



# A validation of Illumina EPIC array system with bisulfite-based amplicon sequencing

Alexandra J. Noble<sup>1</sup>, John F. Pearson<sup>2</sup>, Joseph M. Boden<sup>3</sup>, L. John Horwood<sup>3</sup>, Neil J. Gemmell<sup>4</sup>, Martin A. Kennedy<sup>2</sup> and Amy J. Osborne<sup>1</sup>

<sup>1</sup>School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

<sup>2</sup>Department of Pathology and Biomedical Sciences, University of Otago, Christchurch, New Zealand

<sup>3</sup>Department of Psychological Medicine, University of Otago, Christchurch, New Zealand

<sup>4</sup>Department of Anatomy, University of Otago, Dunedin, New Zealand

## ABSTRACT

The Illumina Infinium<sup>®</sup> MethylationEPIC BeadChip system (hereafter EPIC array) is considered to be the current gold standard detection method for assessing DNA methylation at the genome-wide level. EPIC arrays are often used for hypothesis generation or pilot studies, the natural conclusion to which is to validate methylation candidates and expand these in a larger cohort, in a targeted manner. As such, an accurate smaller-scale, targeted technique, that generates data at the individual CpG level that is equivalent to the EPIC array, is needed. Here, we tested an alternative DNA methylation detection technique, known as bisulfite-based amplicon sequencing (BSAS), to determine its ability to validate CpG sites detected in EPIC array studies. BSAS was able to detect differential DNA methylation at CpG sites to a degree which correlates highly with the EPIC array system at some loci. However, BSAS correlated less well with EPIC array data in some instances, and most notably, when the magnitude of change via EPIC array was greater than 5%. Therefore, our data suggests that BSAS can be used to validate EPIC array data, but each locus must be compared on an individual basis, before being taken forward into large scale screening. Further, BSAS does offer advantages compared to the probe-based EPIC array; BSAS amplifies a region of the genome (~500 bp) around a CpG of interest, allowing analyses of other CpGs in the region that may not be present on the EPIC array, aiding discovery of novel CpG sites and differentially methylated regions of interest. We conclude that BSAS offers a valid investigative tool for specific regions of the genome that are currently not contained on the array system.

Submitted 4 August 2020

Accepted 22 December 2020

Published 10 February 2021

Corresponding author

Alexandra J. Noble,  
alexandra.noble@pg.canterbury.ac.nz

Academic editor

Santosh Patnaik

Additional Information and  
Declarations can be found on  
page 12

DOI 10.7717/peerj.10762

© Copyright  
2021 Noble et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

**Subjects** Biochemistry, Biotechnology, Genetics, Genomics, Molecular Biology

**Keywords** DNA methylation, Illumina EPIC array, Bisulfite based amplicon sequencing

## INTRODUCTION

Epigenetic modifications, such as DNA methylation, play a vital role in regulating gene expression (*Hackett & Surani, 2013*) and have the potential to induce phenotypic changes (*Dolinoy, Huang & Jirtle, 2007; Sinclair et al., 2007; Kucharski et al., 2008; Gertz et al., 2012; Wang et al., 2012*). DNA methylation occurs when a methyl group is covalently transferred to the C5 position of the cytosine ring of a DNA molecule by a methyltransferase enzyme,

with the resulting modified cytosine then termed 5-methylcytosine (5mC) (Mitchell, Schneper & Notterman, 2016). In mammals, most DNA methylation occurs at CpG dinucleotides. CpG sites themselves can be defined as a singular modified cytosine residue which are found throughout the genome, but are particularly dense in promoter regions (Takai & Jones, 2002).

DNA methylation is heavily influenced by the surrounding environment; factors such as tobacco smoking (Breton et al., 2009; Zeilinger et al., 2013; Ambatipudi et al., 2016a; Ambatipudi et al., 2016b; Osborne et al., 2020), alcohol (Philibert et al., 2012; Liu et al., 2016), nutrition (Rampersaud et al., 2000; Delgado-Cruzata et al., 2015), stress (Murgatroyd et al., 2009) and aging (Horvath et al., 2012; Marioni et al., 2015) can all impact on DNA methylation at CpG sites. Alterations to DNA methylation are associated with changes in phenotype and also, in some instances, methylation changes contribute to disease pathology (Kim et al., 2010; Mastroeni et al., 2010; Rakyan et al., 2011; De Jager et al., 2014).

As a result of these relatively recent observations, the assessment of differential DNA methylation in humans, and in particular, epigenome-wide association studies (EWAS), is a burgeoning field. High-throughput array technologies are a popular choice for EWAS, due to their robustness and accuracy (Pidsley et al., 2016). The Illumina Infinium<sup>®</sup> MethylationEPIC array (hereafter 'EPIC array') quantifies methylation at 850,000 different CpG sites (Zhou, Laird & Shen, 2017), and although this is still a small proportion of the total number of CpG sites in the genome (~28 million Lövkvist et al., 2016) it represents a broad distribution of sites that give a specific and robust measurement of methylation at those sites.

Further, the goal of many whole-genome studies of DNA methylation is often a pilot or scoping study to capture a range of targets that may be associating with, e.g., a particular environmental exposure. As such, once the genome has been investigated in a number of samples, a whole-genome approach is not always necessary if the user simply requires follow up and/or validation of identified loci in a larger cohort. To undertake further analyses and to validate methylation array-based experiments, several different methods exist that that rely on bisulfite treatment of DNA: bisulfite-based amplicon sequencing (BSAS), bisulfite pyrosequencing and methylation-specific PCR (MS-PCR) are methods which can specifically target a predetermined area of interest in the genome at a low cost and higher sample throughput, compared to arrays. An informative study conducted by the BLUEPRINT consortium evaluated 27 predefined genomic regions, across 32 reference samples amongst 18 laboratories using six assays (Bock et al., 2016). Good agreement was observed across methods, with amplicon bisulfite sequencing, and bisulfite pyrosequencing showing the best concordance (Bock et al., 2016). A similar study also assessed bisulfite pyrosequencing, observing congruence to EPIC array analysis (Roessler et al., 2012). However, pyrosequencing is known to have quantitative flaws due to the output of sequences generated through fluorescence methods (França, Carrilho & Kist, 2002). MS-PCR is a method often used in clinical settings (Herman et al., 1996), however it has a high false positive rate (Claus et al., 2012). By contrast, BSAS detects cytosine methylation to base-pair scale resolution without reliance on light detection methods for sequencing (Masser, Stanford & Freeman, 2015). BSAS is a multiplex procedure that can quantitatively

assess each CpG site within numerous target regions at the same time (*Masser, Berg & Freeman, 2013*).

Thus, given the limitations of pyrosequencing and MS-PCR, here we examine whether BSAS can also accurately validate EPIC array data, and be used as a replication, and/or expansion tool for targeted DNA methylation analyses, similar to what has been shown using pyrosequencing. Further, we wish to assess the multiple other CpG sites residing within the targeted amplicon region, to investigate differential methylated regions, which would not be able to be explored via EPIC array.

To answer the question, we used EPIC array data generated from individuals from the Christchurch Health and Development Study (CHDS) which evaluated differential DNA methylation in response to regular cannabis use (*Osborne et al., 2020*). The CHDS is a longitudinal study of a birth cohort of 1265 children born in 1977 in Christchurch, New Zealand, who have been studied on 24 occasions from birth to the age of 40 ( $n = 904$  at age 40). Of this, a total 96 individuals were selected, and arrays were performed in two separate batches in consecutive years ( $n = 48$  per year).

For validation analysis we selected individuals with EPIC array data ( $n = 14$ ), as well as new individuals ( $n = 82$ ), to serve as a validation and expansion cohort for the differential DNA methylation identified via EPIC array (*Osborne et al., 2020*). Specifically, we asked whether BSAS, after determination of the most appropriate normalisation method, produced the same average methylation values as EPIC arrays, when comparing case data to control data.

While both EPIC array and BSAS are readily used as standalone experiments, they would provide robust evidence if carried out together. Thus, given the rising popularity of studies investigating DNA methylation, establishing a better understanding of how differential DNA methylation differs between regions within the genome, such as evaluating concordance between methods and then further assessing resultant CpG sites within a designated region, is valuable to the scientific community.

## MATERIALS & METHODS

### Cohort selection and DNA extraction - EPIC arrays

EPIC array data used in this study has previously been published (*Osborne et al., 2020*). Briefly, in this study we use DNA from human participants who are partitioned into three groups: (i) regular cannabis users, who had never used tobacco (“cannabis-only”); those who consumed both cannabis and tobacco (“cannabis plus tobacco”), and; (ii) controls, who consumed neither cannabis nor tobacco. Controls were matched as closely as possible for sex, ethnicity and parental socioeconomic status (data and methods described in *Osborne et al., 2020*).

### Bioinformatics analysis—processing and normalisation of raw EPIC array data

For this study, analysis was carried out using R statistical software (Version 3.5.2) (*Team, 2013*). Quality control was first performed on the raw data; sex chromosomes and 150 failed probes (detection  $P$  value greater than 0.01 in at least 50% of samples) were excluded

from analysis. Furthermore, potentially problematic CpGs with adjacent single nucleotide polymorphisms (SNPs), or that did not map to a unique location in the genome (Pidsley *et al.*, 2016) were also excluded. The raw data were then normalised using Noob (Fortin, Triche Jr & Hansen, 2017) in the minfi package (Aryee *et al.*, 2014). Normalisation was then checked by observing density plots as well as multidimensional scaling plots of the 5,000 most variable CpG sites.

### Cohort selection and DNA extraction—BSAS experiments

BSAS analysis was carried out on two groups: cannabis plus tobacco users ( $n = 44$ ) and controls ( $n = 38$ ), who had never used cannabis. In contrast to the EPIC array analysis, no cannabis-only participants were used in BSAS; this is a consequence of the small number of individuals who use cannabis but who do not also use tobacco. Cannabis users were all selected on the basis that they either met DSM-IV diagnostic criteria (American Psychiatric Association, 1994) for cannabis dependence or had reported using cannabis consumption on a daily basis for a minimum of three years prior to age 28. Participants were matched as closely as possible for the following variables, sex, ethnicity, and parental socioeconomic status (Table S1). All participants in this birth cohort were enrolled across a four month period so they are all of a similar age. Collection and analysis of DNA in the Christchurch Health and Development Study was approved by Southern Health and Disability Ethics Committee (CTB/04/11/234/AM10). DNA extraction protocols were previously described in (Noble *et al.*, 2020). Specifically, DNA was extracted from whole blood samples using a Kingfisher Flex System (Thermo Scientific, Waltham, MA USA) and quantified via nanodrop (Thermo Scientific, Waltham, MA USA). DNA was bisulfite treated using the EZ DNA Methylation-Gold kit (Zymo Research, USA) as per the manufacturer's instructions.

### CpG site selection, primer design and amplification—BSAS

A total of 15 CpG sites, representing 15 individual probes from the Illumina EPIC array, were chosen based on their differential methylation status in cannabis plus tobacco users compared to controls (Table 1). A range of probes at differing levels of significance (not significant, nominally significant, significant after  $P$ -value adjustment) were chosen to reflect the range of data provided by the EPIC arrays. Primers to amplify bisulfite-treated DNA were designed using the online tool BiSearch (Arányi *et al.*, 2006) to amplify a  $\sim 250$  base pair region which spanned the CpG site (Table S2). At the 5' end of each primer sequence, an Illumina overhang (33 base pair sequence) was included to ensure the ability to pool the amplicons and barcode them for high-throughput sequencing. All product lengths were all between 226 and 340 base pairs. To ensure primer specificity, Delta G's were designed to be no lower than  $-9$  kcal/mol for efficiently, using the tool OligoAnalyzer (IDT<sup>®</sup>). A total of 30 primer pairs were initially designed for this experiment, and 15 of these are discussed here, as these were the primer pairs which performed efficiently at first usage. PCRs were undertaken as per (Noble *et al.*, 2020).

Following the PCR, DNA was cleaned up with Agencourt<sup>®</sup> AMPure<sup>®</sup> XP beads (Beckman Coulter) and washed with 80% ethanol and allowed to air-dry. DNA was then eluted with 52.5  $\mu$ l of 10 mM Tris pH 8.5 before being placed back into the magnetic stand.

**Table 1** CpG site differences from EPIC array and the BSAS methods at the 15 loci of differing levels of significance (not significant, nominally significant, and significant after *P*-value adjustment).

	Cg/Gene	Position in genome	Illumina ID	Illumina EPIC array			BSAS			Difference between methods $\beta$ difference
				$\beta$ difference	<i>P</i> value	FDR Adjusted <i>P</i> value	$\beta$ difference	<i>P</i> value	FDR Adjusted <i>P</i> value	
1	AHRR	Chr5, GB	cg05575921	-0.233	5.33E-12	3.7E-06	-0.041	0.006	0.245	-0.192
2	cg11977356 <sup>a</sup>	Chr19	cg11977356	-0.040	0.474	0.999	-0.004	0.406	0.959	-0.036
3	ITPR1	Chr3, GB	cg08987995	-0.001	0.572	0.999	0.005	0.820	0.822	-0.006
4	MAGI	Chr7, GB	cg21121803	-0.008	0.572	0.999	-0.007	0.809	0.959	-0.0004
5	EHMT2	Chr6, GB	cg07829740	0.005	0.037	0.999	-0.015	0.071	0.579	0.020
6	PPM1L	Chr3, GB	cg26406186	-0.006	0.818	0.999	0.011	0.904	0.963	-0.017
7	cg00571101 <sup>a</sup>	Chr12	cg00571101	0.004	0.368	0.999	-0.004	0.813	0.952	0.008
8	cg09078959 <sup>a</sup>	Chr5	cg09078959	-0.001	0.893	0.999	-0.005	0.001	0.245	0.004
9	cg01614625 <sup>a</sup>	Chr7	cg01614625	-0.009	0.370	0.999	-0.006	0.569	0.952	-0.004
10	DP10	Chr2, GB	cg05868547	0.006	0.077	0.999	-0.003	0.713	0.952	0.009
11	cg11293828 <sup>a</sup>	Chr12	cg11293828	-0.014	0.665	0.999	0.032	0.735	0.952	-0.045
12	CHD7	Chr5, 5'UTR	cg19926587	-0.007	0.960	0.999	-0.006	0.429	0.959	-0.001
13	NIPAL4	Chr5, TSS1500	cg17695979	-0.007	0.714	0.999	-0.003	0.106	0.713	-0.004
14	PRDM5	Chr4, GB	cg01118724	-0.004	0.734	0.999	0.005	0.116	0.713	-0.009
15	SLC17A7	Chr19, GB	cg02624701	-0.043	0.312	0.999	0.018	0.646	0.952	-0.061

**Notes.**<sup>a</sup>When a cg number is listed, then there is no known gene associated with that CpG site. GB-Gene Body.

Once the supernatant had cleared, 50  $\mu$ l of supernatant was transferred into a fresh 96-well plate. DNA samples were quantified using the Quant-iT™ PicoGreen™ dsDNA Assay kit (Thermo Fisher) using the FLUROstar® Omega (BMG Labtech). Sequence libraries were prepared using the Illumina MiSeq™ 500 cycle Kit V2, and sequenced on an Illumina MiSeq™ system at Massey Genome Services (Palmerston North, New Zealand).

### Bioinformatic and statistical analysis—BSAS data

Illumina MiSeq™ sequences were trimmed using SolexaQA++ software and aligned to FASTA bisulfite converted reference sequences using the package Bowtie2 (version 2.3.4.3). Each individual read was then aligned to all reference sequences (GRCh37/hg19) using the methylation-specific package Bismark ([Krueger & Andrews, 2011](#)). Bismark produced aligned mapped reads with counts for methylated and unmethylated cytosines at each CpG site, thus BSAS returns additional CpG sites to the intended validation target, as each sequencing read contains multiple CpG sites. Cytosine proportion is calculated based upon the number of cytosines divided by the number of cytosines with the additions of the number of thymines present ( $C/(C_1 + T)$ ). This gave the average methylation  $\beta$  values for each individual at each given CpG site. These  $\beta$  values ranged between 0 - 1, with a  $\beta$  equal to 1 indicating 100% methylation at that CpG site across all sequencing reads. These data were imported into R Studio (RStudio version 3.3.0) and the edgeR package ([Chen et al., 2017](#)) was used to determine differential DNA methylation between cannabis users and controls; coverage level was set to greater or equal to “8” across unmethylated and methylated counts. This was also set at 50 and 100 reads and no differences were seen between the results at any of these thresholds, so “8” was used for the continuation of BSAS calling under the recommendations of [Chen et al. \(2017\)](#) whereby the conservative rule of thumb is total count (both methylated and unmethylated) is at least “8” in every sample. Within the data set 96.5% of the reads were above a methylation coverage of 50 ([Data S1](#)). A negative binomial generalised model was used to fit the counts (methylated and unmethylated reads) in regards to the two variable groups. Summary tables compiled of the CpG sites of interest with nominal  $P$  value significance and post multiple testing using false discovery rate (FDR) of less than 0.05 were considered to be statistically significant. A scatter plot including a linear regression line with adjusted  $R^2$  values was generated in R Studio to quantify the correlation between  $\beta$  values produced with EPIC array and BSAS. Adjusted  $R^2$  values were calculated for: (i) BSAS cases versus EPIC cases, and; (ii) BSAS controls versus EPIC controls. A Bland Altman analysis ([Bland & Altman, 1986](#)) was used to compare the agreement of the two techniques. Means were log transformed and lower and upper levels of agreement with 95% confidence intervals were calculated. Welch two sample t-tests were carried out on each of the loci (cases and controls separately) to assess differences between the two methods. All graphs were constructed using the R package ggplot2 ([Wickham, 2016](#)).

## RESULTS

### Validation and replication of EPIC array data using BSAS

The differences between average methylation ( $\beta$  values) of cannabis plus tobacco users (cases) and controls were calculated for each method (EPIC array and BSAS, [Table 1](#)).

When comparing case vs control data from EPIC and BSAS individually, no significant difference in average methylation between case and control was observed for either detection method, with the exception of cg05575921 in *AHRR* and the site cg09078959. *AHRR* was significant in both BSAS and EPIC ( $P = 0.006$ ,  $P = 5.33 \times 10^{-12}$ ), and cg05575921 was found to only be significant under BSAS ( $P = 0.001$ ,  $P = 0.665$ ).

Correlations between BSAS and EPIC were plotted individually for cases and controls. BSAS versus EPIC cases resulted in an adjusted  $R^2$  of 0.8878 and BSAS versus EPIC controls gave an adjusted  $R^2$  of 0.8683 ([Fig. S1](#)).

### Bland Altman correlation

A Bland Altman analysis was carried out on the loci investigated by BSAS and compared to data for the same loci produced using the Illumina EPIC array. [Figure 1A](#) shows cannabis users (cases) measured by BSAS and the EPIC array on the  $X$  axis, while the  $Y$  axis represents the log differences between the measurements. The observed differences between loci in cannabis cases (EPIC and BSAS) fall within the lines of agreement. [Figure 1B](#) shows the control group differences plotted for the same loci for BSAS and the EPIC array methods. Similar to above, all data points fall within the lower and upper lines of agreement.

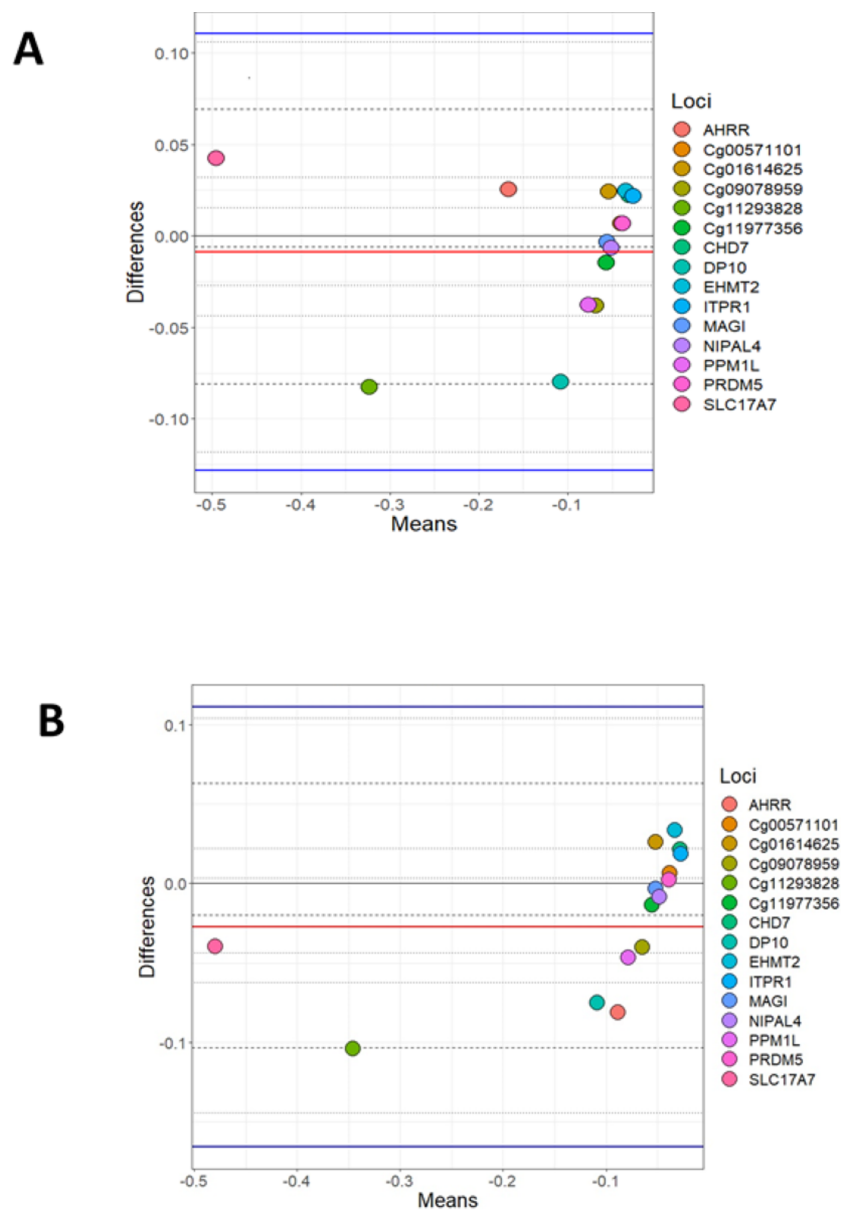
Mean methylation values for each individual were plotted for each of the 15 loci, and these were then compared between BSAS and EPIC, for cases ([Fig. 2](#)) and controls ([Fig. 3](#)). Loci displaying a significant shift in average methylation between the methods of detection are indicated with an \* when using a Welsh two sample comparison. The following loci were found to display differences between BSAS and EPIC array for cases were: *AHRR*, cg09078959, cg11293828, *CHD7*, *DP10*, *EHMT2* and *ITPR1*, and for controls: *AHRR*, cg09078959, cg11293828, *CHD7*, *DP10*, *EHMT2*, *ITPR1*, *NIPAL4* and *PPM1L*.

### Assessing amplicon regions

Multiple CpG sites residing within an amplicon can be sequenced using BSAS, providing information about a larger region of interest, rather than just a single CpG site. [Fig. S2](#) displays the multiple CpG sites found across each of the 15 amplicons in this study. A total of 9 of the 15 amplicons contained more than one CpG site. All CpG sites within the amplicons remained non-differentially methylated between cases and controls, except one site in *AHRR*. The amplicon from *SLC17A7* sequenced here contained a total of 15 CpG sites within the 250 base pairs.

## DISCUSSION

High throughput array technologies have facilitated the next step in assessing associations between DNA methylation changes in response to a known environmental exposure at a genome-wide level. The EPIC array (as well as the predecessor 27k and 450k arrays) is one such platform that allows for the characterisation of these DNA methylation changes.



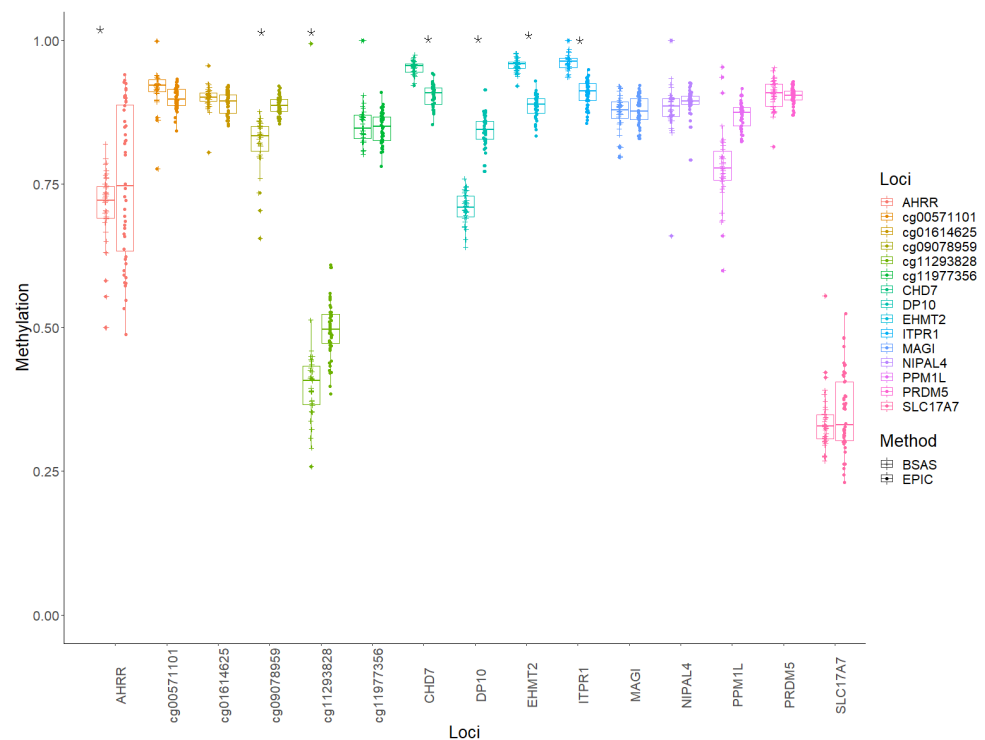
**Figure 1** Bland Altman of BSAS vs EPIC for cases and controls. Bland Altman plots showing the log mean differences between DNA methylation as measured by EPIC array vs. the same CpG sites measured using BSAS. (A) Data from cannabis users, gathered using BSAS and the EPIC array (Cases); (B) the control subjects used in BSAS and the EPIC array. Each of the 15 points represents the CpG sites investigated. Dotted lines represent the limits of agreement, red the mean, blue in the 95% confidence intervals.

Full-size [DOI: 10.7717/peerj.10762/fig-1](https://doi.org/10.7717/peerj.10762/fig-1)

Through these approaches, various studies have furthered our understanding of how DNA methylation can play a role in response to different environmental exposures.

We selected the orthogonal method BSAS to determine its applicability as a validation, replication and/or expansion tool for EPIC array. BSAS is often used as a standalone method for assessing differential DNA methylation at specific CpG sites, usually because it is more

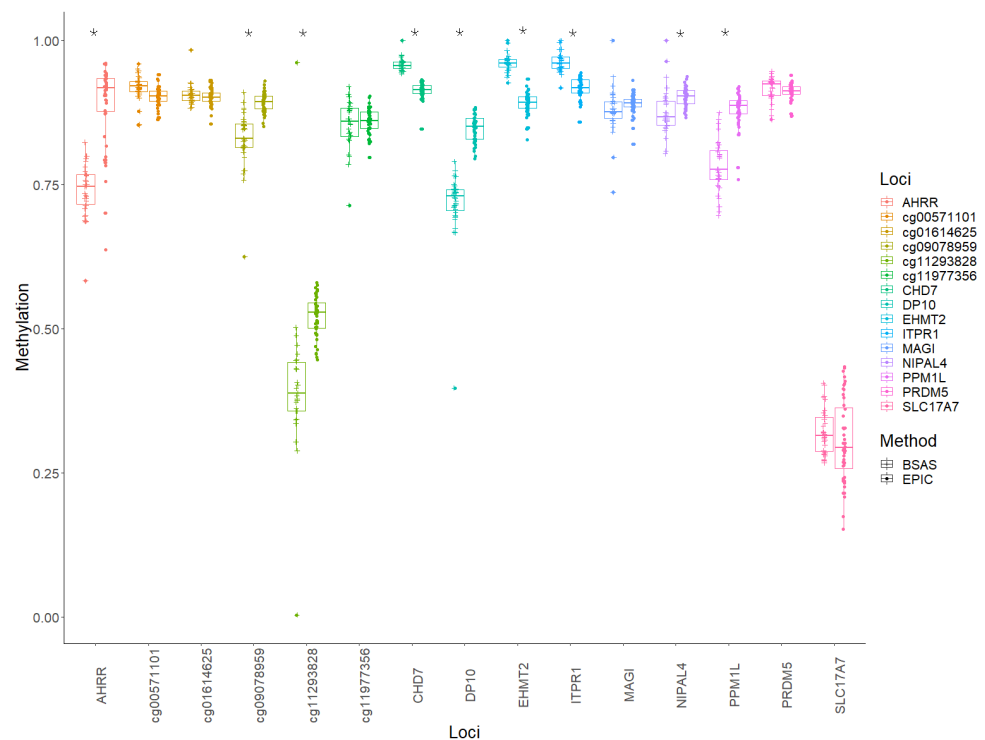




**Figure 2** Individual methylation across the 15 loci for cases. Average methylation for case individuals only across the 15 loci assessed using EPIC and BSAS. As asterisk (\*) represents those loci with significant differences in average methylation between EPIC and BSAS.

Full-size DOI: [10.7717/peerj.10762/fig-2](https://doi.org/10.7717/peerj.10762/fig-2)

cost-effective than EPIC arrays, and allows analysis of many samples at once, in multiplex. It returns data for all CpGs within a targeted region of interest (~250 base pairs) with results providing base pair-level specificity (Masser, Stanford & Freeman, 2015). Overall, when considering average methylation between cases and controls as determined via BSAS or EPIC individually, we did not detect significant differences in average methylation for each detection method; the biological results are discussed elsewhere (Osborne *et al.*, 2020), however, it was expected that the smaller sample set used here would not have the statistical power to detect effects found in the larger cohort. The intent of this study was to compare average methylation as determined via BSAS, to that determined by EPIC array. We show here that the estimation of differential DNA methylation observed using BSAS correlated with differential methylation determined via EPIC array. However, although the data correlates between the methods (adjusted  $R^2$  cases, 0.8878 and adjusted  $R^2$  controls, 0.8683), we urge caution when interpreting this correlation as proof that BSAS will be a suitable independent validation of EPIC array data in every experiment. This is because while the data presented here correlated between BSAS and EPIC array as a whole dataset, some sites showed larger differences between average methylation estimated using BSAS vs. EPIC array. Most notably, where the differential methylation on EPIC array was greater than 5% between cases and controls, BSAS was unable to detect this differential DNA methylation to the same magnitude as EPIC array. Further, a total of 9/15 loci had observed



**Figure 3** Average methylation at the 15 loci for controls. Average methylation for control individuals across the 15 loci assessed using EPIC and BSAS. As asterisk (\*) represents those loci with significant differences in average methylation between EPIC and BSAS.

Full-size [DOI: 10.7717/peerj.10762/fig-3](https://doi.org/10.7717/peerj.10762/fig-3)

*P* value significance when carrying out a Welch two sample *t*-test on control data, and 7/15 on case data, implying there were differences between the methylation values for the methods. For instance, *AHRR* exhibited a 4% difference in methylation between cases and controls when assessed using BSAS (the highest value detected in using BSAS in this study), compared to 23% using EPIC array. Thus, while a strong correlation between EPIC array data and BSAS data was found across the 15 CpG sites investigated, which itself implies an association between the average methylation at each CpG for the two techniques, each locus must be validated on a case by case basis before being taken forward into high-throughput or large scale screening, to ensure it produces results that are equivalent to EPIC. In addition, further work on CpG sites with higher magnitude changes is needed to determine whether BSAS is limited by the magnitude of differential methylation it is able to detect. However, it is worth noting that most studies of differential methylation report modest (<5%) significant differential methylation observations, suggesting that BSAS may prove useful, given inclusion of rigorous controls of known differential methylation to ensure accuracy of results.

Due to the sequence-based nature of BSAS data (compared to the probe-based nature of EPIC arrays) BSAS, as a standalone method, offers some advantages that are not applicable to EPIC arrays. For instance, BSAS has the potential to determine novel differentially methylated CpGs which may be near (in the same targeted region) but not the initial

pre-determined CpG site of interest. This is possible because all CpGs within an, e.g., 500bp region are returned using BSAS data, only one of which may be on an EPIC array. Further, via this targeted sequencing process, BSAS may reveal novel differentially methylated regions (DMRs). DMRs are described as areas which exhibit multiple successive methylated CpG sites which may have biological impact within the genome. For example, they have been implicated in the development and progression of disease ([Hotta et al., 2018](#)). Therefore targeting more than a single CpG site may provide further insight into genes and regions of interest. Consequently, while here we have used BSAS technology to replicate/validate differential methylation identified via EPIC array, given that BSAS outputs can correlate with EPIC data, equally, BSAS could be used as a “discovery-based tool”; if significantly differentially methylated CpGs are identified via BSAS, this would serve to justify further investigation using a robust and more expensive high throughput method.

The EPIC array still remains the most reproducible way to measure DNA methylation ([Bibikova et al., 2009](#)). This is because the probe-based nature of the method frequently produces comparable results across research groups and arrays. For example, detection of differential methylation using the EPIC array found a difference of 23% in cannabis plus tobacco users, compared to controls, at *AHRR* (cg05575921, [Table 1](#)), a result that is supported by other studies in tobacco smokers using EPIC array ([Zeilinger et al., 2013](#); [Guida et al., 2015](#); [Ambatipudi et al., 2016a](#); [Ambatipudi et al., 2016b](#); [Su et al., 2016](#); [Li et al., 2018](#)). *AHRR* has an important role in controlling a range of different physiological functions; it contributes to regulation of cell growth, regulation of apoptosis and contributes to vascular and immune responses ([Trombino et al., 2000](#); [Allan & Sherr, 2005](#); [Lahvis et al., 2005](#); [Marlowe et al., 2008](#)).

BSAS and EPIC array rely upon different chemistries and methods to detect DNA methylation. This may account for the majority of the variation found between the two methods. BSAS relies upon PCR amplification of DNA that is treated with sodium bisulfite. When DNA is treated, unmethylated cytosine residues are converted into uracil via hydrolytic deamination. Amplification of uracil nucleotides during this process are replaced by thymine during replication and the 5-methylcytosines are left unreactive throughout the deamination process and then are amplified as cytosines. It then becomes possible to ‘read’ values of methylation for each cytosine in an amplicon via DNA sequencing ([Booth et al., 2013](#)). The ability to treat DNA with sodium bisulfite has led to the expansion of research undertaken within this field ([Frommer et al., 1992](#)). However, it is important that we ensure the validity of the results are not limited by the manner in which the data was produced. Ensuring that we limit these discrepancies between technologies will allow for better validation of data. There is potential for errors to occur at this step, because incomplete bisulfite conversion cannot be distinguished from 5-methylcytosine, this can possibly introduce false positive methylation calls at this point ([Richards et al., 2018](#); [Krueger et al., 2012](#)). Although both techniques rely upon bisulfite treatment, it is this source of error followed by the PCR amplification that might explain the differences in results we have observed. Refining these sources of error may provide much more comparable results between the two methods.

## CONCLUSIONS

In conclusion, we chose to validate EPIC array data by using the alternative method, BSAS, to detect differential methylation at CpG sites. While BSAS validated EPIC array data at some loci, and correlated across all loci as a whole, some individual loci did not validate. This implies that each locus must be investigated individually before determining its utility in a large-scale analysis. Further, it is possible that BSAS may be unable to reproduce the magnitude of changes that are shown in the EPIC array system, which may be a consequence of lack of specificity and addition error rate through PCR amplification. It does however, have the ability to assess differentially methylated regions, rather than individual CpG sites. As some regions of the genome are more susceptible to methylation change than others, BSAS could detect swathes of correlated differential methylation at neighbouring CpG sites in certain areas of the genome. From the results shown here, BSAS has the potential to be able to detect methylation marks which maybe hallmarks for disease later on in life. Finally, although BSAS does not generate the same significance level as the EPIC array in some instances, we demonstrate that BSAS can be used as an investigative tool for specific regions of the genome.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This study was funded by University of Otago Research Grant to Martin Kennedy, The Carney Centre for Pharmacogenomics. Christchurch Health and Development Study (CHDS) was funded by the Health Research Council of New Zealand (Programme Grant 16/600) and the Canterbury Medical Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

University of Otago Research.

The Carney Centre for Pharmacogenomics.

The Health Research Council of New Zealand: Programme Grant 16/600.

The Canterbury Medical Research Foundation.

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Alexandra Noble and Amy J. Osborne conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- J.F. Pearson conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

- Joseph Boden, John Horwood, Neil J. Gemmell and Martin Kennedy conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

### Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

Collection and analysis of DNA in the Christchurch Health and Development Study was approved by Southern Health and Disability Ethics Committee (CTB/04/11/234/AM10).

### Data Availability

The following information was supplied regarding data availability:

R script is available at GitHub:

<https://github.com/alexnoable1/Validation-of-EPIC-using-BSAS/blob/master/R%20script>.

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.10762#supplemental-information>.

## REFERENCES

- Allan LL, Sherr DH. 2005. Constitutive activation and environmental chemical induction of the aryl hydrocarbon receptor/transcription factor in activated human B lymphocytes. *Molecular Pharmacology* 67(5):1740–1750 DOI 10.1124/mol.104.009100.
- Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Calvez-Kelm FL, Kaaks R, Barrdahl M, Boeing H, Aleksandrova K, Trichopoulou A, Lagiou P, Naska A, Palli D, Krogh V, Polidoro S, Tumino R, Panico S, Bueno-de Mesquita B, Peeters PH, Quirós JR, Navarro C, Ardanaz E, Dorronsoro M, Key T, Vineis P, Murphy N, Riboli E, Romieu I, Herceg Z. 2016a. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. *Epigenomics* 8(5):599–618 DOI 10.2217/epi-2016-0001.
- Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, Barrdahl M, Boeing H, Aleksandrova K, Trichopoulou A. 2016b. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. *Epigenomics* 8(5):599–618 DOI 10.2217/epi-2016-0001.
- American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders. 4th Edition*. Washington, D.C.: American Psychiatric Association.
- Arányi T, Váradi A, Simon I, Tusnády GE. 2006. The BiSearch web server. *BMC Bioinformatics* 7(1):431 DOI 10.1186/1471-2105-7-431.
- Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. 2014. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30(10):1363–1369 DOI 10.1093/bioinformatics/btu049.

- Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL. 2009.** Genome-wide DNA methylation profiling using Infinium<sup>®</sup> assay. *Epigenomics* 1(1):177–200 DOI 10.2217/epi.09.14.
- Bland J, Altman D. 1986.** Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 327(8476):307–310 DOI 10.1016/S0140-6736(86)90837-8.
- Bock C, Halbritter F, Carmona FJ, Tierling S, Datlinger P, Assenov Y, Berdasco M, Bergmann AK, Booher K, Busato F, Campan M, Dahl C, Dahmcke CM, Diep D, Fernández AF, Gerhauser C, Haake A, Heilmann K, Holcomb T, Hussmann D, Ito M, Kläver R, Kreutz M, Kulis M, Lopez V, Nair SS, Paul DS, Plongthongkum N, Qu W, Queirós AC, Reinicke F, Sauter G, Schlomm T, Statham A, Stirzaker C, Strogantsev R, Urtinguio RG, Walter K, Weichenhan D, Weisenberger DJ, Beck S, Clark SJ, Esteller M, Ferguson-Smith AC, Fraga MF, Guldborg P, Hansen LL, Laird PW, Martín-Subero JI, Nygren AOH, Peist R, Plass C, Shames DS, Siebert R, Sun X, Tost J, Walter J, Zhang K, The BC. 2016.** Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. *Nature Biotechnology* 34(7):726–737 DOI 10.1038/nbt.3605.
- Booth MJ, Ost TWB, Beraldi D, Bell NM, Branco MR, Reik W, Balasubramanian S. 2013.** Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nature Protocols* 8:1841–1851 DOI 10.1038/nprot.2013.115.
- Breton CV, Byun H-M, Wenten M, Pan F, Yang A, Gilliland FD. 2009.** Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. *American Journal of Respiratory and Critical Care Medicine* 180(5):462–467 DOI 10.1164/rccm.200901-0135OC.
- Chen Y, Pal B, Visvader JE, Smyth GK. 2017.** Differential methylation analysis of reduced representation bisulfite sequencing experiments using edgeR. *F1000Research* 6:2055–2055 DOI 10.12688/f1000research.13196.1.
- Claus R, Wilop S, Hielscher T, Sonnet M, Dahl E, Galm O, Jost E, Plass C. 2012.** A systematic comparison of quantitative high-resolution DNA methylation analysis and methylation-specific PCR. *Epigenetics* 7(7):772–780 DOI 10.4161/epi.20299.
- De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, Eaton ML, Keenan BT, Ernst J, McCabe C, Tang A, Raj T, Replogle J, Brodeur W, Gabriel S, Chai HS, Younkin C, Younkin SG, Zou F, Szyf M, Epstein CB, Schneider JA, Bernstein BE, Meissner A, Ertekin-Taner N, Chibnik LB, Kellis M, Mill J, Bennett DA. 2014.** Alzheimer’s disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nature Neuroscience* 17(9):1156–1163 DOI 10.1038/nn.3786.
- Delgado-Cruzata L, Zhang W, McDonald JA, Tsai WY, Valdovinos C, Falci L, Wang Q, Crew KD, Santella RM, Hershman DL, Greenlee H. 2015.** Dietary modifications, weight loss, and changes in metabolic markers affect global DNA methylation in Hispanic, African American, and Afro-Caribbean breast cancer survivors. *Journal of Nutrition* 145(4):783–790 DOI 10.3945/jn.114.202853.

- Dolinoy DC, Huang D, Jirtle RL. 2007.** Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proceedings of the National Academy of Sciences of the United States of America* **104**(32):13056–13061 DOI [10.1073/pnas.0703739104](https://doi.org/10.1073/pnas.0703739104).
- Fortin J-P, Triche Jr TJ, Hansen KD. 2017.** Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**(4):558–560 DOI [10.1093/bioinformatics/btw691](https://doi.org/10.1093/bioinformatics/btw691).
- França LT, Carrilho E, Kist TB. 2002.** A review of DNA sequencing techniques. *Quarterly Reviews of Biophysics* **35**(2):169–200 DOI [10.1017/S0033583502003797](https://doi.org/10.1017/S0033583502003797).
- Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. 1992.** A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America* **89**(5):1827–1831 DOI [10.1073/pnas.89.5.1827](https://doi.org/10.1073/pnas.89.5.1827).
- Gertz J, Reddy TE, Varley KE, Garabedian MJ, Myers RM. 2012.** Genistein and bisphenol A exposure cause estrogen receptor 1 to bind thousands of sites in a cell type-specific manner. *Genome Research* **22**(11):2153–2162 DOI [10.1101/gr.135681.111](https://doi.org/10.1101/gr.135681.111).
- Guida F, Sandanger TM, Castagné R, Campanella G, Polidoro S, Palli D, Krogh V, Tumino R, Sacerdote C, Panico S, Severi G, Kyrtopoulos SA, Georgiadis P, Vermeulen RCH, Lund E, Vineis P, Chadeau-Hyam M. 2015.** Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Human Molecular Genetics* **24**(8):2349–2359 DOI [10.1093/hmg/ddu751](https://doi.org/10.1093/hmg/ddu751).
- Hackett JA, Surani MA. 2013.** DNA methylation dynamics during the mammalian life cycle. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**(1609):20110328 DOI [10.1098/rstb.2011.0328](https://doi.org/10.1098/rstb.2011.0328).
- Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. 1996.** Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences of the United States of America* **93**(18):9821–9826 DOI [10.1073/pnas.93.18.9821](https://doi.org/10.1073/pnas.93.18.9821).
- Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, Van Eijk K, Van den Berg LH, Ophoff RA. 2012.** Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biology* **13**(10):R97 DOI [10.1186/gb-2012-13-10-r97](https://doi.org/10.1186/gb-2012-13-10-r97).
- Hotta K, Kitamoto A, Kitamoto T, Ogawa Y, Honda Y, Kessoku T, Yoneda M, Imajo K, Tomeno W, Saito S, Nakajima A. 2018.** Identification of differentially methylated region (DMR) networks associated with progression of nonalcoholic fatty liver disease. *Scientific Reports* **8**(1):13567 DOI [10.1038/s41598-018-31886-5](https://doi.org/10.1038/s41598-018-31886-5).
- Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW. 2010.** DNA methylation as a biomarker for cardiovascular disease risk. *PLOS ONE* **5**(3):e9692 DOI [10.1371/journal.pone.0009692](https://doi.org/10.1371/journal.pone.0009692).
- Krueger F, Andrews SR. 2011.** Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**(11):1571–1572 DOI [10.1093/bioinformatics/btr167](https://doi.org/10.1093/bioinformatics/btr167).

- Krueger F, Kreck B, Franke A, Andrews SR. 2012. DNA methylome analysis using short bisulfite sequencing data. *Nature Methods* 9:145–151 DOI 10.1038/nmeth.1828.
- Kucharski R, Maleszka J, Foret S, Maleszka R. 2008. Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319(5871):1827–1830 DOI 10.1126/science.1153069.
- Lahvis GP, Pyzalski RW, Glover E, Pitot HC, McElwee MK, Bradfield CA. 2005. The aryl hydrocarbon receptor is required for developmental closure of the ductus venosus in the neonatal mouse. *Molecular Pharmacology* 67(3):714–720 DOI 10.1124/mol.104.008888.
- Li S, Wong EM, Bui M, Nguyen TL, Joo J-HE, Stone J, Dite GS, Giles GG, Saffery R, Southey MC, Hopper JL. 2018. Causal effect of smoking on DNA methylation in peripheral blood: a twin and family study. *Clinical Epigenetics* 10:18–18 DOI 10.1186/s13148-018-0452-9.
- Liu C, Marioni RE, Hedman ÅK, Pfeiffer L, Tsai PC, Reynolds LM, Just AC, Duan Q, Boer CG, Tanaka T, Elks CE, Aslibekyan S, Brody JA, Kühnel B, Herder C, Almlı LM, Zhi D, Wang Y, Huan T, Yao C, Mendelson MM, Joehanes R, Liang L, Love SA, Guan W, Shah S, McRae AF, Kretschmer A, Prokisch H, Strauch K, Peters A, Visscher PM, Wray NR, Guo X, Wiggins KL, Smith AK, Binder EB, Ressler KJ, Irvin MR, Absher DM, Hernandez D, Ferrucci L, Bandinelli S, Lohman K, Ding J, Trevisi L, Gustafsson S, Sandling JH, Stolk L, Uitterlinden AG, Yet I, Castillo-Fernandez JE, Spector TD, Schwartz JD, Vokonas P, Lind L, Li Y, Fornage M, Arnett DK, Wareham NJ, Sotoodehnia N, Ong KK, van Meurs JBJ, Conneely KN, Baccarelli AA, Deary IJ, Bell JT, North KE, Liu Y, Waldenberger M, London SJ, Ingelsson E, Levy D. 2016. A DNA methylation biomarker of alcohol consumption. *Molecular Psychiatry* 23:422–433.
- Lövkvist C, Dodd IB, Sneppen K, Haerter JO. 2016. DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic Acids Research* 44(11):5123–5132 DOI 10.1093/nar/gkw124.
- Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, Murphy L, Martin NG, Montgomery GW, Feinberg AP, Fallin MD, Multhaup ML, Jaffe AE, Joehanes R, Schwartz J, Just AC, Lunetta KL, Murabito JM, Starr JM, Horvath S, Baccarelli AA, Levy D, Visscher PM, Wray NR, Deary IJ. 2015. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biology* 16(1):25 DOI 10.1186/s13059-015-0584-6.
- Marlowe JL, Fan Y, Chang X, Peng L, Knudsen ES, Xia Y, Puga A. 2008. The aryl hydrocarbon receptor binds to E2F1 and inhibits E2F1-induced apoptosis. *Molecular Biology of the Cell* 19(8):3263–3271 DOI 10.1091/mbc.e08-04-0359.
- Masser DR, Berg AS, Freeman WM. 2013. Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenetics & Chromatin* 6(1):33 DOI 10.1186/1756-8935-6-33.
- Masser DR, Stanford DR, Freeman WM. 2015. Targeted DNA methylation analysis by next-generation sequencing. *Journal of Visualized Experiments* 96:e52488.



- Mastroeni D, Grover A, Delvaux E, Whiteside C, Coleman PD, Rogers J. 2010.** Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. *Neurobiology of Aging* 31(12):2025–2037 DOI [10.1016/j.neurobiolaging.2008.12.005](https://doi.org/10.1016/j.neurobiolaging.2008.12.005).
- Mitchell C, Schnepfer LM, Notterman DA. 2016.** DNA methylation, early life environment, and health outcomes. *Pediatric Research* 79(1–2):212–219 DOI [10.1038/pr.2015.193](https://doi.org/10.1038/pr.2015.193).
- Murgatroyd C, Patchev AV, Wu Y, Micale V, Bockmühl Y, Fischer D, Holsboer F, Wotjak CT, Almeida OFX, Spengler D. 2009.** Dynamic DNA methylation programs persistent adverse effects of early-life stress. *Nature Neuroscience* 12:1559–1566 DOI [10.1038/nn.2436](https://doi.org/10.1038/nn.2436).
- Noble AJ, Pearson JF, Noble AD, Boden JM, Horwood LJ, Kennedy MA, Osborne AJ. 2020.** Epigenetic signatures associated with the observed link between maternal tobacco use during pregnancy, and offspring conduct problems in childhood and adolescence. *bioRxiv*. 2020.2007.2002.183285.
- Osborne AJ, Pearson JF, Noble AJ, Gemmell NJ, Horwood LJ, Boden JM, Benton MC, Macartney-Coxson DP, Kennedy MA. 2020.** Genome-wide DNA methylation analysis of heavy cannabis exposure in a New Zealand longitudinal cohort. *Translational Psychiatry* 10(1):114 DOI [10.1038/s41398-020-0800-3](https://doi.org/10.1038/s41398-020-0800-3).
- Philibert RA, Plume JM, Gibbons FX, Brody GH, Beach SRH. 2012.** The impact of recent alcohol use on genome wide DNA methylation signatures. *Frontiers in Genetics* 3:54–54.
- Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Dijk S, Muhlhausler B, Stirzaker C, Clark SJ. 2016.** Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology* 17(1):208–208 DOI [10.1186/s13059-016-1066-1](https://doi.org/10.1186/s13059-016-1066-1).
- Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S, Aden D, Daunay A, Busato F, Mein CA, Manfras B, Dias K-RM, Bell CG, Tost J, Boehm BO, Beck S, Leslie RD. 2011.** Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLOS Genetics* 7(9):e1002300 DOI [10.1371/journal.pgen.1002300](https://doi.org/10.1371/journal.pgen.1002300).
- Rampersaud GC, Kauwell GP, Hutson AD, Cerda JJ, Bailey LB. 2000.** Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. *The American Journal of Clinical Nutrition* 72(4):998–1003 DOI [10.1093/ajcn/72.4.998](https://doi.org/10.1093/ajcn/72.4.998).
- Richards R, Patel J, Stevenson K, Harbison S. 2018.** Evaluation of massively parallel sequencing for forensic DNA methylation profiling. *Electrophoresis* 39(21):2798–2805 DOI [10.1002/elps.201800086](https://doi.org/10.1002/elps.201800086).
- Roessler J, Ammerpohl O, Gutwein J, Hasemeier B, Anwar SL, Kreipe H, Lehmann U. 2012.** Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. *BMC Research Notes* 5(1):210 DOI [10.1186/1756-0500-5-210](https://doi.org/10.1186/1756-0500-5-210).
- Sinclair KD, Allegrucci C, Singh R, Gardner DS, Sebastian S, Bispham J, Thurston A, Huntley JF, Rees WD, Maloney CA. 2007.** DNA methylation, insulin resistance, and

- blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status. *Proceedings of the National Academy of Sciences of the United States of America* **104**(49):19351–19356 DOI [10.1073/pnas.0707258104](https://doi.org/10.1073/pnas.0707258104).
- Su D, Wang X, Campbell MR, Porter DK, Pittman GS, Bennett BD, Wan M, Englert NA, Crowl CL, Gimple RN, Adamski KN, Huang Z, Murphy SK, Bell DA. 2016.** Distinct epigenetic effects of tobacco smoking in whole blood and among leukocyte subtypes. *PLOS ONE* **11**(12):e0166486 DOI [10.1371/journal.pone.0166486](https://doi.org/10.1371/journal.pone.0166486).
- Takai D, Jones PA. 2002.** Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the National Academy of Sciences of the United States of America* **99**(6):3740–3745 DOI [10.1073/pnas.052410099](https://doi.org/10.1073/pnas.052410099).
- Team RC. 2013.** R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- Trombino AF, Near RI, Matulka RA, Yang S, Hafer LJ, Toselli PA, Kim DW, Rogers AE, Sonenshein GE, Sherr DH. 2000.** Expression of the aryl hydrocarbon receptor/transcription factor (AhR) and AhR-regulated CYP1 gene transcripts in a rat model of mammary tumorigenesis. *Breast Cancer Research and Treatment* **63**(2):117–131 DOI [10.1023/A:1006443104670](https://doi.org/10.1023/A:1006443104670).
- Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R. 2012.** Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Research* **22**(9):1680–1688 DOI [10.1101/gr.136101.111](https://doi.org/10.1101/gr.136101.111).
- Wickham H. 2016.** *ggplot2: elegant graphics for data analysis*. Springer-Verlag New York, USA.
- Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, Strauch K, Waldenberger M, Illig T. 2013.** Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLOS ONE* **8**(5):e63812 DOI [10.1371/journal.pone.0063812](https://doi.org/10.1371/journal.pone.0063812).
- Zhou W, Laird PW, Shen H. 2017.** Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Research* **45**(4):e22–e22.