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

© 2016 American Heart Association, Inc. Circulation Research is available at http://circres.ahajournals.org 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,* Golareh Agha,* Andrea A. Baccarelli

Abstract: Epidemiological studies have demonstrated that genetic, environmental, behavioral, and clinical factors contribute to cardiovascular disease development. How these risk factors interact at the cellular level to cause cardiovascular disease 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 cause and biological implications. In this review, we provide a brief overview of the biology and mechanisms of DNA methylation and its role in cardiovascular disease. 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. (Circ Res. 2016;118:119-131. DOI: 10.1161/CIRCRESAHA.115.305206.)

Key Words: 5-methylcytosine cardiovascular diseases DNA methylation epidemiology risk factors
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 the external environmental influences, thus they might provide better mechanistic insight into the cause of CVD. Epigenetic changes can occur via DNA methylation, histone modification, and post-transcriptional silencing mediated by micro-RNAs (miRNAs). Although different forms of epigenetic regulation have been investigated in relation to cardiovascular biology and pathology, DNA methylation has been most studied in large population studies because it is most feasible to measure in an epidemiological setting. 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.

### Biology of DNA Methylation

DNA methylation, a pretranscriptional 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). In somatic cells, DNA methylation occurs predominantly on cytosine residues of the dinucleotide sequence CpG. CpG dinucleotides are distributed unevenly throughout the genome, and in normal, healthy cells ≈80% of CpGs are methylated.

The early postconceptional period is a critical window during which DNA methylation patterns are established. 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.

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). This modification is considered to be stable through cell differentiation and is retained during cell division. Of the 3 DNMTs, DNMT1 is required to maintain established methylation patterns, whereas DNMT3a and DNMT3b are required for de novo methylation. The functions of DNMTs are discussed in detail elsewhere.

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 because of inhibited or absent DNMT1 during DNA replication.

### DNA Methylation in CVD

Comprehensive reviews and studies suggest that changes in DNA methylation states contribute to the regulation of biological processes underlying CVD, such as atherosclerosis, hypertension, and inflammation. For example, mice with a hypomethylated genome had elevated expression of inflammatory markers, and DNA hypomethylation was shown to precede the formation of aortic fatty streaks. In atherosclerosis-prone ApoE-null mice, DNA methylation changes occurred in both peripheral blood leukocytes and the aorta before the formation of vascular lesions. 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.

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. Specifically, elevated methylation of ALU in leukocytes was associated with prevalence of CVD and obesity in Chinese individuals. In the Normative Aging Study conducted in the Boston area, subjects with prevalent ischemic heart disease and stroke had lower blood LINE-1 methylation, whereas in longitudinal analyses, lower LINE-1 methylation was associated with higher risk of ischemic heart disease and stroke. Patients with heart failure were found to have altered promoter methylation in genes relevant to myocyte apoptosis, fibrosis, and altered contractility. 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.

Many gaps still remain in understanding the mechanisms by which epigenetic changes may contribute to atherosclerotic CVD, and inconsistencies exist in the evidence to date, as the studies discussed above suggest. Studies of global DNA methylation in CVD patients report both increased and decreased DNA methylation. 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.

### Lifestyle, Environment, Aging, and DNA Methylation

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

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. 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. 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. Some of that evidence is discussed below (Major findings discussed in the remainder of this section are available in the Online Table I).

**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 and 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. 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 than the control animals.

Protein restriction in pregnant rats reduced DNA methylation in the promoter of key metabolic regulatory genes in first- and second-generation progeny. 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 than the control animals.

Protein restriction in pregnant rats reduced DNA methylation in the promoter of key metabolic regulatory genes in first- and second-generation progeny. 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 than the control animals.

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. Famine during fetal development was also related to DNA methylation changes in genes implicated in growth and metabolic disease (Online Table I). 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 although a study of serum folate levels did not find evidence of such associations (Online Table I). Flavonoids, which may exert protective effects against heart disease, have been shown to decrease gene-specific methylation, particularly by affecting DNMT.

**Smoking**

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

In a study of maternal plasma cotinine level (an objective biomarker of smoking), maternal cotinine displayed a dose-dependent association with lower methylation level in the aryl-hydrocarbon 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.89 A recent monozygotic twin study60 found differential methylation between the smoking and the nonsmoking twins in 8 loci that had been reported in previous studies, including loci in AHRR and F2RL3. Differential DNA methylation has also been observed with respect to current versus former smoking, cumulative smoking exposure (eg, 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 and long-term risk of heart disease and stroke.66 Studies have also shown that ambient particulate matter,56,69 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 were far more likely to be methylated with age than promoters of non–polycomb group protein target genes, regardless of sex, tissue type, disease state, and methylation assay used.80

Recently, Horvath84,85 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 patients with cancer, morbidly obese individuals, and patients with HIV, 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 previous studies.57–69 In observational studies, lifestyle and environmental factors (eg, 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 previous epigenetic epidemiologic studies are not ideal,56,67,70,81 as conclusions about directionality or causality cannot be drawn. In addition, studies have typically measured DNA methylation status at only 1 time point.46–48,50,51,54,55 Because natural variation in epigenetic states occurs over time, epigenetic patterns measured at 1 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 size46,47,50,51,54,56,60,81; larger sample sizes are required to ensure adequate statistical power.57–59

Given these challenges, longitudinal cohort studies, particularly birth cohorts, are important for understanding epigenetic contributions to health and disease.58,85 The incorporation of epigenome-wide DNA methylation data into such cohort studies is now occurring in large cohorts in the United States and Europe.66 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.46–50,51,54,56,68,72,80,81,86 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 mellitus, and atherosclerosis.84 However, more studies are starting to use targeted tissue (adipose, skeletal, and heart) or other sources of easily accessible tissues (eg, saliva and 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 body mass index 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 data sets, along with basic phenotypic information (age, sex, and body mass index) 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.98

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 when compared with those that change rapidly. Although studies are indeed beginning to use multiple tissues in their analyses,59,90,92 some human tissues are not readily available (eg, brain, heart, and 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,88–90 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 (eg, smoking, pollutants) or disease states (eg, immune cell activation and inflammation). Consequently, an association observed between DNA methylation and an exposure or outcome of interest could be because of 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 data sets of cell-type distribution, and although reference data sets 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 data sets.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 is often used as a reference to determine the distribution of major white blood cell types, including neutrophils, lymphocytes, monocytes, eosinophils, and basophils. However, complete blood counts can be performed only on fresh blood. In addition, a complete blood count does not cover all types of white blood cells 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 white blood cells. Therefore, adjustment for white blood cells 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 2 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 (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, it does not provide comprehensive representation 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.100,102 This platform is currently being modified to extend coverage from ≈450000 CpGs sites to 850000. 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 posthybridization 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 (ie, 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 (eg, 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 toward the null hypothesis (ie, 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 (eg, 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 because of chance. Controlling the familywise error rate, or the false discovery rate, must be performed to address the multiple comparisons problem. On the contrary, 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 because of mixing 2 different technological assays (type I and II) in the Infinium 450K,108 and dye bias because of 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 before 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 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 recently formed an epigenetic consortium to also study DNA methylation in their studies.118 Similarly, several birth cohorts (eg, Project VIVA119 and ALSPAC [Avon Longitudinal Study of Parents and Children]120) 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. 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.\textsuperscript{121,122} For example, single nucleotide polymorphisms 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.\textsuperscript{88} Loci harboring genetic variants that influence methylation state have been termed methylation quantitative trait loci,\textsuperscript{122} and integration of EWAS and GWAS data in methylation quantitative trait loci studies is advancing.\textsuperscript{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). In an internal replication design, discovery and replication components are conducted within a single study cohort by using a 2-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 ≥2 independent but comparable cohorts. This 2-phase approach not only validates the strongest associations (or not) but also ensures the generalizability of the results. Ideally, external replication should be performed by an independent group of researchers using a different laboratory technique, or even different study design,\textsuperscript{88} as technical problems with the assay or inadequate statistical analysis methods could introduce the same bias into both the discovery and the replication analyses.\textsuperscript{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 randomization\textsuperscript{123} to epigenetic studies to strengthen causation. For Mendelian randomization, a germline genetic variant is used as an instrumental variable 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 2-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).\textsuperscript{124} Genetic proxies for smoking behavior,\textsuperscript{125} alcohol consumption,\textsuperscript{126} body mass index,\textsuperscript{127} lipid profiles,\textsuperscript{128} or inflammation markers\textsuperscript{129} have been used in Mendelian randomization studies and could be applied to assess whether those factors have a causal impact not only on disease outcome but also on DNA methylation signatures. Epigenetic studies using Mendelian randomization are beginning to be undertaken\textsuperscript{130} as the number of cohort studies that investigate both genetic and epigenetic phenomena 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 \textit{FTO} gene, which has been shown to have DNA demethylase activity, would not be a suitable choice.\textsuperscript{131} Good candidate genetic proxies for DNA methylation include \textit{cis}-single nucleotide polymorphisms that are correlated with DNA methylation sites in close proximity. In addition, single nucleotide polymorphisms 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 and Davey Smith\textsuperscript{124} for detailed discussion on applying Mendelian randomization for epigenetic studies.

Another useful approach that can strengthen causality in epigenomic epidemiology studies is mediation analysis,\textsuperscript{132} a statistical method that is widely used 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.\textsuperscript{71,133,134} For example, a recent study demonstrated that decreased methylation of the \textit{F3} gene mediated the effects of black carbon on increased fibrinogen in circulation.\textsuperscript{71} Tarantini et al\textsuperscript{134} also showed that hypomethylation of the nitric oxide synthase-3 (\textit{NOS3}) or endothelin-1 (\textit{EDN1}) gene is an intermediate mechanism for air pollution–related coagulation effects.

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 folate acid, vitamin B\textsubscript{9}, B\textsubscript{12}, methionine, betaine, and choline. In addition, DNA methylation
is sensitive to nutrients such as zinc, flavonoid, vitamin C, and niacin. 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. 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. 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.

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. 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). 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.

### Table. Common Discovery-Replication Study Designs in Epigenome-Wide Association Studies

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Discovery Only</th>
<th>Internal Replication</th>
<th>External Replication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study samples for analysis</td>
<td>1</td>
<td>≥2</td>
<td>≥2</td>
</tr>
<tr>
<td>Study population</td>
<td>Original study population</td>
<td>Original study population</td>
<td>Original study population</td>
</tr>
<tr>
<td>Discovery</td>
<td>N/A</td>
<td>N/A</td>
<td>At least 1 other independent study</td>
</tr>
<tr>
<td>Analytical strategy</td>
<td>Epigenome-wide scan</td>
<td>Candidate gene/locus analysis</td>
<td>Epigenome-wide scan or candidate gene/locus analysis</td>
</tr>
<tr>
<td>Advantages</td>
<td>Relatively low cost</td>
<td>Easy to operate</td>
<td>Ensures result validation</td>
</tr>
<tr>
<td>Challenges</td>
<td>False-positive findings</td>
<td>Does not augment generalizability</td>
<td>Cost and limited availability of target tissue samples</td>
</tr>
<tr>
<td>N/A indicates not applicable.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Step 1**

**Figure 3. Two-step epigenetic Mendelian randomization.** In Mendelian randomization, step 1 uses a genetic proxy for the modifiable exposure 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 the disease outcome. SNP indicates single nucleotide polymorphism. Adapted from Relton and Davey Smith with permission of the publisher. Copyright ©2012, Oxford University Press.
Opportunities and Future Directions

Immunoeigenetics and Immunophenotyping
Growing evidence indicates that systemic inflammation not only is an essential mechanistic pathway linking environmental exposures with adverse cardiovascular events but also exacerbates disturbances of autonomic modulation during external stressors. Thus, the role of the immune system is critical in immune-mediated cardiovascular pathology. DNA methylation has been shown to be a key mechanism modulating the activity of inflammatory genes after exposure to air pollution. In turn, inflammation and oxidative stress can influence both global and gene-specific methylation. Taken together, these findings illustrate the dynamic cross talk 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 et al used 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.

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 3 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.

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-mC can be oxidized to 5-hydroxymethylcytosine in mammalian DNA. Whereas 5-mC is often considered an epigenetic marker that suppresses gene expression, 5-hydroxymethylcytosine has been associated with actively transcribed genes. However, the biological mechanisms behind these findings and their implications for CVD are not known. Some evidence suggests that 5-hydroxymethylcytosine might be an intermediate step in DNA demethylation pathways. 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 (eg, histone modification post-transcriptional silencing by micro-RNAs, long and short noncoding RNA, and mitochondrial DNA methylation) and should further address how the interplay between different epigenetic mechanisms relates to CVD risk factors or outcomes. Similarly, additional studies need to investigate more thoroughly whether observed DNA methylation changes in fact translate to downstream biological changes (eg, changes in chromatin accessibility, mRNA, and protein expression).

Conclusions
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 on 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. Additional 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.

Sources of Funding

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

Disclosures

None.

References

DNA Methylation in Cardiovascular Epidemiology


DNA Methylation in Cardiovascular Epidemiology


The Role of DNA Methylation in Cardiovascular Risk and Disease: Methodological Aspects, Study Design, and Data Analysis for Epidemiological Studies
Jia Zhong, Golareh Agha and Andrea A. Baccarelli

Circ Res. 2016;118:119-131
doi: 10.1161/CIRCRESAHA.115.305206

Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2016 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/118/1/119

An erratum has been published regarding this article. Please see the attached page for:
/content/118/3/e30.full.pdf

Data Supplement (unedited) at:
http://circres.ahajournals.org/content/suppl/2016/01/12/CIRCRESAHA.115.305206.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org//subscriptions/

In the Figure 1 legend, the expansion of “MBP” has been corrected to “methyl CpG-binding protein.”

The compositor apologizes for the error, and the correction appears in the online version of the article, which is available at http://circres.ahajournals.org/content/118/1/119.full
Supplementary Table I: Lifestyle, environment, aging, and DNA methylation: major findings from epidemiological studies

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Details</th>
<th>Tissue Analyzed</th>
<th>Gene Studied</th>
<th>Observation</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diet</td>
<td>Prenatal famine</td>
<td>Peripheral blood (offspring)</td>
<td>IGF2</td>
<td>Lower methylation</td>
<td>46</td>
</tr>
<tr>
<td>Diet</td>
<td>Prenatal famine</td>
<td>Peripheral blood (offspring)</td>
<td>15 candidate genes</td>
<td>Higher methylation: IL10, GNASAS, LEP, MEG3, ABCA1; Lower methylation: INSIGF</td>
<td>47</td>
</tr>
<tr>
<td>Diet</td>
<td>Prenatal famine</td>
<td>Peripheral blood (offspring)</td>
<td>IGF2, H19</td>
<td>Lower methylation: H19</td>
<td>48</td>
</tr>
<tr>
<td>Diet</td>
<td>Prenatal folic acid intake</td>
<td>Umbilical cord blood</td>
<td>IGF2</td>
<td>Higher methylation</td>
<td>50</td>
</tr>
<tr>
<td>Diet</td>
<td>Prenatal folic acid intake</td>
<td>Umbilical cord blood</td>
<td>IGF2</td>
<td>Lower methylation: IL1</td>
<td>48</td>
</tr>
<tr>
<td>Diet</td>
<td>Serum folate</td>
<td>Maternal blood</td>
<td>IGF2</td>
<td>Higher methylation</td>
<td>50</td>
</tr>
<tr>
<td>Diet</td>
<td>Serum B12</td>
<td>Maternal blood</td>
<td>IGF2</td>
<td>No association</td>
<td>51</td>
</tr>
<tr>
<td>Diet</td>
<td>Prudent diet</td>
<td>Peripheral blood</td>
<td>LINE-1</td>
<td>Lower methylation of LINE-1</td>
<td>54</td>
</tr>
<tr>
<td>Smoking</td>
<td>Maternal smoking</td>
<td>Buccal cells (offspring)</td>
<td>LINE-1, Alu, and 8 candidate genes</td>
<td>More hypomethylated CpGs</td>
<td>55</td>
</tr>
<tr>
<td>Smoking</td>
<td>Maternal smoking</td>
<td>Umbilical cord blood</td>
<td>Epigenome-wide</td>
<td>Lower methylation: F2RL3</td>
<td>56</td>
</tr>
<tr>
<td>Smoking</td>
<td>Maternal plasma cotinine level</td>
<td>Umbilical cord blood</td>
<td>Epigenome-wide</td>
<td>Differential methylation: AHRR, CYP1A1, GF11</td>
<td>58</td>
</tr>
<tr>
<td>Smoking</td>
<td>Smoking</td>
<td>Lymphoblasts and lung alveolar macrophage</td>
<td>Epigenome-wide</td>
<td>Lower methylation: AHRR</td>
<td>59</td>
</tr>
<tr>
<td>Air Pollution</td>
<td>Particulate matter</td>
<td>Peripheral blood</td>
<td>LINE-1, Alu</td>
<td>Lower methylation: LINE-1</td>
<td>68</td>
</tr>
<tr>
<td>Air Pollution</td>
<td>Particulate matter, ozone</td>
<td>Peripheral blood</td>
<td>5 candidate genes</td>
<td>Lower methylation: F3, TLR2, ICAM1; Higher methylation: IFNγ, IL6</td>
<td>71</td>
</tr>
<tr>
<td>Aging</td>
<td>Particulate matter</td>
<td>Peripheral blood</td>
<td>LINE-1, Alu, and iNOS</td>
<td>Lower methylation</td>
<td>72</td>
</tr>
<tr>
<td>Aging</td>
<td>CD4+ T cells</td>
<td>Peripheral blood</td>
<td>LINE-1, Alu</td>
<td>Lower methylation: Alu</td>
<td>73</td>
</tr>
<tr>
<td>Aging</td>
<td>Epigenome-wide</td>
<td>Epigenome-wide</td>
<td>More hypomethylated CpGs</td>
<td>Lower methylation: Alu</td>
<td>74</td>
</tr>
</tbody>
</table>

Note: Ref indicates the corresponding reference ID in the manuscript.