yousefi et al. 2013; Dedeurwaerder et al. 2014). Type II bias is the foremost issue originating from the two different assay designs, Infinium I and Infinium II, used in HM450 and EPIC. The methylation values derived from these two designs were shown to exhibit different distributions. Infinium II probes showed a lower dynamic range compared with Infinium I probes (Dedeurwaerder et al. 2011; Bibikova et al. 2011). Furthermore, methylation values of the Infinium II probes were shown to be biased and less reproducible in comparison with bisulfite pyrosequencing data (Dedeurwaerder et al. 2011). This issue has led to the development of a variety of algorithms to adjust for type II bias (reviewed in Morris and Beck 2015).

Step 3: Between-Array Normalization

Batch effects represent nonbiological, technical variations observed across different batches of experiments and are common among high-throughput data including HumanMethylation BeadChip data (Leek et al. 2010; Sun et al. 2011). Differences in experiment dates, reagent lots, laboratory conditions, and personnel can cause batch effects. In large-scale studies, in which BeadChip experiments need to be performed with many arrays at different times, batch effects may not always be avoidable. Batch effects are also expected when data from multiple institutes are analyzed together. Several algorithms are available for batch effect correction for HumanMethylation BeadChip data including ComBat, surrogate variable analysis (SVA), RUVm, functional normalization, and BEclear (see the references in Akulenko et al. 2016).

Lie and Siegmund (2016) evaluated the effects of nine processing methods including both within- and between-array normalization methods and found that within-array normalization using Noob (Triche et al. 2013) and BMIQ (Teschendorff et al. 2013) consistently improved signal sensitivity and that RUVm (Maksimovic et al. 2015) outperformed the other batch effect correction algorithms in a combination analysis using Noob and BMIQ. Inclusion of batch effect correction depends on the degree of methylation differences expected between the groups compared; the process is facilitated if variations stemming from a technical origin outnumber those stemming from biological signals, but in the opposite situation, some batch effect correction tools may remove biological signals (Liu and Siegmund 2016).

In addition to the options for probe filtering and within- and between-array normalization, many of the current software packages (Table 2) include options for correcting for cell heterogeneity, another potential confounding factor associated with methylome studies using human population samples. Whole blood is frequently chosen as a source of genomic DNA in EWAS. Because blood is a heterogeneous collection of different cell types, the associations detected in such studies may be confounded by cellular heterogeneity (Jaffe and Irizarry 2014).

After appropriate data preprocessing, differentially methylated probes (DMP) or regions (DMR) between groups of interest can be searched. Probe annotations such as those of Price et al. (2013) and Zhou et al. (2017) are useful for the biological interpretation of identified DMPs/DMRs. The functional annotation provided by Zhou et al. (2017) includes the positions of HM27/HM450/EPIC probes relative to imprinted differentially methylated regions (Court et al. 2014), transcription factor binding sites identified by the ENCODE project, and chromatin states identified by the Roadmap Epigenomics Project (<http://www.roadmapepigenomics.org/>).

Conclusions

The HumanMethylationEPIC array represents a significant improvement over the HM450 array in terms of genome coverage of regulatory regions, consistent reproducibility, and reliability. Therefore, this platform will serve as a valuable tool in large-scale human methylome studies requiring identification of disease biomarkers and/or etiology and assessment of the impact of environmental exposures. Many integrated analysis packages have been developed as computational solutions for technical confounders originating in the underlying principle and design of HumanMethylation BeadChip as well as biological cofounders inherent in human biology.

Dictionary of Terms

- **Methylome** – The genome-wide profile of DNA methylation.
- **EWAS** – Epigenome-wide association study for uncovering epigenetic variants underlying common diseases.
- **Bisulfite conversion** – Sodium bisulfite conversion of genomic DNA to differentiate and detect unmethylated versus methylated cytosine.
- **Infinium assay** – Illumina's SNP genotyping technology based on on-bead single-base extension and fluorescent detection.
- **Methylation beta value** – $\text{Beta} = \frac{\text{methylated intensity}}{\text{unmethylated intensity} + \text{methylated intensity} + 100}$.
- **Detection p-value** – Threshold value for filtering out probes that failed to hybridize in HumanMethylation BeadChip experiments.
- **Batch effects** – Nonbiological experimental variation observed across multiple batches of microarray experiments.

Key Facts of HumanMethylation BeadChip

- HumanMethylation BeadChip is an array platform for highly multiplexed measurement of DNA methylation at individual CpG loci in the human genome based on Illumina's bead technology.
- The platform measures the DNA methylation level of individual CpG sites by quantitative genotyping of C/T polymorphisms generated in bisulfite-converted and amplified genomic DNA.
- The current version, HumanMethylation EPIC, measures the DNA methylation level of >850,000 CpG sites, while previous versions, HumanMethylation450 and HumanMethylation27, measured that of >480,000 and >27,000 CpG sites, respectively.
- HumanMethylation BeadChip requires only 4 days to produce methylome profiles of human samples using 250–500 ng of genomic DNA as a starting material.
- Because of its time and cost efficiency, high sample output, and overall quantitative accuracy and reproducibility, HumanMethylation450 was the most widely used means of large-scale methylation profiling of human samples in recent years.
- Considering the substantial increase of the coverage of regulatory regions along with the advantages inherited from HumanMethylation450, HumanMethylation EPIC is expected to maintain its popularity as a platform for epigenome-wide association studies for the foreseeable future.

Summary Points

- Illumina HumanMethylation BeadChip is a popular platform for obtaining DNA methylome data from human samples.
- The current version, HumanMethylationEPIC, measures the DNA methylation levels of over 850,000 CpG sites by highly multiplexed quantitative genotyping of C/T polymorphisms in bisulfite-converted genomic DNA.
- HumanMethylation450 was very widely used in EWAS projects, and its successor, EPIC, is expected to remain the preferred tool for the foreseeable future in studies with large sample numbers due to its cost and time effectiveness and high sample throughput.
- Considering potential confounders originating in the technical limitations of HumanMethylation BeadChip such as cross-reactive probes, SNP-affected probes, within-array bias (Infinium I and II bias), and between-array bias (batch effects) is of particular importance, especially when subtle methylation differences need to be detected by statistical tests between large numbers of cases and controls.
- Many integrated analysis packages have been developed by the epigenetics research community as computational solutions for technical and biological confounders associated with HumanMethylation BeadChip data.

References

- Akulenko R, Merl M, Helms V (2016) BEclear: batch effect detection and adjustment in DNA methylation data. *PLoS One* 11:e0159921
- Almeida D, Skov I, Lund J et al (2016) Jllumina - a comprehensive java-based API for statistical Illumina Infinium HumanMethylation450 and MethylationEPIC data processing. *J Integr Bioinform* 13:294
- Almeida D, Skov I, Silva A et al (2017) Efficient detection of differentially methylated regions using DiMmeR. *Bioinformatics* 33:549–551
- Bibikova M, Le J, Barnes B et al (2009) Genome-wide DNA methylation profiling using Infinium[®] assay. *Epigenomics* 1:177–200
- Bibikova M, Barnes B, Tsan C et al (2011) High density DNA methylation array with single CpG site resolution. *Genomics* 98:288–295
- Carless MA (2015) Determination of DNA methylation levels using Illumina human methylation BeadChips. In: Chellappan SP (ed) *Chromatin protocols, Methods in molecular biology*, vol 1288. Springer, New York, pp 143–192
- Chen YA, Lemire M, Choufani S et al (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 8:203–209
- Cotton AM, Price EM, Jones MJ et al (2015) Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Hum Mol Genet* 24:1528–1539
- Court F, Tayama C, Romanelli V et al (2014) Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. *Genome Res* 24:554–569
- Dedeurwaerder S, Defrance M, Bizet M et al (2014) A comprehensive overview of Infinium HumanMethylation450 data processing. *Brief Bioinform* 15:929–941
- Dedeurwaerder S, Defrance M, Calonne E et al (2011) Evaluation of the Infinium methylation 450K technology. *Epigenomics* 3:771–784
- de Mello VD, Pulkkinen L, Lalli M et al (2014) DNA methylation in obesity and type 2 diabetes. *Ann Med* 46:103–113
- Dick KJ, Nelson CP, Tsaprouni L et al (2014) DNA methylation and body-mass index: a genome-wide analysis. *Lancet* 383:1990–1998
- Eckhardt F, Lewin J, Cortese R et al (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 38:1378–1385
- Fan JB, Hu SX, Craumer WC et al (2005) BeadArray-based solutions for enabling the promise of pharmacogenomics. *BioTechniques* 39:583–588
- Florath I, Butterbach K, Müller H et al (2014) Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet* 23:1186–1201
- Florath I, Butterbach K, Heiss J et al (2016) Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. *Diabetologia* 59:130–138
- Gunderson KL, Kruglyak S, Graige MS et al (2004) Decoding randomly ordered DNA arrays. *Genome Res* 14:870–877
- Gu J, Stevens M, Xing X et al (2016) Mapping of variable DNA methylation across multiple cell types defines a dynamic regulatory landscape of the human genome. *G3 (Bethesda)* 6:973–986
- Hayatsu H (2008) Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis – a personal account. *Proc Jpn Acad Ser B Phys Biol Sci* 84:321–330
- Jaffe AE, Irizarry RA (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 15:R31
- Kelly TK, De Carvalho DD, Jones PA (2010) Epigenetic modifications as therapeutic targets. *Nat Biotechnol* 28:1069–1078
- Kent WJ, Sugnet CW, Furey TS et al (2002) The human genome browser at UCSC. *Genome Res* 12:996–1006

- Leek JT, Scharpf RB, Bravo HC et al (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 11:733–739
- Liu J, Siegmund KD (2016) An evaluation of processing methods for HumanMethylation450 BeadChip data. *BMC Genomics* 17:469
- Liu Y, Aryee MJ, Padyukov L et al (2013) Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* 31:142–147
- Miyata K, Miyata T, Nakabayashi K et al (2015) DNA methylation analysis of human myoblasts during in vitro myogenic differentiation: de novo methylation of promoters of muscle-related genes and its involvement in transcriptional down-regulation. *Hum Mol Genet* 24:410–423
- Morris TJ, Beck S (2015) Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods* 72:3–8
- Maksimovic J, Gagnon-Bartsch JA, Speed TP et al (2015) Removing unwanted variation in a differential methylation analysis of Illumina HumanMethylation450 array data. *Nucleic Acids Res* 43:e106
- Petersen AK, Zeilinger S, Kastenmüller G et al (2014) Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. *Hum Mol Genet* 23:534–545
- Pidsley R, Zotenko E, Peters TJ et al (2016) Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* 17:208
- Price ME, Cotton AM, Lam LL et al (2013) Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* 6:4
- Rask-Andersen M, Martinsson D, Ahsan M, et al (2016) Epigenome-wide association study reveals differential DNA methylation in individuals with a history of myocardial infarction. *Hum Mol Genet* (Epub ahead of print) PMID: 27634651
- Relton CL, Gaunt T, McArdle W, Ho K, Duggirala A, Shihab H, Woodward G, Lyttleton O, Evans DM, Reik W, Paul YL, Ficz G, Ozanne SE, Wipat A, Flanagan K, Lister A, Heijmans BT, Ring SM, Davey SG (2015) Data resource profile: accessible resource for integrated Epigenomic studies (ARIES). *Int J Epidemiol* 44:1181–1190
- Sun Z, Chai HS, Wu Y et al (2011) Batch effect correction for genome-wide methylation data with Illumina Infinium platform. *BMC Med Genet* 4:84
- Shoemaker R, Deng J, Wang W, Zhang K (2010) Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res* 20:883–889
- Stirzaker C, Taberlay PC, Statham AL, Clark SJ (2014) Mining cancer methylomes: prospects and challenges. *Trends Genet* 30:75–84
- Triche TJ Jr, Weisenberger DJ, Van Den Berg D et al (2013) Low-level processing of Illumina Infinium DNA methylation BeadArrays. *Nucleic Acids Res* 41:e90
- Teschendorff AE, Marabita F, Lechner M et al (2013) A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 29:189–196
- Wen KX, Milić J, El-Khodori B et al (2016) The role of DNA methylation and histone modifications in neurodegenerative diseases: a systematic review. *PLoS One* 11:e0167201
- Yousefi P, Huen K, Aguilar Schall R et al (2013) Considerations for normalization of DNA methylation data by Illumina 450K BeadChip assay in population studies. *Epigenetics* 8:1141–1152
- Zhang Z, Zhang R (2015) Epigenetics in autoimmune diseases: pathogenesis and prospects for therapy. *Autoimmun Rev* 14:854–863
- Zhou W, Laird PW, Shen H (2017) Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* 45:e22
- Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A, Meissner A (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500:477–481