

## RESEARCH ARTICLE

# Accurate Measurement of 5-Methylcytosine and 5-Hydroxymethylcytosine in Human Cerebellum DNA by Oxidative Bisulfite on an Array (OxBS-Array)

Sarah F. Field<sup>1</sup>, Dario Beraldi<sup>1</sup>, Martin Bachman<sup>1</sup>, Sabrina K. Stewart<sup>2</sup>, Stephan Beck<sup>2</sup>, Shankar Balasubramanian<sup>1,3\*</sup>

**1** Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, United Kingdom, **2** Department of Cancer Biology, UCL Cancer Institute, University College London, London WC1E 6BT, United Kingdom, **3** Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, United Kingdom

\* [sb10031@cam.ac.uk](mailto:sb10031@cam.ac.uk)

 OPEN ACCESS

**Citation:** Field SF, Beraldi D, Bachman M, Stewart SK, Beck S, Balasubramanian S (2015) Accurate Measurement of 5-Methylcytosine and 5-Hydroxymethylcytosine in Human Cerebellum DNA by Oxidative Bisulfite on an Array (OxBS-Array). *PLoS ONE* 10(2): e0118202. doi:10.1371/journal.pone.0118202

**Academic Editor:** Esteban Ballestar, Bellvitge Biomedical Research Institute (IDIBELL), SPAIN

**Received:** September 1, 2014

**Accepted:** January 8, 2015

**Published:** February 23, 2015

**Copyright:** © 2015 Field et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are available at GEO accession: GSE63179 without reservation.

**Funding:** The Balasubramanian group received core institutional funding from CRUK. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** S. Balasubramanian is a shareholder of Cambridge Epigenetix, Ltd. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

## Abstract

The Infinium 450K Methylation array is an established tool for measuring methylation. However, the bisulfite (BS) reaction commonly used with the 450K array cannot distinguish between 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC). The oxidative-bisulfite assay disambiguates 5mC and 5hmC. We describe the use of oxBS in conjunction with the 450K array (oxBS-array) to analyse 5hmC/5mC in cerebellum DNA. The “methylation” level derived by the BS reaction is the combined level of 5mC and 5hmC at a given base, while the oxBS reaction gives the level of 5mC alone. The level of 5hmC is derived by subtracting the oxBS level from the BS level. Here we present an analysis method that distinguishes genuine positive levels of 5hmC at levels as low as 3%. We performed four replicates of the same sample of cerebellum and found a high level of reproducibility (average  $r$  for BS = 98.3, and average  $r$  for oxBS = 96.8). In total, 114,734 probes showed a significant positive measurement for 5hmC. The range at which we were able to distinguish 5hmC occupancy was between 3% and 42%. In order to investigate the effects of multiple replicates on 5hmC detection we also simulated fewer replicates and found that decreasing the number of replicates to two reduced the number of positive probes identified by > 50%. We validated our results using qPCR in conjunction with glucosylation of 5hmC sites followed by MspI digestion and we found good concordance with the array estimates ( $r = 0.94$ ). This experiment provides a map of 5hmC in the cerebellum and a robust dataset for use as a standard in future 5hmC analyses. We also provide a novel method for validating the presence of 5hmC at low levels, and highlight some of the pitfalls associated with measuring 5hmC and 5mC.

## Introduction

DNA methylation has a crucial role in gene expression and thus in differentiation and development[1]. In light of this many tools have been developed to investigate methylation patterns across the genome[2]. However, recently it has been discovered that there are several different types of base modifications in mammalian DNA, which have previously been indistinguishable. The first type of modification identified was 5-methylcytosine (5mC)[3], but recently other modifications have been found which include: 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[4–6]. Proposed models of demethylation suggest that there may be two pathways; either it may occur passively, during DNA replication, or actively through enzymatic action[7–9]. The TET enzymes catalyse the sequential oxidation of 5mC to 5hmC, then 5fC and finally 5caC [6,7]. It has been proposed that this TET mediated oxidation can ultimately lead to demethylation via base excision repair of one or more oxidised intermediates via a process involving thymidine glycosylase (TDG) or other repair enzymes[10].

It also remains to be addressed whether 5hmC, 5fC and 5caC are functional modifications in their own right, as opposed to simply being intermediates of active demethylation. Patterns of 5hmC distribution have been found to vary from 5mC distribution both during development and in adult cells[10–12]. Changes in 5hmC have been associated with adult onset diseases including hypertension and Alzheimer's, and with neuronal maintenance[13–15]. Functional studies of the binding domain of MeCP2 (a transcription factor implicated in Rett syndrome), have shown that mutations associated with the disease preferentially impact binding to 5hmC, as opposed to 5mC[16].

To investigate the role of 5mC and 5hmC it is necessary to be able to accurately detect and quantitate the levels of these modifications at single base resolution. The most common method for interrogating DNA methylation, at the base pair level, is to react the DNA with bisulfite (BS). BS treatment deaminates C bases to uracil, which are read as thymines (T) in downstream assays[17,18]. Both 5mC and 5hmC are resistant to deamination by BS and are read as C in downstream assays [17,18]. Thus, the untransformed C bases quantified at single base resolution by BS treatment actually represent the sum of 5mC and 5hmC levels at that base. Our laboratory developed an oxidative bisulfite (oxBS) treatment that allows the disambiguation of 5mC from 5hmC[19,20]. It was observed that 5fC is deaminated to uracil by bisulfite treatment; therefore, by selectively oxidising 5hmC to 5fC prior to bisulfite treatment, only 5mC remains unconverted by bisulfite treatment. By comparing the results of sequencing or array probes for a BS and oxBS treated sample it becomes possible to accurately quantify both 5mC and 5hmC.

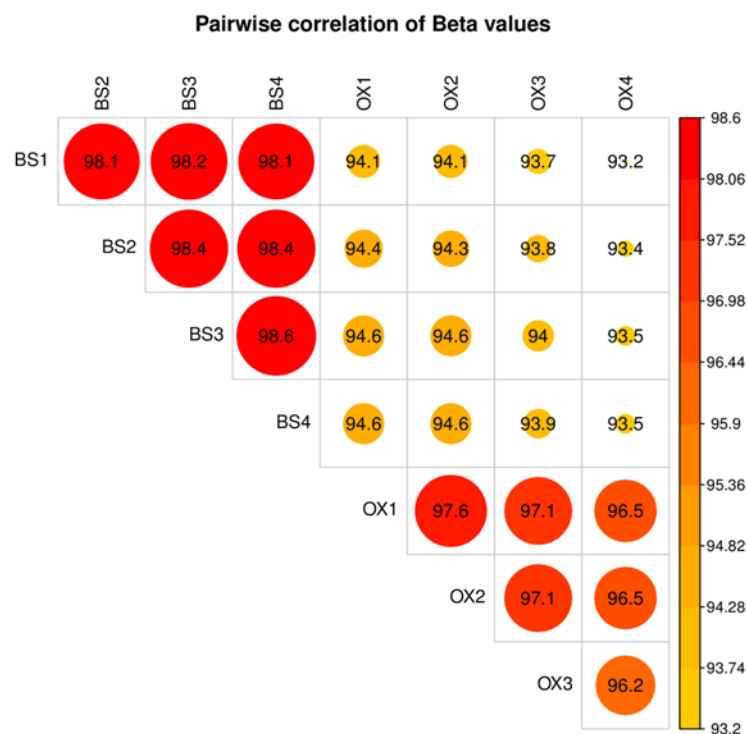
The Infinium 450K methylation array targets cytosine-phosphate-guanine (CpG) nucleotides (although non-CpG sites are also included), in conjunction with BS treatment, it can be used to interrogate methylation at > 450,000 sites across the human genome[2]. Many studies have made use of this technology and the array has been validated by both targeted and whole genome sequencing[2]. In order to quantify both 5mC and 5hmC at these sites we treated the sample both with BS and with oxBS. The methylation level, as measured by untransformed C, on the oxBS treated sample is the true level of 5mC, while the level measured on the BS treated sample is the actual level of 5mC plus 5hmC, therefore, by subtracting the oxBS level from the BS level we were able to ascertain the level of 5hmC alone.

In general, the levels of total 5hmC observed across the genome are approximately 10 fold lower than those of 5mC, although these levels vary across tissue types[21]. In our sample the levels of 5mC, as measured by liquid chromatography-mass spectrometry (LC-MS), were four fold higher than the levels of 5hmC. At individual sites the level of 5mC may be considerably higher than the level of 5hmC resulting in a small difference between untransformed C measured by BS and by oxBS. Thus when doing the BS—oxBS subtraction it is imperative to be

able to distinguish between a genuine difference, resulting from the presence of 5hmC, and the noise inherent in the assay. Our goals in this experiment are to evaluate the reproducibility of this method, to validate the levels of 5hmC measured, and to determine the lower limit of 5hmC detectability empirically, rather than by choosing an arbitrary cut off point. To this end, we have chosen a single, commercially available sample of cerebellum DNA since the highest levels of 5hmC in mammals are observed in adult neuronal tissues [10,21]. This sample might be used as a control in future investigations of 5hmC (GEO accession: GSE63179).

## Results

We performed four replicates of BS and oxBS, which showed a high degree of reproducibility, with the beta values for BS replicates having an average correlation coefficient ( $r$ ) of 98.3 (98.1–98.6), and the oxBS replicates having an average  $r$  of 96.8 (96.2–97.6). Comparing the beta values for BS replicates to the oxBS replicates gave an average  $r$  of 94.0 (range: 93.4–94.6; for the distribution of beta values see S1 Fig.; for pairwise correlations see Fig. 1 and S2 Fig.). Principal components analysis and hierarchical clustering of the arrays showed a clear separation of BS and oxBS treated samples with no clear outliers (see S1 Document and S2 and S3 Figs.). In our analysis we measured 5mC and 5hmC in all four replicates and then used the average of these replicates as our final measurement of 5mC and 5hmC. In this paper we refer to probes as “positive” where we reliably observe 5hmC at any level above 0%. In total 114,734 probes showed a significant positive measurement for 5hmC at a false discovery rate (FDR)  $<0.01$  (see S5 and S6 Figs.). In order to investigate the effects of multiple replicates on 5hmC detection we also simulated fewer replicates by analysing replicates 1 and 3 separately from replicates 2



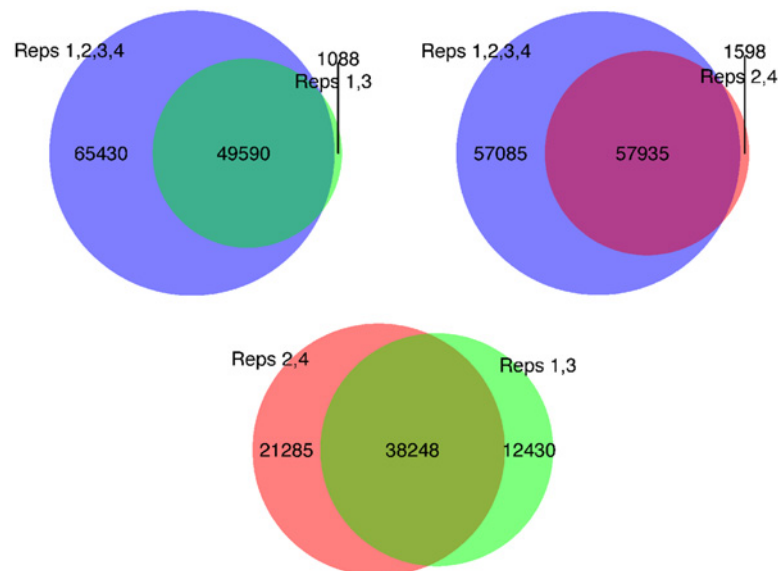
**Fig 1. Pairwise correlation in beta values between BS and oxBS arrays.** The same sample was divided into four BS and four oxBS replicates. Beta values are generated by the minfi analysis package and represent the percentage of unconverted C’s at any given locus. The average correlation of BS replicates is 98.3 and of the oxBS replicates is 96.8. The average correlation between the two methods is 94.0.

doi:10.1371/journal.pone.0118202.g001

and 4. For replicates 1 and 3 we were able to detect 50,678 probes with significant difference between BS and oxBS (either positive or negative), while for replicates 2 and 4 we were able to detect 59,533 probes. The overlap between the probe subsets was 38,248 probes. In addition it should be noted that analysing these subsets yielded probes which did not pass the significance test when all four replicates were analysed together; 1,088 in the 1 and 3 subset and 1,598 in the 2 and 4 subset (see Fig. 2). These values are consistent with the 1% FDR threshold. The median difference in effect size between BS and oxBS in four replicates is 14.0% and between two replicates is 19.4 and 19.8% respectively (see S6 Fig.).

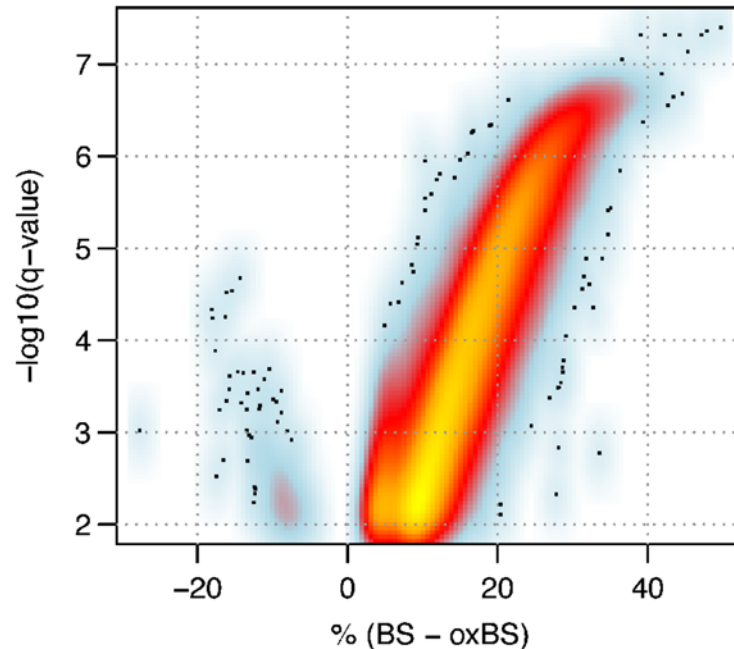
At any given site a mixture of states will be observed, with some molecules methylated, hydroxymethylated or unmethylated. The levels of 5hmC we observe, at any given C, fall between 2.5% and 49.7% with the median at 14.1%. In the replicate subsets the lowest limits were 4.3% and 4.2% respectively (see S7 and S8 Figs.). It should be noted that the detection of 5hmC relies on the difference between BS and oxBS read-outs, therefore it is expected that at sites with zero or low levels of 5hmC this difference may be negative as a consequence of inevitable random noise. In our data we do, indeed, see “negative” values that pass the threshold for significance. However, the number of these “negative” calls also falls within the 1% FDR expected (see Fig. 3 and S6 Fig., see S4 Fig. for distribution of p-values in positive and negative differences between BS and oxBS).

We measured global levels of 5mC and 5hmC in the sample by LC-MC; 5mC was found at 4.6% (averaged from three replicates: 4.24, 4.55 and 5.19%) of all Cs, while 5hmC was found at 1.14% (averaged from three replicates: 1.02, 1.10 and 1.25%). The distribution of positive 5hmC probes was between 13% and 47% of all 450K array probes for any given region, with the highest number of probes observed at downstream, exonic and intronic regions. Here upstream and downstream regions were defined as 1 kb from the transcription start site (TSS) and transcription termination site of each gene, respectively. The lowest number of positive probes was observed in 5'UTRs (13%; see Fig. 4). The average levels of 5hmC were quite similar



**Fig 2. Effect of number of replicates on 5hmC detection.** Top left: Number of probes detected as differentially methylated, i.e. as hydroxymethylated (having a positive 5hmC signal), with all four replicates and with only replicates 1 and 3. Top right: Using only replicates 2 and 4. The number of probes observed in two replicates but not in four fall within the expected 1% FDR. Bottom: Probes with 5hmC in pair 1 and 3 compared to pair 2 and 4. Only 38,246 probes were observed in both pairs.

doi:10.1371/journal.pone.0118202.g002

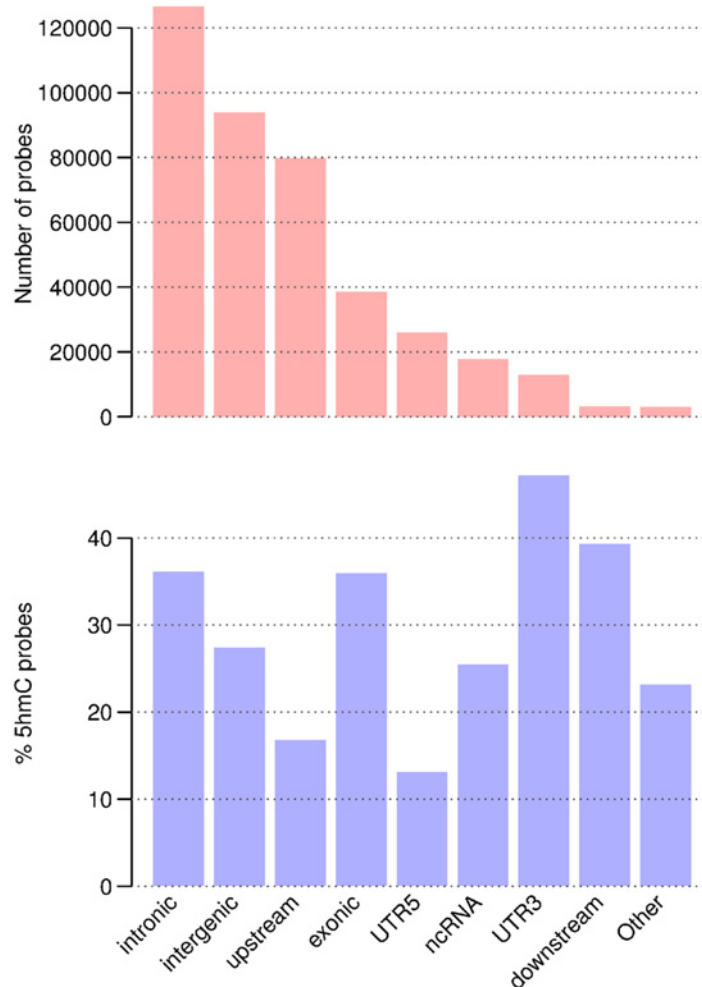


**Fig 3. Distribution of 5hmC levels.** Only probes which were significantly differentially methylated between BS and oxBS are shown hence the lack of probes around zero. For the plot of all probes see [S5 Fig](#). Negative values fall within the expected 1% FDR.

doi:10.1371/journal.pone.0118202.g003

across all genomic regions (mean 14%, [Fig. 5](#)). The lowest average levels were found upstream of genes and in the 5'UTRs (mean 10%), this correlates with the low percentage of probes in this region found to have a positive 5hmC signal. These regions also showed the lowest levels of 5mC (see [Fig. 5](#)). The highest average 5hmC levels were found in introns and 3'UTRs. However, at individual Cs, levels of 5hmC as high as 42% were observed in gene bodies (particularly in introns) (see [Fig. 5](#)). The highest average levels of 5mC were observed in exons (see [Fig. 5](#)). Average levels of both 5hmC and 5mC were lower at CpG islands than at the CpG shores and shelves (regions flanking CpG islands by 0 to 2 kb, and 2 to 4 kb respectively, [Fig. 6a](#)). Transcription start sites also show a marked decrease in 5mC and 5hmC levels, which rise steeply as one moves away from them (see [Fig. 6b](#)). These regional levels correlate with those found in other studies and by other methods[22] (see [S10 Fig](#)). In general the highest levels of 5hmC occurred where 5mC was at ~ 50%, and there was a greater correlation of 5mC-5hmC levels at the lowest levels of 5mC and 5hmC than at the highest (See [Fig. 7](#)). GO analysis did not reveal any particular enrichment of terms associated with 5hmC positive probes in this sample.

We validated these results using an orthogonal assay that employs glucosylation of 5hmC sites followed by MspI digestion, which is specific for CCGG sites, and qPCR. 5hmC glucosyl-transferase can be used to modify 5hmC, yielding glucosyl-5hmC, which is not susceptible to digestion by MspI[23]. Therefore by performing qPCR on glucosylated, unglucosylated and undigested DNA it is possible to ascertain the levels of 5hmC at a given site. We chose 27 sites that had a range of 5hmC values according to the oxBS array data from 0% to 40% and including apparently “negative” values. Sites were selected first by 5hmC level then by the availability of a CCGG motif at that site (see [S3 Table](#)). The correlation coefficient between the 450K results and the qPCR results was 0.94. Sites with negative values by the oxBS array clustered around zero on the qPCR (See [Fig. 8](#) and [S2 Table](#)) demonstrating that these “negative” values are indeed an artefact resulting from the random distribution around 0. The correlation



**Fig 4. Top: the number of probes in each genomic region in the entire 450K array.** The highest number of probes are in introns, followed by intergenic regions. Bottom: the proportion of probes in each region found to have a positive 5hmC signal. While only a small number of probes are designed in 3'UTR's (12984), this region has the highest proportion of probes with a positive 5hmC signal (47%). Regions labeled "Other" are those that do not fit into any of the previous categories.

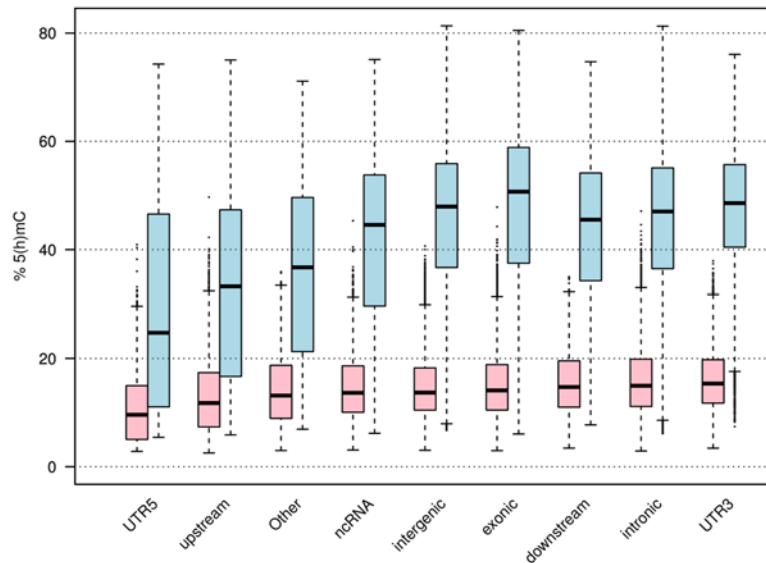
doi:10.1371/journal.pone.0118202.g004

between the oxBS array and the qPCR validation was maintained across the dynamic range of the array experiment from 3% to 40% (see Fig. 8). Two recently published studies have also measured 5hmC in brain tissue using arrays. Stewart et al. [24] used the oxBS array assay while Chopra et al. [22] used TAB array. Our data shows consistency with the dataset from Stewart et al. (see S10 Fig.) whereas the consistency was less good with the TAB array data, suggesting the possibility of method-based differences (see S9 Fig.). However, we must be cautious in drawing conclusions about the comparisons, as the samples were different. The independent validation of our method by a distinct assay (glucosylation and MspI digestion) across the full dynamic range of our data, provides us with confidence in the accuracy of our measurements.

## Discussion

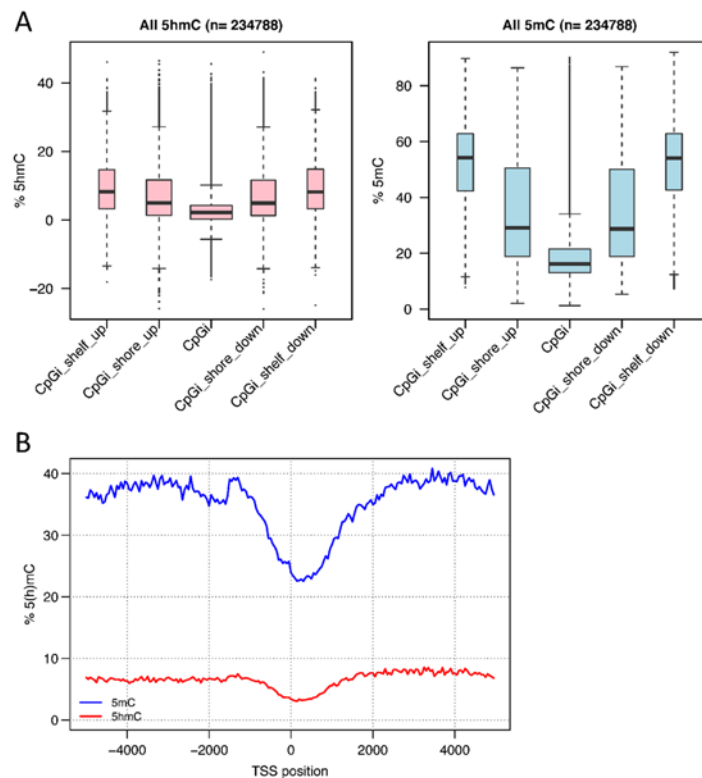
In this paper we have presented 5hmC and 5mC data for a single, commercially available, cerebellum DNA sample. We have also presented an analysis technique that allows validation of a





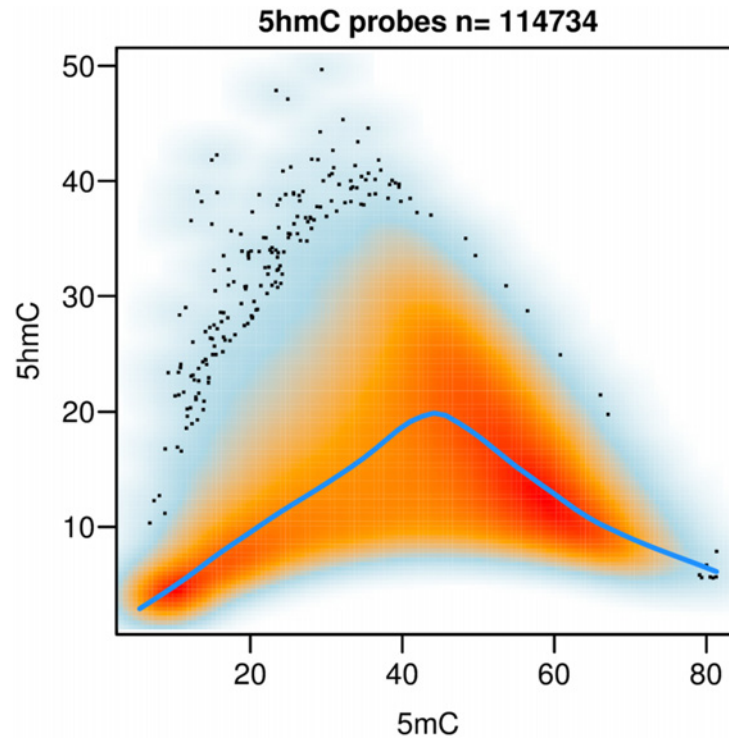
**Fig 5. Distribution of 5mC (blue) and 5hmC (pink) levels across regions.** The lowest median levels of both are in 5'UTR's, however, beyond that there is little correlation between 5mC and 5hmC levels. Median 5hmC levels are similar across all regions, while 5mC levels show greater variation.

doi:10.1371/journal.pone.0118202.g005



**Fig 6. Genomic distribution of 5mC and 5hmC.** 6a: Distribution of 5hmC and 5mC levels at CpG islands, shores (0–2 kb) and shelves (2–4 kb). 5mC levels show a steep rise moving away from CpG sites, 5hmC levels are lowest at CpG sites but show less variation between shores and shelves. The width of the boxes is proportional to the number of underlying probes. 6b: Distribution of 5mC and 5hmC levels around the TSS. Levels of both types of methylation rise steeply moving away from TSS's.

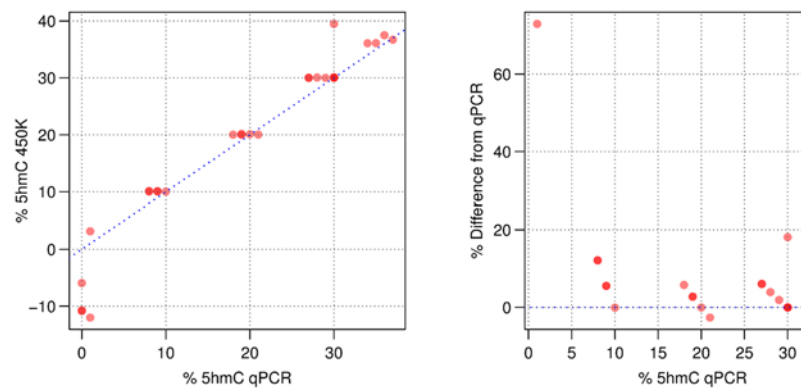
doi:10.1371/journal.pone.0118202.g006



**Fig 7. Scatter plot of 5hmC vs 5mC.** 5hmC levels are highest at intermediate levels of 5mC (~40%).

doi:10.1371/journal.pone.0118202.g007

genuinely positive 5hmC signal at low levels. We have demonstrated the value of multiple replicates. While valuable information can be gained from fewer than four replicates, researchers should be aware of the level of information that may be missed. We found that four replicates allowed the identification of 98%–132% more 5hmC positive probes than only two replicates. It must be stressed that with the methylation array, as with expression arrays, a lack of positive signal cannot be interpreted to mean that there is no 5hmC at that position; only that either the level of 5hmC is below the detection threshold, or that the call quality of the assay at that site was insufficient. However, given that we observed (and validated) 5hmC levels below 5% it is



**Fig 8. Comparison between oxBS arrays and qPCR.** Right: Correlation between 5hmC detected by qPCR vs array probes. Left: Discrepancy between qPCR and arrays quantified as percentage difference relative to qPCR (See also [S1 Document](#)). Data points having the same coordinates appear in darker red.

doi:10.1371/journal.pone.0118202.g008



likely that probes with levels below the detection threshold are very close to zero. The amount of 5hmC at any given probe was estimated by subtracting the level of methylation detected in the oxBS arrays, i.e. 5mC only, from the methylation in the BS arrays, i.e. sum of 5mC and 5hmC. This approach does not explicitly consider that the percentage of C, 5mC and 5hmC must sum to 1. Future work aimed at modeling this constraint would be beneficial to provide more robust estimates of cytosine modification.

The aim of demonstrating the accuracy, and highlighting the pitfalls, of this method can be met using a single sample with multiple replicates, in order to answer the biological question of the role of 5hmC many more samples will be required. Alternative methods have and are being developed[22,24] In every sequencing experiment a trade-off must be made between the number of sites interrogated and the depth of sequencing. In methylation experiments this trade-off becomes even more pertinent as the percentage of modified C's at any given site may be very low (as we have observed in this study), in order to accurately call a site with 10% 5hmC a sequencing depth of hundreds of reads would be required. The 450K array is therefore a particularly valuable tool for case/control regional association studies, as it provides wide coverage and can accurately call 5hmC levels <5%. By making the data for this commercial sample available we hope to provide the field with a resource in the form of a control sample that may be analysed alongside experimental samples. This control would permit greater confidence of 5hmC calls and give an estimate of the variance of the particular assay used.

## Methods

We used a single sample of cerebellum DNA from a 25 year old, healthy male (amsbio ltd. Catalogue number: D1234039-DC, Lot No: A807003). The DNA was taken from fresh tissue. In order to measure the reproducibility of the method, this sample was split into eight 1.2 Kg aliquots, and each aliquot was separately treated with the CEGX BS or oxBS reagents according to the manufacturers specifications ([http://www.cambridge-epigenetix.com/uploads/files/TrueMethyl24\\_UGuide.pdf](http://www.cambridge-epigenetix.com/uploads/files/TrueMethyl24_UGuide.pdf)). The DNA was first sheared (using Covaris G tubes) to 10 kb, and then purified on P6 columns (New England Biolabs). Four of the aliquots were subjected to oxidation and four to mock oxidation prior to the BS treatment (see CEGX protocol). Following the BS reaction the samples were re-quantified. The recovered volume was between 12 and 18  $\mu$ l, and the concentration was between 30 and 60 ng/ $\mu$ l for the oxBS treated aliquots and between 40 and 60 ng/ $\mu$ l for the BS treated aliquots. Since we started with 1.2 Kg, the average loss of material (as measured by Qubit after bisulfite treatment, according to the manufacturers specifications) for oxBS was 44% and for BS was 38%. The samples were then run on the Infinium 450K chip by Genome Services at the Department of Pathology (Cambridge University). The only modification to the standard 450K loading protocol was that 7 Kl of sample were loaded with 1  $\mu$ l NaOH, as opposed to the standard protocol which calls for 4  $\mu$ l of sample and 4  $\mu$ l of NaOH; in this way we ensured that a minimum of 210 ng of sample were loaded on to each array.

Total levels of 5hmC and 5mC in the sample were measured by LC-MS. Global levels of mC and hmC were analyzed as described in Booth *et al.*[19] but using a Q-Exactive mass spectrometer (Thermo). The levels are expressed as percentage of total cytosines[19]. The array results were validated by glycosylation followed by qPCR using the Zymo Quest 5-hmC detection kit and the Zymo QuestTaq qPCR PreMix respectively according to the manufacturers specifications (<http://www.zymoresearch.com/downloads/dl/file/id/126/e2026i.pdf>).

## Analysis

The raw idat files produced by the Illumina platform were analysed with the R/Bioconductor package minfi[25]. The applied normalization method was the subset-quantile within array

normalisation (SWAN)[26]. Probe qualities within each array were assessed using the function `detectionP` in the `minfi` package and probes with detection p-value  $< 0.01$  in more than one array were excluded from further analysis. Probes differentially methylated between BS and oxBS were identified by F-test with variance shrinkage fitting the treatment “BS” or “oxBS” as categorical variable (function `dmpFinder` with `shrinkVar = TRUE`). Raw p-values were converted to false discovery rate (FDR) to control for multiple testing and probes with  $FDR < 0.01$  were deemed as significantly differentially methylated. The beta values were used as methylation percentage to interpret the data. To summarize methylation of each probe within BS and within oxBS, the average of beta values of the 4 replicates was taken. [S1 Document](#) includes full details of the analyses and should enable the reader to reproduce the findings presented here (see also <http://github.com/dariober/Illumina-450K-oxBS> for the code in the [S1 Document](#)).

## Ethics statement

This DNA sample was collected under the BioChain Institute Inc. Institutional Review Board (IRB). This IRB is registered with the Office for Human Research Protections (OHRP), registration number IRB00008283 and has been issued with Federal Wide Assurance (FWA), FWA00017355, for the Protection of Human Subjects for Institutions within the United States by OHRP of the Department of Health and Human Service (DHHS), USA.

## Supporting Information

### **S1 Document.**

(PDF)

### **S1 Fig. Density plot of beta values after having removed failed probes and after SWAN normalization.**

(TIFF)

### **S2 Fig. Principal components analysis based on probe beta values.** In blue the “BS” arrays, in red the “oxBS” arrays.

(TIFF)

### **S3 Fig. Hierarchical clustering of arrays on beta values.**

(TIFF)

### **S4 Fig. Histogram of pvalues for the evidence of BS—oxBS $< 0$ (grey bars) and BS—oxBS $> 0$ (red bars).**

(TIFF)

### **S5 Fig. Probes differentially methylated between BS and oxBS arrays.** Left: All the probes passing quality control ( $n = 401866$ ). Right: Probes passing quality control and with $FDR < 0.01$ for difference between BS and oxBS ( $n = 115020$ , $n$ positive = 114734).

(TIFF)

### **S6 Fig. Effect size, i.e. significant difference in beta values between BS and oxBS, detected with four and two replicates.**

(TIFF)

### **S7 Fig. Difference between BS and oxBS beta values for probes detected significantly different in both the 4 replicates and in replicates 1 and 3.** The shaded grey band marks the diagonal plus and minus 5, i.e. probes outside the grey band differ between 4 and 2 replicates by more than 5% in detected beta value.

(TIFF)

**S8 Fig. Screenshot from UCSC genome browser used to extract the positions of transcription start sites.**

(TIFF)

**S9 Fig. Comparison in estimated 5hmC between oxBS and TAB arrays [3].**

(TIFF)

**S10 Fig. Comparison in estimated 5hmC between the current study and the 5hmC reported in Stewart et al. 2014 [7].**

(TIFF)

**S1 Table. Sample sheet for array design.**

(PDF)

**S2 Table. 5hmC percentages detected by qPCR and 450k array.**

(PDF)

**S3 Table. Details of primers used for qPCR validation.**

(PDF)

## Acknowledgments

S. Balasubramanian is a Senior Investigator of The Wellcome Trust and the Balasubramanian group is core-funded by Cancer Research UK. We would like to thank Tobias Ost and Christine Clark of Cambridge Epigenetix Ltd. for valuable discussions and development of the method.

## Author Contributions

Conceived and designed the experiments: SFF DB MB SKS S. Beck S. Balasubramanian. Performed the experiments: SFF MB. Analyzed the data: DB. Contributed reagents/materials/analysis tools: S. Balasubramanian. Wrote the paper: SFF DB S. Balasubramanian.

## References

1. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. *Genes Dev* 25: 1010–1022. doi: [10.1101/gad.2037511](https://doi.org/10.1101/gad.2037511) PMID: [21576262](https://pubmed.ncbi.nlm.nih.gov/21576262/)
2. Pan H, Chen L, Dogra S, Teh AL, Tan JH, et al. (2012) Measuring the methylome in clinical samples: improved processing of the Infinium Human Methylation450 BeadChip Array. *Epigenetics* 7: 1173–1187. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3469459&tool=pmcentrez&rendertype=abstract>. doi: [10.4161/epi.22102](https://doi.org/10.4161/epi.22102) PMID: [22964528](https://pubmed.ncbi.nlm.nih.gov/22964528/)
3. Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, et al. (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* 10: 2709–2721. PMID: [7079182](https://pubmed.ncbi.nlm.nih.gov/7079182/)
4. Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324: 929–930. doi: [10.1126/science.1169786](https://doi.org/10.1126/science.1169786). PMID: [19372393](https://pubmed.ncbi.nlm.nih.gov/19372393/)
5. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324: 930–935. doi: [10.1126/science.1170116](https://doi.org/10.1126/science.1170116) PMID: [19372391](https://pubmed.ncbi.nlm.nih.gov/19372391/)
6. Ito S, Shen L, Dai Q, Wu SC, Collins LB, et al. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333: 1300–1303. doi: [10.1126/science.1210597](https://doi.org/10.1126/science.1210597). PMID: [21778364](https://pubmed.ncbi.nlm.nih.gov/21778364/)
7. Shen L, Wu H, Diep D, Yamaguchi S, D'Alessio AC, et al. (2013) Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* 153: 692–706. doi: [10.1016/j.cell.2013.04.002](https://doi.org/10.1016/j.cell.2013.04.002) PMID: [23602152](https://pubmed.ncbi.nlm.nih.gov/23602152/)
8. Hashimoto H, Hong S, Bhagwat AS, Zhang X, Cheng X (2012) Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for

- active DNA demethylation. *Nucleic Acids Res* 40: 10203–10214. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22962365>. doi: [10.1093/nar/gks845](https://doi.org/10.1093/nar/gks845) PMID: [22962365](https://pubmed.ncbi.nlm.nih.gov/22962365/)
9. Maiti A, Drohat AC (2011) Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: POTENTIAL IMPLICATIONS FOR ACTIVE DEMETHYLATION OF CpG SITES. *J Biol Chem* 286: 35334–35338. doi: [10.1074/jbc.C111.284620](https://doi.org/10.1074/jbc.C111.284620). PMID: [21862836](https://pubmed.ncbi.nlm.nih.gov/21862836/)
  10. Ficiz G, Branco MR, Seisenberger S, Santos F, Krueger F, et al. (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 473: 398–402. doi: [10.1038/nature10008](https://doi.org/10.1038/nature10008). PMID: [21460836](https://pubmed.ncbi.nlm.nih.gov/21460836/)
  11. Lister R, Mukamel E, Nery JR, Urich M, Puddifoot C, et al. (2013) Global epigenomic reconfiguration during mammalian brain development. *Science* 341: 1237905. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3785061&tool=pmcentrez&rendertype=abstract>. doi: [10.1126/science.1237905](https://doi.org/10.1126/science.1237905) PMID: [23828890](https://pubmed.ncbi.nlm.nih.gov/23828890/)
  12. Branco MR, Ficiz G, Reik W (2012) Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet* 13: 7–13. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22083101>. Accessed 2014 Feb 21. doi: [10.1038/nrg3080](https://doi.org/10.1038/nrg3080) PMID: [22083101](https://pubmed.ncbi.nlm.nih.gov/22083101/)
  13. Liu S, Wang J, Su Y, Guerrero C, Zeng Y, et al. (2013) Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Res* 41: 6421–6429. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3711458&tool=pmcentrez&rendertype=abstract>. doi: [10.1093/nar/gkt360](https://doi.org/10.1093/nar/gkt360) PMID: [23658232](https://pubmed.ncbi.nlm.nih.gov/23658232/)
  14. Mastroeni D, Grover A, Delvaux E, Whiteside C, Coleman PD, et al. (2010) Epigenetic changes in Alzheimer's disease: Decrements in DNA methylation. *Neurobiol Aging* 31: 2025–2037. doi: [10.1016/j.neurobiolaging.2008.12.005](https://doi.org/10.1016/j.neurobiolaging.2008.12.005) PMID: [19117641](https://pubmed.ncbi.nlm.nih.gov/19117641/)
  15. Rudenko A, Dawlaty MM, Seo J, Cheng AW, Meng J, et al. (2013) Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* 79: 1109–1122. Available: <http://www.sciencedirect.com/science/article/pii/S0896627313007149> <http://www.ncbi.nlm.nih.gov/pubmed/24050401>. doi: [10.1016/j.neuron.2013.08.003](https://doi.org/10.1016/j.neuron.2013.08.003) PMID: [24050401](https://pubmed.ncbi.nlm.nih.gov/24050401/)
  16. Mellén M, Ayata P, Dewell S, Kriaucionis S, Heintz N (2012) MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151: 1417–1430. doi: [10.1016/j.cell.2012.11.022](https://doi.org/10.1016/j.cell.2012.11.022) PMID: [23260135](https://pubmed.ncbi.nlm.nih.gov/23260135/)
  17. Shapiro R, DiFate V, Welcher M (1974) Deamination of cytosine derivatives by bisulfite. Mechanism of the reaction. *J Am Chem Soc* 96: 906–912. doi: [10.1021/ja00810a043](https://doi.org/10.1021/ja00810a043). PMID: [4814744](https://pubmed.ncbi.nlm.nih.gov/4814744/)
  18. Hayatsu H, Wataya Y, Kai K, Iida S (1970) Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry* 9: 2858–2865. doi: [10.1021/ja00706a062](https://doi.org/10.1021/ja00706a062). PMID: [5459538](https://pubmed.ncbi.nlm.nih.gov/5459538/)
  19. Booth MJ, Branco MR, Ficiz G, Oxley D, Krueger F, et al. (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336: 934–937. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22539555>. Accessed 2014 Feb 21. doi: [10.1126/science.1220671](https://doi.org/10.1126/science.1220671) PMID: [22539555](https://pubmed.ncbi.nlm.nih.gov/22539555/)
  20. Booth MJ, Ost TWB, Beraldi D, Bell NM, Branco MR, et al. (2013) Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat Protoc* 8: 1841–1851. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24008380>. doi: [10.1038/nprot.2013.115](https://doi.org/10.1038/nprot.2013.115) PMID: [24008380](https://pubmed.ncbi.nlm.nih.gov/24008380/)
  21. Ruzov A, Tsenkina Y, Serio A, Dudnakova T, Fletcher J, et al. (2011) Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Res* 21: 1332–1342. doi: [10.1038/cr.2011.113](https://doi.org/10.1038/cr.2011.113). PMID: [21747414](https://pubmed.ncbi.nlm.nih.gov/21747414/)
  22. Chopra P, Papale L a, White AT, Hatch A, Brown RM, et al. (2014) Array-based assay detects genome-wide 5-mC and 5-hmC in the brains of humans, non-human primates, and mice. *BMC Genomics* 15: 131. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24524199>. Accessed 2014 Feb 20. doi: [10.1186/1471-2164-15-131](https://doi.org/10.1186/1471-2164-15-131) PMID: [24524199](https://pubmed.ncbi.nlm.nih.gov/24524199/)
  23. Robertson AB, Dahl JA, Vågbo CB, Tripathi P, Krokan HE, et al. (2011) A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* 39: e55. doi: [10.1093/nar/gkr051](https://doi.org/10.1093/nar/gkr051) PMID: [21300643](https://pubmed.ncbi.nlm.nih.gov/21300643/)
  24. Stewart SK, Morris TJ, Guilhamon P, Bulstrode H, Bachman M, et al. (n.d.) oxBS-450K: A method for analysing hydroxymethylation using 450K BeadChips. *Epigenomics*. doi: [10.2217/epi.14.89](https://doi.org/10.2217/epi.14.89) PMID: [25590359](https://pubmed.ncbi.nlm.nih.gov/25590359/)
  25. Aryee M, Jaffe A, Corrada-Bravo H, Ladd-Acosta C, Feinberg A, et al. (2014) Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays. *Bioinformatics* 1–8. Available: <http://bioinformatics.oxfordjournals.org/content/early/2014/01/28/bioinformatics.btu049.short>. doi: [10.1186/1471-2105-15-1](https://doi.org/10.1186/1471-2105-15-1) PMID: [25189783](https://pubmed.ncbi.nlm.nih.gov/25189783/)
  26. Makismovic J, Gordon L, Oshlack A (2012) SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. *Genome Biol* 13: R44. doi: [10.1186/gb-2012-13-6-r44](https://doi.org/10.1186/gb-2012-13-6-r44). PMID: [22703947](https://pubmed.ncbi.nlm.nih.gov/22703947/)