CHAPTER

12

Best practices for epigenome-wide DNA modification data collection and analysis

Joseph Kochmanski[a,b] **and Alison I. Bernstein**[c,d]

[a]*Rancho BioSciences, San Diego, CA, United States,* [b]*Department of Translational Neuroscience, Michigan State University, East Lansing, MI, United States,* [c]*Department of Pharmacology and Toxicology, Rutgers University, Piscataway, NJ, United States,* [d]*Environmental and Occupational Health Science Institute, Rutgers University, Piscataway, NJ, United States*

Introduction

In response to a series of analyses that revealed very low reproducibility of studies in life sciences, many initiatives have been launched to assist, establish, and improve standards for rigor and reproducibility in life sciences research [1]. In 2012, Landis and colleagues published a core set of guidelines to establish standards for reporting methods and data [2]. These include standards for randomization, blinding, sample-size planning, and data handling. In response to these guidelines, the FAIR Guiding Principles for data management and stewardship were published to clearly define good data management practices [3], and new repositories have been launched to facilitate sharing of data, analysis methods, and codes, including https://fairsharing.org [4]. In addition, as discussed later in this chapter, preregistration is becoming more common to prevent statistical negligence [5]. These issues are particularly relevant to the field of epigenetics, where adherence to these principles is made more difficult by the underlying biology of epigenetic marks, rapidly evolving methods, and statistical challenges. As this field continues to grow, it is critical that researchers adhere to existing recommendations as they plan, execute, and report their studies. In this chapter, we provide an overview of a specific class of epigenetic marks, DNA modifications. We discuss why these marks are important, how to measure them, and best practices for ensuring rigor and reproducibility in DNA modification research, with a focus on epigenome-wide association studies (EWAS).

DNA modifications

Epigenetics generally refers to a set of mechanisms that affect gene expression without altering the DNA sequence, including DNA modifications, histone modifications, noncoding RNAs, and alterations in chromatin accessibility [6–8]. The initiation and maintenance of marks play critical roles in many aspects of mammalian development, including stem cell self-renewal, stem cell differentiation,

Rigor and Reproducibility in Genetics and Genomics. https://doi.org/10.1016/B978-0-12-817218-6.00006-1

and neurodevelopment [9–11]. Dysregulation of epigenetic mechanisms has been identified in a variety of disorders and it is thought that these mechanisms, in particular DNA modifications, are especially important for the development and function of the central nervous system (CNS) [10–16]. In addition, epigenetic marks are sensitive to the environment, established during cellular differentiation, and regulate gene expression throughout the lifespan [17,18]. Given these unique traits, the epigenome has been recognized as a mediator of the relationship between environmental exposures and disease [6,19].

Over the past two decades, a rapidly growing number of epigenome-wide association studies have investigated the associations between DNA modifications and disease. As this field has exploded, researchers have begun to realize the need to ensure rigor and reproducibility in EWAS, specifically related to the biological features of DNA modifications, as well as the technical and statistical aspects of data analysis. In this chapter, we provide a brief overview of DNA modifications and the methods used to measure them and then focus on issues related to rigor and reproducibility in EWAS.

DNA modifications are covalent modifications of DNA that do not alter the genetic sequence itself. The most widely studied is methylation of the fifth position of cytosine (5-methylcytosine, 5-mC), which most commonly occurs at cytosine-phospho-guanine (CpG) dinucleotides. However, DNA methylation can also occur at non-CpG sites, including cytosine-phospho-adenine (CpA), cytosine-phospho-cytosine (CpC), and cytosine-phospho-thymine (CpT) [15]. Methylation of cytosines is catalyzed by DNA methyltransferase (DNMT) enzymes, which either carry out de novo methylation or maintain methylation through DNA replication [20]. Once established, 5-mC is recognized by methylation-specific DNA-binding proteins, including methyl-CpG binding protein 2 (MeCp2) and the methyl-CpG binding domain proteins 1-6 (MDB1-6) [21,22]. DNA methylation has been demonstrated to play an important role in transcriptional regulation, maintenance of chromatin structure, genomic imprinting, X-chromosome inactivation, and genomic stability [9,23,24].

While 5-mC is a fairly stable modification, it can be enzymatically oxidized to 5-hydroxymethylcytosine (5-hmC) and eventually to 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) through the activity of ten-eleven translocation (Tet) enzymes [25–36] (Fig. 12.1).

DNA modifications in health and disease

In the past two decades, an explosion of research has investigated the potential role of DNA modifications in human health and disease. These studies have documented clear associations between DNA modifications and disease risk in multiple contexts [37–41]. This is most well-established in the cancer field, such that for some cancers, research findings have been translated to clinical practice using DNA modifications as clinical biomarkers and treatments [42–49].

DNA modifications are particularly relevant to work on neurodegenerative diseases [38,50–52], as a growing body of research supports a critical role for DNA modifications, particularly 5-hydroxymethylcytosine (5-hmC), in neurological function and in the response to environmental exposures [38,40,53–55]. The specific enrichment and the temporal and the spatial patterns of 5-mC and 5-hmC in the mammalian brain suggest that these modifications are critical for brain development and function [32,56–58]. Changes in DNA methylation and hydroxymethylation have been associated with a range of neurodevelopmental and neurological disorders, including but not limited to Rett syndrome,

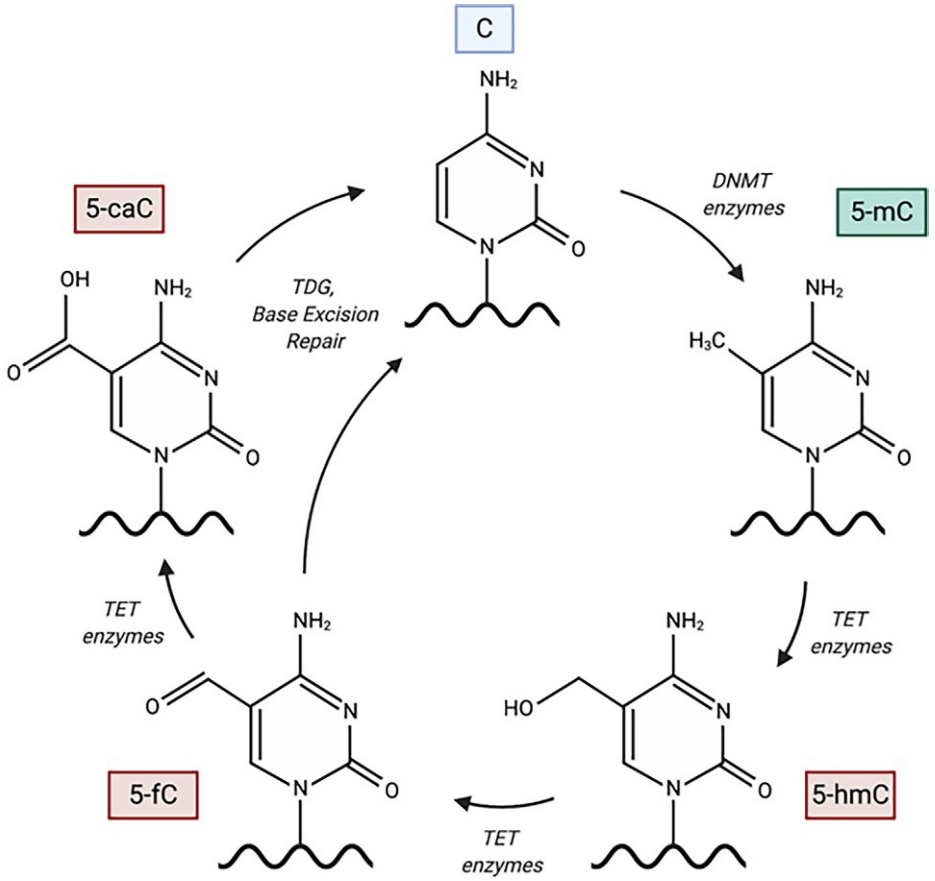

FIG. 12.1

DNA modification pathways. Cytosine (C) is methylated to 5-methylcytosine (5-mC) through activity of DNA methyltransferase (DNMT) enzymes. From there, 5-mC can be further oxidized to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC), and 5-carboxylcytosine (5-caC) through the activity of ten-eleven translocation (TET) enzymes. 5-fC and 5-caC can then be converted back to an unmodified cytosine through thymine DNA glycosylase (TDG)-mediated base excision repair.

autism spectrum disorders, schizophrenia, Alzheimer's disease, Huntington's disease, and Parkinson's disease [38,40,48,52,59–64].

In addition, the epigenome is recognized as a potential mediator of the relationship between exposures and disease. Epigenetic marks are sensitive to the environment, established during cellular differentiation, and regulate gene expression throughout the lifespan [17,18]. Of note, epigenetic marks can change in response to lifestyle changes, diet, and environmental exposures, both during early development and into adulthood [6,65–72]. Given this, DNA modifications have emerged as an important area of research in understanding the mechanisms of disease etiology for a wide range of human diseases.

Methods for detection of DNA modifications

Over the past two decades, studies have interrogated genome-wide DNA modifications using multiple methods to measure modifications. These methods operate at three different genomic resolutions: global, regional, and base-pair. The pros and cons of these methods are comprehensively reviewed [73–75]. In general, the field is moving away from global and regional methods, as they generate data that is difficult to interpret in functional terms. Given this trend, for purposes of this chapter, we will focus only on the widely used base-pair resolution methods that are suitable for EWAS. EWAS are used to find associations between genome-wide epigenetic marks and phenotypes, much as genome-wide association studies are used to identify SNPs associated with phenotypes.

Bisulfite conversion methods

Base-pair resolution methods quantify DNA modifications at individual cytosine bases. Traditionally, these methods have relied on the bisulfite (BS) conversion of DNA. BS deaminates unmodified cytosines, 5-caC, and 5-fC to uracil, but not to 5-mC or 5-hmC [76–78]. Converted uracils are then read as thymine during PCR amplification, sequencing, or detection on a methylation array. From this data, the ratio of cytosines to thymines at a given cytosine is used to estimate a beta value, which represents the proportion of cytosines at a given locus that are modified. All of the methods described below produce beta values as their output; in EWAS, these beta values are then compared between groups.

The gold standard method in the field is whole-genome bisulfite sequencing (WGBS), which pairs BS conversion with next-generation sequencing to provide single base-pair resolution beta values at cytosines across the entire sequenced genome [79–81]. While WGBS produces comprehensive data, cost remains a huge barrier for most labs, even with declining sequencing prices [82]. Two common genome-wide alternatives with lower costs are reduced representation bisulfite sequencing (RRBS) and Illumina methylation arrays.

RRBS reduces the number of nucleotides sequenced by incorporating a methylation-insensitive restriction enzyme digestion step prior to BS conversion [83–85]. Typically, *Msp*I is used, which cuts at C/CGG recognition sites, regardless of the methylation status. When combined with size selection of restriction fragments, this generates a "reduced representation" of the genome, enriching for fragments containing potentially modified cytosines. RRBS data typically produces beta value data for approximately 2–4 million cytosines.

Illumina DNA methylation arrays combine sodium bisulfite treatment, whole-genome amplification, and sequence-specific probe hybridization using BeadChip technology to estimate DNA methylation at a preselected set of cytosines across the genome [86]. The current version of this technology, the Illumina Methylation EPIC array, covers over 850,000 CpG sites. Unfortunately, as of this writing, these arrays are only available for human samples, so their application to animal models is limited. Readers interested in detailed comparisons of RRBS, methylation arrays, and WGBS should refer to publications on the topic [74,75,87].

Another, newer option for reducing cost is capture hybridization [88], which allows for targeted selection of specific sequences prior to BS conversion and sequencing. A disadvantage of this method is that it requires prior knowledge of which regions to select, and has not been widely adopted as of this writing.

A major drawback of BS-based methods is that BS treatment does not differentiate 5-mC from 5-hmC [89]. Both 5-mC and 5-hmC are protected from standard BS-mediated deamination and read the same in BS-based sequencing or array output [90].

Alternatives to BS-based methods

Oxidative BS (oxBS) conversion allows for the detection of only 5-mC without 5-hmC [91]. In oxBS, the first step is the selective, chemical-mediated oxidation of 5-hmC to 5-fC by potassium perruthenate ($KuRO_4$). The oxidized 5-fC is then sensitive to BS treatment and is converted like an unmodified cytosine in a BS conversion step to uracil [91–96].

In addition to comeasurement of 5-mC and 5-hmC, methods have also been developed to address the high input requirements for BS-based methods and to allow interrogation of DNA modifications from low amounts of input DNA. One of these methods, enzymatic methyl-sequencing (EM-Seq), has been commercialized in the NEBNext Enzymatic Methyl-seq Kit (New England BioLabs). This is a two-step enzymatic conversion with no chemical conversion step, allowing for successful detection of DNA modifications from low input DNA [97,98]. In this method, TET2 and T4-phage β glucosyltransferase (T4-βGT) convert 5-mC and 5-hmC to 5-caC and 5-(β-glucosyloxymethyl) cytosine (5-gmC), respectively. Then APOBEC3A deaminates unmodified cytosines to uracils. EM-seq output mimics the output of BS conversion (Fig. 12.2).

Other methods, like apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A (APOBEC)-coupled epigenetic sequencing (ACE-seq), and TET-assisted pyridine borane sequencing (TAPS), also remove the BS conversion step. Unlike EM-seq, these alternative methods each generate slightly different output, detecting either only 5-hmC (ACE-Seq) or both marks (TAPS) (summarized in Refs. [90,99].

Challenges to reproducibility in DNA modification EWAS research

While methodological advancements provide exciting opportunities in the field, the unique combination of complicated biology and the massive scale of modern sequencing data present a number of challenges for EWAS. These challenges fall into several categories: biology, methodology, statistics, experimental planning, and data and method reporting. Overcoming these challenges for epigenome-wide data hinges upon rigorous and reproducible methodology and analysis. Here, we highlight the challenges listed above and present some solutions for addressing them in EWAS research.

Biology

Tissue heterogeneity and cell-type specificity

A major challenge standing in the way of producing rigorous, reproducible epigenome-wide results is the fact that different cell types exhibit distinct epigenetic profiles [100–102]. As a result, studies that use tissue comprised of a mixture of cell types introduce noise into their data, washing out potentially important results with unwanted variability. Recent work has begun to address this issue using both experimental and computational tools [103,104]. For example, bioinformatics tools have been developed to estimate cell type composition from reference, bulk-tissue epigenetic data, allowing for correction during statistical analysis [105]. Some methods even allow for reference-free cell type estimation, enabling researchers to estimate

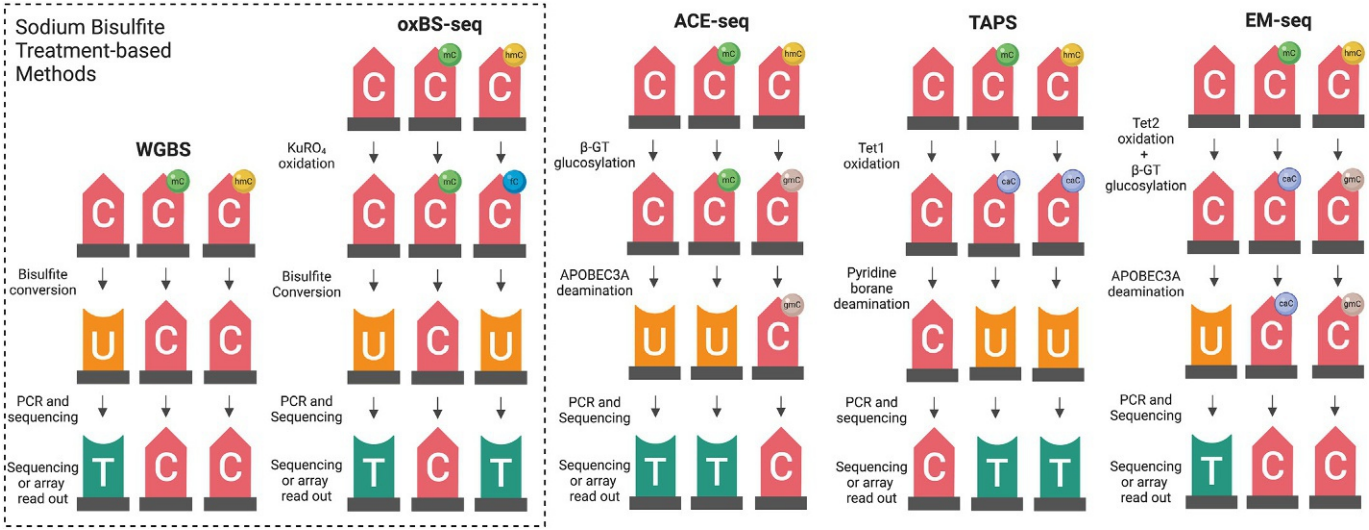

FIG. 12.2

Conversion methods to measure base-pair resolution DNA modifications. Over the past two decades, the gold standard method to measure base-pair resolution DNA methylation has been whole-genome bisulfite sequencing (WGBS). While reliable, this method involves a sodium bisulfite conversion step, which damages DNA, and does not distinguish between 5-mC and 5-hmC. To avoid these issues, researchers have used a variety of different enzymatic and chemical treatments to develop improved methods. In some cases, these methods allow for specific measurement of either 5-mC (oxBS-seq) or 5-hmC (ACE-seq), or simply improve on the damaging effects of sodium bisulfite conversion through the use of enzymatic deamination via APOBEC3A enzymes (TAPS and EM-seq). In the diagram, different cytosine modifications are indicated by colored bubbles; *green*, 5mC; *yellow*, 5-hmC; *blue*, 5-fC; *purple*, 5-caC; *gray*, 5-(β-glucosyloxymethyl)cytosine (5-gmC). Abbreviations: *WGBS*, whole-genome bisulfite sequencing; *oxBS-seq*, oxidative bisulfite sequencing; *ACE-seq*, APOBEC-coupled epigenetic sequencing; *TAPS*, TET-assisted pyridine borane sequencing; *EM-seq*, enzymatic methyl-sequencing; *KuRO4*, potassium perruthenate; *PCR*, polymerase chain reaction; β-*GT*, b-glucosyltranferase; *Tet1/2*, ten-eleven translocase ½; *APOBEC3A*, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A.

cell types without reference data from their tissue of interest [106,107]. Meanwhile, at the benchtop, cell sorting methods, including flow-assisted cell sorting and magnetic-assisted cell sorting [108,109] can be used to enrich specific cell populations prior to isolating DNA for downstream epigenetic follow-up.

Another enrichment method is laser capture microdissection (LCM), which allows for targeted dissection of individual cells or tissue regions prior to DNA isolation and measurement of DNA modifications [110]. The LCM method is less applicable to genome-wide methods that require high DNA inputs, as is often the case for DNA modifications, but as technology improves, this technology will become more feasible to pair with EWAS. These types of laboratory-based enrichment tools are critical for reducing random variability in EWAS studies, allowing for not only improved biological interpretability but also increased reproducibility. The issue of cell-type heterogeneity is well-recognized in epigenetics research, and in an ideal world, all EWAS studies would include techniques to isolate specific cell populations for analysis. However, universal implementation of cell enrichment approaches will require improved compatibility with low amounts of input DNA. Furthermore, to ensure reproducibility, researchers must not only report their cell type enrichment and/or estimation methods, but also share any code or protocols related to these efforts, as they will benefit the entire field.

Temporal changes in DNA modifications

DNA modifications can change with age [111–115], and preliminary evidence suggests that early-life exposures can modify rates of epigenetic aging [116]. As a result, the age at which tissue collection takes place could affect DNA modification signatures, complicating interpretation of EWAS results. Luckily, there are some simple ways to account for this variability during experimental planning:

- In human studies, researchers can control for the effects of age on DNA modifications by either age-matching human samples or including age as a covariate in statistical modeling.
- In animal studies, where greater experimental control is available, animals should always be age-matched during tissue collection.

Sex-related changes in DNA modifications

In addition to age, sex affects levels of DNA modification across the genome [117–119]. Of note, evidence shows that sex is not only associated with differences in baseline DNA methylation at specific genes, but also the epigenetic response to the environment [51,120–123]. As a result, studies in humans and animals that include both males and females must account for sex in their analysis. This can be done in a similar fashion to age:

- In human studies, researchers can control for the effects of sex on DNA modifications by either sex-matching human samples or including sex as a covariate in statistical modeling.
- In animal studies, where there is greater experimental control, statistical analyses should be stratified by sex.

Methodology

Another major challenge facing the field is the ever-changing landscape of methods to interrogate the epigenome [73,74,87,124]. Technologies are changing and updating so quickly that grants often include methods that are outdated by the time grants are funded. In addition to new conversion methods,

there has also been a flood of new, improved bioinformatics tools for data processing and analysis [125–127]. While these new methods and tools present exciting new possibilities in the field, the selection of the correct analysis pipeline is not always clear, as there are subtle differences between many of the statistical approaches employed by these methods.

There are also differences in the structure and scale of EWAS data that affect downstream statistical inference. A useful example is a comparison of the analysis pipelines for RRBS data and Illumina EPIC array data. These two methods cover similar numbers of CpG sites across the genome (RRBS: ~2–4 million CpGs; EPIC array: ~850,000 CpGs). At first glance, one might think that similar methods could be used to analyze both of these datasets. However, due to methodological differences, the pipelines used to analyze RRBS and EPIC data are vastly different. RRBS produces sequencing reads in the form of .fastq files, which must be aligned to the genome using specific software settings prior to analysis [128]. Meanwhile, the EPIC array produces probe-based red and green channel.idat files, which require a specialized quality control, processing, and normalization pipeline [129]. Furthermore, when running statistical tests on these large datasets, the effect of controlling for multiple testing will vary based on the number of sites being analyzed. As a result, one chromosomal position that shows a false discovery rate < 0.05 in an EPIC dataset may not in an RRBS dataset, even if the effect size and sample size are the same. This is because more comparisons are being performed in the RRBS dataset, therefore limiting the power to detect changes when correcting for multiple hypothesis testing. These methodological and statistical differences between EWAS datasets lead to a lack of reproducibility in the field. While some of these differences are inherent to a study's data structure, epigenetics researchers must work to establish standard, method-specific pipelines for genome-wide 5-mC and 5-hmC data analysis. Readers interested in the specific recommendations regarding methodology in EWAS research, including ways to avoid bias in tissue processing and statistical analysis, can find additional information in a number of recent articles [82,130–133]. By incorporating existing methodological guidelines, the field of epigenetics research could improve its rigor and reproducibility, overcoming the uncertainty that has plagued its data for decades.

Statistics

Sample size planning and power analysis

The most important statistical consideration for any EWAS is sample-size planning. This process should be performed during the initial planning stages of an experiment, and involves analysis of statistical power, or the probability that a statistical test will correctly reject the null hypothesis.

Statistical power analysis requires four main parameters:

1. Estimated effect size.
2. Estimated data variability.
3. Sample size.
4. Selected alpha significance level (e.g., $P < .05$), and the desired statistical power (e.g., 0.8)

As such, for a scientist to estimate the required sample size for an experiment at a given power, they will need to input their estimated effect size, data variability, and alpha significance level. This is a foundational power analysis as it is typically applied to a simple statistical test; for example, a Student's T-test.

However, when analyzing genome-wide data, researchers are typically conducting tests on thousands of gene regions at once. In this larger-scale scenario, power analysis must consider additional parameters, including the false discovery rate to correct for multiple testing, the distribution of read counts, and the dispersion coefficients of analyzed genes [134]. To address these additional issues, recent work has used existing RNA-sequencing data as a frame of reference to create bioinformatics tools to perform power analysis for EWAS. Examples include the *RNASeqPower* [134], *PROPER* [135], *powsimR* [136], and *RnaSeqSampleSize* [137], among others. The sequencing data produced in RNA-sequencing experiments are similar in structure to those generated by RRBS or WGBS, meaning the tools developed for RNA-seq experiments can provide some guidance for EWAS studies. However, the data dispersion and statistical assumptions used to analyze RRBS or WGBS data are not the same as those used for RNA-seq, meaning it would be better to use specialized techniques.

Unfortunately, as of this writing, there are only a few existing tools to specifically performs power analysis on WGBS or Illumina Beadchip array data. For WGBS data, power analysis can be performed using the *Bisulfite sequencing simulator/power analysis* shiny app developed by the Tung Lab [138]. Meanwhile, for Illumina Beadchip arrays, there are two existing tools in the literature: *pwrEWAS*, an R shiny app [139], and an interactive web application specifically for conducting power calculations for Illumina EPIC array datasets [130]. Along with this web application, the authors of the second tool also provide a suggested *P*-value significance cutoff for EPIC array data that accounts for both multiple testing and correlation between probes: $P < 9 \times 10^{-8}$. They also provide recommendations regarding sample size for studies that utilize this methodology. Based on their simulations, at a sample size of N=200 (n=100 case, n=100 control), approximately 85% of analyzed sites will have >80% power to detect a beta value difference of 0.05 or greater [130]. Utilization of the highlighted sample size planning tools is critical for ensuring statistical rigor in future EPIC array studies.

Human epidemiological studies

Human epidemiological cohorts are heterogeneous and provide little control over environmental and lifestyle factors that can affect the epigenome, including diet, stress, chemical exposures, and physical activity level [18,69,70,140–144]. As such, it can be difficult to determine the effect of a single variable, e.g., exposure to a toxicant, on DNA modifications in a human cohort. Generally, epidemiological studies attempt to control for potential confounders by including them as covariates in statistical modeling. In this way, it is possible to account for the heterogeneity of human populations while examining DNA modifications as a primary outcome measure. There is certainly value in this approach, but variable selection must be performed carefully. Adding too many additional covariates can inflate model estimates and lead to multicollinearity, thereby impacting the ability of a model to accurately detect significant changes [145]. Of note, these issues are exacerbated by small sample sizes, as often occurs in EWAS research. As a result, careful model building is incredibly important for producing reliable results for human EWAS. Strategies for variable selection are varied, but two of the most common approaches are testing the significance of covariate coefficients or testing whether the magnitude of changes in the outcome due to covariate inclusion are biologically meaningful (e.g., above a predetermined threshold) [145]. Unfortunately, specific recommendations are difficult to come by, as selection of covariate inclusion criteria will vary based on study design and sample size. Despite this uncertainty, it should be emphasized that EWAS investigators must always consider variable selection during model building, as inclusion/exclusion of covariates will have a significant effect on results.

Hypothesis testing in EWAS

While rigorous and reproducible analysis of epigenome-wide datasets is certainly possible, it is not simple or straightforward, and epigenome-wide association studies from different labs typically use very different methods to test their hypotheses, making it difficult to reconcile data between groups. As of now, there are some existing guidelines from the NIH Roadmap Epigenomics Program regarding methodology, data reporting, and analysis for epigenome-wide association studies [146]. Unfortunately, these guidelines provide only general suggestions regarding data handling and reporting, e.g., investigators report all data handling steps prior to read mapping, and do not give researchers guidance regarding specific analysis parameters. Other large epigenomics consortia, including the NIEHS TaRGET II Consortium [147], provide protocol files for their bioinformatics pipelines, which is extremely helpful for the field. However, the establishment of best practice guidelines for EWAS analysis is not straightforward and represents an area of continuing development.

A major issue in establishing standardized analysis practices is that both 5-mC and 5-hmC beta values are heteroscedastic in their data distributions, with most cytosines showing either very high or very low beta values [96]. Given this unique distribution, statistical tests to test for differential DNA modifications must utilize statistical methods that can account for this unique beta value distribution. Although the field has not settled on a single best practice to address this heteroscedasticity, there are several methods that are commonly used during hypothesis testing in EWAS.

1. *Logit transformation*
   - **a.** One of the most common method uses logit transformation to generate *m*-values from the beta values [148]. Under this approach, *m*-values exhibit a homoscedastic distribution and can be analyzed using standard linear regression methods.
   - **b.** This type of analysis is often implemented in pipelines that rely upon the *limma* and *BSmooth* R packages [127]. Importantly, linear regression requires logit transformation to *m*-values prior to modeling because linear regression assumes homoscedasticity.
   - **c.** A weakness of this method is that *m*-values are less interpretable than beta values and are not recommended for data visualization [148,149]. As such, we only recommend using this method, if investigators have no intention of visualizing their data after detection of differential DNA modifications.
2. *Regression modeling*
   - **a.** There are several analysis approaches that do not rely upon the generation of *m*-values; these include beta regression, rank-based regression, and beta-binomial regression modeling [138,150–152].
   - **b.** These types of alternative regression methods, which are more appropriate for analyzing raw beta values, are implemented in a number of existing R packages, including *DSS, BiSeq, MethylSig*, and *MOABS*, among others [127].
   - **c.** The exact choice of method will depend on data distribution and the desired modeling flexibility, but in general, we recommend using one of these alternative regression-based methods when investigators want to keep their untransformed beta values for improved data interpretation and visualization (Box 12.1).

Even when using these more appropriate analysis methods, the validity and sensitivity of genome-wide analyses of DNA modifications can break down when test assumptions do not match data distributions. **To ensure that investigators are using the most appropriate methods for their data, they should**

Box 12.1 Heteroscedasticity

Although, it sounds like a difficult concept to understand, heteroscedasticity (or "different dispersion") is actually relatively simple. It refers to data in which the variability of an outcome variable (i.e., DNA modification beta value) is unequal across the range of values for a second, independent variable (i.e., experimental treatment). Homoscedasticity, meanwhile, refers to data in which the variability of an outcome variable is equal across the range of values for a second variable. Linear regression assumes homoscedastic data, but DNA modification data often shows inherently unequal variance across its own distribution, with the majority of values close to 0 or 1. As such, it is important to account for the potential effects of heteroscedasticity when applying regression methods to DNA modification data.

always double-check their modeling output to determine the rigor of their experiment. For example, when conducting a large number of hypotheses tests in EWAS, the authors recommend visualization of the *P*-value distribution to diagnose potential issues with error rates [153]. By comparing an experimental *P*-value distribution shape to the expected uniform distribution under the null hypothesis, it is possible to determine whether an analysis is problematic. When *P*-value histograms do not show a peak near zero and/or broadly deviate from uniformity, investigators should consider adjusting their model, stratifying data to control for uncontrolled confounders, or revisiting power analysis to determine whether their data is adequately testing the question of interest. While it is difficult to provide specific recommendations given that each experiment will vary in its hypothesis testing, we strongly recommend examining *P*-value histograms for all EWAS research.

Reconciling paired 5-mC and 5-hmC data

Recent technological advances have made it possible to measure both 5-mC and 5-hmC from the same tissues, providing opportunities to tease apart associations between these two marks and gene regulation. However, there are not yet standard statistical methods for co-analyzing 5-mC and 5-hmC data. Recent studies have either examined the distribution of 5-hmC alone across the genome [93,154,155] or treated 5-mC and 5-hmC as independent outcomes, analyzing each as a separate epigenetic mark [94,156,157]. While these approaches are not wrong, they ignore the biological interdependence of 5-mC and 5-hmC.

In a recent publication, the authors proposed a new statistical approach for reconciling combined 5-mC and 5-hmC data [96]. In this streamlined analysis pipeline, rather than treating 5-mC and 5-hmC as separate outcomes, a mixed-effects (ME) modeling approach is proposed that treats the two epigenetic marks as “repeated” measures of a single outcome variable. An interaction term is used to determine whether the relationship between the outcome (i.e., beta value) and experimental treatment varies by DNA modification. One limitation of this approach is that it is specific to sites with measurable beta values for both 5-mC and 5-hmC; for those sites that show only 5-mC or 5-hmC, separate analyses for each mark are still needed. Implementation of mixed modeling for 5-mC and 5-hmC data is still in its infancy, and as of now, there is no specialized software to perform only this analysis method. However, it can be achieved using mixed-effects regression modeling in the existing *gamlss* R package [158]. To improve the co-analysis of paired 5-mC and 5-hmC data, future bioinformatics tools should work to streamline this process, implementing a mixed-effects modeling approach on a genome-wide scale into an R package or shiny application. We also hope that other groups will take up this challenge and develop additional methods for coanalyzing 5-mC and 5-hmC data.

Experimental planning & reporting: Methods, code, and data

Preregistration

Another way to improve experimental planning in general is through preregistration. This is particularly important in epigenetics, as methodologies and analysis pipelines are continually changing. **Preregistration refers to the process of defining one's research questions and proposed analyses prior to data collection** [5]. By preregistering a research project, it is possible to distinguish data generated as part of predictive hypothesis testing as opposed to exploratory analyses.

A variety of services are available online to preregister studies; as one example, the Open Science Framework (OSF) is a free, open-access service that provides workflows for preregistration [159]. This service is maintained by the Center for Open Science, a nonprofit organization committed to improving rigor and reproducibility in scientific research. By completing a preregistration, researchers can avoid hindsight bias and *P*-value hacking, common analysis pitfalls that can occur when working with large-scale datasets that present many possible avenues for statistical analysis.

In vivo animal colony reporting

A crucial component of in vivo EWAS studies is careful reporting of animal breeding schemes. Many EWAS studies utilize animal models to examine the link between developmental exposures and the later-life epigenome in offspring. This type of intergenerational and/or transgenerational research is critical to improving our understanding of the Developmental Origins of Health and Disease (DOHaD) hypothesis [19,160,161], but the field has been hampered by a lack of reproducibility [162]. One of the major reasons for the poor reproducibility in these types of studies is a lack of transparency regarding animal breeding schemes. In particular, many studies fail to account for litter effects [163,164].

To address these persistent deficiencies, researchers recently released the animal research: reporting of in vivo experiments (ARRIVE) guidelines, a checklist of items to include in animal model publications [165]. These guidelines include reporting suggestions for study design, sample size, inclusion/exclusion criteria, randomization, and more. It is recommended that investigators utilize these guidelines both during experimental planning and reporting. By adhering to these criteria for publication, scientists who use animals in their EWAS research can improve the transparency of their projects, thereby increasing rigor and reproducibility.

Sharing data through repositories

Another major area of continued improvement is in data publication. Recently, there has been a significant push toward data publication in public, online repositories. Some well-known examples include the Gene Expression Omnibus (GEO), Sequence Read Archives (SRA), and BioStudies, which are the repositories maintained by funding agencies. More recently, independent repositories have been created, including Dryad and Mendeley Data. Specific to epigenetics data, the National Center for Biotechnology Information (NCBI) recently established an Epigenomics database [166]. **Given these readily available resources, no published EWAS should ever use the term "data not shown."** Rather, all data, both raw and processed, should be provided in an online repository as part of a manuscript's publication. Furthermore, all data tables resulting from statistical analysis should be shared as supplementary tables. This will ensure that all aspects of a study's results are reported, allowing for verification of findings by other researchers.

Code sharing

Given the evolving EWAS methodology and analysis tools highlighted earlier in this chapter, **code sharing is critical for ensuring rigor and reproducibility** in EWAS. Many research groups use their own custom pipelines to analyze data, yet they do not publish code alongside their publications. There is nothing inherently wrong with using an in-house analysis pipeline, especially given the number of choices and methods available, but doing so limits the interoperability of data, decreasing the potential for shared procedures and results. In essence, an anticollaborative culture is created not by intention but by necessity. Groups only know how to process their own data, which reduces the chances for cooperation that might spur innovation in the field.

To combat this trend, it is critical for researchers to share their code along with their publications so that their analysis is clear. When shared, code should be usable and editable; preferably in a format that can be easily rerun. Related to this, it is recommended that researchers avoid simply sharing raw code, and instead provide annotations or comments that explain the parameter choices that were made during initial analysis. In many cases, default software settings are not used, so the selection of specific parameters must be clear for users. We also recommend that all quality control figures and metrics produced in data processing be shared publicly. The exact implementation of these recommendations will vary based on the selected analysis software, but they could be implemented through the Github platform or through the use of R Notebooks. This will increase a future user's ability to recreate past analyses, thereby improve the reproducibility of a bioinformatics pipeline.

FAIR principles for data sharing

Data repositories and code sharing are certainly steps in the right direction, but data files and code provided by the authors are often difficult to interpret, with obtuse, nondescriptive file names, and little annotation available to recreate analyses. In response to this lack of standardization, a diverse group of scientists recently came together to design the **FAIR Data Principles**, which act as a set of guidelines for improving the reusability of data after publication [3]. Under these principles, publicly hosted data are suggested to be four things: Findable, Accessible, Interoperable, and Reusable (FAIR). Under this framework,

Findable—refers to registration in a searchable resource along with rich metadata, including a unique identifier variable.
Accessible—refers to the data being retrievable through a free, user-friendly communications portal or protocol.
Interoperable—refers to the inclusion formal, shared language (e.g., code), clear references to related data, and a consistent vocabulary. Finally,
Reusable—refers to providing a rich metadata and linking data to a usage license.

By adhering to these principles, scientists can not only increase the believability of their results, but also provide clear opportunities for rigorous data validation.

Existing data repositories in the life sciences have variable implementations of FAIR principles, putting much of the responsibility for accurate and complete data reporting on users [167,168]. Furthermore, the lack of standardized language in these repositories can make it difficult to utilize publicly available data. For example, specific information regarding experimental samples may be lacking, including tissue heterogeneity, age, and collection methods. To help investigators identify their best options for data sharing, https://fairsharing.org is a publicly available resource that provides data

and metadata standards as well as ratings for existing repositories [4]. To improve the interoperability of public EWAS data, investigators must aim to adhere to FAIR principles when submitting their data to a repository.

Conclusion

In this chapter, we describe a number of existing methods for analyzing DNA modification data. We also presented challenges related to modern sequencing datasets, including biology, methodology, statistics, experimental planning, and data and method reporting. Overcoming these challenges for epigenome-wide data hinges on rigorous and reproducible methodology and analysis. Related to this point, we highlighted a number of specific recommendations for EWAS research.

To address the biological heterogeneity of the epigenome across cell types, we recommend cell type enrichment methods (e.g., flow-assisted cell sorting) and/or computational estimation. For statistical analysis, we recommend specific tools for analysis planning and power analysis in EWAS, including *pwrEWAS*. We also discuss covariate inclusion criteria in human EWAS, methods for addressing beta value heteroscedasticity, *P*-value histograms as a diagnostic tool, and mixed effects regression methods for coanalyzing paired 5-mC and 5-hmC data. After that, we provided recommendations regarding reporting in EWAS research. In particular, we highlight areas of potential improvement in the field: preregistration to avoid p-hacking, adherence to the ARRIVE principles for in vivo studies, data reporting via public repositories, code and parameter sharing, and the FAIR principles for data sharing.

Through careful implementation of the provided recommendations, we believe the field can begin to produce epigenome-wide association studies that are both rigorous and reproducible.

References

[1] M. Baker, D. Penny, Is there a reproducibility crisis? Nature 533 (2016) 452–454, https://doi.org/10.1038/533452A.

[2] S.C. Landis, S.G. Amara, K. Asadullah, C.P. Austin, R. Blumenstein, E.W. Bradley, R.G. Crystal, R.B. Darnell, R.J. Ferrante, H. Fillit, R. Finkelstein, M. Fisher, H.E. Gendelman, R.M. Golub, J.L. Goudreau, R.A. Gross, A.K. Gubitz, S.E. Hesterlee, D.W. Howells, J. Huguenard, K. Kelner, W. Koroshetz, D. Krainc, S.E. Lazic, M.S. Levine, M.R. MacLeod, J.M. McCall, R.T.M. Iii, K. Narasimhan, L.J. Noble, S. Perrin, J.D. Porter, O. Steward, E. Unger, U. Utz, S.D. Silberberg, A call for transparent reporting to optimize the predictive value of preclinical research, Nature 490 (2012) 187–191, https://doi.org/10.1038/nature11556.

[3] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. Van Der Lei, E. Van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, Comment: The FAIR guiding principles for scientific data management and stewardship, Sci. Data 3 (2016), https://doi.org/10.1038/sdata.2016.18.

[4] S.A. Sansone, P. McQuilton, P. Rocca-Serra, A. Gonzalez-Beltran, M. Izzo, A.L. Lister, M. Thurston, FAIRsharing as a community approach to standards, repositories and policies, Nat. Biotechnol. 37 (2019) 358–367, https://doi.org/10.1038/s41587-019-0080-8.

[5] B.A. Nosek, C.R. Ebersole, A.C. DeHaven, D.T. Mellor, The preregistration revolution, Proc. Natl. Acad. Sci. USA 115 (2018) 2600–2606, https://doi.org/10.1073/pnas.1708274114.
[6] A.J. Bernal, R.L. Jirtle, Epigenomic disruption: the effects of early developmental exposures, Birth Defects Res. A Clin. Mol. Teratol. 88 (2010) 938–944, https://doi.org/10.1002/bdra.20685.
[7] B.E. Bernstein, A. Meissner, E.S. Lander, The mammalian epigenome, Cell 128 (2007) 669–681, https://doi.org/10.1016/j.cell.2007.01.033.
[8] G. Egger, G. Liang, A. Aparicio, P.A. Jones, Epigenetics in human disease and prospects for epigenetic therapy, Nature 429 (2004) 457–463, https://doi.org/10.1038/nature02625.
[9] E. Li, Chromatin modification and epigenetic reprogramming in mammalian development, Nat. Rev. Genet. 3 (2002) 662–673, https://doi.org/10.1038/nrg887.
[10] Z.D. Smith, A. Meissner, DNA methylation: roles in mammalian development, Nat. Rev. Genet. 14 (2013) 204–220, https://doi.org/10.1038/nrg3354.
[11] B. Yao, P. Jin, Cytosine modifications in neurodevelopment and diseases, Cell Mol. Life Sci. 71 (2014) 405–418, https://doi.org/10.1007/s00018-013-1433-y.
[12] J. Feng, G. Fan, The role of DNA methylation in the central nervous system and neuropsychiatric disorders, Int. Rev. Neurobiol. 89 (2009), https://doi.org/10.1016/S0074-7742(09. 89004-1.
[13] K. Gapp, B.T. Woldemichael, J. Bohacek, I.M. Mansuy, Epigenetic regulation in neurodevelopment and neurodegenerative diseases, Neuroscience 264 (2014) 99–111, https://doi.org/10.1016/j.neuroscience.2012.11.040.
[14] M. Jakovcevski, S. Akbarian, Epigenetic mechanisms in neurological disease, Nat. Med. 18 (2012) 1194–1204, https://doi.org/10.1038/nm.2828.
[15] H.S. Jang, W.J. Shin, J.E. Lee, J.T. Do, CpG and non-CpG methylation in epigenetic gene regulation and brain function, Genes 8 (2017) 2–20, https://doi.org/10.3390/genes8060148.
[16] A. Rudenko, L.H. Tsai, Epigenetic modifications in the nervous system and their impact upon cognitive impairments, Neuropharmacology 80 (2014) 70–82, https://doi.org/10.1016/j.neuropharm.2014.01.043.
[17] C.D. Allis, T. Jenuwein, The molecular hallmarks of epigenetic control, Nat. Rev. Genet. 17 (2016) 487–500, https://doi.org/10.1038/nrg.2016.59.
[18] C. Faulk, D.C. Dolinoy, Timing is everything: the when and how of environmentally induced changes in the epigenome of animals, Epigenetics 6 (2011) 791–797, https://doi.org/10.4161/epi.6.7.16209.
[19] T. Bianco-Miotto, J.M. Craig, Y.P. Gasser, S.J. Van Dijk, S.E. Ozanne, Epigenetics and DOHaD: from basics to birth and beyond, J. Dev. Orig. Health Dis. 8 (2017) 513–519, https://doi.org/10.1017/S2040174417000733.
[20] H. Gowher, A. Jeltsch, Mammalian DNA methyltransferases: new discoveries and open questions, Biochem. Soc. Trans. 46 (2018) 1191–1202, https://doi.org/10.1042/BST20170574.
[21] O. Bogdanović, G.J.C. Veenstra, DNA methylation and methyl-CpG binding proteins: developmental requirements and function, Chromosoma 118 (2009) 549–565, https://doi.org/10.1007/s00412-009-0221-9.
[22] Q. Du, P.L. Luu, C. Stirzaker, S.J. Clark, Methyl-CpG-binding domain proteins: readers of the epigenome, Epigenomics 7 (2015) 1051–1073, https://doi.org/10.2217/epi.15.39.
[23] S. Gopalakrishnan, B.O. Van Emburgh, K.D. Robertson, DNA methylation in development and human disease, Mutat. Res. Fundam. Mol. Mech. Mutagen. 647 (2008) 30–38, https://doi.org/10.1016/j.mrfmmm.2008.08.006.
[24] M.M. Suzuki, A. Bird, DNA methylation landscapes: provocative insights from epigenomics, Nat. Rev. Genet. 9 (2008) 465–476, https://doi.org/10.1038/nrg2341.
[25] S. Ito, L. Shen, Q. Dai, S.C. Wu, L.B. Collins, J.A. Swenberg, C. He, Y. Zhang, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, Science 333 (2011) 1300–1303, https://doi.org/10.1126/science.1210597.
[26] R.M. Kohli, Y. Zhang, TET enzymes, TDG and the dynamics of DNA demethylation, Nature 502 (2013) 472–479, https://doi.org/10.1038/nature12750.

[27] M. Tahiliani, K.P. Koh, Y. Shen, W.A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L.M. Iyer, D.R. Liu, L. Aravind, A. Rao, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1, Science 324 (2009) 930–935, https://doi.org/10.1126/science.1170116.
[28] H. Wu, Y. Zhang, Mechanisms and functions of Tet proteinmediated 5-methylcytosine oxidation, Genes Dev. 25 (2011) 2436–2452, https://doi.org/10.1101/gad.179184.111.
[29] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, 5-Hydroxymethylcytosine is a predominantly stable DNA modification, Nat. Chem. 6 (2014) 1049–1055, https://doi.org/10.1038/nchem.2064.
[30] M.A. Hahn, P.E. Szabó, G.P. Pfeifer, 5-Hydroxymethylcytosine: a stable or transient DNA modification? Genomics 104 (2014) 314–323, https://doi.org/10.1016/j.ygeno.2014.08.015.
[31] C.X. Song, C. He, Potential functional roles of DNA demethylation intermediates, Trends Biochem. Sci. 38 (2013) 480–484, https://doi.org/10.1016/j.tibs.2013.07.003.
[32] Y. Chen, N.P. Damayanti, J. Irudayaraj, K. Dunn, F.C. Zhou, Diversity of two forms of DNA methylation in the brain, Front. Genet. 5 (2014), https://doi.org/10.3389/fgene.2014.00046.
[33] T. Khare, S. Pai, K. Koncevicius, M. Pal, E. Kriukiene, Z. Liutkeviciute, M. Irimia, P. Jia, C. Ptak, M. Xia, R. Tice, M. Tochigi, S. Moréra, A. Nazarians, D. Belsham, A.H.C. Wong, B.J. Blencowe, S.C. Wang, P. Kapranov, R. Kustra, V. Labrie, S. Klimasauskas, A. Petronis, 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary, Nat. Struct. Mol. Biol. 19 (2012) 1037–1044, https://doi.org/10.1038/nsmb.2372.
[34] C.G. Spruijt, F. Gnerlich, A.H. Smits, T. Pfaffeneder, P.W.T.C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H.C. Eberl, A. Mensinga, A.B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. Van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives, Cell 152 (2013) 1146–1159, https://doi.org/10.1016/j.cell.2013.02.004.
[35] H. Stroud, S. Feng, S. Morey Kinney, S. Pradhan, S.E. Jacobsen, 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells, Genome Biol. 12 (2011), https://doi.org/10.1186/gb-2011-12-6-r54.
[36] A.A. Sérandour, S. Avner, F. Oger, M. Bizot, F. Percevault, C. Lucchetti-Miganeh, G. Palierne, C. Gheeraert, F. Barloy-Hubler, C.L. Péron, T. Madigou, E. Durand, P. Froguel, B. Staels, P. Lefebvre, R. Métivier, J. Eeckhoute, G. Salbert, Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers, Nucleic Acids Res. 40 (2012) 8255–8265, https://doi.org/10.1093/nar/gks595.
[37] M.J. Armstrong, Y. Jin, E.G. Allen, P. Jin, Diverse and dynamic DNA modifications in brain and diseases, Hum. Mol. Genet. 28 (2019) R241–R253, https://doi.org/10.1093/hmg/ddz179.
[38] Y. Cheng, A. Bernstein, D. Chen, P. Jin, 5-Hydroxymethylcytosine: a new player in brain disorders? Exp. Neurol. 268 (2015) 3–9, https://doi.org/10.1016/j.expneurol.2014.05.008.
[39] Z. Jin, Y. Liu, DNA methylation in human diseases, Genes Dis. 5 (2018) 1–8, https://doi.org/10.1016/j.gendis.2018.01.002.
[40] R. Lardenoije, E. Pishva, K. Lunnon, D.L. van den Hove, Neuroepigenetics of aging and age-related neurodegenerative disorders, in: Progress in Molecular Biology and Translational Science, Elsevier B.V, Netherlands, 2018, pp. 49–82, https://doi.org/10.1016/bs.pmbts.2018.04.008.
[41] A. Portela, M. Esteller, Epigenetic modifications and human disease, Nat. Biotechnol. 28 (2010) 1057–1068, https://doi.org/10.1038/nbt.1685.
[42] N. Ahuja, A.R. Sharma, S.B. Baylin, Epigenetic therapeutics: a new weapon in the war against cancer, Annu. Rev. Med. 67 (2016) 73–89, https://doi.org/10.1146/annurev-med-111314-035900.
[43] K. Kamińska, E. Nalejska, M. Kubiak, J. Wojtysiak, Ł. Żołna, J. Kowalewski, M.A. Lewandowska, Prognostic and predictive epigenetic biomarkers in oncology, Mol. Diagn. Ther. 23 (2019) 83–95, https://doi.org/10.1007/s40291-018-0371-7.
[44] M. Kulis, M. Esteller, DNA methylation and cancer, Spain (2010), https://doi.org/10.1016/B978-0-12-380866-0.60002-2.

[45] W.J. Locke, D. Guanzon, C. Ma, Y.J. Liew, K.R. Duesing, K.Y.C. Fung, J.P. Ross, DNA methylation cancer biomarkers: translation to the clinic, Front. Genet. 10 (2019), https://doi.org/10.3389/fgene.2019.01150.
[46] A. Nebbioso, F.P. Tambaro, C. Dell'Aversana, L. Altucci, Cancer epigenetics: moving forward, PLoS Genet. 14 (2018), https://doi.org/10.1371/journal.pgen.1007362.
[47] A. Roberti, A.F. Valdes, R. Torrecillas, M.F. Fraga, A.F. Fernandez, Epigenetics in cancer therapy and nanomedicine, Clin. Epigenetics 11 (2019), https://doi.org/10.1186/s13148-019-0675-4.
[48] M.L. Thomas, P. Marcato, Epigenetic modifications as biomarkers of tumor development, therapy response, and recurrence across the cancer care continuum, Cancer 10 (2018), https://doi.org/10.3390/cancers10040101.
[49] S. Virani, S. Virani, J.A. Colacino, J.H. Kim, L.S. Rozek, Cancer epigenetics: a brief review, ILAR J. 53 (2012) 359–369, https://doi.org/10.1093/ilar.53.3-4.359.
[50] A.I. Bernstein, Y. Lin, R.C. Street, L. Lin, Q. Dai, L. Yu, H. Bao, M. Gearing, J.J. Lah, P.T. Nelson, C. He, A.I. Levey, J.G. Mullé, R. Duan, P. Jin, 5-Hydroxymethylation-associated epigenetic modifiers of Alzheimer's disease modulate Tau-induced neurotoxicity, Hum. Mol. Genet. 25 (2016) 2437–2450, https://doi.org/10.1093/hmg/ddw109.
[51] J. Kochmanski, S.E. Vanoeveren, J.R. Patterson, A.I. Bernstein, Developmental dieldrin exposure alters DNA methylation at genes related to dopaminergic neuron development and Parkinson's disease in mouse midbrain, Toxicol. Sci. 169 (2019) 593–607, https://doi.org/10.1093/toxsci/kfz069.
[52] J. Kochmanski, N.C. Kuhn, A.I. Bernstein, Parkinson's disease-associated, sex-specific changes in DNA methylation at PARK7 (DJ-1), SLC17A6 (VGLUT2), PTPRN2 (IA-2β), and NR4A2 (NURR1) in cortical neurons, NPJ Parkinsons Dis. 8 (1) (2022) 120, https://doi.org/10.1038/s41531-022-00355-2.
[53] T. Dao, R.Y.S. Cheng, M.P. Revelo, W. Mitzner, W.Y. Tang, Hydroxymethylation as a novel environmental biosensor, Curr. Environ. Health Rep. 1 (2014), https://doi.org/10.1007/s40572-013-0005-5.
[54] O.A. Efimova, A.S. Koltsova, M.I. Krapivin, A.V. Tikhonov, A.A. Pendina, Environmental epigenetics and genome flexibility: focus on 5-hydroxymethylcytosine, Int. J. Mol. Sci. 21 (2020), https://doi.org/10.3390/ijms21093223.
[55] J. Kochmanski, A.I. Bernstein, The impact of environmental factors on 5-hydroxymethylcytosine in the brain, Curr. Environ Health Rep. 7 (2020) 109–120, https://doi.org/10.1007/s40572-020-00268-3.
[56] A. Madrid, P. Chopra, R.S. Alisch, Species-specific 5 mC and 5 hmC genomic landscapes indicate epigenetic contribution to human brain evolution, Front. Mol. Neurosci. 11 (2018), https://doi.org/10.3389/fnmol.2018.00039.
[57] H. Spiers, E. Hannon, L.C. Schalkwyk, N.J. Bray, J. Mill, 5-Hydroxymethylcytosine is highly dynamic across human fetal brain development, BMC Genomics 18 (2017), https://doi.org/10.1186/s12864-017-4091-x.
[58] L. Wen, X. Li, L. Yan, Y. Tan, R. Li, Y. Zhao, Y. Wang, J. Xie, Y. Zhang, C. Song, M. Yu, X. Liu, P. Zhu, X. Li, Y. Hou, H. Guo, X. Wu, C. He, R. Li, F. Tang, J. Qiao, Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain, Genome Biol. 15 (2014), https://doi.org/10.1186/gb-2014-15-3-r49.
[59] S. Kriaucionis, A. Bird, DNA methylation and Rett syndrome, Hum. Mol. Genet. 12 (2003) R221–R227, https://doi.org/10.1093/hmg/ddg286.
[60] S. Marques, T.F. Outeiro, Epigenetics in Parkinson's and Alzheimer's diseases, Subcell. Biochem. 61 (2013) 507–525, https://doi.org/10.1007/978-94-007-4525-4_22.
[61] E. Miranda-Morales, K. Meier, A. Sandoval-Carrillo, J. Salas-Pacheco, P. Vázquez-Cárdenas, O. Arias-Carrión, Implications of DNA methylation in Parkinson's disease, Front. Mol. Neurosci. 10 (2017), https://doi.org/10.3389/fnmol.2017.00225.
[62] L.K. Pries, S. Gülöksüz, G. Kenis, DNA methylation in schizophrenia, in: Advances in Experimental Medicine and Biology, Springer, New York LLC, Netherlands, 2017, pp. 211–236, https://doi.org/10.1007/978-3-319-53889-1_12.

[63] S. Rangasamy, S.R. D'Mello, V. Narayanan, Epigenetics, autism spectrum, and neurodevelopmental disorders, Neurotherapeutics 10 (2013) 742–756, https://doi.org/10.1007/s13311-013-0227-0.
[64] E.A. Thomas, DNA methylation in Huntington's disease: implications for transgenerational effects, Neurosci. Lett. 625 (2016) 34–39, https://doi.org/10.1016/j.neulet.2015.10.060.
[65] O.S. Anderson, M.S. Nahar, C. Faulk, T.R. Jones, C. Liao, K. Kannan, C. Weinhouse, L.S. Rozek, D.C. Dolinoy, Epigenetic responses following maternal dietary exposure to physiologically relevant levels of bisphenol A, Environ. Mol. Mutagen. 53 (2012) 334–342, https://doi.org/10.1002/em.21692.
[66] M.J. Essex, W. Thomas Boyce, C. Hertzman, L.L. Lam, J.M. Armstrong, S.M.A. Neumann, M.S. Kobor, Epigenetic vestiges of early developmental adversity: childhood stress exposure and DNA methylation in adolescence, Child Dev. 84 (2013) 58–75, https://doi.org/10.1111/j.1467-8624.2011.01641.x.
[67] J.J. Kochmanski, E.H. Marchlewicz, R.G. Cavalcante, B.P.U. Perera, M.A. Sartor, D.C. Dolinoy, Longitudinal effects of developmental bisphenol a exposure on epigenome-wide DNA hydroxymethylation at imprinted loci in mouse blood, Environ. Health Perspect. 126 (2018), https://doi.org/10.1289/EHP3441.
[68] F.A.D. Leenen, C.P. Muller, J.D. Turner, DNA methylation: conducting the orchestra from exposure to phenotype? Clin. Epigenetics 8 (2016), https://doi.org/10.1186/s13148-016-0256-8.
[69] C.J. Marsit, Influence of environmental exposure on human epigenetic regulation, J. Exp. Biol. 218 (2015) 71–79, https://doi.org/10.1242/jeb.106971.
[70] B.P.U. Perera, C. Faulk, L.K. Svoboda, J.M. Goodrich, D.C. Dolinoy, The role of environmental exposures and the epigenome in health and disease, Environ. Mol. Mutagen. 61 (2020) 176–192, https://doi.org/10.1002/em.22311.
[71] G. Singh, V. Singh, Z.X. Wang, G. Voisin, F. Lefebvre, J.M. Navenot, B. Evans, M. Verma, D.W. Anderson, J.S. Schneider, Effects of developmental lead exposure on the hippocampal methylome: influences of sex and timing and level of exposure, Toxicol. Lett. 290 (2018) 63–72, https://doi.org/10.1016/j.toxlet.2018.03.021.
[72] R.O. Wright, J. Schwartz, R.J. Wright, V. Bollati, L. Tarantini, S.K. Park, H. Hu, D. Sparrow, P. Vokonas, A. Baccarelli, Biomarkers of lead exposure and DNA methylation within retrotransposons, Environ. Health Perspect. 118 (2010) 790–795, https://doi.org/10.1289/ehp.0901429.
[73] M.F. Fraga, M. Esteller, DNA methylation: a profile of methods and applications, BioTechniques 33 (2002) 632–649, https://doi.org/10.2144/02333rv01.
[74] S. Kurdyukov, M. Bullock, DNA methylation analysis: choosing the right method, Biology 5 (2016), https://doi.org/10.3390/biology5010003.
[75] W.S. Yong, F.M. Hsu, P.Y. Chen, Profiling genome-wide DNA methylation, Epigenetics Chromatin 9 (2016), https://doi.org/10.1186/s13072-016-0075-3.
[76] S. Beck, V.K. Rakyan, The methylome: approaches for global DNA methylation profiling, Trends Genet. 24 (2008) 231–237, https://doi.org/10.1016/j.tig.2008.01.006.
[77] S.J. Clark, A. Statham, C. Stirzaker, P.L. Molloy, M. Frommer, DNA methylation: bisulphite modification and analysis, Nat. Protoc. 1 (2006) 2353–2364, https://doi.org/10.1038/nprot.2006.324.
[78] T. Rein, M.L. DePamphilis, H. Zorbas, Identifying 5-methylcytosine and related modifications in DNA genomes, Nucleic Acids Res. 26 (1998) 2255–2264, https://doi.org/10.1093/nar/26.10.2255.
[79] S.J. Cokus, S. Feng, X. Zhang, Z. Chen, B. Merriman, C.D. Haudenschild, S. Pradhan, S.F. Nelson, M. Pellegrini, S.E. Jacobsen, Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning, Nature 452 (2008) 215–219, https://doi.org/10.1038/nature06745.
[80] R. Lister, M. Pelizzola, R.H. Dowen, R.D. Hawkins, G. Hon, J. Tonti-Filippini, J.R. Nery, L. Lee, Z. Ye, Q.M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A.H. Millar, J.A. Thomson, B. Ren, J.R. Ecker, Human DNA methylomes at base resolution show widespread epigenomic differences, Nature 462 (2009) 315–322, https://doi.org/10.1038/nature08514.
[81] M. Suzuki, W. Liao, F. Wos, A.D. Johnston, J. DeGrazia, J. Ishii, T. Bloom, M.C. Zody, S. Germer, J.M. Greally, Whole-genome bisulfite sequencing with improved accuracy and cost, Genome Res. 28 (2018) 1364–1371, https://doi.org/10.1101/gr.232587.117.

[82] V.K. Rakyan, T.A. Down, D.J. Balding, S. Beck, Epigenome-wide association studies for common human diseases, Nat. Rev. Genet. 12 (2011) 529–541, https://doi.org/10.1038/nrg3000.
[83] A. Akalin, F.E. Garrett-Bakelman, M. Kormaksson, J. Busuttil, L. Zhang, I. Khrebtukova, T.A. Milne, Y. Huang, D. Biswas, J.L. Hess, C.D. Allis, R.G. Roeder, P.J.M. Valk, B. Löwenberg, R. Delwel, H.F. Fernandez, E. Paietta, M.S. Tallman, G.P. Schroth, C.E. Mason, A. Melnick, M.E. Figueroa, Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia, PLoS Genet. 8 (2012), https://doi.org/10.1371/journal.pgen.1002781.
[84] H. Gu, Z.D. Smith, C. Bock, P. Boyle, A. Gnirke, A. Meissner, Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling, Nat. Protoc. 6 (2011) 468–481, https://doi.org/10.1038/nprot.2010.190.
[85] A. Meissner, A. Gnirke, G.W. Bell, B. Ramsahoye, E.S. Lander, R. Jaenisch, Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis, Nucleic Acids Res. 33 (2005) 5868–5877, https://doi.org/10.1093/nar/gki901.
[86] R. Pidsley, E. Zotenko, T.J. Peters, M.G. Lawrence, G.P. Risbridger, P. Molloy, S. Van Djik, B. Muhlhausler, C. Stirzaker, S.J. Clark, Critical evaluation of the illumina methylationEPIC BeadChip microarray for whole-genome DNA methylation profiling, Genome Biol. 17 (2016), https://doi.org/10.1186/s13059-016-1066-1.
[87] D. Barros-Silva, C.J. Marques, R. Henrique, C. Jerónimo, Profiling DNA methylation based on next-generation sequencing approaches: new insights and clinical applications, Genes 9 (2018), https://doi.org/10.3390/genes9090429.
[88] J. Wang, H. Jiang, G. Ji, F. Gao, M. Wu, J. Sun, H. Luo, J. Wu, R. Wu, X. Zhang, High resolution profiling of human exon methylation by liquid hybridization capture-based bisulfite sequencing, BMC Genomics 12 (2011), https://doi.org/10.1186/1471-2164-12-597.
[89] S.G. Jin, S. Kadam, G.P. Pfeifer, Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine, Nucleic Acids Res. 38 (2010) e125, https://doi.org/10.1093/nar/gkq223.
[90] E.K. Schutsky, J.E. Denizio, P. Hu, M.Y. Liu, C.S. Nabel, E.B. Fabyanic, Y. Hwang, F.D. Bushman, H. Wu, R.M. Kohli, Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase, Nat. Biotechnol. 36 (2018) 1083–1090, https://doi.org/10.1038/nbt.4204.
[91] M.J. Booth, T.W.B. Ost, D. Beraldi, N.M. Bell, M.R. Branco, W. Reik, S. Balasubramanian, Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine, Nat. Protoc. 8 (2013) 1841–1851, https://doi.org/10.1038/nprot.2013.115.
[92] Z. Xu, J.A. Taylor, Y.K. Leung, S.M. Ho, L. Niu, OxBS-MLE: an efficient method to estimate 5-methylcytosine and 5-hydroxymethylcytosine in paired bisulfite and oxidative bisulfite treated DNA, Bioinformatics 32 (2016) 3667–3669, https://doi.org/10.1093/bioinformatics/btw527.
[93] J.R. Hernandez Mora, M. Sanchez-Delgado, P. Petazzi, S. Moran, M. Esteller, I. Iglesias-Platas, D. Monk, Profiling of oxBS-450K 5-hydroxymethylcytosine in human placenta and brain reveals enrichment at imprinted loci, Epigenetics 13 (2018) 182–191, https://doi.org/10.1080/15592294.2017.1344803.
[94] A.R. Smith, R.G. Smith, E. Pishva, E. Hannon, J.A.Y. Roubroeks, J. Burrage, C. Troakes, S. Al-Sarraj, C. Sloan, J. Mill, D.L. Van Den Hove, K. Lunnon, Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease, Clin. Epigenetics 11 (2019), https://doi.org/10.1186/s13148-019-0636-y.
[95] S.K. Stewart, T.J. Morris, P. Guilhamon, H. Bulstrode, M. Bachman, S. Balasubramanian, S. Beck, OxBS-450K: a method for analysing hydroxymethylation using 450K BeadChips, Methods 72 (2015) 9–15, https://doi.org/10.1016/j.ymeth.2014.08.009.
[96] J. Kochmanski, C. Savonen, A.I. Bernstein, A novel application of mixed effects models for reconciling base-pair resolution 5-methylcytosine and 5-hydroxymethylcytosine data in neuroepigenetics, Front. Genet. 10 (2019), https://doi.org/10.3389/fgene.2019.00801.

[97] A. Hoppers, L. Williams, V.K.C. Ponnaluri, B. Sexton, L. Saleh, M. Campbell, K. Marks, M. Samaranayake, L. Ettwiller, S. Guan, H. Church, B. Langhorst, Z. Sun, T.C. Evans, R. Vaisvila, E. Dimalanta, F. Stewart, Enzymatic methyl-seq: next generation methylomes, J. Biomol. Tech. 31 (2020) S15.

[98] R. Vaisvila, V.K. Chaithanya Ponnaluri, Z. Sun, B.W. Langhorst, L. Saleh, S. Guan, N. Dai, M.A. Campbell, B. Sexton, K. Marks, M. Samaranayake, J.C. Samuelson, H.E. Church, E. Tamanaha, I.R. Corrêa, S. Pradhan, E.T. Dimalanta, T.C. Evans, L. Williams, T.B. Davis, EM-seq: detection of DNA methylation at single base resolution from picograms of DNA, bioRxiv (2019), https://doi.org/10.1101/2019.12.20.884692.

[99] Y. Liu, P. Siejka-Zielińska, G. Velikova, Y. Bi, F. Yuan, M. Tomkova, C. Bai, L. Chen, B. Schuster-Böckler, C.X. Song, Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution, Nat. Biotechnol. 37 (2019) 424–429, https://doi.org/10.1038/s41587-019-0041-2.

[100] H. Ji, L.I.R. Ehrlich, J. Seita, P. Murakami, A. Doi, P. Lindau, H. Lee, M.J. Aryee, R.A. Irizarry, K. Kim, D.J. Rossi, M.A. Inlay, T. Serwold, H. Karsunky, L. Ho, G.Q. Daley, I.L. Weissman, A.P. Feinberg, Comprehensive methylome map of lineage commitment from haematopoietic progenitors, Nature 467 (2010) 338–342, https://doi.org/10.1038/nature09367.

[101] L.E. Reinius, N. Acevedo, M. Joerink, G. Pershagen, S.E. Dahlén, D. Greco, C. Söderhäll, A. Scheynius, J. Kere, Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility, PLoS One 7 (2012), https://doi.org/10.1371/journal.pone.0041361.

[102] B. Zhang, Y. Zhou, N. Lin, R.F. Lowdon, C. Hong, R.P. Nagarajan, J.B. Cheng, D. Li, M. Stevens, H.J. Lee, X. Xing, J. Zhou, V. Sundaram, G. Elliott, J. Gu, T. Shi, P. Gascard, M. Sigaroudinia, T.D. Tlsty, T. Kadlecek, A. Weiss, H. O'Geen, P.J. Farnham, C.L. Maire, K.L. Ligon, P.A.F. Madden, A. Tam, R. Moore, M. Hirst, M.A. Marra, B. Zhang, J.F. Costello, T. Wang, Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm, Genome Res. 23 (2013) 1522–1540, https://doi.org/10.1101/gr.156539.113.

[103] S.A. Smallwood, H.J. Lee, C. Angermueller, F. Krueger, H. Saadeh, J. Peat, S.R. Andrews, O. Stegle, W. Reik, G. Kelsey, Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity, Nat. Methods 11 (2014) 817–820, https://doi.org/10.1038/nmeth.3035.

[104] A.J. Titus, R.M. Gallimore, L.A. Salas, B.C. Christensen, Cell-type deconvolution from DNA methylation: a review of recent applications, Hum. Mol. Genet. 26 (2017) R216–R224, https://doi.org/10.1093/hmg/ddx275.

[105] E. Rahmani, R. Schweiger, B. Rhead, L.A. Criswell, L.F. Barcellos, E. Eskin, S. Rosset, S. Sankararaman, E. Halperin, Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology, Nat. Commun. 10 (2019), https://doi.org/10.1038/s41467-019-11052-9.

[106] E.A. Houseman, M.L. Kile, D.C. Christiani, T.A. Ince, K.T. Kelsey, C.J. Marsit, Reference-free deconvolution of DNA methylation data and mediation by cell composition effects, BMC Bioinform. 17 (2016), https://doi.org/10.1186/s12859-016-1140-4.

[107] M. Scherer, P.V. Nazarov, R. Toth, S. Sahay, T. Kaoma, V. Maurer, N. Vedeneev, C. Plass, T. Lengauer, J. Walter, P. Lutsik, Reference-free deconvolution, visualization and interpretation of complex DNA methylation data using DecompPipeline, MeDeCom and FactorViz, Nat. Protoc. 15 (2020) 3240–3263, https://doi.org/10.1038/s41596-020-0369-6.

[108] S. Basu, H.M. Campbell, B.N. Dittel, A. Ray, Purification of specific cell population by fluorescence activated cell sorting (FACS), J. Vis. Exp. (2010), https://doi.org/10.3791/1546.

[109] B. Schmitz, A. Radbruch, T. Kümmel, C. Wickenhauser, H. Korb, M.L. Hansmann, J. Thiele, R. Fischer, Magnetic activated cell sorting (MACS)—a new immunomagnetic method for megakaryocytic cell isolation: comparison of different separation techniques, Eur. J. Haematol. 52 (1994) 267–275, https://doi.org/10.1111/j.1600-0609.1994.tb00095.x.

[110] L. Hackler, T. Masuda, V.F. Oliver, S.L. Merbs, D.J. Zack, Use of laser capture microdissection for analysis of retinal mRNA/miRNA expression and DNA methylation, Methods Mol. Biol. 884 (2012) 289–304, https://doi.org/10.1007/978-1-61779-848-1_21.

[111] S. Horvath, K. Raj, DNA methylation-based biomarkers and the epigenetic clock theory of ageing, Nat. Rev. Genet. 19 (2018) 371–384, https://doi.org/10.1038/s41576-018-0004-3.
[112] A.A. Johnson, K. Akman, S.R.G. Calimport, D. Wuttke, A. Stolzing, J.P. De Magalhães, The role of DNA methylation in aging, rejuvenation, and age-related disease, Rejuvenation Res. 15 (2012) 483–494, https://doi.org/10.1089/rej.2012.1324.
[113] N.D. Johnson, L. Huang, R. Li, Y. Li, Y. Yang, H.R. Kim, C. Grant, H. Wu, E.A. Whitsel, D.P. Kiel, A.A. Baccarelli, P. Jin, J.M. Murabito, K.N. Conneely, Age-related DNA hydroxymethylation is enriched for gene expression and immune system processes in human peripheral blood, Epigenetics 15 (2020) 294–306, https://doi.org/10.1080/15592294.2019.1666651.
[114] M.J. Jones, S.J. Goodman, M.S. Kobor, DNA methylation and healthy human aging, Aging Cell 14 (2015) 924–932, https://doi.org/10.1111/acel.12349.
[115] J. Kochmanski, E.H. Marchlewicz, R.G. Cavalcante, M.A. Sartor, D.C. Dolinoy, Age-related epigenome-wide DNA methylation and hydroxymethylation in longitudinal mouse blood, Epigenetics 13 (2018) 779–792, https://doi.org/10.1080/15592294.2018.1507198.
[116] J. Kochmanski, L. Montrose, J.M. Goodrich, D.C. Dolinoy, Environmental deflection: the impact of toxicant exposures on the aging epigenome, Toxicol. Sci. 156 (2017) 325–335, https://doi.org/10.1093/toxsci/kfx005.
[117] E. Hall, P. Volkov, T. Dayeh, J.L.S. Esguerra, S. Salö, L. Eliasson, T. Rönn, K. Bacos, C. Ling, Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets, Genome Biol. 15 (2014) 522, https://doi.org/10.1186/s13059-014-0522-z.
[118] J. Liu, M. Morgan, K. Hutchison, V.D. Calhoun, A study of the influence of sex on genome wide methylation, PLoS One 5 (2010) e10028, https://doi.org/10.1371/journal.pone.0010028.
[119] H. Xu, F. Wang, Y. Liu, Y. Yu, J. Gelernter, H. Zhang, Sex-biased methylome and transcriptome in human prefrontal cortex, Hum. Mol. Genet. 23 (2014) 1260–1270, https://doi.org/10.1093/hmg/ddt516.
[120] S.W. Curtis, S.A. Gerkowicz, D.O. Cobb, V. Kilaru, M.L. Terrell, M.E. Marder, D.B. Barr, C.J. Marsit, M. Marcus, K.N. Conneely, A.K. Smith, Sex-specific DNA methylation differences in people exposed to polybrominated biphenyl, Epigenomics 12 (2020) 757–770, https://doi.org/10.2217/epi-2019-0179.
[121] R. Massart, Z. Nemoda, M.J. Suderman, S. Sutti, A.M. Ruggiero, A.M. Dettmer, S.J. Suomi, M. Szyf, Early life adversity alters normal sex-dependent developmental dynamics of DNA methylation, Dev. Psychopathol. 28 (2016) 1259–1272.
[122] E.W. Tobi, L.H. Lumey, R.P. Talens, D. Kremer, H. Putter, A.D. Stein, P.E. Slagboom, B.T. Heijmans, DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific, Hum. Mol. Genet. 18 (2009) 4046–4053, https://doi.org/10.1093/hmg/ddp353.
[123] K. Wang, S. Liu, L.K. Svoboda, C.A. Rygiel, K. Neier, T.R. Jones, J.A. Colacino, D.C. Dolinoy, M.A. Sartor, Tissue- and sex-specific DNA methylation changes in mice perinatally exposed to lead (Pb), Front. Genet. 11 (2020) 840, https://doi.org/10.3389/fgene.2020.00840.
[124] P.W. Laird, Principles and challenges of genome-wide DNA methylation analysis, Nat. Rev. Genet. 11 (2010) 191–203, https://doi.org/10.1038/nrg2732.
[125] I. Rauluseviciute, F. Drabløs, M.B. Rye, DNA methylation data by sequencing: experimental approaches and recommendations for tools and pipelines for data analysis, Clin. Epigenetics 11 (2019), https://doi.org/10.1186/s13148-019-0795-x.
[126] A. Shafi, C. Mitrea, T. Nguyen, S. Draghici, A survey of the approaches for identifying differential methylation using bisulfite sequencing data, Brief. Bioinform. 19 (2018) 737–753, https://doi.org/10.1093/bib/bbx013.
[127] K. Wreczycka, A. Gosdschan, D. Yusuf, B. Grüning, Y. Assenov, A. Akalin, Strategies for analyzing bisulfite sequencing data, J. Biotechnol. 261 (2017) 105–115, https://doi.org/10.1016/j.jbiotec.2017.08.007.

[128] X. Sun, Y. Han, L. Zhou, E. Chen, B. Lu, Y. Liu, X. Pan, A.W. Cowley, M. Liang, Q. Wu, Y. Lu, P. Liu, A comprehensive evaluation of alignment software for reduced representation bisulfite sequencing data, Bioinformatics 34 (2018) 2715–2723, https://doi.org/10.1093/bioinformatics/bty174.

[129] J.P. Fortin, T.J. Triche, K.D. Hansen, Preprocessing, normalization and integration of the Illumina Human MethylationEPIC array with minfi, Bioinformatics 33 (2017) 558–560, https://doi.org/10.1093/bioinformatics/btw691.

[130] G. Mansell, T.J. Gorrie-Stone, Y. Bao, M. Kumari, L.S. Schalkwyk, J. Mill, E. Hannon, Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array, BMC Genomics 20 (2019), https://doi.org/10.1186/s12864-019-5761-7.

[131] K.B. Michels, A.M. Binder, S. Dedeurwaerder, C.B. Epstein, J.M. Greally, I. Gut, E.A. Houseman, B. Izzi, K.T. Kelsey, A. Meissner, A. Milosavljevic, K.D. Siegmund, C. Bock, R.A. Irizarry, Recommendations for the design and analysis of epigenome-wide association studies, Nat. Methods 10 (2013) 949–955, https://doi.org/10.1038/nmeth.2632.

[132] J. Mill, B.T. Heijmans, From promises to practical strategies in epigenetic epidemiology, Nat. Rev. Genet. 14 (2013) 585–594, https://doi.org/10.1038/nrg3405.

[133] C.L. Relton, G. Davey Smith, Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment, PLoS Med. 7 (2010) e1000356, https://doi.org/10.1371/journal.pmed.1000356.

[134] S.N. Hart, T.M. Therneau, Y. Zhang, G.A. Poland, J.P. Kocher, Calculating sample size estimates for RNA sequencing data, J. Comput. Biol. 20 (2013) 970–978, https://doi.org/10.1089/cmb.2012.0283.

[135] H. Wu, C. Wang, Z. Wu, PROPER: comprehensive power evaluation for differential expression using RNA-seq, Bioinformatics 31 (2015) 233–241, https://doi.org/10.1093/bioinformatics/btu640.

[136] B. Vieth, C. Ziegenhain, S. Parekh, W. Enard, I. Hellmann, powsimR: power analysis for bulk and single cell RNA-seq experiments, Bioinformatics (Oxford, England) 33 (2017) 3486–3488, https://doi.org/10.1093/bioinformatics/btx435.

[137] S. Zhao, C.I. Li, Y. Guo, Q. Sheng, Y. Shyr, RnaSeqSampleSize: real data based sample size estimation for RNA sequencing, BMC Bioinform. 19 (2018), https://doi.org/10.1186/s12859-018-2191-5.

[138] A.J. Lea, T.P. Vilgalys, P.A.P. Durst, J. Tung, Maximizing ecological and evolutionary insight in bisulfite sequencing data sets, Nat. Ecol. Evol. 1 (2017) 1074–1083, https://doi.org/10.1038/s41559-017-0229-0.

[139] S. Graw, R. Henn, J.A. Thompson, D.C. Koestler, PwrEWAS: a user-friendly tool for comprehensive power estimation for epigenome wide association studies (EWAS), BMC Bioinform. 20 (2019), https://doi.org/10.1186/s12859-019-2804-7.

[140] O.S. Anderson, K.E. Sant, D.C. Dolinoy, Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation, J. Nutr. Biochem. 23 (2012) 853–859, https://doi.org/10.1016/j.jnutbio.2012.03.003.

[141] J. Bakusic, W. Schaufeli, S. Claes, L. Godderis, Stress, burnout and depression: a systematic review on DNA methylation mechanisms, J. Psychosom. Res. 92 (2017) 34–44, https://doi.org/10.1016/j.jpsychores.2016.11.005.

[142] E. Grazioli, I. Dimauro, N. Mercatelli, G. Wang, Y. Pitsiladis, L. Di Luigi, D. Caporossi, Physical activity in the prevention of human diseases: role of epigenetic modifications, BMC Genomics 18 (2017), https://doi.org/10.1186/s12864-017-4193-5.

[143] N. Matosin, C. Cruceanu, E.B. Binder, Preclinical and clinical evidence of DNA methylation changes in response to trauma and chronic stress, Chronic Stress 1 (2017), https://doi.org/10.1177/2470547017710764.

[144] S. Voisin, N. Eynon, X. Yan, D.J. Bishop, Exercise training and DNA methylation in humans, Acta Physiol. 213 (2015) 39–59, https://doi.org/10.1111/apha.12414.

[145] S. Greenland, N. Pearce, Statistical foundations for model-based adjustments, Annu. Rev. Public Health 36 (2015) 89–108, https://doi.org/10.1146/annurev-publhealth-031914-122559.

[146] L.H. Chadwick, The NIH roadmap epigenomics program data resource, Epigenomics 4 (2012) 317–324, https://doi.org/10.2217/epi.12.18.
[147] T. Wang, E.C. Pehrsson, D. Purushotham, D. Li, X. Zhuo, B. Zhang, H.A. Lawson, M.A. Province, C. Krapp, Y. Lan, C. Coarfa, T.A. Katz, W.Y. Tang, Z. Wang, S. Biswal, S. Rajagopalan, J.A. Colacino, Z.T.Y. Tsai, M.A. Sartor, K. Neier, D.C. Dolinoy, J. Pinto, R.B. Hamanaka, G.M. Mutlu, H.B. Patisaul, D.L. Aylor, G.E. Crawford, T. Wiltshire, L.H. Chadwick, C.G. Duncan, A.E. Garton, K.A. McAllister, M.S. Bartolomei, C.L. Walker, F.L. Tyson, The NIEHS TaRGET II Consortium and environmental epigenomics, Nat. Biotechnol. 36 (2018) 225–227, https://doi.org/10.1038/nbt.4099.
[148] P. Du, X. Zhang, C.C. Huang, N. Jafari, W.A. Kibbe, L. Hou, S.M. Lin, Comparison of beta-value and M-value methods for quantifying methylation levels by microarray analysis, BMC Bioinform. 11 (2010), https://doi.org/10.1186/1471-2105-11-587.
[149] C. Xie, Y.K. Leung, A. Chen, D.X. Long, C. Hoyo, S.M. Ho, Differential methylation values in differential methylation analysis, Bioinformatics 35 (2019) 1094–1097, https://doi.org/10.1093/bioinformatics/bty778.
[150] E. Dolzhenko, A.D. Smith, Using beta-binomial regression for high-precision differential methylation analysis in multifactor whole-genome bisulfite sequencing experiments, BMC Bioinform. 15 (2014), https://doi.org/10.1186/1471-2105-15-215.
[151] M. Saadati, A. Benner, Statistical challenges of high-dimensional methylation data, Stat. Med. 33 (2014) 5347–5357, https://doi.org/10.1002/sim.6251.
[152] W.J. Seow, A.C. Pesatori, E. Dimont, P.B. Farmer, B. Albetti, A.S. Ettinger, V. Bollati, C. Bolognesi, P. Roggieri, T.I. Panev, T. Georgieva, D.F. Merlo, P.A. Bertazzi, A.A. Baccarelli, Urinary benzene biomarkers and DNA methylation in bulgarian petrochemical workers: study findings and comparison of linear and beta regression models, PLoS One 7 (2012), https://doi.org/10.1371/journal.pone.0050471.
[153] P. Breheny, A. Stromberg, J. Lambert, P-value histograms: inference and diagnostics, High-Throughput 7 (2018), https://doi.org/10.3390/HT7030023.
[154] B.B. Green, E.A. Houseman, K.C. Johnson, D.J. Guerin, D.A. Armstrong, B.C. Christensen, C.J. Marsit, Hydroxymethylation is uniquely distributed within term placenta, and is associated with gene expression, FASEB J. 30 (2016) 2874–2884, https://doi.org/10.1096/fj.201600310R.
[155] K.C. Johnson, E.A. Houseman, J.E. King, K.M. Von Herrmann, C.E. Fadul, B.C. Christensen, 5-Hydroxymethylcytosine localizes to enhancer elements and is associated with survival in glioblastoma patients, Nat. Commun. 7 (2016), https://doi.org/10.1038/ncomms13177.
[156] W.K. Glowacka, H. Jain, M. Okura, A. Maimaitiming, Y. Mamatjan, R. Nejad, H. Farooq, M.D. Taylor, K. Aldape, P. Kongkham, 5-Hydroxymethylcytosine preferentially targets genes upregulated in isocitrate dehydrogenase 1 mutant high-grade glioma, Acta Neuropathol. 135 (2018) 617–634, https://doi.org/10.1007/s00401-018-1821-3.
[157] Z. Xue, C. Xiaoting, M.T. Weirauch, X. Zhang, J.D. Burleson, E.B. Brandt, H. Ji, Diesel exhaust and house dust mite allergen lead to common changes in the airway methylome and hydroxymethylome, Environ. Epigen. (2018), https://doi.org/10.1093/eep/dvy020.
[158] R.A. Rigby, D.M. Stasinopoulos, Generalized additive models for location, scale and shape (with discussion), J. R. Stat. Soc.: Ser. C: Appl. Stat. 54 (2005) 507–554, https://doi.org/10.1111/j.1467-9876.2005.00510.x.
[159] M.E.D. Foster, M.A. Deardorff, Open science framework (OSF), J. Med. Libr. Assoc. (2017), https://doi.org/10.5195/jmla.2017.88.
[160] J.J. Heindel, L.N. Vandenberg, Developmental origins of health and disease: a paradigm for understanding disease cause and prevention, Curr. Opin. Pediatr. 27 (2015) 248–253, https://doi.org/10.1097/MOP.0000000000000191.
[161] B. Janis, J. Chandni, B. Mary, F. Caroline, H. Mark, H. Nicholas, I. Hazel, K. Kalyanaraman, C. Cyrus, Developmental origins of health and disease: a lifecourse approach to the prevention of non-communicable diseases, Healthcare 14 (2017), https://doi.org/10.3390/healthcare5010014.

[162] J. Bohacek, I.M. Mansuy, A guide to designing germline-dependent epigenetic inheritance experiments in mammals, Nat. Methods 14 (2017) 243–249, https://doi.org/10.1038/nmeth.4181.
[163] S.E. Lazic, L. Essioux, Improving basic and translational science by accounting for litter-to-litter variation in animal models, BMC Neurosci. 14 (2013), https://doi.org/10.1186/1471-2202-14-37.
[164] I. Plewis, Pesticides and transgenerational inheritance of pathologies: designing, analysing and reporting rodent studies, PLoS One 15 (2020), https://doi.org/10.1371/journal.pone.0228762.
[165] N.P. du Sert, V. Hurst, A. Ahluwalia, S. Alam, M.T. Avey, M. Baker, W.J. Browne, A. Clark, I.C. Cuthill, U. Dirnagl, M. Emerson, P. Garner, S.T. Holgate, D.W. Howells, N.A. Karp, S.E. Lazic, K. Lidster, C.J. MacCallum, M. Macleod, E.J. Pearl, O.H. Petersen, F. Rawle, P. Reynolds, K. Rooney, E.S. Sena, S.D. Silberberg, T. Steckler, H. Würbel, The arrive guidelines 2.0: updated guidelines for reporting animal research, PLoS Biol. 18 (2020), https://doi.org/10.1371/journal.pbio.3000410.
[166] I.M. Fingerman, L. McDaniel, X. Zhang, W. Ratzat, T. Hassan, Z. Jiang, R.F. Cohen, G.D. Schuler, N.C.B.I. Epigenomics, A new public resource for exploring epigenomic data sets, Nucleic Acids Res. 39 (2011) D908–D912, https://doi.org/10.1093/nar/gkq1146.
[167] A. Ammar, S. Bonaretti, L. Winckers, J. Quik, M. Bakker, D. Maier, I. Lynch, J. van Rijn, E. Willighagen, A semi-automated workflow for fair maturity indicators in the life sciences, Nanomaterials 10 (2020) 1–14, https://doi.org/10.3390/nano10102068.
[168] M. Corpas, N.V. Kovalevskaya, A. McMurray, F.G.G. Nielsen, A FAIR guide for data providers to maximise sharing of human genomic data, PLoS Comput. Biol. 14 (2018), https://doi.org/10.1371/journal.pcbi.1005873.