Published in final edited form as:

Circ Res. 2016 January 8; 118(1): 119–131. doi:10.1161/CIRCRESAHA.115.305206.

The Role of DNA Methylation in Cardiovascular Risk and Disease: Methodological Aspects, Study Design, and Data Analysis for Epidemiological Studies

Jia Zhong^{#1}, Golareh Agha^{#1}, and Andrea A. Baccarelli^{1,†}

¹Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

[#] These authors contributed equally to this work.

Abstract

Epidemiological studies have demonstrated that genetic, environmental, behavioral, and clinical factors contribute to cardiovascular disease (CVD) development. How these risk factors interact at the cellular level to cause CVD is not well known. Epigenetic epidemiology enables researchers to explore critical links between genomic coding, modifiable exposures, and manifestation of disease phenotype. One epigenetic link, DNA methylation, is potentially an important mechanism underlying these associations. In the past decade, there has been a significant increase in the number of epidemiological studies investigating cardiovascular risk factors and outcomes in relation to DNA methylation, but many gaps remain in our understanding of the underlying etiology and biological implications. In this review we provide a brief overview of the biology and mechanisms of DNA methylation and its role in CVD. In addition, we summarize the current evidence base in epigenetic epidemiology studies relevant to cardiovascular health and disease, and discuss the limitations, challenges, and future directions of the field. Finally, we provide guidelines for well-designed epigenetic epidemiology studies, with particular focus on methodological aspects, study design, and analytical challenges.

Keywords

DNA methylation; Epidemiology; Cardiovascular Diseases; 5-methylcytosine; risk factors

Introduction

Over the past few decades, major strides have been made in identifying environmental, behavioral, and clinical risk factors for cardiovascular disease (CVD).¹ A considerable decline in CVD mortality has been achieved due to management of clinical risk factors, secondary and tertiary prevention, as well as improvements in quality of care and treatment.² However, primary prevention remains a challenge.² In particular, many gaps remain in

[†]Correspondence to: Andrea A. Baccarelli, MD, PhD, FAHA, Department of Environmental Health, Harvard T.H. Chan School of Public Health, Bldg 1, Rm G5, 665 Huntington Ave, Boston, MA 02115, abaccare@hsph.harvard.edu, Phone: (617) 432-2095, Fax: (617) 432-6913.

Disclosures: None

understanding how genetic and environmental factors interact to contribute to CVD development. Epigenetics plays a central role in developmental processes, such as X-chromosome inactivation, parental genomic imprinting, and cellular differentiation. Recently, epigenetic mechanisms have been shown to link the internal genetic landscape and external environmental influences; thus they might provide better mechanistic insight into the etiology of CVD.³ Epigenetic changes can occur via DNA methylation, histone modification, and post-transcriptional silencing mediated by micro-RNAs (miRNAs).⁴ Although different forms of epigenetic regulation have been investigated in relation to cardiovascular biology and pathology,⁵⁻⁷ DNA methylation has been most studied in large population studies because it is most feasible to measure in an epidemiological setting.⁸ Consequently, this review will focus on DNA methylation, particularly with respect to how it relates to CVD and its risk factors. For detailed accounts of the other forms of epigenetic regulation, the reader is referred elsewhere.⁹⁻¹³

Biology of DNA Methylation

DNA methylation, a pre-transcriptional modification characterized by the addition of methyl groups to specific nucleotides, regulates the stability of gene expression states and maintains genome integrity by collaborating with proteins that modify nucleosomes (Figure 1).^{4, 14} In somatic cells, DNA methylation occurs predominantly on cytosine residues of the dinucleotide sequence CpG.^{15, 16} CpG dinucleotides are distributed unevenly throughout the genome, and in normal, healthy cells approximately 80% of CpGs are methylated.^{17, 18}

The early postconceptional period is a critical window during which DNA methylation patterns are established.¹⁶⁻¹⁸ Specifically, during gametogenesis rapid demethylation of the entire parental genome occurs, followed by genome-wide *de novo* methylation after implantation, leading to tissue-specific DNA methylation patterns that then influence cellular differentiation in the developing organism.¹⁹

DNA methyltransferases (DNMTs) are enzymes that have highly conserved catalytic motifs and play a pivotal role in DNA methylation: they catalyze the addition of methyl groups to cytosine and generate 5-methylcytosine (5-mC).^{4, 14, 20} This modification is considered to be stable through cell differentiation and is retained during cell division.¹⁶ Of the three DNMTs, DNMT1 is required to maintain established methylation patterns, whereas DNMT3a and DNMT3b are required for *de novo* methylation.¹⁶ The functions of DNMTs are discussed in detail elsewhere.²⁻⁴ Conversely, the process of demethylation removes methyl groups. Active DNA demethylation is the enzymatic process that removes the methyl group from 5-mC, whereas passive DNA demethylation is characterized as the loss of the methyl group from 5-mC due to inhibited or absent DNMT1 during DNA replication.¹⁹

DNA Methylation in CVD

Comprehensive reviews and studies also suggest that changes in DNA methylation states contribute to the regulation of biological processes underlying CVD, such as atherosclerosis, hypertension, and inflammation.^{3, 21, 22} For example, mice with a hypomethylated genome had elevated expression of inflammatory markers,²³ and DNA hypomethylation was shown to precede the formation of aortic fatty streaks.²⁴ In atherosclerosis-prone apolipoprotein E

(ApoE)-null mice, DNA methylation changes occurred in both peripheral blood leukocytes and the aorta prior to the formation of vascular lesions.²⁵ Such changes were shown to contribute to inflammation and to promote atherosclerosis. When normal heart tissue was compared with tissue from patients with heart failure, the diseased tissue showed differential methylation in angiogenesis-related genes, as well as differential expression.²⁶

In recent years, research on cardiovascular epigenetics has begun to expand rapidly from biological and animal studies to epidemiological studies. In blood, methylation of repetitive sequences, such as long-interspersed nucleotide repetitive elements-1 (*LINE-1*) and *ALU* have been associated with CVD.^{27, 28} Specifically, elevated methylation of *ALU* in leukocytes was associated with prevalence of CVD and obesity in Chinese individuals.²⁸ In the Normative Aging Study conducted in the Boston area, subjects with prevalent ischemic heart disease and stroke had lower blood *LINE-1* methylation, while in longitudinal analyses, lower *LINE-1* methylation was associated with higher risk of ischemic heart disease and stroke.²⁷ Heart failure patients were found to have altered promoter methylation in genes relevant to myocyte apoptosis, fibrosis, and altered contractility.²⁹ DNA methylation in the *GNAS* antisense RNA 1 (*GNASAS1*) and insulin (*INS*) genes, which were previously shown to be sensitive to prenatal nutrition, were found to be associated with 3-year risk of myocardial infarction among elderly women.³⁰

Many gaps still remain in understanding the mechanisms by which epigenetic changes may contribute to atherosclerotic CVD, and inconsistencies exist in the evidence thus far, as the studies discussed above suggest. Studies of global DNA methylation in CVD patients report both increased and decreased DNA methylation.^{2, 27, 28} In addition, DNA isolated from atherosclerotic tissue has been observed to be hypomethylated globally; however, DNA hypermethylation was observed in the promoter regions of genes associated with atherosclerosis.²²

Lifestyle, Environment, Aging, and DNA methylation

Although CVD has a heritable component, a large body of epidemiological research showed that lifestyle and environmental risk factors can impact the cardiovascular health of individuals and populations.³¹⁻³⁵

Epigenetic mechanisms might be involved in the pathways by which environmental and lifestyle factors contribute to the development of CVD. This is of particular interest within the framework of the developmental origins of health and disease hypothesis, which proposes that risk factors operating during fetal life, such as maternal exposures, can program an individual for disease risk during adulthood.³⁶ As mentioned, epigenetic mechanisms are crucial during organismal development. The *in utero* period therefore represents a vulnerable time frame during which external environmental stimuli can have considerable influence on long-term risks.³⁷ There is also evidence that early and midlife exposures can lead to DNA methylation changes, and such evidence is highly useful for understanding the underlying mechanisms by which CVD and other chronic diseases develop throughout the life course and manifest late in life.³¹ Some of that evidence is

discussed below (See Supplemental Table I for major findings discussed in the remainder of this section).

Diet

Studies in animals and humans suggest that diet-induced DNA methylation changes likely contribute to CVD. In animal studies, manipulation of nutrition, particularly with foods that donate methyl groups, has led to DNA methylation modifications as well as clinically relevant phenotypic changes. A well-known example is the effect of maternal dietary genistein on DNA methylation in the agouti metastable epiallele and the consequent phenotypic changes in fur color and obesity in the offspring.³⁸ In sheep, a maternal diet deficient in vitamin B₁₂, folate, and methionine led to widespread CpG island DNA methylation changes in the fetal liver. The offspring were also heavier, more insulin resistant, and had higher adulthood blood pressure, compared with the control animals.³⁹

Protein restriction in pregnant rats reduced DNA methylation in the promoter of key metabolic regulatory genes in first- and second-generation progeny.^{40, 41} Feeding maternal rodents a high-fat diet also resulted in hyperphagia, adiposity, hypertension, and insulin resistance in the offspring, as well as hypomethylation both globally and in appetite-regulatory genes.⁴²

In humans, the Dutch Famine Birth Cohort provided seminal evidence suggesting an important mechanistic role for epigenetics in CVD. *In utero* exposure to famine was shown to be related to numerous cardiovascular outcomes decades later, including coronary heart disease, obesity, elevated lipid levels, and altered clotting.⁴³⁻⁴⁵ Famine during fetal development was also related to DNA methylation changes in genes implicated in growth and metabolic disease (Supplemental Table I).^{46, 47}

Multiple nutrients from the prudent diet (e.g., methyl donors and antioxidants) in combination with low fat intake may interact to modulate DNA methylation. Studies demonstrate that maternal folate intake led to DNA methylation changes in the insulin-like growth factor 2 (*IGF2*) gene in the offspring,⁴⁸⁻⁵⁰ although a study of serum folate levels did not find evidence of such associations (Supplemental Table I).⁵¹ Flavonoids, which may exert protective effects against heart disease, have been shown to decrease gene-specific methylation, particularly by affecting DNMT.⁵²⁻⁵⁴ In the North Texas Health Study, participants with high vegetable and fruit intake had higher *LINE-1* methylation levels than those who consumed a diet of mostly meats and carbohydrates.⁵⁴

Smoking

There is concrete evidence that smoking leads to epigenetic alterations in several tissues. Maternal exposure to tobacco smoke led to global as well as gene-specific DNA methylation changes in the child's buccal cell DNA.⁵⁵ Lower methylation was observed in the coagulation factor II receptor-like 3 gene (*F2RL3*) when comparing adult smokers compared to nonsmokers, and this association was independently replicated.⁵⁶ Intriguingly, the gene's protein product, protease-activated receptor-4 gene (*PAR4*) has been shown to introduce platelet activation and affect intimal hyperplasia and inflammation.⁵⁷

In a study of maternal plasma cotinine level (an objective biomarker of smoking),⁵⁸ maternal cotinine displayed a dose-dependent association with lower methylation level in the arylhydrocarbon receptor repressor (*AHRR*) and higher levels in the cytochrome P450 isoform (*CYP1A1*) gene. This finding has important biological significance, given that both *AHRR* and *CYP1A1* play important roles in the detoxification of components in tobacco smoke, such as polycyclic aromatic hydrocarbons. In adult smokers, an association with DNA methylation in the *AHRR* gene was also observed in lymphoblasts and alveolar macrophages.⁵⁹ A recent monozygotic twin study⁶⁰ found differential methylation between the smoking and nonsmoking twins in eight loci that had been reported in prior studies, including loci in *AHRR* and *F2RL3*. Differential DNA methylation has also been observed with respect to current vs. former smoking, cumulative smoking exposure (e.g., lifetime pack-years of smoking), and time since quitting (among former smokers).^{61, 62} Such dynamic DNA methylation changes in response to smoking correspond to associations observed between smoking and CVD,⁶³⁻⁶⁵ suggesting a potential mechanistic role for DNA methylation.

Environmental Toxins

Environmental exposures, such as air pollution, heavy metals, and arsenic can induce oxidative stress and trigger inflammation, which might contribute to atherogenesis via increased systemic cytokine-mediated inflammation, endothelial dysfunction, thrombosis, and arrhythmias.³¹ Particulate air pollution has been associated with onset of myocardial infarction as well as long-term risk of heart disease and stroke.^{66, 67} Studies have also shown that ambient particulate matter,^{68, 69} metals, and endocrine-disrupting toxicants alter DNA methylation in repetitive elements, as well as in specific genes.⁷⁰⁻⁷²

Methylation and Aging

At the genome-wide level, aging has been associated with decreased DNA methylation or longitudinal decline in genomic methylation over time.^{73, 74} This evidence is in line with the age-related decrease in DNA methylation that has been observed in animal studies.^{75, 76} However, studies targeting specific genomic regions showed either increased or decreased methylation with age, depending on the loci in question.⁷⁷⁻⁸³ For example, promoters of polycomb group protein target genes (PCGTs) were far more likely to be methylated with age than promoters of non-PCGTs, regardless of sex, tissue type, disease state, and methylation assay employed.⁸⁰

Recently, Steve Horvath introduced the hypothesis of an epigenetic clock. Using DNA methylation data from thousands of samples from different tissues, he developed an algorithm that uses 353 CpGs in the genome to measure one's DNA methylation-based age. In otherwise healthy individuals, this measure of DNA methylation-based age can predict chronological age with astonishing accuracy. However, its great potential lies in its ability to serve as a biomarker of accelerated aging and disease, when the predicted DNA methylation-based age is greater than an individual's chronological age. For example, significant age acceleration was observed in samples from cancer patients, morbidly obese individuals, and HIV patients,^{84, 85} and age acceleration was associated with all-cause mortality.⁸⁶ DNA methylation-based aging remains to be investigated in CVD.

Although there is a growing evidence base for the association between DNA methylation and many environmental factors that are known to influence CVD, there is now a need for studies to better elucidate how environmentally induced DNA methylation changes contribute to clinically manifest disease.

Methods in Epigenetic Epidemiology: Current Limitations, Solutions, and Future Directions

Epidemiological study design and selection of study population

The studies described above have provided rich evidence of DNA methylation in relation to cardiovascular health and disease; however, many inconsistencies in the results and limitations in the research still exist. Because epigenetic states are highly dynamic, they can be altered by environmental factors and can vary over time; therefore, measurement error, confounding, and reverse causation are all significant limitations in many prior studies.⁸⁷⁻⁸⁹ In observational studies, lifestyle and environmental factors (e.g., smoking, age, and pollution) can directly confound an epigenetic association by affecting both the epigenetic patterns and the exposure or outcome of interest. Despite statistical adjustment for confounders, residual confounding can still occur. The case-control and cross-sectional study designs in prior epigenetic epidemiologic studies are not ideal,^{46, 47, 50, 81} as conclusions about directionality or causality cannot be drawn. In addition, studies have typically measured DNA methylation status at only one time point.^{46-48, 50, 51, 54, 55} Because natural variation in epigenetic states occurs over time, epigenetic patterns measured at one time point do not necessarily indicate when the observed patterns were established, whether they changed over time, and how such changes might relate to the exposure or outcome of interest. Furthermore, many studies to date have had a small sample size^{46, 47, 50, 51, 54-56, 80, 81}; larger sample sizes are required to ensure adequate statistical power.⁸⁷⁻⁸⁹

Given these challenges, longitudinal cohort studies, particularly birth cohorts, are important for understanding epigenetic contributions to health and disease.^{88, 89} The incorporation of epigenome-wide DNA methylation data into such cohort studies is now occurring in large cohorts in the United States and Europe.⁸⁶ However, given that longitudinal cohort studies are expensive and require lengthy follow-up, a feasible compromise is to undertake case-control studies nested within an existing longitudinal cohort.⁸⁷

Monozygotic twin studies are also a useful study design and were commonly used early in epigenetic epidemiology research.⁸⁸ By studying twins, a genetic cause can be ruled out for any association observed in cases where the twins are discordant for a disease or phenotype of interest.⁸³ However, it can be difficult to recruit a large number of discordant monozygotic twins, and unless longitudinal data are available for epigenetic measures and other factors, confounding and reverse causality can still pose problems. Similar to twin studies, intergenerational studies that collect epigenetic data on parents and siblings of the study subjects can also have particular value.⁸⁷

Epigenetic Samples for Analyses: considerations for sample collection, tissue-specificity, and cell type mixtures

The tissue specificity of DNA methylation patterns is important to consider when investigating the epigenetic mechanisms of disease, given that environmental factors or disease states might affect DNA methylation changes differently across different tissues. Most epidemiological studies to date have used blood samples because they are easy to obtain.^{28, 46-48, 50, 51, 54, 56, 68, 72, 80, 81, 86} Although blood might be relevant for immune-mediated diseases and blood-based cancers, other targeted tissues, such as adipose or heart tissue, might be more appropriate for other diseases, such as obesity, diabetes, and atherosclerosis.⁸⁸ However, more studies are starting to utilize targeted tissue (adipose, skeletal, heart) or other sources of easily accessible tissues (e.g., saliva, buccal cells) for epigenetic analyses.^{26, 90-95} For example, in a 6-month exercise intervention study among 23 men, adipose tissue samples were used for genome-wide DNA methylation analyses.⁹⁶ Similarly, another large study included analyses of DNA methylation in subcutaneous adipose tissue in relation to BMI in a large group of female participants.⁹²

Recent sequencing and genome-wide array studies have made available detailed information on DNA methylation across multiple tissues. For example, many genome-wide human methylation datasets, along with basic phenotypic information (age, sex, BMI) are now publicly available on Gene Expression Omnibus⁹⁷ for numerous tissues, including liver, adipose, muscle, and blood. Such data repositories are being used to generate exciting epigenetic findings, such as the epigenetic molecular-aging clock discussed previously.^{84, 85}

Given the increasing availability of epigenetic measures in different tissues, an important next step is to better understand the differences and similarities in DNA methylation patterns, both those that occur naturally or in response to environmental insults, as well as those that persist over time compared with those that change rapidly. Although studies are indeed beginning to employ multiple tissues in their analyses,^{59, 90, 92} some human tissues are not readily available (e.g., brain, heart, nerve); therefore, the National Institute of Environmental Health Sciences recently put forth a request for application that called for studies to investigate whether epigenetic changes in accessible surrogate tissues accurately reflect the epigenetic changes in target tissues.

In addition to tissue specificity of DNA methylation, cell-type specificity of DNA methylation adds another layer of complexity. All tissues are composed of a heterogeneous mixture of cells,^{88, 98} each with a distinct DNA methylation signature.⁹⁹ This has important consequences for analysis, as cell-type composition is related to DNA methylation signatures and can change in response to exposures (e.g., smoking, pollutants) or disease states (e.g., immune cell activation and inflammation). Consequently, an association observed between DNA methylation and an exposure or outcome of interest could be due to changes in the cell composition of the tissue analyzed.¹⁰⁰ It is now widely acknowledged that cell-type mixture should be taken into account when conducting epigenetic analyses,^{101, 102} and the development of statistical deconvolution techniques enables researchers to adjust for cell-type mixture in epigenetic association studies.⁹⁸ Until recently, the standard statistical technique relied on reference datasets of cell-type distributions, and although reference datasets are available for blood, they are not available for other tissue

types. Fortunately, ‘reference-free’ statistical methods have been proposed, which allow for cell-mixture adjustment without the need for such reference datasets.^{103, 104} Another common approach is to determine the distribution of major cell types in the target tissue and make statistical adjustments. For example, the complete blood count (CBC) is often used as a reference to determine the distribution of major white blood cell (WBC) types, including neutrophils, lymphocytes, monocytes, eosinophils, and basophils. However, CBCs can be performed only on fresh blood. In addition, a CBC does not cover all types of WBCs and fails to provide information on important subtypes, such as natural killer cells, T cells (including T helper cells, cytotoxic T cells, memory T cells, and regulatory T cells), and B cells, which represent a large proportion of total WBCs. Therefore, adjustment for WBCs does not completely eliminate potential bias due to cell-type distribution.

Technologies for DNA methylation analyses

Easily accessible techniques for DNA methylation analysis, such as bisulfite sequencing, methylation-specific polymerase chain reaction, and pyrosequencing have developed rapidly in the past two decades. Detailed features of the different analytic techniques were reviewed in depth elsewhere,²⁰ and the potential sources of bias associated with each technique were compared (see tables 1 and 2 of Laird²⁰). When selecting an appropriate technology, the goal is to balance genome coverage, resolution, accuracy, specificity, throughput, and cost. Here, we will focus on high-throughput sequencing and microarray-based methods.

The advantages of sequencing-based methods include comprehensive and highly accurate assessment of DNA methylation, as well as assessment of allele-specific and repetitive element DNA methylation states. Whole-genome bisulfite sequencing generates genome-wide quantitative DNA methylation data with nucleotide resolution. However, it can be cost-prohibitive for epidemiologic studies, which prompted the development of “survey approaches” that test a subset of the cytosines throughout the genome. For example, reduced-representation bisulfite sequencing is cost-efficient for large sample sizes. However, reduced-representation bisulfite sequencing is not a comprehensive presentation of the entire genome, as it preferentially covers CpG-rich regions such as promoters.¹⁰⁵

Array-based approaches are currently a more feasible and popular choice, allowing for the analysis of a large number of samples at an affordable cost. The Illumina Infinium Human Methylation 450 BeadChip¹⁰⁶ (450K array)—a platform that allows the assessment of approximately half a million CpG sites across 99% of RefSeq genes within the genome—offers a good compromise between coverage and cost,^{88, 102} and this platform is currently being modified to extend coverage from approximately 450,000 CpGs sites to 850,000. The adaptation of microarray hybridization techniques has facilitated the growth of epigenome-wide studies.^{107, 108} In the Infinium methylation assays,^{109, 110} DNA is treated with bisulfite to convert unmethylated cytosines to uracil, which reveals the location of methylated cytosines.¹⁰⁹ Bisulfite conversion is followed by whole-genome amplification and enzymatic fragmentation; samples are then applied to BeadChips (which contain probes that recognize the methylated and unmethylated DNA), and analyzed by microarray. Finally, the fluorescent signals that are generated reflect the methylation status of the specific target sequence. However, the fluorescent signal can be affected by technical conditions, such as

the hybridization temperature and post-hybridization washing. Therefore, proper experimental design is crucial for accurately assessing methylation intensities, which provide the measures of DNA methylation levels. These technicalities can introduce sources of error that must be minimized. Techniques are described below.

Minimizing batch effect and plate effect

In epidemiologic studies, array-based analyses usually require multiple BeadChips and plates to analyze all the samples in a study. Consequently, technical variation between the different batches (i.e., plates and BeadChips) can be a common problem. More specifically, between-plate differences, between-chip differences, and row-to-row variation can lead to batch effects that bias microarray-based studies. Therefore, proper experimental workflow and experimental design are critical. For example, because measurement of DNA methylation is known to vary by plate and batch, it is important to ensure that the variables of interest (e.g., disease status) are independent of the plating scheme; this can be achieved by randomizing samples from controls and study subjects across plates. Thus, the measurement error of DNA methylation can be assumed to be nondifferential and therefore likely to bias the results towards the null hypothesis (i.e., underestimation of the association) rather than to produce false-positive findings. Statistical analyses can account for known batch effects by adjusting for batch variables in the analytic model or by applying statistical methods (e.g., ComBat using R¹¹¹) to remove sources of variation due to batch effects from the methylation dataset. However, these approaches can be imperfect or incomplete, particularly if the initial experimental design and plating scheme of a study were flawed.

Processing of methylation data, statistical approaches, and analyses

Following in the footsteps of genome-wide association studies (GWAS), epigenome-wide association studies (EWAS) allow systematic assessment of many CpG sites across the genome in relation to the phenotype or disease of interest.^{88, 102} Similar to GWAS, the complexity of high-dimensional epigenome-wide data poses additional analytic challenges.

For example, as the number of loci increases, significant associations can be generated due to chance. Controlling the familywise error rate, or the false discovery rate, must be performed to address the multiple comparisons problem. On the other hand, many CpGs are consistently methylated or unmethylated in human genome; therefore they might not be biologically informative. Including these uninformative CpGs will increase the risk of false positive discoveries and reduce statistical power after multiple testing adjustment. Filtering of CpG sites based on between subject variability has been proposed as a viable strategy to boost statistical power and help detect biologically relevant methylation markers.¹¹² However, currently, there is no consensus on filtering strategies and—as a result—methods for site filtering are not yet standardized. This is an area of methodological work that may bring fruitful development to streamline data analysis and strengthen EWAS findings.

Additional technical issues that are unique to the current technology for EWAS analyses include background fluorescence signal (“background noise”) that can distort measurement of methylation intensities,¹¹³ probe-design-type bias due to mixing two different technological assays (type I and II) in the Infinium 450K,¹¹⁰ and dye bias due to the

measurement technique for type II assay methylation signals.¹¹⁴ In response to these technical problems, several processing and normalization approaches have been proposed and are routinely applied to DNA methylation data prior to analysis.^{106, 115-117} Figure 2 outlines a general algorithm for processing and analyzing methylation data in an array-based EWAS. In addition, the reader is referred to Wilhelm-Benartzi *et al.*⁹⁵ for more technical and computational details.

Similar to the coordinated efforts that have standardized and streamlined GWAS analyses, good practice and accepted standards are essential for EWAS in order to foster the generation of reproducible and comparable results. In the United States and Europe, epidemiologic researchers participating in the Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) recently formed an epigenetic consortium to also study DNA methylation in their studies.¹¹⁸ Similarly, several birth cohorts (e.g., Project VIVA¹¹⁹, ALSPAC¹²⁰) have also succeeded in obtaining methylation measures on cord blood or other samples collected later in life, and many of these cohorts are now participating in the Pregnancy and Child Epigenetics Consortium (PACE). Therefore, EWAS will be able to recapitulate the large meta-analytic efforts that have been typical of GWAS.

Integration of EWAS and GWAS is an emerging avenue of research that will likely provide further biological insight on health and disease. Studies have revealed that the underlying genotype can influence epigenetic variation.^{121, 122} For example, single nucleotide polymorphisms (SNPs) that affect both DNA methylation and gene expression have been identified. Therefore, EWAS can be used to investigate genetic predispositions that exert their function through epigenetic mechanisms.⁸⁸ Loci harboring genetic variants that influence methylation state have been termed methylation quantitative trait loci,¹²² and integration of EWAS and GWAS data in methylation quantitative trait loci studies is advancing.^{88, 121}

Strengthening causality: Discovery and replication, Mendelian randomization, and mediation analyses

Given the challenges involved in EWAS analyses, replication of findings is an important component of epigenetic studies—to better elucidate causality, increase generalizability, and reduce false-positive findings. Discovery-replication study designs can include an internal or external replication component, or both (Table 1). In an internal replication design, discovery and replication components are conducted within a single study cohort by using a two-stage, split-sample design. A proportion of the study population (discovery set) is used to identify the associations between exposure, methylation status, and outcome, and the other proportion (replication set) is used to confirm or validate the findings from the discovery set. In an external replication approach, the discovery-phase and replication-phase analyses are conducted in two or more independent but comparable cohorts. This two-phase approach not only validates the strongest associations (or not), but also ensures the generalizability of the results. Ideally, external replication should be carried out by an independent group of researchers using a different laboratory technique, or even different study design,⁸⁸ as technical problems with the assay or inadequate statistical analysis methods could introduce the same bias into both the discovery and replication analyses.¹⁰²

Because epigenetic studies are susceptible to confounding and reverse causation, it is necessary to address those issues with the appropriate statistical approaches. One emerging strategy is to apply the epidemiological tool of Mendelian randomization¹²³ to epigenetic studies in order to strengthen causation. For Mendelian randomization, a germline genetic variant is used as an instrumental variable in order to establish a causal relationship between an environmentally modifiable exposure (for which the genetic variant serves as a proxy) and a disease outcome of interest. This method can be extended to epigenetic epidemiology studies. For example, a two-step epigenetic Mendelian randomization approach can be applied in a setting where DNA methylation is proposed as a mediator in the causal pathway between a modifiable exposure and disease (Figure 3).¹²⁴ Genetic proxies for smoking behavior,¹²⁵ alcohol consumption,¹²⁶ BMI,¹²⁷ lipid profiles,¹²⁸ or inflammation markers¹²⁹ have been used in Mendelian randomization studies and could be applied to assess whether those factors have a causal impact not just on disease outcome but also on DNA methylation signatures. Epigenetic studies using Mendelian randomization are beginning to be undertaken¹³⁰ as the number of cohort studies that investigate both genetic and epigenetic phenomena is increase.

Mendelian randomization, however, requires well-defined assumptions. The genetic proxy chosen for the environmental exposure or disease outcome should be associated with DNA methylation only through the exposure and should not directly influence DNA methylation itself. For example, the use of genetic variants of the *FTO* gene, which has been shown to have DNA demethylase activity, would not be a suitable choice.¹³¹ Good candidate genetic proxies for DNA methylation include *cis*-SNPs that are correlated with DNA methylation sites in close proximity. In addition, SNPs and DNA methylation status are not robust instrumental variables; therefore large sample sizes are necessary to ensure statistic power and robust findings. Taken together, caution is needed in interpreting the results from these studies. Enthusiasm for an elegant approach such as Mendelian Randomization should not overtake the—often obvious—limitations of applying it to real life data. Readers are referred to Relton *et al.*¹²⁴ for detailed discussion on applying Mendelian randomization for epigenetic studies.

Another useful approach that can strengthen causality in epigenetic epidemiology studies is mediation analysis,¹³² a statistical method that is widely employed in standard epidemiology. Instead of hypothesizing a direct causal relationship between the exposure and the outcome, a mediational model deconstructs a total effect into direct and indirect effects (via the mediator) and clarifies the nature of the relationship between the exposure, mediator, and outcome. By adapting mediation analysis to epigenetic epidemiology, it is possible to investigate whether DNA methylation is a component of the pathway linking the exposure/risk factor and the health outcome under the causal framework.^{71, 133, 134} For example, a recent study demonstrated that decreased methylation of the *F3* gene mediated the effects of black carbon on increased fibrinogen in circulation.⁷¹ Tarantini and coauthors also showed that hypomethylation of the nitric oxide synthase-3 (*NOS3*) or endothelin-1 (*EDNI*) gene is an intermediate mechanism for air pollution-related coagulation effects.¹³⁴

In sum, we outlined several important sequential steps that need to be carefully considered and undertaken in the design and analysis of an epidemiological DNA methylation study. These steps have been summarized in Figure 4.

Potential for Human Intervention and CVD Prevention

By understanding the epigenetic underpinnings of CVD, particularly when the epigenome can serve as a modifiable target for intervention, physicians could directly translate such knowledge into practice. For example, given that DNA methylation depends on a biochemical cycle that requires dietary methyl donors, targeted dietary interventions could consist of consuming foods containing folic acid, vitamin B₆, B₁₂, methionine, betaine, and choline. In addition, DNA methylation is sensitive to nutrients such as zinc,¹³⁵ flavonoid,³⁸ vitamin C,¹³⁶ and niacin.¹³⁷ Therefore, baseline epigenetic analysis can be used to identify individuals who might benefit from personalized nutritional interventions. Both animal and human studies have demonstrated the plasticity of DNA methylation in response to dietary intervention. Among postmenopausal women, for example, intake of a folate-depleted diet for several weeks led to hypomethylation of lymphocyte DNA, and this hypomethylation was reversed when folate intake was subsequently increased.¹³⁸ Another intervention study found that folate supplementation among patients with colorectal adenomatous polyps led to a 31% increase in leukocyte DNA methylation and a 25% increase in methylation of DNA from the colonic mucosa.¹³⁹ Epigenetic changes in response to physical activity interventions have also been observed. Among 23 healthy men with low levels of physical activity at baseline, genome-wide DNA methylation changes were observed after a 6-month exercise intervention. Significantly altered genes included 18 obesity-related genes that were identified by a GWAS.⁹⁶

Intervention on potential epigenetic targets might also be able to reverse the effects of harmful environmental exposures. For example, there is evidence that hypomethylation of the Toll-like receptor 4 (*TLR4*) gene in leukocytes mediates a part of the effect of particulate matter air pollution on blood pressure elevation.¹³³ Furthermore, a recent observational study in a cohort of older men demonstrated that those with higher dietary flavonoid intake were less susceptible to the adverse effects of particulate air pollution on heart rate variability. This finding might be explained by the fact that high flavonoid intake was associated with decreased methylation of an important immune-regulatory gene, Toll-like receptor 2 (*TLR2*).¹⁴⁰ Thus, increased knowledge and biological insight into specific epigenetic responses to environmental exposures could pave the way for targeted epigenetic interventions. However, most available evidence derives either from animal models or from observational human studies. More randomized, controlled human trials incorporating epigenetic measures are needed to open new avenues for applying nutraceutical or pharmaceutical interventions and targeted strategies to treat CVD.

Opportunities and Future Directions

Immunoepigenetics and immunophenotyping

Growing evidence indicates that systemic inflammation is not only an essential mechanistic pathway linking environmental exposures with adverse cardiovascular events,^{67, 141} but it

also exacerbates disturbances of autonomic modulation during external stressors.³ Thus, the role of the immune system is critical in immune-mediated cardiovascular pathology.⁶⁷ DNA methylation has been shown to be a key mechanism modulating the activity of inflammatory genes after exposure to air pollution.^{68, 72, 142-144} In turn, inflammation and oxidative stress can influence both global and gene-specific methylation. Taken together, these findings illustrate the dynamic crosstalk between the epigenome and the immune system and the potential opportunity to target the epigenome to prevent or treat immune-mediated adverse cardiovascular effects. Furthermore, we can use DNA methylation profiles to infer cell-type distribution, or even a specific cell type, which reflect immune processes that might be relevant to exposures or outcomes of interest. This approach might help overcome the barriers posed by expensive and laborious cell sorting and quantifying technologies. Recently, Koestler and coauthors utilized cell mixture deconvolution to infer blood cell counts and demonstrated that *in utero* exposure to inorganic arsenic was associated with a higher proportion of CD8+ T lymphocytes in cord blood.¹⁴⁵

Combine the genome, methylome, and transcriptome: An Integrated approach

The genome, methylome, and transcriptome are functionally interrelated; therefore, creating a comprehensive network is essential to shed light on the health impact of DNA methylation. A few studies have adapted this integrated approach in a gene-specific manner. For example, Movassagh *et al.* found altered methylation profiles at three angiogenesis-related genetic loci in cardiomyopathic heart tissues, and further strengthened their findings by demonstrating that the expression levels of these genes are also modulated. Their finding suggested a causative role of DNA methylation at those loci in the progression of heart failure.²⁶

The rapid development of genome-wide analytical technologies has strengthened our capability to fully exploit the power of EWASs. Studying the genome, methylome, and transcriptome in concert will enhance our understanding of the complex and multidimensional nature of CVDs, and ultimately aid the advance of genomic cardiovascular medicine.

Moving beyond DNA methylation

Recent research found that 5-methylcytosine can be oxidized to 5-hydroxymethylcytosine (5-hmC) in mammalian DNA. Whereas 5-methylcytosine is often considered an epigenetic marker that suppresses gene expression, 5-hmC has been associated with actively transcribed genes. However, the biological mechanisms behind these findings and their implications for CVD are not known.¹⁴⁶⁻¹⁴⁸ Some evidence suggests that 5-hmC might be an intermediate step in DNA demethylation pathways.^{149, 150} Taken together, these emerging mechanisms of epigenetic regulation might open up new avenues to understanding the pathogenesis of CVD and the development of novel therapies.

Although much of the epidemiological research to date has focused on DNA methylation, future epidemiological research needs more focus on other forms of epigenetic regulation (e.g., histone modification post-transcriptional silencing by microRNAs, long and short noncoding RNA, mitochondrial DNA methylation) and should further address how the

interplay between different epigenetic mechanisms relates to CVD risk factors or outcomes. Similarly, future studies need to investigate more thoroughly whether observed DNA methylation changes in fact translate to downstream biological changes (e.g., changes in chromatin accessibility, mRNA, and protein expression).

Conclusion

The contribution of epigenetic regulation to the development of CVD is an active and exciting yet intricate and complex field of research. From proposal and design to analysis and interpretation within a study, important details regarding DNA methylation need careful consideration, including tissue heterogeneity, cell-type heterogeneity, temporal variation, error in measurement of DNA methylation, and statistical implications related to analyzing a large number of correlated CpGs in the genome. Characteristics of a well-designed and well-conducted epigenetic epidemiology study include accurate measurement of DNA methylation (using the most appropriate technological assay), adequate quality control measures, robust statistical techniques tailored for high-dimensional data, and incorporation of validation or replication steps. Further steps, such as Mendelian randomization and mediation analysis, can help strengthen the level of causality. An increasing number of such high-quality studies are now beginning to contribute significantly to the evidence base in cardiovascular epigenetics. Future studies are warranted to better understand the biological and clinical implications of epigenetic mechanisms in the pathophysiology of CVD and the interplay between different forms of epigenetic regulation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of Funding

The authors were supported by grants from the National Institutes of Health to Dr. Baccarelli (R01ES02083603; R01NR013945; R21ES020984; R01ES021357; R01ES021733; P30ES000002).

Nonstandard Abbreviations and Acronyms

CVD	cardiovascular disease
DNMTs	DNA methyltransferases
5-mC	5-methylcytosine
ApoE	apolipoprotein E
PCGTs	polycomb group protein target genes
CBC	complete blood count
WBC	white blood cell
GWAS	genome-wide association studies

EWAS	epigenome-wide association studies
SNPs	single nucleotide polymorphisms
5-hmC	5-hydroxymethylcytosine

References

1. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: The Framingham study. *Am J Public Health Nations Health*. 1951; 41:279–281. [PubMed: 14819398]
2. Ordovas JM, Smith CE. Epigenetics and cardiovascular disease. *Nat Rev Cardiol*. 2010; 7:510–519. [PubMed: 20603647]
3. Baccarelli A, Rienstra M, Benjamin EJ. Cardiovascular epigenetics: Basic concepts and results from animal and human studies. *Circ Cardiovasc Genet*. 2010; 3:567–573. [PubMed: 21156932]
4. Jaenisch R, Bird A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat Genet*. 2003; 33(Suppl):245–254. [PubMed: 12610534]
5. Barrington KG, Zamore PD. MicroRNAs: Regulating a change of heart. *Circulation*. 2009; 119:2217–2224. [PubMed: 19398677]
6. Pons D, de Vries FR, van den Elsen PJ, Heijmans BT, Quax PH, Jukema JW. Epigenetic histone acetylation modifiers in vascular remodelling: New targets for therapy in cardiovascular disease. *Eur Heart J*. 2009; 30:266–277. [PubMed: 19147603]
7. Silvestri P, Di Russo C, Rigattieri S, Fedele S, Todaro D, Ferraiuolo G, Altamura G, Loschiavo P. MicroRNAs and ischemic heart disease: Towards a better comprehension of pathogenesis, new diagnostic tools and new therapeutic targets. *Recent Pat Cardiovasc Drug Discov*. 2009; 4:109–118. [PubMed: 19519553]
8. Epigenetic epidemiology. Springer Netherlands. 2012
9. Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. *Cell*. 2007; 128:669–681. [PubMed: 17320505]
10. Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001; 293:1074–1080. [PubMed: 11498575]
11. Li E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet*. 2002; 3:662–673. [PubMed: 12209141]
12. Morgan HD, Santos F, Green K, Dean W, Reik W. Epigenetic reprogramming in mammals. *Hum Mol Genet*. 2005; 14:R47–58. Spec No 1. [PubMed: 15809273]
13. Zaratiegui M, Irvine DV, Martienssen RA. Noncoding rnas and gene silencing. *Cell*. 2007; 128:763–776. [PubMed: 17320512]
14. Wolffe AP, Matzke MA. Epigenetics: Regulation through repression. *Science*. 1999; 286:481–486. [PubMed: 10521337]
15. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev*. 2002; 16:6–21. [PubMed: 11782440]
16. Wu H, Zhang Y. Reversing DNA methylation: Mechanisms, genomics, and biological functions. *Cell*. 2014; 156:45–68. [PubMed: 24439369]
17. Pinnick KE, Karpe F. DNA methylation of genes in adipose tissue. *Proc Nutr Soc*. 2011; 70:57–63. [PubMed: 21144123]
18. Smith ZD, Meissner A. DNA methylation: Roles in mammalian development. *Nat Rev Genet*. 2013; 14:204–220. [PubMed: 23400093]
19. Wu SC, Zhang Y. Active DNA demethylation: Many roads lead to rome. *Nat Rev Mol Cell Biol*. 2010; 11:607–620. [PubMed: 20683471]
20. Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet*. 2010; 11:191–203. [PubMed: 20125086]

21. Friso S, Pizzolo F, Choi SW, Guarini P, Castagna A, Ravagnani V, Carletto A, Pattini P, Corrocher R, Olivieri O. Epigenetic control of 11 beta-hydroxysteroid dehydrogenase 2 gene promoter is related to human hypertension. *Atherosclerosis*. 2008; 199:323–327. [PubMed: 18178212]
22. Turunen MP, Aavik E, Yla-Herttuala S. Epigenetics and atherosclerosis. *Biochim Biophys Acta*. 2009; 1790:886–891. [PubMed: 19233248]
23. Makar KW, Wilson CB. DNA methylation is a nonredundant repressor of the Th2 effector program. *J Immunol*. 2004; 173:4402–4406. [PubMed: 15383570]
24. Chen Z, Karaplis AC, Ackerman SL, et al. Mice deficient in methylenetetrahydrofolate reductase exhibit hyperhomocysteinemia and decreased methylation capacity, with neuropathology and aortic lipid deposition. *Hum Mol Genet*. 2001; 10:433–443. [PubMed: 11181567]
25. Lund G, Andersson L, Lauria M, Lindholm M, Fraga MF, Villar-Garea A, Ballestar E, Esteller M, Zaina S. DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein e. *J Biol Chem*. 2004; 279:29147–29154. [PubMed: 15131116]
26. Movassagh M, Choy MK, Goddard M, Bennett MR, Down TA, Foo RS. Differential DNA methylation correlates with differential expression of angiogenic factors in human heart failure. *PLoS One*. 2010; 5:e8564. [PubMed: 20084101]
27. Baccarelli A, Wright R, Bollati V, Litonjua A, Zanobetti A, Tarantini L, Sparrow D, Vokonas P, Schwartz J. Ischemic heart disease and stroke in relation to blood DNA methylation. *Epidemiology*. 2010; 21:819–828. [PubMed: 20805753]
28. Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW. DNA methylation as a biomarker for cardiovascular disease risk. *PLoS One*. 2010; 5:e9692. [PubMed: 20300621]
29. Movassagh M, Vujic A, Foo R. Genome-wide DNA methylation in human heart failure. *Epigenomics*. 2011; 3:103–109. [PubMed: 22126157]
30. Talens RP, Jukema JW, Trompet S, Kremer D, Westendorp RG, Lumey LH, Sattar N, Putter H, Slagboom PE, Heijmans BT, Group P. Hypermethylation at loci sensitive to the prenatal environment is associated with increased incidence of myocardial infarction. *Int J Epidemiol*. 2012; 41:106–115. [PubMed: 22101166]
31. Baccarelli A, Ghosh S. Environmental exposures, epigenetics and cardiovascular disease. *Curr Opin Clin Nutr Metab Care*. 2012; 15:323–329. [PubMed: 22669047]
32. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, McQueen M, Budaj A, Pais P, Varigos J, Lisheng L, Investigators IS. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the interheart study): Case-control study. *Lancet*. 2004; 364:937–952. [PubMed: 15364185]
33. Ford ES, Bergmann MM, Kroger J, Schienkiewitz A, Weikert C, Boeing H. Healthy living is the best revenge: Findings from the european prospective investigation into cancer and nutrition-potsdam study. *Arch Intern Med*. 2009; 169:1355–1362. [PubMed: 19667296]
34. Kaplan GA, Keil JE. Socioeconomic factors and cardiovascular disease: A review of the literature. *Circulation*. 1993; 88:1973–1998. [PubMed: 8403348]
35. van Dam RM, Li T, Spiegelman D, Franco OH, Hu FB. Combined impact of lifestyle factors on mortality: Prospective cohort study in us women. *BMJ*. 2008; 337:a1440. [PubMed: 18796495]
36. Gluckman PD, Hanson MA. Developmental and epigenetic pathways to obesity: An evolutionary-developmental perspective. *Int J Obes (Lond)*. 2008; 32(Suppl 7):S62–71. [PubMed: 19136993]
37. Fleisch AF, Wright RO, Baccarelli AA. Environmental epigenetics: A role in endocrine disease? *J Mol Endocrinol*. 2012; 49:R61–67. [PubMed: 22798698]
38. Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL. Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environ Health Perspect*. 2006; 114:567–572. [PubMed: 16581547]
39. Sinclair KD, Allegrucci C, Singh R, Gardner DS, Sebastian S, Bispham J, Thurston A, Huntley JF, Rees WD, Maloney CA, Lea RG, Craigmiles J, McEvoy TG, Young LE. DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional b vitamin and methionine status. *Proc Natl Acad Sci U S A*. 2007; 104:19351–19356. [PubMed: 18042717]
40. Lillycrop KA, Phillips ES, Torrens C, Hanson MA, Jackson AA, Burdge GC. Feeding pregnant rats a protein-restricted diet persistently alters the methylation of specific cytosines in the hepatic ppar alpha promoter of the offspring. *Br J Nutr*. 2008; 100:278–282. [PubMed: 18186951]

41. Lillycrop KA, Slater-Jefferies JL, Hanson MA, Godfrey KM, Jackson AA, Burdge GC. Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced DNA methyltransferase-1 expression is involved in impaired DNA methylation and changes in histone modifications. *Br J Nutr.* 2007; 97:1064–1073. [PubMed: 17433129]
42. Vucetic Z, Kimmel J, Totoki K, Hollenbeck E, Reyes TM. Maternal high-fat diet alters methylation and gene expression of dopamine and opioid-related genes. *Endocrinology.* 2010; 151:4756–4764. [PubMed: 20685869]
43. Lumey LH, Stein AD, Kahn HS, van der Pal-de Bruin KM, Blauw GJ, Zybert PA, Susser ES. Cohort profile: The Dutch hunger winter families study. *Int J Epidemiol.* 2007; 36:1196–1204. [PubMed: 17591638]
44. Painter RC, Roseboom TJ, Bleker OP. Prenatal exposure to the Dutch famine and disease in later life: An overview. *Reprod Toxicol.* 2005; 20:345–352. [PubMed: 15893910]
45. Roseboom T, de Rooij S, Painter R. The dutch famine and its long-term consequences for adult health. *Early Hum Dev.* 2006; 82:485–491. [PubMed: 16876341]
46. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE, Lumey LH. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A.* 2008; 105:17046–17049. [PubMed: 18955703]
47. Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, Slagboom PE, Heijmans BT. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet.* 2009; 18:4046–4053. [PubMed: 19656776]
48. Hoyo C, Murtha AP, Schildkraut JM, Jirtle RL, Demark-Wahnefried W, Forman MR, Iversen ES, Kurtzberg J, Overcash F, Huang Z, Murphy SK. Methylation variation at IGF2 differentially methylated regions and maternal folic acid use before and during pregnancy. *Epigenetics.* 2011; 6:928–936. [PubMed: 21636975]
49. Mathers JC, Strathdee G, Relton CL. Induction of epigenetic alterations by dietary and other environmental factors. *Adv Genet.* 2010; 71:3–39. [PubMed: 20933124]
50. Steegers-Theunissen RP, Obermann-Borst SA, Kremer D, Lindemans J, Siebel C, Steegers EA, Slagboom PE, Heijmans BT. Periconceptional maternal folic acid use of 400 microg per day is related to increased methylation of the igf2 gene in the very young child. *PLoS One.* 2009; 4:e7845. [PubMed: 19924280]
51. Ba Y, Yu H, Liu F, Geng X, Zhu C, Zhu Q, Zheng T, Ma S, Wang G, Li Z, Zhang Y. Relationship of folate, vitamin B12 and methylation of insulin-like growth factor-II in maternal and cord blood. *Eur J Clin Nutr.* 2011; 65:480–485. [PubMed: 21245875]
52. Fang MZ, Wang Y, Ai N, Hou Z, Sun Y, Lu H, Welsh W, Yang CS. Tea polyphenol (-)-epigallocatechin-3-gallate inhibits DNA methyltransferase and reactivates methylation-silenced genes in cancer cell lines. *Cancer Res.* 2003; 63:7563–7570. [PubMed: 14633667]
53. Lim U, Song MA. Dietary and lifestyle factors of DNA methylation. *Methods Mol Biol.* 2012; 863:359–376. [PubMed: 22359306]
54. Zhang FF, Morabia A, Carroll J, Gonzalez K, Fulda K, Kaur M, Vishwanatha JK, Santella RM, Cardarelli R. Dietary patterns are associated with levels of global genomic DNA methylation in a cancer-free population. *J Nutr.* 2011; 141:1165–1171. [PubMed: 21525250]
55. Breton CV, Byun HM, Wenten M, Pan F, Yang A, Gilliland FD. Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. *Am J Respir Crit Care Med.* 2009; 180:462–467. [PubMed: 19498054]
56. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet.* 2011; 88:450–457. [PubMed: 21457905]
57. Leger AJ, Covic L, Kuliopulos A. Protease-activated receptors in cardiovascular diseases. *Circulation.* 2006; 114:1070–1077. [PubMed: 16952995]
58. Joubert BR, Haberg SE, Nilsen RM, et al. 450k epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect.* 2012; 120:1425–1431. [PubMed: 22851337]

59. Monick MM, Beach SR, Plume J, Sears R, Gerrard M, Brody GH, Philibert RA. Coordinated changes in ahrh methylation in lymphoblasts and pulmonary macrophages from smokers. *Am J Med Genet B Neuropsychiatr Genet.* 2012; 159B:141–151. [PubMed: 22232023]
60. Allione A, Marcon F, Fiorito G, Guarrera S, Siniscalchi E, Zijno A, Crebelli R, Matullo G. Novel epigenetic changes unveiled by monozygotic twins discordant for smoking habits. *PLoS One.* 2015; 10:e0128265. [PubMed: 26043106]
61. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, Agusti A, Anderson W, Lomas DA, Demeo DL. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum Mol Genet.* 2012; 21:3073–3082. [PubMed: 22492999]
62. Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H. F2RL3 methylation as a biomarker of current and lifetime smoking exposures. *Environ Health Perspect.* 2014; 122:131–137. [PubMed: 24273234]
63. McElduff P, Dobson A, Beaglehole R, Jackson R. Rapid reduction in coronary risk for those who quit cigarette smoking. *Aust N Z J Public Health.* 1998; 22:787–791. [PubMed: 9889444]
64. Rosenberg L, Palmer JR, Shapiro S. Decline in the risk of myocardial infarction among women who stop smoking. *N Engl J Med.* 1990; 322:213–217. [PubMed: 2294448]
65. Tverdal A, Thelle D, Stensvold I, Leren P, Bjartveit K. Mortality in relation to smoking history: 13 years' follow-up of 68,000 norwegian men and women 35–49 years. *J Clin Epidemiol.* 1993; 46:475–487. [PubMed: 8501474]
66. Brook RD, Franklin B, Cascio W, Hong Y, Howard G, Lipsett M, Luepker R, Mittleman M, Samet J, Smith SC Jr, Tager I, Expert Panel on P, Prevention Science of the American Heart Association. Air pollution and cardiovascular disease: A statement for healthcare professionals from the expert panel on population and prevention science of the American Heart Association. *Circulation.* 2004; 109:2655–2671. [PubMed: 15173049]
67. Brook RD, Rajagopalan S, Pope CA 3rd, et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the american heart association. *Circulation.* 2010; 121:2331–2378. [PubMed: 20458016]
68. Baccarelli A, Wright RO, Bollati V, Tarantini L, Litonjua AA, Suh HH, Zanobetti A, Sparrow D, Vokonas PS, Schwartz J. Rapid DNA methylation changes after exposure to traffic particles. *Am J Respir Crit Care Med.* 2009; 179:572–578. [PubMed: 19136372]
69. Bind MA, Baccarelli A, Zanobetti A, Tarantini L, Suh H, Vokonas P, Schwartz J. Air pollution and markers of coagulation, inflammation, and endothelial function: Associations and epigenvironment interactions in an elderly cohort. *Epidemiology.* 2012; 23:332–340. [PubMed: 22237295]
70. Baccarelli A, Bollati V. Epigenetics and environmental chemicals. *Curr Opin Pediatr.* 2009; 21:243–251. [PubMed: 19663042]
71. Bind MA, Lepeule J, Zanobetti A, Gasparrini A, Baccarelli A, Coull BA, Tarantini L, Vokonas PS, Koutrakis P, Schwartz J. Air pollution and gene-specific methylation in the normative aging study: Association, effect modification, and mediation analysis. *Epigenetics.* 2014; 9:448–458. [PubMed: 24385016]
72. Tarantini L, Bonzini M, Apostoli P, Pegoraro V, Bollati V, Marinelli B, Cantone L, Rizzo G, Hou L, Schwartz J, Bertazzi PA, Baccarelli A. Effects of particulate matter on genomic DNA methylation content and iNOS promoter methylation. *Environ Health Perspect.* 2009; 117:217–222. [PubMed: 19270791]
73. Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, Sparrow D, Vokonas P, Baccarelli A. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev.* 2009; 130:234–239. [PubMed: 19150625]
74. Heyn H, Li N, Ferreira HJ, et al. Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A.* 2012; 109:10522–10527. [PubMed: 22689993]
75. Vanyushin BF, Nemirovsky LE, Klimenko VV, Vasiliev VK, Belozersky AN. The 5-methylcytosine in DNA of rats. Tissue and age specificity and the changes induced by hydrocortisone and other agents. *Gerontologia.* 1973; 19:138–152. [PubMed: 4763637]

76. Wilson VL, Smith RA, Ma S, Cutler RG. Genomic 5-methyldeoxycytidine decreases with age. *J Biol Chem.* 1987; 262:9948–9951. [PubMed: 3611071]
77. Bjornsson HT, Sigurdsson MI, Fallin MD, et al. Intra-individual change over time in DNA methylation with familial clustering. *JAMA.* 2008; 299:2877–2883. [PubMed: 18577732]
78. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, van den Berg LH, Ophoff RA. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.* 2012; 13:R97. [PubMed: 23034122]
79. Rodriguez-Rodero S, Fernandez-Morera JL, Fernandez AF, Menendez-Torre E, Fraga MF. Epigenetic regulation of aging. *Discov Med.* 2010; 10:225–233. [PubMed: 20875344]
80. Teschendorff AE, Menon U, Gentry-Maharaj A, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.* 2010; 20:440–446. [PubMed: 20219944]
81. Bell JT, Tsai PC, Yang TP, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.* 2012; 8:e1002629. [PubMed: 22532803]
82. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, Sugarbaker DJ, Yeh RF, Wiencke JK, Kelsey KT. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet.* 2009; 5:e1000602. [PubMed: 19680444]
83. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P, Spector TD. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res.* 2010; 20:434–439. [PubMed: 20219945]
84. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013; 14:R115. [PubMed: 24138928]
85. Horvath S, Erhart W, Brosch M, et al. Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci U S A.* 2014; 111:15538–15543. [PubMed: 25313081]
86. Marioni RE, Shah S, McRae AF, et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* 2015; 16:25. [PubMed: 25633388]
87. Ng JW, Barrett LM, Wong A, Kuh D, Smith GD, Relton CL. The role of longitudinal cohort studies in epigenetic epidemiology: Challenges and opportunities. *Genome Biol.* 2012; 13:246. [PubMed: 22747597]
88. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet.* 2011; 12:529–541. [PubMed: 21747404]
89. Relton CL, Davey Smith G. Is epidemiology ready for epigenetics? *Int J Epidemiol.* 2012; 41:5–9. [PubMed: 22422447]
90. Agha G, Houseman EA, Kelsey KT, Eaton CB, Buka SL, Loucks EB. Adiposity is associated with DNA methylation profile in adipose tissue. *Int J Epidemiol.* 2014
91. Bouchard L, Rabasa-Lhoret R, Faraj M, Lavoie ME, Mill J, Perusse L, Vohl MC. Differential epigenomic and transcriptomic responses in subcutaneous adipose tissue between low and high responders to caloric restriction. *Am J Clin Nutr.* 2010; 91:309–320. [PubMed: 19939982]
92. Dick KJ, Nelson CP, Tsaprouni L, et al. DNA methylation and body-mass index: A genome-wide analysis. *Lancet.* 2014
93. Ma B, Wilker EH, Willis-Owen SA, Byun HM, Wong KC, Motta V, Baccarelli AA, Schwartz J, Cookson WO, Khabbaz K, Mittleman MA, Moffatt MF, Liang L. Predicting DNA methylation level across human tissues. *Nucleic Acids Res.* 2014; 42:3515–3528. [PubMed: 24445802]
94. Movassagh M, Choy MK, Knowles DA, Cordeddu L, Haider S, Down T, Siggins L, Vujic A, Simeoni I, Penkett C, Goddard M, Lio P, Bennett MR, Foo RS. Distinct epigenomic features in end-stage failing human hearts. *Circulation.* 2011; 124:2411–2422. [PubMed: 22025602]
95. Wilhelm-Benartzi CS, Koestler DC, Karagas MR, Flanagan JM, Christensen BC, Kelsey KT, Marsit CJ, Houseman EA, Brown R. Review of processing and analysis methods for DNA methylation array data. *Br J Cancer.* 2013; 109:1394–1402. [PubMed: 23982603]
96. Ronn T, Volkov P, Davegardh C, Dayeh T, Hall E, Olsson AH, Nilsson E, Tornberg A, Dekker Nitert M, Eriksson KF, Jones HA, Groop L, Ling C. A six months exercise intervention influences

- the genome-wide DNA methylation pattern in human adipose tissue. *PLoS Genet.* 2013; 9:e1003572. [PubMed: 23825961]
97. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002; 30:207–210. [PubMed: 11752295]
  98. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012; 13:86. [PubMed: 22568884]
  99. Wu HC, Delgado-Cruzata L, Flom JD, Kappil M, Ferris JS, Liao Y, Santella RM, Terry MB. Global methylation profiles in DNA from different blood cell types. *Epigenetics.* 2011; 6:76–85. [PubMed: 20890131]
  100. Lowe R, Rakyan VK. Correcting for cell-type composition bias in epigenome-wide association studies. *Genome Med.* 2014; 6:23. [PubMed: 25031617]
  101. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* 2014; 15:R31. [PubMed: 24495553]
  102. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Grealley JM, Gut I, Houseman EA, Izzi B, Kelsey KT, Meissner A, Milosavljevic A, Siegmund KD, Bock C, Irizarry RA. Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods.* 2013; 10:949–955. [PubMed: 24076989]
  103. Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics.* 2014; 30:1431–1439. [PubMed: 24451622]
  104. Zou J, Lippert C, Heckerman D, Aryee M, Listgarten J. Epigenome-wide association studies without the need for cell-type composition. *Nat Methods.* 2014; 11:309–311. [PubMed: 24464286]
  105. Ulahannan N, Grealley JM. Genome-wide assays that identify and quantify modified cytosines in human disease studies. *Epigenetics Chromatin.* 2015; 8:5. [PubMed: 25788985]
  106. Lowe R, Morris T. Report on the 2nd annual Infinium humanmethylation450 array workshop: 15 april 2013 qmul, london, uk. *Epigenetics.* 2013; 8:1123–1124. [PubMed: 23949542]
  107. Balog RP, de Souza YE, Tang HM, DeMasellis GM, Gao B, Avila A, Gaban DJ, Mittelman D, Minna JD, Luebke KJ, Garner HR. Parallel assessment of CpG methylation by two-color hybridization with oligonucleotide arrays. *Anal Biochem.* 2002; 309:301–310. [PubMed: 12413464]
  108. Hatada I, Kato A, Morita S, Obata Y, Nagaoka K, Sakurada A, Sato M, Horii A, Tsujimoto A, Matsubara K. A microarray-based method for detecting methylated loci. *J Hum Genet.* 2002; 47:448–451. [PubMed: 12181645]
  109. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL. Genome-wide DNA methylation profiling using Infinium(r) assay. *Epigenomics.* 2009; 1:177–200. [PubMed: 22122642]
  110. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium methylation 450k technology. *Epigenomics.* 2011; 3:771–784. [PubMed: 22126295]
  111. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics.* 2007; 8:118–127. [PubMed: 16632515]
  112. Meng H, Joyce AR, Adkins DE, Basu P, Jia Y, Li G, Sengupta TK, Zedler BK, Murrelle EL, van den Oord EJ. A statistical method for excluding non-variable CpG sites in high-throughput DNA methylation profiling. *BMC Bioinformatics.* 2010; 11:227. [PubMed: 20441598]
  113. Triche TJ Jr, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA methylation beadarrays. *Nucleic Acids Res.* 2013; 41:e90. [PubMed: 23476028]
  114. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics.* 2013; 14:293. [PubMed: 23631413]
  115. Marabita F, Almgren M, Lindholm ME, Ruhmann S, Fagerstrom-Billai F, Jagodic M, Sundberg CJ, Ekstrom TJ, Teschendorff AE, Tegner J, Gomez-Cabrero D. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina humanmethylation450 beadchip platform. *Epigenetics.* 2013; 8:333–346. [PubMed: 23422812]

116. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450K DNA methylation data. *Bioinformatics*. 2013; 29:189–196. [PubMed: 23175756]
117. Wu MC, Joubert BR, Kuan PF, Haberg SE, Nystad W, Peddada SD, London SJ. A systematic assessment of normalization approaches for the Infinium 450K methylation platform. *Epigenetics*. 2014; 9:318–329. [PubMed: 24241353]
118. Psaty BM, Sitlani C. The cohorts for heart and aging research in genomic epidemiology (charge) consortium as a model of collaborative science. *Epidemiology*. 2013; 24:346–348. [PubMed: 23549178]
119. Rich-Edwards J, Krieger N, Majzoub J, Zierler S, Lieberman E, Gillman M. Maternal experiences of racism and violence as predictors of preterm birth: Rationale and study design. *Paediatr Perinat Epidemiol*. 2001; 15(Suppl 2):124–135. [PubMed: 11520405]
120. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort profile: The 'children of the 90s'--the index offspring of the avon longitudinal study of parents and children. *Int J Epidemiol*. 2013; 42:111–127. [PubMed: 22507743]
121. Bell CG, Finer S, Lindgren CM, et al. Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the fto type 2 diabetes and obesity susceptibility locus. *PLoS One*. 2010; 5:e14040. [PubMed: 21124985]
122. Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, Craig DW, Redman M, Gershon ES, Liu C. Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet*. 2010; 86:411–419. [PubMed: 20215007]
123. Ebrahim S, Davey Smith G. Mendelian randomization: Can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet*. 2008; 123:15–33. [PubMed: 18038153]
124. Relton CL, Davey Smith G. Two-step epigenetic mendelian randomization: A strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol*. 2012; 41:161–176. [PubMed: 22422451]
125. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, Poole C. Illustrating bias due to conditioning on a collider. *Int J Epidemiol*. 2010; 39:417–420. [PubMed: 19926667]
126. Chen L, Smith GD, Harbord RM, Lewis SJ. Alcohol intake and blood pressure: A systematic review implementing a mendelian randomization approach. *PLoS Med*. 2008; 5:e52. [PubMed: 18318597]
127. Timpson NJ, Harbord R, Davey Smith G, Zacho J, Tybjaerg-Hansen A, Nordestgaard BG. Does greater adiposity increase blood pressure and hypertension risk?: Mendelian randomization using the FTO/MC4R genotype. *Hypertension*. 2009; 54:84–90. [PubMed: 19470880]
128. Benn M, Tybjaerg-Hansen A, Stender S, Frikke-Schmidt R, Nordestgaard BG. Low-density lipoprotein cholesterol and the risk of cancer: A mendelian randomization study. *J Natl Cancer Inst*. 2011; 103:508–519. [PubMed: 21285406]
129. Davey Smith G, Harbord R, Ebrahim S. Fibrinogen, C-reactive protein and coronary heart disease: Does mendelian randomization suggest the associations are non-causal? *QJM*. 2004; 97:163–166. [PubMed: 14976273]
130. Binder AM, Michels KB. The causal effect of red blood cell folate on genome-wide methylation in cord blood: A mendelian randomization approach. *BMC Bioinformatics*. 2013; 14:353. [PubMed: 24305512]
131. Gerken T, Girard CA, Tung YC, et al. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science*. 2007; 318:1469–1472. [PubMed: 17991826]
132. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992; 3:143–155. [PubMed: 1576220]
133. Bellavia A, Uch B, Speck M, Brook RD, Scott JA, Albeti B, Behbod B, North M, Valeri L, Bertazzi PA, Silverman F, Gold D, Baccarelli AA. DNA hypomethylation, ambient particulate matter, and increased blood pressure: Findings from controlled human exposure experiments. *J Am Heart Assoc*. 2013; 2:e000212. [PubMed: 23782920]

134. Tarantini L, Bonzini M, Tripodi A, Angelici L, Nordio F, Cantone L, Apostoli P, Bertazzi PA, Baccarelli AA. Blood hypomethylation of inflammatory genes mediates the effects of metal-rich airborne pollutants on blood coagulation. *Occup Environ Med*. 2013; 70:418–425. [PubMed: 23476046]
135. Dreosti IE. Zinc and the gene. *Mutat Res*. 2001; 475:161–167. [PubMed: 11295161]
136. Halliwell B. Vitamin c and genomic stability. *Mutat Res*. 2001; 475:29–35. [PubMed: 11295151]
137. Hageman GJ, Stierum RH. Niacin, poly(adp-ribose) polymerase-1 and genomic stability. *Mutat Res*. 2001; 475:45–56. [PubMed: 11295153]
138. Jacob RA, Gretz DM, Taylor PC, James SJ, Pogribny IP, Miller BJ, Henning SM, Swendseid ME. Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. *J Nutr*. 1998; 128:1204–1212. [PubMed: 9649607]
139. Pufulete M, Al-Ghnaniem R, Khushal A, Appleby P, Harris N, Gout S, Emery PW, Sanders TA. Effect of folic acid supplementation on genomic DNA methylation in patients with colorectal adenoma. *Gut*. 2005; 54:648–653. [PubMed: 15831910]
140. Zhong J, Colicino E, Lin X, et al. Cardiac autonomic dysfunction: Particulate air pollution effects are modulated by epigenetic immunoregulation of Toll-like receptor 2 and dietary flavonoid intake. *J Am Heart Assoc*. 2015; 4:e001423. [PubMed: 25628407]
141. Stone PH, Godleski JJ. First steps toward understanding the pathophysiologic link between air pollution and cardiac mortality. *Am Heart J*. 1999; 138:804–807. [PubMed: 10539808]
142. Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, Cavallo D, Byun HM, Jiang J, Marinelli B, Pesatori AC, Bertazzi PA, Yang AS. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. *Cancer Res*. 2007; 67:876–880. [PubMed: 17283117]
143. Lund G, Zaina S. Atherosclerosis risk factors can impose aberrant DNA methylation patterns: A tale of traffic and homocysteine. *Curr Opin Lipidol*. 2009; 20:448–449. [PubMed: 19741342]
144. Madrigano J, Baccarelli A, Mittleman MA, Wright RO, Sparrow D, Vokonas PS, Tarantini L, Schwartz J. Prolonged exposure to particulate pollution, genes associated with glutathione pathways, and DNA methylation in a cohort of older men. *Environ Health Perspect*. 2011; 119:977–982. [PubMed: 21385671]
145. Koestler DC, Avissar-Whiting M, Houseman EA, Karagas MR, Marsit CJ. Differential DNA methylation in umbilical cord blood of infants exposed to low levels of arsenic in utero. *Environ Health Perspect*. 2013; 121:971–977. [PubMed: 23757598]
146. Branco MR, Ficuz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet*. 2012; 13:7–13. [PubMed: 22083101]
147. Pastor WA, Pape UJ, Huang Y, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*. 2011; 473:394–397. [PubMed: 21552279]
148. Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev*. 2011; 25:679–684. [PubMed: 21460036]
149. Shirodkar AV, Marsden PA. Epigenetics in cardiovascular disease. *Curr Opin Cardiol*. 2011; 26:209–215. [PubMed: 21415727]
150. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by mll partner tet1. *Science*. 2009; 324:930–935. [PubMed: 19372391]

Figure 1. DNA methylation is usually associated with inactive transcription and suppressed gene expression. The figure depicts the molecular mechanism linking DNA methylation and inactive transcription. Cytosine is methylated to 5-methylcytosine by DNMT. Binding of a methyl CpG-binding protein to methylated sequences prevents access to this sequence by TFs, thereby suppressing transcription. DNMT = DNA methyltransferase. SAM = S-adenosylmethionine. SAH = S-adenosylhomocysteine. TF = transcription factor. MBP = methyl CpG-binding protein.

Figure 2.
Processing and analysis pipeline for an array-based DNA methylation study.

Step 1

Step 2

Figure 3.

Two-step epigenetic Mendelian randomization. In Mendelian randomization, step 1 uses a genetic proxy for the modifiable exposure in order to assess the causal impact of the modifiable exposure on DNA methylation (the mediator). In step 2, a genetic proxy for DNA methylation is needed to evaluate the causal relationship between the methylation mediator and disease outcome (Figure adapted from Relton and Davey Smith, *Int J Epidemiol* 2012).

Figure 4.
Design and analysis of an epidemiological DNA methylation study.

Table 1
Common discovery-replication study designs in epigenome-wide association studies

Study Design				
	Discovery Only	Internal Replication	External Replication	
Study Samples for Analysis	1	2	2	
Study Population	Discovery	Original study population	Original study population	Original study population
	Replication	Original study population	At least one other independent study	
Analytical Strategy	Discovery	Epigenome-wide scan	Epigenome-wide scan	Epigenome-wide scan
	Replication	Candidate gene/locus analysis	Epigenome-wide scan or candidate gene/locus analysis	
Advantages	1. Relatively low cost 2. Easy to operate	1. Easy to operate 2. Minimize false-positive findings	1. Ensures result validation 2. Enhances generalizability	
Challenges	1. False-positive findings 2. Lack of generalizability	1. Does not augment generalizability 2. Should not be used when epigenome-wide data are available in both the discovery set(s) and replication set(s) (loss of power)	1. Cost and limited availability of target tissue samples 2. Requires collaborations across different studies	