



Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data



Tiffany J. Morris*, Stephan Beck

UCL Cancer Institute, University College London, London WC1E 6BT, UK

ARTICLE INFO

Article history:

Received 6 June 2014

Revised 30 July 2014

Accepted 5 August 2014

Available online 16 September 2014

Keywords:

HumanMethylation450

Epigenetics

Analysis pipelines

EWAS

DNA methylation

450k BeadChip

ABSTRACT

The Illumina HumanMethylation450 BeadChip has become a popular platform for interrogating DNA methylation in epigenome-wide association studies (EWAS) and related projects as well as resource efforts such as the International Cancer Genome Consortium (ICGC) and the International Human Epigenome Consortium (IHEC). This has resulted in an exponential increase of 450k data in recent years and triggered the development of numerous integrated analysis pipelines and stand-alone packages. This review will introduce and discuss the currently most popular pipelines and packages and is particularly aimed at new 450k users.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

1. Introduction

DNA methylation (DNAm) is an epigenetic modification that plays an important role in the regulation of gene expression and has become an important avenue of research in our quest to gain a better understanding of human development and disease. A number of lab techniques are available for interrogating DNAm [1] including the popular Infinium HumanMethylation450 BeadChip (450k array) [2,3]. Although not as comprehensive as sequencing-based methods, it is more affordable and much simpler to analyse and interpret. These characteristics make the 450k array an ideal choice for epigenome-wide association studies (EWAS) involving hundreds or even thousands of cohort samples [4–6] and also for identifying methylation signatures as biomarkers of disease state and progression [7,8].

The 450k array contains 485,512 probes covering 99% of RefSeq genes. The probes interrogate 19,755 unique CpG islands with additional coverage in shore regions and miRNA promoters as well as 3091 probes at non-CpG sites. For each probe sequence, a median of 14 beads is randomly distributed on the array. Each of these beads contains hundreds of thousands of oligonucleotides. This provides a unique set of internal technical replication on each array. The 450k array is an extension of an earlier BeadChip, the HumanMethylation27 array (27k array). The 27k array was

designed using the Infinium I probe chemistry, which has two beads per probe, one in the red channel and one in the green channel. To create the more comprehensive array platform a second assay type, Infinium II, was used in addition to the Infinium I assay. The Infinium II assay only uses one bead per probe in the red or green channel representing methylated or unmethylated respectively. The Infinium II assay enabled more probes to fit on the array but due to particular probe characteristics (particularly CpG density within the probe) 30% of the probes retained the Infinium I assay creating the two-assay design of the 450k array. The two probe types display a slightly different dynamic range potentially leading to a type II bias during analysis [9].

This two-assay design has led to the development of a variety of specialised within-array normalisation algorithms [9–14] to adjust for potential type II bias. This assay adjustment combined with other basic steps for preprocessing make up the general 450k array analysis workflow reviewed in detail by Dedeurwaerder et al. [15]. Briefly, raw data (IDAT files) is imported using the illuminaio [16] tool implemented in *minfi*. Then a number of quality control metrics are examined to determine the success of the bisulphite conversion and subsequent array hybridisation. Probe filtering is performed to remove probes that have failed to hybridise (detection *p*-value) and that are not represented by a minimum of 3 beads on the array. It may also be of interest to remove probes that overlap with single nucleotide polymorphisms (SNPs), that cross hybridise to multiple genomic locations or are on the sex chromosomes. After filtering it is important to perform within-array normalisation. Methods

* Corresponding author.

E-mail address: tiffany.morris@ucl.ac.uk (T.J. Morris).

include background correction, dye-bias adjustment [17] and the aforementioned type II adjustment. Following this it is important to analyse data for potential batch effects. Batch effects [18] are a common source of variation in high-throughput experiments. They represent measurements related to conditions that are not the biological or scientific variables in the study (i.e. date of experiment, chip or instrument used, batch of reagents, technician running samples, etc.). Batch effects can usually be avoided with careful study design and the ability to correct for them is dependent on the degree of confounding in the particular dataset. Popular methods for batch effect correction are the supervised correction methods *ComBat* [19], surrogate variable analysis (SVA) [20] and independent surrogate variable analysis (ISVA) [21]. Also an unsupervised between-array normalisation method (i.e. functional normalisation [22]) may be effective in removing batch effects. Depending on the study, one may want to consider correction for cell heterogeneity [23]. After these preprocessing steps, the calculation of differentially methylated positions (DMPs) between groups of interest is the first step in the analysis process. This is most easily done between two groups, however, more complex analyses can be designed to look at time-course studies, twin designs [24–27], multiple groups and paired samples. To focus analysis and narrow results calling differentially methylated regions (DMRs) is the next step [28]. Following this, additional steps in the 450k array analysis workflow include copy number alteration (CNA) analysis, integration with external data resources, data visualisation and data interpretation. There are a number of analysis packages that include many or all of these modules that make up the 450k workflow. The purpose of this review is to describe in detail five comprehensive, freely available 450k software packages (*methylumi*, *minfi*, *wateRmelon*, *ChAMP* and *RnBeads*). Fig. 1 shows the modules within these five analysis

packages colour-coded to indicate which module is in which package. We will also highlight a number of stand-alone packages for additional downstream analysis (*Marmal-aid*, *RefFreeEWAS*, *EWasher*) that can be used in addition to the more comprehensive packages. The Illumina Genome Studio software only offers basic preprocessing and analysis options and requires the purchase of a license so will not be discussed in this review. Table 1 shows all free 450k software packages available to date.

2. Comprehensive 450K analysis packages

Here we introduce five comprehensive packages developed for 450k array data analysis: *methylumi*, *minfi*, *wateRmelon*, *ChAMP* and *RnBeads*. The first four are available through Bioconductor while *RnBeads* is a stand-alone R-package. All five packages allow the user to import raw IDAT files or tabular methylation values. However, the availability of the IDAT files enables access to more functionality in each package (particularly quality control metrics).

2.1. Methylumi

The *methylumi* [29] Bioconductor package provides R classes for holding and manipulating 450k array data. These classes enable the access of MIAME (Minimum Information about a Microarray Experiment) information including sample details, feature information and multiple matrices of data. *Methylumi* enables the user to perform quality control interrogation, background correction and normalisation. In addition to 450k analysis *methylumi* has methods that work with GoldenGate and 27k array data.

Methylumi implements an effective normalisation option called noob (normal-exponential convolution using out-of-band probes)

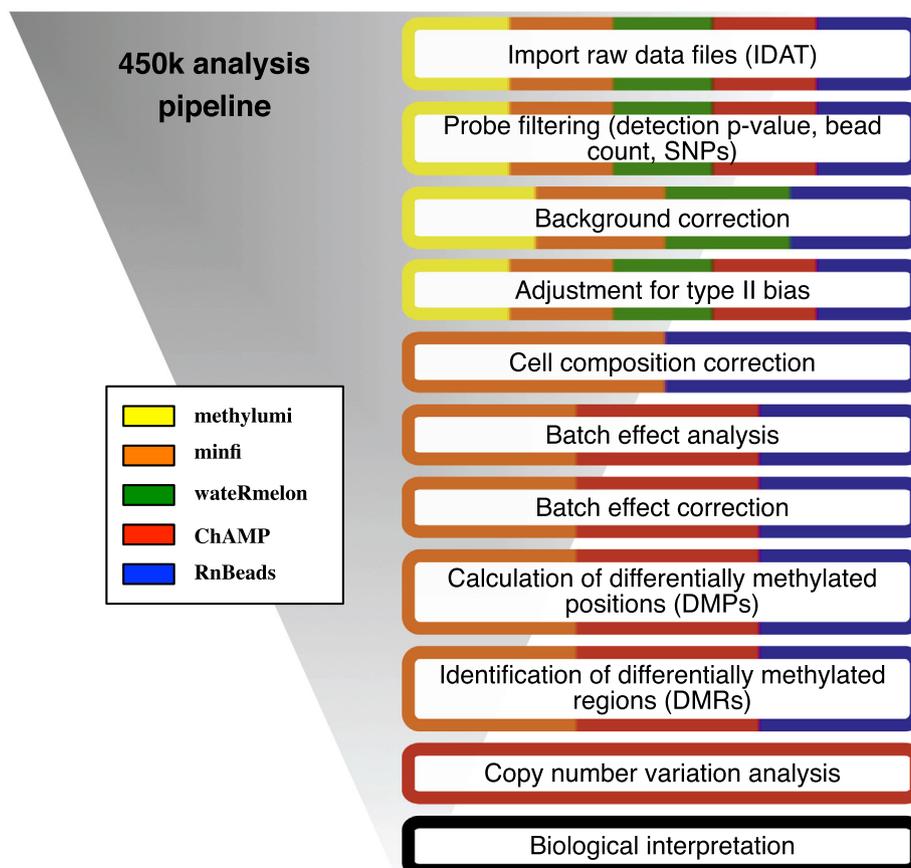


Fig. 1. Popular 450k analysis pipelines with their respective module options.

Table 1
Freely available packages for Infinium 450k data analysis.

Package	Use	References
<i>ChAMP</i>	Comprehensive suite of functions; automated pipeline	[35]
<i>COHCAP</i>	CpG island analysis and gene expression data integration	[49]
<i>Comb-p</i>	DMR calling	[47]
<i>DMRcate</i>	DMR calling	[45]
Epigenetic clock	Predictor of sample age	[54]
<i>EWasher</i>	Reference-free cell composition correction	[53]
<i>FastDMA</i>	Quantile normalisation and DMP/DMR calling	[46]
<i>IMA</i>	Preprocessing including normalisation methods; Pipeline option	[60]
<i>Lumi</i>	Background correction, general normalisation	[61]
<i>Marmal-aid</i>	450k database for data integration	[48]
<i>MethylAid</i>	Interface for interactive sample QC	[39]
<i>Methylumi</i>	Comprehensive suite of functions	[29]
<i>Minfi</i>	Comprehensive suite of functions	[31]
<i>NIMBL</i>	Matlab code for QC and DMP calling	[62]
<i>RefFreeEWAS</i>	Reference-free cell composition correction	[37]
<i>RnBeads</i>	Comprehensive suite of functions	[36]
<i>shinyMethyl</i>	Interface for interactive sample QC	[38]
<i>wateRmelon</i>	Preprocessing including performance metrics and numerous normalisation methods	[12]

[30]. This method includes dye-bias equalisation to control for the different average intensities in the red and green channels and background correction to remove technical variation. Three models for background correction are included in *methylumi* and are used together with the intensities of either negative controls or out-of-band controls within each colour channel. Out-of-band probes are those that provide signals in the opposite channel from the probe design. There are 135,501 out-of-band probes on the 450k array. The noob normalisation uses normal-exponential convolution to subtract the background mean intensity (estimated from the out-of-band control probes) from the foreground intensity. Although noob was not specifically designed to correct for type II bias it is effective in doing so.

Methylumi has methods that convert its R classes to work with other 450k analysis packages allowing users to pick and choose modules from various packages that work best for a particular application.

2.2. Minfi

The *minfi* [31] Bioconductor package is a comprehensive package that like *methylumi* was developed early on and has its own set of R classes for holding and manipulating 450k array data. Although *minfi* does not provide a single function to run the entire pipeline it does offer all the modules required for a full workflow. It is frequently updated to offer methods for the newest analysis options available to 450k users. These include both DMR calling and block finding modules, a module for estimating cell-type composition of whole blood using a reference dataset [32] and a new between-array normalisation algorithm, termed functional normalisation [22].

Minfi has included the Genome Studio normalisation and preprocessing methods for users familiar with those methods and interested in using them. For quality control analysis a number of images are included and an HTML QC report can be created that includes visualisation of the array's internal controls. It also has a useful function for predicting sample gender.

Two different functions in *minfi* estimate DMRs depending on the genomic region. The function *bumphunte()* implements the Bioconductor package *bumphunter* [33] and focuses on short-range (1–2 kb) DNAm changes, e.g. around gene promoters. To find blocks of differential methylation [34], the function *cpgCollapse()* covers long-range changes in DNAm status as represented by the 170,000 open sea probes on the 450k array.

For estimation of cell-type composition of whole blood samples and correction for this difference in composition the *estimateCellCounts()* function implements an algorithm that utilises a reference dataset of different blood cell types to estimate cell composition in any dataset of whole blood [32].

The most recent update to *minfi* is the functional normalisation algorithm, *preprocessFunnorm()* [22]. This utilises control probes on the bead chip (similar to noob in *methylumi*) to remove technical variation. The control probes act as surrogates representing the unwanted variation, which may include batch effects. In fact, the authors show that replication between experiments are improved even with a batch effect present. The method is expected to be most useful in datasets with large differences between samples (e.g. cancer vs. normal). The authors show that the functional normalisation outperforms within array type II adjustment normalisations by improving the ability to replicate findings between experiments.

2.3. wateRmelon

The *wateRmelon* Bioconductor package was developed to provide access to the 15 normalisation methods and three performance metrics developed and described by Pidsley et al. [12]. The package provides the functions for loading data in text files or IDAT files and for filtering based on detection *p*-value and bead count per probe.

wateRmelon takes advantage of known DNAm patterns that have been associated with genomic imprinting, X-chromosome inactivation (XCI) and also the 65 single nucleotide polymorphisms (SNPs) present on the array to create three independent metrics which can be used to test methods of correction and normalisation. Discrete imprinted differentially methylated regions (iDMRs) are expected to have monoallelic methylation ($\beta = 0.5$). XCI causes male–female differences on the X-chromosome that lead to females showing at least 50% methylation and males substantially less. Finally, the signal from the SNP probes is expected to cluster into three distinct genotype groups (i.e. AA, AB, BB) and could be used to provide an indication of technical variation between samples. These three natural control sets have been used to create the performance metrics *dmrse_row()*, *genki()* and *seabi()* respectively. Of the 15 normalisation functions implemented in *wateRmelon*, the authors show their function, *dasen()*, to be the best performing.

To enable users to take advantage of the performance metrics and normalisation methods *wateRmelon* works with the R classes needed for the other packages.

2.4. ChAMP

The Chip Analysis Methylation Pipeline, *ChAMP* [35], is a Bioconductor package that streamlines and automates many of the *minfi* functions for a more inexperienced R user. As such it offers eight functions that can be manipulated by changing parameters if the user wishes to use options other than the default values. One useful feature of the *ChAMP* package is the option to run the entire pipeline with a single function *champ.process()*. The caveats with this function are the time it takes to run and the fact that not all intermediate tables are saved. However, once the user is familiar with the package this offers a quick option for subsequent analyses. The default normalisation in *ChAMP* is BMIQ [9] and there is also the option to run *ComBat* [19] to correct for batch effect related to slide number (Sentrix ID) as long as the study design is not confounded by slide number (an example of this would be all control samples on one slide and test samples on another).

ChAMP offers a few modules that are currently unique to *ChAMP*. The DMR hunting algorithm ‘Probe Lasso’ is run using the *champ.lasso()* function. Probe lasso requires differential methylation *p*-values for all probes after filtering (even those that are above the significance threshold) and uses a dynamic window based on genomic features to capture DMRs. This method is described in more detail in this issue (Butcher and Beck). In addition, *ChAMP* includes a module for copy number alteration analysis using the function *champ.CNA()* [19]. This module utilises the intensity values so users must have either the IDAT files or a table of intensity values.

2.5. RnBeads

RnBeads [36] is another comprehensive package that can analyse sequencing based DNAm data, in addition to 27k and 450k data. Like *ChAMP*, *RnBeads* offers a full pipeline option that runs through all the modules with a single function *rnbeads.run.analysis()*. Alternatively the workflow can be run in steps with a number of normalisation options (including SWAN, BMIQ, *wateRmelon* methods and *methyumi*’s *noob* method. A benefit of *RnBeads* is the detailed HTML report that describes the analysis that was done along with results and images including a number of quality control plots.

RnBeads includes a module for ‘Annotation Inference’ that includes functions to infer gender of samples using an internal classifier based on signal intensity or identify hidden confounders using SVA. This module also implements both a reference based cell type correction method for blood [32] (also implemented in *minfi*) and a reference free method (also available as a stand-alone package described below) [37]. A separate module allows the user to inspect data for sample traits that might lead to batch effects with a few different plots (principal component analysis, PCA and multidimensional scaling, MDS). *RnBeads* also offers functions for interactions with UCSC to allow ENCODE annotation and data visualisation.

RnBeads provides images for inspecting methylation profiles and also conducts differential methylation analysis between sample groups using linear modelling (*limma* package) or regular *t*-tests to compute *p*-values. There is an option for paired analysis. The function for calculating DMRs simply calculates the mean methylation across all probes in the genomic region.

3. Additional resources

In addition to the comprehensive packages reviewed here we thought it would be worthwhile to highlight other stand-alone packages that can be incorporated into any analysis pipeline.

3.1. Quality control

Although all the packages above include modules for assessing quality control two new packages worth mentioning are *shinyMethyl* [38] and *MethylAid* [39]. Both of these packages are interactive Shiny applications (<http://www.rstudio.com/shiny/>) that enable users to investigate sample quality through an interactive user interface. The packages each provide slightly different diagnostic plots of the internal control probes present on the array, but both allow users to select thresholds and observe the potential impact of batch effects.

3.2. Probe annotation

As described above it is important to filter out probes that have failed during hybridisation. In many studies it may also be beneficial to filter out X- and Y-chromosomes. A few resources are now available that annotate probes hybridising to multiple genomic locations (non-specific probes) or those that overlap with SNPs particularly at the probed CpG dinucleotide [40–42]. Additionally, the Bioconductor package *Illumina450ProbeVariants.db* [43] is a curated collection of data from the 1000 genomes project [44] that includes minor allele frequency of SNPs derived from four populations (American, African, European or Asian) with details on the SNPs distance from the CpG dinucleotide.

3.3. DMR calling methods

In addition to the DMR hunting methods included in packages reviewed here (i.e. probe lasso and *bumphunter*) two stand-alone packages for calling DMRs have been developed recently.

DMRcate [45] is a Bioconductor package that uses a kernel function to aggregate groups of significant probes where the distance to the next consecutive probe is less than the designated number of nucleotides.

FastDMA [46] is a package written in C++/Python that includes a function to perform basic quantile normalisation before analysing samples for DMPs or DMRs. The DMP analysis uses ANCOVA (a generalised linear model combining ANOVA and regression) to compare two linear regression models, one assuming an overall mean across all study groups while the other assumes different means. A *p*-value is calculated and then the false discovery rate is computed to identify significantly differentially methylated probes. The DMR caller then identifies DMRs containing several probes with uniform methylation in predefined regions (promoter or CpG Island) or in arbitrary regions based on a sliding window that looks for more than two significant probes.

Another package *comb-p* [47] has been extended to work with 450k data. It combines previously calculated *p*-values that are spatially correlated resulting in a list of regions with an aggregated and corrected *p*-value assigned to each region.

3.4. Data integration

With the large number of EWAS being conducted and the vast amount of publicly available data it is important to come up with methods and tools for data integration. This is an area of analysis that needs more emphasis as data integration can be challenging.

Marmal-aid [48] is an R-package and database that is frequently updated and currently includes more than 11,000 450k array samples from nearly 200 different tissues and approximately 100 different diseases. This is a powerful tool for integrating smaller datasets to ask different biological questions, to validate results or to observe trends across different samples.

COHCAP [49] is a software package that takes methylation data (450k or BS-Seq) that has been preprocessed and identifies CpG

islands that show a consistent methylation pattern among CpG sites. This package also integrates with gene expression data to identify CpG islands that potentially regulate downstream gene expression.

3.5. Visualisation

All the pipelines and packages discussed here produce a number of images. However, custom high quality images and plots can be created using R/Bioconductor packages including *ggplot2* [50] and *Gviz* [51]. Analysis output from any of the packages mentioned here could also be imported into UCSC Genome Browser or IGV (Integrative Genomics Viewer) for data visualisation.

4. Emerging 450K analyses

As reviewed here, there are a number of comprehensive pipelines for 450k array analysis for those with varying skills in R programming and statistics as well as several packages for downstream data analysis. In addition, a number of new packages have recently been released. Here we briefly highlight some of the most recent developments.

4.1. Cell composition correction

An important issue in methylation studies is cellular heterogeneity [23]. Whole blood, like many other tissues, is made up of a number of different cell types each with different methylation profiles that can vary in proportion with age or disease status. Houseman et al. [32] developed a statistical method using reference data to accurately estimate relative proportions of differing cell types in whole blood. This method has been incorporated into methods in *minfi* and *RnBeads*. Guintivano et al. [52] have implemented a similar method for particular brain cell types. A reference dataset for each cell type would be the ideal solution and the International Human Epigenome Consortium (IHEC) is attempting to generate this data. Unfortunately it is costly, time consuming and in some cases unrealistic to come up with reference sets for each cell type. As such two methods have recently been developed to deal with this issue without the need for a reference dataset. *EWasher* [53] is a package written in Perl and *RefFreeEWAS* [37] is an R-package. Both can be used alongside any of the other analysis packages. It is important to consider the top down systems view of the study being analysed before applying a cell-composition correction as it may correct away the methylation differences that are of interest. Cell-composition correction may be of different importance in studies with large methylation differences (cancer vs. normal) compared to those with small methylation differences (common disease).

4.2. Estimating sample age

A significant amount of work is going into finding DNAm profiles for particular disease states. However one condition every sample has in common is age. Horvath [54] took advantage of this fact by using 8000 samples from 82 27k or 450k datasets to define a DNAm profile consisting of 353 CpG sites that can act as a predictor of age. It is not clear whether these clock CpGs are simply a marker for age or relate to an effector of aging. However, they do highlight interesting questions that may give us more information on tissue aging. For one, the clock systematically overestimates the age of female breast tissue, and conversely heart tissue age seems to be underestimated while the prediction of brain tissue age is extremely accurate. In addition, work with monozygotic and dizygotic twins showed that age acceleration is extremely heritable. The epigenetic clock script is written in R and utilises the *WGCNA*

R-package to measure pure age effects. Details about using the epigenetic clock can be found in the supplementary files of the paper.

4.3. Hydroxymethylation analysis

The recent development of chemical [55] and enzymatic [56] assays for oxidation of 5-methylcytosine (5mC) enables the discrimination between 5mC and 5-hydroxymethylcytosine (5hmC) and their quantitative measurements in a single workflow. In addition to sequencing-based applications, this has created a new use for the 450k array which has already been successfully translated for the chemical assay into a commercially available kit (TrueMethyl) by CEGX [57]. First attempts indicate that both assays work on the 450k platform (Field et al., submitted, [58]) but as yet no methods specific to 5hmC analysis have been developed. However, judging by the speed this community has addressed such needs in the past, it will not be long before appropriate methods become available and integrated into the pipelines discussed in this review.

5. Limitations and conclusions

Despite its popularity and obvious success, the 450k platform has some limitations which cannot be addressed by computational approaches. The coverage of total CpG sites, for instance, is low (only around 2%) which means that some features such as enhancers are only barely or not at all covered. The design of the Infinium assay does not allow the detection of allele-specific DNAm that is important in the context of imprinting and other parent-of-origin effects. Furthermore, the platform is human-specific and thus not suitable for comparative analysis with model organisms such as mice. The mouse reference genome matches only 13,715 probes on the 450k array [59].

Looking forward, however, the 450k platform is likely to remain popular for some time to come and it will be the platform of choice for EWAS for the foreseeable future. This is largely due to the innovative and freely available computational solutions that the community has been developing to streamline the analysis and address the many technical and biological confounders discussed above. Judged on publications indexed in PubMed and data submissions to GEO (www.ncbi.nlm.nih.gov/geo/), the public repository for this type of data, 450k use and analysis was exponentially growing at the time of writing this review.

Acknowledgements

We wish to acknowledge support from the Wellcome Trust (99148), a Royal Society Wolfson Research Merit Award (WM100023) and EU-FP7 projects IDEAL (259679), EPIGENESYS (257082) and BLUEPRINT (282510).

References

- [1] C. Bock, *Nat. Rev. Genet.* 13 (2012) 705–719.
- [2] M. Bibikova et al., *Genomics* 98 (2011) 288–295.
- [3] J. Sandoval et al., *Epigenetics* 6 (2011) 692–702.
- [4] V.K. Rakyan et al., *Nat. Rev. Genet.* 12 (2011) 529–541.
- [5] K.B. Michels et al., *Nat. Methods* 10 (2013) 949–955.
- [6] Y. Liu et al., *Nat. Biotechnol.* 31 (2013) 142–147.
- [7] P.W. Laird, *Nat. Rev. Cancer* 3 (2003) 253–266.
- [8] V.V. Levenson, *Expert Rev. Mol. Diagn.* 10 (2010) 481–488.
- [9] A.E. Teschendorff et al., *Bioinformatics* 29 (2013) 189–196.
- [10] J. Maksimovic et al., *Genome Biol.* 13 (2012) R44.
- [11] S. Dedeurwaerder et al., *Epigenomics* 3 (2011) 771–784.
- [12] R. Pidsley et al., *BMC Genomics* 14 (2013) 293.
- [13] F. Marabita et al., *Epigenetics* 8 (2013) 333–346.
- [14] N. Touleimat, J. Tost, *Epigenomics* 4 (2012) 325–341.
- [15] S. Dedeurwaerder et al., *Brief Bioinform.* (2013), <http://dx.doi.org/10.1093/bib/bbt054>.

- [16] M.L. Smith, *F1000Res.* 2 (2013) 264, <http://dx.doi.org/10.12688/f1000research.2-264.v1>.
- [17] P. Yousefi et al., *Epigenetics* 8 (2013) 1141–1152.
- [18] J.T. Leek et al., *Nat. Rev. Genet.* 11 (2010) 733–739.
- [19] W.E. Johnson et al., *Biostatistics* 8 (2007) 118–127.
- [20] J.T. Leek et al., *Bioinformatics* 28 (2012) 882–883.
- [21] A.E. Teschendorff et al., *Bioinformatics* 27 (2011) 1496–1505.
- [22] J.-P. Fortin et al., *bioRxiv* (2014), <http://dx.doi.org/10.1101/002956>.
- [23] A.E. Jaffe, R.A. Irizarry, *Genome Biol.* 15 (2014) R31.
- [24] D. Martino et al., *Genome Biol.* 14 (2013) R42.
- [25] E. Grundberg et al., *Am. J. Hum. Genet.* 93 (2013) 876–890.
- [26] H. Heyn et al., *Carcinogenesis* 34 (2013) 102–108.
- [27] J.T. Bell, T.D. Spector, *Trends Genet.* 27 (2011) 116–125.
- [28] U. Baron et al., *Epigenetics* 1 (2006) 55–60.
- [29] P. Du, S. Davis, S. Bilke, T. Triche Jr., M. Bootwalla (2014). methylumi: Handle Illumina methylation data. R package version 2.10.0.
- [30] T.J. Triche Jr. et al., *Nucleic Acids Res.* 41 (2013) e90.
- [31] M.J. Aryee et al., *Bioinformatics* 30 (2014) 1363–1369.
- [32] E.A. Houseman et al., *BMC Bioinformatics* 13 (2012) 86.
- [33] A.E. Jaffe et al., *Int. J. Epidemiol.* 41 (2012) 200–209.
- [34] K.D. Hansen et al., *Nat. Genet.* 43 (2011) 768–775.
- [35] T.J. Morris et al., *Bioinformatics* 30 (2014) 428–430.
- [36] Y. Assenov, F. Müller, P. Lutsik, et al., *Nature Methods*, in press.
- [37] E.A. Houseman et al., *Bioinformatics* 30 (2014) 1431–1439.
- [38] J. Fortin et al., *F1000Res.* 3 (2014) 175, <http://dx.doi.org/10.12688/f1000research.4680.2>.
- [39] M. Irterson et al., *Bioinformatics* (2014) 1–3, <http://dx.doi.org/10.1093/bioinformatics/btu566>.
- [40] M.E. Price et al., *Epigenetics Chromatin* 6 (2013) 4.
- [41] Y.A. Chen et al., *Epigenetics* 8 (2013).
- [42] J. Roessler et al., *BMC Res. Notes* 5 (2012) 210.
- [43] L. Butcher, T. Morris, *Illumina450ProbeVariants.db*: annotation package combining variant data from 1000 Genomes Project for Illumina HumanMethylation450 Bead Chip probes. R package version 1.0.0.
- [44] G.P. Consortium et al., *Nature* 491 (2012) 56–65.
- [45] T. Peters, M. Buckley, *DMRcate*: Illumina 450K methylation array spatial analysis methods, R package version 1.0.2. (2014).
- [46] D. Wu et al., *PLoS One* 8 (2013) e74275.
- [47] B.S. Pedersen et al., *Bioinformatics* 28 (2012) 2986–2988.
- [48] R. Lowe, V.K. Rakyan, *BMC Bioinformatics* 14 (2013) 359.
- [49] C.D. Warden et al., *Nucleic Acids Res.* 41 (2013) e117.
- [50] H. Wickham, *Elegant Graphics for Data Analysis*, Springer, New York, 2009.
- [51] F. Hahne, S. Durinck, R. Ivanek, A. Mueller, S. Lianoglou, G. Tan, *Gviz*: Plotting data and annotation information along genomic coordinates. R package version 1.8.4.
- [52] J. Guintivano et al., *Mol. Psychiatry* 19 (2014) 560–567.
- [53] J. Zou et al., *Nat. Methods* 11 (2014) 309–311.
- [54] S. Horvath, *Genome Biol.* 14 (2013) R115.
- [55] M.J. Booth et al., *Science* 336 (2012) 934–937.
- [56] M. Yu et al., *Cell* 149 (2012) 1368–1380.
- [57] P. Chopra et al., *BMC Genomics* 15 (2014) 131.
- [58] S.K. Stewart et al., *Methods* 72 (2015) 9–15, <http://dx.doi.org/10.1016/j.jymeth.2014.08.009>.
- [59] N.C. Wong et al., *Genomics* 102 (2013) 38–46.
- [60] D. Wang et al., *Bioinformatics* 28 (2012) 729–730.
- [61] P. Du et al., *Bioinformatics* 24 (2008) 1547–1548.
- [62] F. Wessely, R.D. Emes, *Front. Genet.* 3 (2012) 161.