The Influences of Genetic and Environmental Factors on Methylome-Wide Association Studies for Human Diseases

Yan V. Sun

Abstract DNA methylation (DNAm) is an essential epigenetic mechanism for normal development, and its variation may be associated with diseases. High-throughput technology allows robust measurement of DNA methylome in population studies. Methylome-wide association studies (MWAS) scan the DNA methylome to detect new epigenetic loci affecting disease susceptibility. MWAS is an emerging approach to unraveling the mechanism linking genetics, environment, and human diseases. Here I review the recent studies of genetic determinants and environmental modifiers of DNAm, and the concept for partitioning genetic and environmental contribution to DNAm. These studies establish the correlation maps between genome and methylome, and enable the interpretation of epigenetic association with disease traits. Recent findings suggested that MWAS was a promising genomic method to identify epigenetic predictors accounting for unexplained disease risk. However, new study designs, analytical methods, and shared resources need to be implemented to address the limitations and challenges in future epigenomic epidemiologic studies.

Keywords Epigenome-wide association study (EWAS) · Methylome-wide association study (MWAS) · Epigenetics · Epigenomics · Epigenetic epidemiology · DNA methylation · Monozygotic twin · Genetic epidemiology · Gene-environment interaction

Introduction

Epigenetics usually refers to the heritable molecular modifications that have effects independent of the primary DNA sequence, that can be modified by environmental exposures at various developmental stages throughout the lifespan [1, 2]. Epigenetic modifications can be inherited across cell generations to exert a long-term impact on the development of chronic diseases. DNA methylation (DNAm) is an essential epigenetic mechanism for normal development and is associated with several key processes linked to chronic diseases. The methyl group is passed from the donor molecule to cytosine catalyzed by DNA methyltransferases (DNMT). Most human diseases are thought to be due to both genetic and environmental factors, and the interplay between genes and environment. Studying the epigenome in the well-characterized samples may enable us to discover novel genes and pathways through which genetic factors and environmental exposures influence disease development, and thereby provide new targets for prevention and treatment [2–4].

Expanding research in epigenetics and epigenomics has resulted in increasing knowledge of the important roles of epigenetics in both normal development and the disease process. After the completion of the Human Genome Project in 2003 [5], multiple consortial studies have emphasized identifying and annotating epigenomic features in humans. National Institutes of Health (NIH) roadmap epigenomics project (http://www.roadmapepigenomics.org/) aims to produce a public resource of human epigenomic data to catalyze basic biology and disease-oriented research. The

Y. V. Sun (✉)
Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Road NE #3049, Atlanta, GA 30322, USA
e-mail: yvsun@emory.edu

Y. V. Sun
Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA, USA

mapping consortium has generated a comprehensive map of the dynamic human DNA methylome [6]. The International Human Epigenome Consortium (IHEC) is a global consortium with the primary goal of providing access to high-resolution reference human epigenome maps for normal and disease cell types to the research community (http://www.ihec-epigenomes.org/). Another major community project, ENCyclopedia Of DNA Elements (ENCODE) has targeted the identification of all functional DNA elements in the human genome including some epigenetic modifications [7, 8]. The mapping of epigenomic features has provided important data sets and information for molecular mechanism linking epigenetic variants and functional outcomes. Additionally, the European Union has committed more than €200 million funding to support over 300 epigenetics projects [9]. The availability of high-throughput genotyping methods, high-density reference panel of human epigenome, and well-characterized samples enables epigenetic association studies at the genomic and population levels. Epigenome-wide association study (EWAS) is an examination of epigenome-wide markers in many individuals to scan for epigenetic markers associated with a trait. Current EWAS in human populations are all based on genome-wide measurement of DNA methylation of cytosine. Thus, I refer these studies as methylome-wide association studies (MWAS) to distinguish them from studies of other epigenetic markers such as histone modifications.

MWAS has emerged as a promising approach to searching for molecular mediators of genetic and environmental factors, and unexplained risks for diseases. Current population level MWAS rely on high-density microarrays [10] or sequencing-based methods [11] following biochemical modifications or enrichments of genomic DNA [12]. Although MWAS only represent a small portion of epigenetic studies, several studies have successfully identified DNAm sites associated with disease traits [13•, 14, 15••, 16]. Using hundreds of subjects, these MWAS have much smaller sample sizes compared to a typical genome-wide association study (GWAS) of the same trait, indicating rather larger effects of identified DNAm sites [15••]. However, this first wave of MWAS aiming to discover associations between epigenomic variation and disease traits is limited by a number of issues such as imperfect technologies, tissue and cell-type specificity, access to biospecimens, sample size, and analytical framework for data pre-processing and statistical modeling [17]. The field has rapidly evolved and equipped us with improved designs, methods, and tools for MWAS. Several recent articles have reviewed the design, analysis and interpretation of EWAS [12, 18–20], especially in the context of epigenetic association study of human disease. In light of these well-covered reviews about EWAS of complex diseases and traits, here I focus on recent reports

on environmental and genetic determinants of the DNA methylome, and discuss their influences on MWAS. In the following sections, I only include recent reports using a methylome-wide approach among *a minimum of 50 subjects* due to power consideration.

Genetic Determinants of DNA Methylome

Genetics, along with environmental factors and stochastic process, are the primary sources of epigenetic variation [21]. Genetic influences can be from the key genes maintaining the epigenetic profile (e.g., DNMT gene family), or from the neighboring genetic variants affecting the epigenetic state (e.g., modification of binding affinity of enzymes). DNA methylation quantitative trait loci (meQTLs) refer to genetic variations within genomic regions associated with a DNAm site. A genome-wide meQTL map can be obtained by correlating the GWAS data with DNA methylomic data from the same samples. Several studies have reported genome-wide meQTLs in peripheral blood, brain, lung, adipose tissue, lymphoblastoid cell lines (LCL), as well as tumor tissues. I summarize thirteen meQTL studies published since 2010 in Table 1 [22–34]. All studies analyzed the *cis*-meQTLs (i.e., SNP and DNAm site collocate in the same genomic region), while a few also reported the *trans*-meQTLs. Because of the large number of SNP-DNAm pairs to test, *trans*-meQTL analyses are more computational intensive and require a more stringent multiple-testing correction. The definition of the *cis*-effect is arbitrary from study to study, ranging from 5 kb [27] to the whole chromosome [33] with majority (8 out of 13) being between 100 kb and 1 Mb from the DNAm site. The heterogeneous definition of *cis*-effects poses a challenge to the direct comparison of the *cis*-meQTL maps between studies, since the non-overlapping sites can simply be a result of the inclusion or exclusion of certain SNP-DNAm pairs. I would recommend 1 Mb flanking region as a standard definition of *cis*-effect in meQTL analysis, being inclusive and not very computationally demanding.

All studies used at least one type of array-based platform to measure the DNA methylome, either the Illumina Infinium HumanMethylation27 (27 K), or the more recent HumanMethylation 450 (450 K) BeadChips (Illumina Inc., San Diego, CA), the latter providing better coverage of genomic regions. Interestingly, all studies excluded the DNAm sites located on the sex chromosomes, although thousands of DNAm sites were measured for their methylation status. Comparing to DNAm sites located on the autosomes, most X chromosome sites are hemimethylated in females, showing a bimodal distribution strongly associated with sex [24, 35]. The sex-specific DNAm sites on X

Table 1 Summary of published studies of meQTL in multiple human tissues and cell types

Study	Tissue/cell type	N	Race	Sex	Mean age	DNAm platform	Region in *cis*	Transcriptome
Gibbs et al. [22]	Cerebellum, frontal cortex, pons & temporal cortex	108, 133, 125, 127	EA	M + F	46.2	27 K	1 Mb	Y
Zhang et al. [23]	Cerebellum	153	EA	M + F	44.3	27 K	1 Mb	N
Bell et al. [24]	LCL	77	Afr.	M + F	NA	27 K	50 kb	Y
Drong et al. [25]	Adipose tissue	38 + 181(F)	EA	M + F	61.1	DMH + 27 K	500 kb	
Zhi et al. [29]	CD4 + T-cell	593	EA	M + F	48	450 K	20 kb	N
Heyn et al. [28]	LCL	288	EA AA Han.	M + F	34.3	450 K	1 Mb	Y
Gutierrez-Arcelus et al. [27]	Fibroblasts, T-cells & LCL from cord blood	204	EA	M = F	Neonate	450 K	5 kb	Y
Grundberg et al. [26]	Adipose tissue, PBL	648, 200	EA	F	59	450 K + BSS	100 kb	Y
Zhang et al. [34]	LCL	133	EA Afr.	M + F	NA	450 K	100 kb	Y
Smith et al. [32]	Peripheral and cord blood, 4 postmortem brain regions [22]	90, 174, 111, 125, 105, 106	EA AA	M + F	Neonate, 43–48	27 K	50 kb	N
Shi et al. [31]	Normal lung tissue	210	EA	M + F	65	450 K	500 kb	Y[a]
Heyn et al. [30]	13 solid cancer types	3,649	EA	M + F	NA	450 K	1 Mb	N
Teh et al. [33]	Umbilical cord blood	237	Asian	M + F	Neonate	450 K	Same chr.	N

LCL lymphoblastoid cell line, *DMH* differential methylation hybridization, *BSS* bisulfite sequencing, *EA* European ancestry, *Afr* African, *AA* African American

[a] RNA-seq data are from The Cancer Genome Atlas (TCGA) study

chromosome are caused by X-chromosome inactivation (XCI), characterized by an X chromosome-wide methylation on the female genome [36]. A new analytical strategy specific to such bimodal distributed DNAm sites would fill the gap of the ignored X chromosome in meQTL study, and benefit the MWAS of human diseases.

Another common theme is that the majority of studies (11 out of 13) scanned for meQTL in samples of European ancestry (EA). In addition to two studies of HapMap LCLs [24, 34] and one study of human variation panel (HVP) LCLs [28], only two studies reported non-EA meQTL maps, one in African Americans (AA) [32] and one in Asians [33] using peripheral and cord blood from individuals. Many genetic associations identified in GWAS of EA are not transferable to other racial groups due to their diverse genetic backgrounds. Similarly, the association between common SNPs and DNAm sites can be different across racial groups. Smith et al. highlighted the overlapping meQTLs between racial groups [32]. However, a large proportion of *cis*-meQTLs are race-specific, and the inter-race consistency of *trans*-meQTLs is unknown. Therefore,

more studies in diverse racial groups are needed to address potentially distinct genetic influence on DNA methylome.

Using over 700,000 genome-wide SNPs and 22,928 autosomal DNAm sites from 460 unrelated AA individuals, I analyzed the genome-wide meQTL of peripheral blood leukocytes (PBLs). Adjusting for age, sex, cell type proportions, and batch effect, I identified 16,320 pairs of significant meQTLs adjusting for multiple testing (Bonferroni corrected alpha level of 0.05). These meQTLs are summarized in Fig. 1, ordered by the chromosomal position of SNPs and CpG sties. Among 16,320 meQTLs representing 6,081 unique SNPs and 2,991 unique CpG sites, 32 % are *cis*-meQTLs (within 1 Mb flanking region) spreading across the entire genome (points along the diagonal line in Fig. 1). Consistent with meQTL studies of other tissues and cell types [22, 24], the genetic contributions to the methylome observed in this AA sample are ubiquitous, both locally (in *cis*) and from long distance (in *trans*). A number of SNPs are associated with multiple DNAm sites, which imply the existence of master genetic regulators for epigenetic markers.
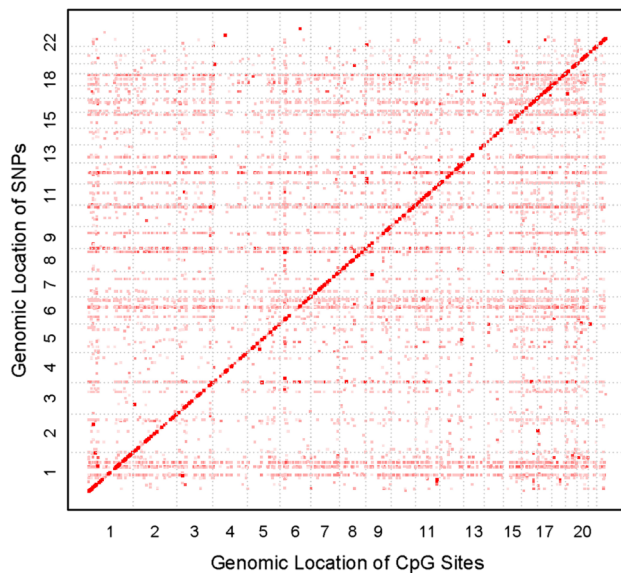
Fig. 1 A genomic map of meQTLs in human peripheral blood lymphocytes

These reported meQTL studies cover both young and old adults, as well as neonates. However, none of the studies involve pediatric samples between newborn and early teens. DNA methylation profile changes dramatically during this critical development stage [37], which is important to understand the genetic effect on methylome and chronic diseases. More importantly, early life exposures may interact with genetic factors to synergistically impact the methylome. Thus, future meQTL studies need to focus on the pediatric age group to complete our understanding of genome-methylome correlation across the entire life-span.

LCLs have been developed to provide epidemiological studies with an 'unlimited' supply of DNA for genetic studies. These immortalized cell lines can be easily obtained to investigate the molecular system involving transcriptome, proteome, and epigenome. LCLs from HapMap [24, 34] and HVP [28] were used to map meQTLs. Although the overall correlation of DNAm profile between LCL and PBLs was high, the transformation process may alter the methylation status of a large number of DNAm sites and increase the inter-individual variation [35]. Therefore, PBLs better reflect the natural DNAm profile of an individual, and are preferred in MWAS over LCLs. Due to tissue and cell type specificity of DNAm, the meQTLs can be substantially different from tissue to tissue [26, 32], cell type to cell type [27]. The functional impact of these tissue-specific meQTLs may be important to understand the pathophysiology of diseases in target tissues. To date, we do not have the meQTL maps of many

targeted tissues and cell types from sizeable population samples. We need to establish a more complete panel of tissue-specific meQTL maps as a publically available resource, to better understand the genetic determinants of DNA methylome, and its role in chronic diseases.

Utilities and Applications of meQTL

Overall, these studies cataloged a large number of meQTLs in multiple tissues and cell types across the human genome, and provide a rich resource not only to understand the genetic regulation of DNAm, but also to illustrate an important molecular mechanism mediating the interaction between genetic variants and environment for chronic diseases. Several recent studies have demonstrated the utilities of meQTL data.

First, as a potentially functional feature of the genome, DNAm sites may mediate the genetic association between SNPs and disease traits. meQTLs have been used to link the genetic associations to disease traits from recent GWAS. Significant GWAS loci are enriched for meQTLs. Over 300 unique GWAS SNPs covering 34 % of reported diseases/traits are meQTLs in LCLs [34]. Among the top susceptibility variants of bipolar disorder, meQTLs are enriched in the cerebellum but not in lymphocytes [38]. Liu et al. reported that rheumatoid arthritis-associated genetic variants may function through the epigenetic mechanism as a potential molecular mediator for gene-environment interaction [13•]. Shi et al. reported that four out of the five established lung cancer loci in people of European ancestry are meQTLs in lung tissue. In aggregate, *cis*-meQTLs in lung tissue are enriched for lung cancer risk [31].

Secondly, meQTL can be used as the instrumental variable in Mendelian randomization (MR) study of DNAm markers [3, 39]. MR was initially proposed as an epidemiologic method to obtain unbiased estimates of the putative casual effects without conducting a randomized trial [40, 41]. The MR approach uses the genetic variant mimicking the biological effects of a modifiable exposure. If the exposure truly alters the disease risk, the genetic variant should also be associated with the disease through the causal pathway. Because the genetic variant is randomly assigned to the offspring during meiosis in a population, the genotype distribution should not be biased by confounding. Only the genetic variant in the causal pathway should be associated with disease outcome by carrying the association through the causal exposure. MR assumes that the genetic variants are independent of the confounders, the genetic variants are reliably associated with the exposure, and there is no direct effect of the genetic variants on disease. In epigenetic epidemiologic study, MR approach can be applied to study (1) the causal environmental factor for epigenetic profile, and (2) the causal

Table 2 Summary of published methylome-wide association studies of cigarette smoking

MWAS study	Source of DNA	Platform	Sample size	Race	Significant smoking-related Loci
Breitling et al. [50••]	PB	27K	177 + 316	EA	F2RL3
Wan et al. [51]	PB	27K	1,085 + 369	EA	F2RL3, GPR15, LRRN3, LIM2
Monick et al. [52]	LCL + LAM	450K	119 + 19	EA	AHRR
Sun et al. [54•]	PB	27K + 450K	972 + 239	AA	F2RL3, GPR15, AHRR, 2q37.1 et al.
Shenker et al. [53]	PB	450K	374 + 180	EA	AHRR, 2q37.1, 6p21.33, F2RL3 et al.
Zeilinger et al. [55]	PB	450K	1,793 + 479	EA	AHRR, ALPPL2, F2RL3 & many more
Dogan et al. [56]	PB	450K	111	AA	AHRR, GPR15
Elliott et al. [57]	PB	450K	192	SA + EA	AHRR, 2q37.1, F2RL3 et al.
Harlid 2014 et al. [58]	Whole blood	450K	908 + 200	EA	7 loci replicated + 2 new loci
Besingi et al. [59•][a]	PB	450K	432	EA	AHRR, 2q37.1 et al.; none for snuffing
Joubert et al. [60•][b]	Cord blood	450K	1,062 + 36	EA	AHRR, CYP1A1, GFI1 et al.
Markunas et al. [61][b]	Whole blood	450K	889	EA	7 loci replicated + 10 new loci

EA European ancestry, AA African American, SA South Asian, PB peripheral blood, LAM lung alveolar macrophage

[a] Tobacco smoking versus tobacco snuffing (smoking-less)

[b] Maternal smoking exposure

epigenetic risk factor for human disease, as detailed in recent publications [3, 39].

Environmental Modifiers of the DNA Methylome

Environmental factors are also important epigenetic modifiers [2, 20]. Epigenetic variation may mediate the environmental risks for human diseases during the entire lifespan, from early embryogenesis, in utero development, childhood to adulthood. Although environment-induced epigenetic changes may vary at different development stages, the cumulative effects can eventually lead to chronic disease in later life. While facing challenges in exposure measurement, epityping technology, availability of biospecimens, study design and analytical methods, genome-wide environmental epigenetics studies also hold the promise to discover epigenetic dysregulation mediating the life-course exposures [20]. Heijmans et al. first showed that early life environment can cause DNAm changes in humans persisting over 60 years after the Dutch Hunger Winter [42]. Individuals who were prenatally exposed to famine had less DNA methylation of IGF2 gene compared with their unexposed, same-sex siblings adjusted for age and relatedness [42]. A number of environmental factors, such as socioeconomic status [43], early life environment [44], traumatic experience [45], pollutants [46], nutrition [47], and physical activity [48•, 49], have been associated with the DNAm profile. These preliminary findings require further replication studies with larger sample sizes and consistent phenotyping. To date, cigarette smoking is the most convincing environmental modifier of the DNA methylome. The strong epigenetic associations with

smoking behavior have been consistently demonstrated in several comprehensive MWAS in different populations. After the first MWAS of cigarette smoking published in 2011 [50••], 11 more MWAS have been reported in sizable population samples. In Table 2, I summarize a total of ten MWAS of smoking in adults [50••, 51, 53, 54•, 55, 56, 57, 58, 59•], and two MWAS of maternal smoking effect on offspring [60•, 61].

Cigarette Smoking-Related DNA Methylation

Cigarette smoking is an environmental risk factor for many chronic diseases including CVD and cancer, the deadliest diseases in the US. Smoking can induce cellular and molecular changes, including epigenetic modification, but the short-term and long-term epigenetic modifications caused by cigarette smoking at the gene level have not been well understood. Recent MWAS studies have identified and replicated smoking-related DNAm sites in samples of European ancestry [50••, 51, 53]. The most significant smoking-related DNAm sites are hypomethylated among smokers [53, 54•, 58]. Several studies also reported an inverse association between pack-years and DNA methylation, and a positive association between time since quitting smoking and DNA methylation [51, 55, 58, 62]. With hundreds of smokers and non-smoking controls, over a dozen of differentially methylated loci have been discovered and replicated in at least two studies. The most replicable smoking-related DNAm loci include F2RL3 (factor II receptor-like 3), AHRR (aryl hydrocarbon receptor repressor), GPR15 (G-protein-coupled receptor 15), 2q37.1, LRRN3 (leucine rich repeat neuronal 3), AKT3 (v-akt murine thymoma viral oncogene homolog 3), LIM2

(lens intrinsic membrane protein 2, 19 kDa), *NCAPD3* (non-SMC condensin II complex, subunit D3), and *CNT-NAP2* (contactin associated protein-like 2). These smoking-related DNAm loci are often replicable across racial groups [54•]. So far, we have limited knowledge of how these genes and their products relate to smoking and physiological function. *F2RL3* plays a role in platelet activation and cell signaling, which may mediate the risk for cardiovascular disease. Aryl hydrocarbon receptor triggers expression of a diverse set of genes, some of which are involved in metabolism of endogenous toxins from cigarette smoke [63, 64].

Joubert et al. conducted an MWAS of maternal smoking and identified 10 loci differently methylated in umbilical cord blood. Interestingly, these loci include reported sites associated with adult smoking (e.g., *AHRR*, *CYP1A1* and *CNTNAP2*) and sites uniquely associated with maternal smoking exposure (e.g., *GFI1*) [60•]. The notable difference in smoking-related DNAm between newborns and adults may indicate distinct effects of direct versus indirect smoking exposure, or of different developmental stages. Markunas et al. replicated 7 previously reported DNAm loci, and identified 10 new regions using 287 mothers smoking during pregnancy and their newborn infants [61]. The DNAm variants in newborns may mediate the effect of in utero smoking exposure on health outcomes in later life. A recent MWAS compared the DNAm changes associated with tobacco smoking to using snuff. Contrary to tobacco smoking, smoke-less tobacco use is not significantly associated with any DNAm site, nor with any enrichment of biological functions and molecular processes [59•]. This interesting observation suggests that the smoking-related DNAm variations are most likely caused by the burnt products of tobacco, but not by its natural components.

Utilities and Applications of Environment-Associated DNAm Marker

The environment-associated DNAm sites may serve as novel biomarkers for both short-term and long-term exposures. In the case of cigarette smoking, DNAm sites located in *F2RL3* and *AHRR* not only predicts current smoking as cotinine does, but also distinguish former-smokers versus never-smokers [65, 66]. The DNAm levels of *F2RL3* sites are associated with the cumulative dose of smoking, as well the time since quitting [65]. Therefore, these DNAm sites are potentially better predictors representing the long-term risk of smoking than the often-flawed self-reported smoking data. Such epigenetic biomarkers of long-term exposures may exist for other environmental risk factors, where the longitudinal profile of the exposure is challenging to measure. For these methylation biomarkers, we can assess if they are more strongly associated with

disease traits than the traditional risk factors in epidemiologic studies. More importantly, we may identify new epigenetic predictors of disease outcomes. Recently, two studies showed that the DNAm sites in *F2RL3* predict secondary cardiovascular events among patients with stable coronary heart disease [67], and predict total mortality after 10-year follow up after adjusting for smoking status [62]. Based on single-marker predictors, we can develop a joint methylation risk score (MRS) for each risk factor with potentially stronger predictive ability. In light of recent MWAS findings of physical activity [48•], [49], BMI [15••] and blood lipids [16], we will be able to examine the predictive ability of the MRS and may improve the prognosis of chronic diseases. The epigenetic plasticity, in combination with the life-course exposure of disease risks, will offer a new window into a more comprehensive explanation for the development of chronic diseases and inter-individual variations.

Partition of Environmental and Genetic Influences on DNA Methylome

The genetic and environmental contribution to a disease is not fixed across the life-span. The proportion of genetic and environmental components of each DNAm site may vary in different studies due to the developmental stage, tissue type of interest, and life-course experience. For chronic diseases, the cumulative environmental effects may dominant the genetic effect in older adults. For age-related chronic diseases such as cardiovascular disease and chronic kidney disease, the epigenetic mechanism may be particularly important in explaining different disease risks among people carrying the same genetic risk alleles. The classical twin design has been used to estimate the contributions of genetic and environmental effects to complex traits. The phenotypic variation in a trait can be decomposed to the genetic variance and environmental variance. These unobserved variance components can be quantitatively estimated from the observed phenotypic covariance in monozygotic (MZ) and dizygotic (DZ) twin pairs.

MZ twins carry identical genetic information from the primary sequence of DNA, but their epigenetic profiles diverge during aging [68]. Fraga et al. showed that MZ twins were epigenetically indistinguishable during the early years of life, and older MZ twins exhibited remarkable differences in DNAm and histone acetylation [68]. The increase of epigenetic differences between MZ twins over time strongly suggests the influence of unshared environment on the epigenome. Rates of discordant disease phenotypes in MZ twins are usually above 50 % even for highly heritable diseases [69–71]. It has been suggested that epigenetics can make a significant contribution to

phenotypic discordance in MZ twins [72, 73]. The epigenomes of discordant MZ twin pairs are, by definition, matched by genetic profile, age and sex, three important predictors of epigenetic profile. In addition, MZ twin design allows controlling for unmeasured confounders such as shared environmental influences. Therefore, using discordant MZ twin pairs is the most powerful design to study epigenomic variation associated with targeted diseases traits [74]. The discordant twin model is particularly useful in detecting moderate effects even with small samples of twin pairs. Several epigenetic studies of MZ discordant twins have investigated DNA methylation profiles in relation to several human diseases such as diabetes [75], psychiatric disorders [76–79], cancer [80], and autoimmune diseases [81–83].

In addition to the discordant twins design, we can also use MZ twin pairs to study the within-twin DNAm difference associated with quantitative traits, which is driven by unshared environment. We are able to estimate both within-pair and between-pair effects in a single regression model [84]. Using 69 pairs of middle-aged MZ twins, we found that the within-twin effect of cg22891070 (a BMI-associated CpG site [15••]) is 7.7×10^{-4} (p value of 0.93), whereas the between-twin effect is 0.031 (p value of 0.07). This observation hints that the epigenetic association with BMI on the *HIF3A* loci is not likely driven by the unshared environment.

Future Opportunities in DNA Methylome Research

The modifiability of epigenetics provides an important molecular mechanism for response to the external and internal environments, but also poses a great challenge in understanding the causality underlying the identified epigenetic association. Many environmental factors can potentially modify the DNAm across the life-span, and many are variable over time. Thus, long-term profiles of DNAm have to be measured and studied to understand their relationship to environment and disease development. In the case of pharmacoepigenetic research, the DNAm profiles have to be captured prior to and after the exposure to establish the causal effect of a given treatment. To improve the accuracy and efficiency of recruiting participants and collecting biospecimens, Sun and Davis developed a novel strategy for pharmacoepigenetic research taking advantage of the real-time data of an electronic medical record system [85•]. The longitudinal studies of DNA methylome and other epigenomic features will be critical to establish the contribution to human diseases in the life-span.

The integrative genomic approach involving transcriptomic and GWAS data has greatly improved our

understanding of DNA methylome. However, the technology for a comprehensive profiling of environmental factors is not available. Exposome is a new concept referring to the totality of human environmental exposures. Although currently we cannot measure or model the exposome, improved metabolomic technology is a promising method to measure thousands of metabolites representing both external and internal environment [86, 87]. A joint metabolome-DNA methylome study surveyed the epigenetic association with 649 blood metabolic traits [88]. The identified associations were driven by an underlying genetic effect, or by environmental and life-style-dependent factors without any underlying genetic signals. These findings extend the role of DNAm in regulating metabolism. Regarding the epityping, both 27K and 450K array, as well as BSS require bisulfite conversion of genomic DNA in order to distinguish methylated from unmethylated cytosines. This chemical conversion does not distinguish 5-methylcytosine (5mC) from 5-hydroxymethylcytosine (5hmC) [89], another type of cytosine modification with potentially different function in cellular processes [90]. Future sequencing-based MWAS can distinguish 5hmC from 5mC, to discovery 5hmC specific loci associated with disease traits.

Conclusions

Epidemiologic and human genetic studies have demonstrated that complex diseases are caused by both genetic and environmental factors. The epigenome including the DNA methylome, links environmental exposures and genetic effects to the development of human diseases. Thus, study of the DNA methylome may lead to a new approach to examining gene-environment interaction, and to identify unexplained risks for complex diseases beyond known environmental and genetic risks. Our knowledge of human DNA methylome has advanced dramatically in recent years. Using high-throughput technologies such as microarray and next-generation sequencing, large sample sizes from human populations can be achieved to unveil the potential impact of DNA methylome on human disease. Recent MWAS involving hundreds to thousands of subjects have successfully identified epigenetic associations with human diseases, environmental, and genetic factors. However, sheer increase of sample size is not sufficient to address the challenges and limitations facing MWAS. To fully recognize the role of DNA methylome in complex diseases and to eventually establish a new epigenetic approach to preventing and treating diseases, we have to carefully consider the characteristics of samples, study design, and key research questions for each MWAS. Therefore, epidemiology methods and designs will play a

very important role in the era of epigenomic epidemiology and MWAS. We will need more well designed MWAS and follow-up studies to continuously unravel the complexity between human epigenome and diseases, and to further understand the epigenetic mechanisms of complex diseases, and to examine the utilities of the DNAm markers.

Disclosure YV Sun declares no conflicts of interest.

Human and Animal Rights and Informed Consent All studies by YV Sun involving animal and/or human subjects were performed after approval by the appropriate institutional review boards. When required, written informed consent was obtained from all participants.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance;

1. Bird A. Perceptions of epigenetics. Nature. 2007;447(7143): 396–8.
2. Foley DL, Craig JM, Morley R, Olsson CA, Dwyer T, Smith K, Saffery R. Prospects for epigenetic epidemiology. Am J Epidemiol. 2009;169(4):389–400.
3. Relton CL, Davey Smith G. Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. PLoS Med. 2010;7(10):e1000356.
4. Jablonka E. Epigenetic epidemiology. Int J Epidemiol. 2004; 33(5):929–35.
5. Collins FS, Green ED, Guttmacher AE, Guyer MS, USNHGR Institute. A vision for the future of genomics research. Nature. 2003;422(6934):835–47.
6. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013;500(7463):477–81.
7. Consortium EP. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004;306(5696):636–40.
8. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.
9. Almouzni G, Altucci L, Amati B, Ashley N, Baulcombe D, Beaujean N, Bock C, Bongcam-Rudloff E, Bousquet J, Braun S, et al. Relationship between genome and epigenome—challenges and requirements for future research. BMC Genome. 2014;15: 487.
10. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450 K technology. Epigenomics. 2011;3(6):771–84.
11. McClay JL, Aberg KA, Clark SL, Nerella S, Kumar G, Xie LY, Hudson AD, Harada A, Hultman CM, Magnusson PK, et al. A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from

blood in over 700 subjects. Hum Mol Genet. 2014;23(5): 1175–85.
12. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12(8):529–41.
13. • Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol. 2013; 31(2):142–147. *This study was among the first to demonstrate the epigenetic mediation of the genetic association.*
14. Hidalgo B, Irvin MR, Sha J, Zhi D, Aslibekyan S, Absher D, Tiwari HK, Kabagambe EK, Ordovas JM, Arnett DK. Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study. Diabetes. 2014;63(2):801–7.
15. •• Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S, Meduri E, Morange PE, Gagnon F, Grallert H et al. DNA methylation and body-mass index: a genome-wide analysis. Lancet. 2014, *This MWAS study demonstrated consistent BMI-associated DNA methylation across multiple tissue types.*
16. Irvin MR, Zhi D, Joehanes R, Mendelson M, Aslibekyan S, Claas SA, Thibeault KS, Patel N, Day K, Waite Jones L, et al. Epigenome-wide association study of fasting blood lipids in the genetics of lipid lowering drugs and diet network study. Circulation. 2014;130:565–72.
17. Heijmans BT, Mill J. Commentary: the seven plagues of epigenetic epidemiology. Int J Epidemiol. 2012;41(1):74–8.
18. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet. 2013;14(8):585–94.
19. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Greally JM, Gut I, Houseman EA, Izzi B, Kelsey KT, Meissner A, et al. Recommendations for the design and analysis of epigenome-wide association studies. Nat Methods. 2013;10(10):949–55.
20. Cortessis VK, Thomas DC, Levine AJ, Breton CV, Mack TM, Siegmund KD, Haile RW, Laird PW. Environmental epigenetics: prospects for studying epigenetic mediation of exposure-response relationships. Hum Genet. 2012;131(10):1565–89.
21. Bjornsson HT, Fallin MD, Feinberg AP. An integrated epigenetic and genetic approach to common human disease. Trends Genet. 2004;20(8):350–8.
22. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, Arepalli S, Dillman A, Rafferty IP, Troncoso J, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet. 2010;6(5): e1000952.
23. 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(3):411–9.
24. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. 2011;12(1):R10.
25. Drong AW, Nicholson G, Hedman AK, Meduri E, Grundberg E, Small KS, Shin SY, Bell JT, Karpe F, Soranzo N, et al. The presence of methylation quantitative trait loci indicates a direct genetic influence on the level of DNA methylation in adipose tissue. PLoS ONE. 2013;8(2):e55923.
26. Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, Busche S, Yuan W, Nisbet J, Sekowska M, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am J Hum Genet. 2013;93(5):876–90.
27. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, Bryois J, Giger T, Romano L, Planchon

A, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. Elife. 2013;2:e00523.

28. Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J, Monk D, Hata K, Marques-Bonet T, Wang L, et al. DNA methylation contributes to natural human variation. Genome Res. 2013;23(9):1363–72.

29. Zhi D, Aslibekyan S, Irvin MR, Claas SA, Borecki IB, Ordovas JM, Absher DM, Arnett DK. SNPs located at CpG sites modulate genome-epigenome interaction. Epigenetics. 2013;8(8):802–6.

30. Heyn H, Sayols S, Moutinho C, Vidal E, Sanchez-Mut JV, Stefansson OA, Nadal E, Moran S, Eyfjord JE, Gonzalez-Suarez E, et al. Linkage of DNA methylation quantitative trait loci to human cancer risk. Cell Rep. 2014;7(2):331–8.

31. Shi J, Marconett CN, Duan J, Hyland PL, Li P, Wang Z, Wheeler W, Zhou B, Campan M, Lee DS, et al. Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. Nat Commun. 2014;5:3365.

32. Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, Tylavsky FA, Conneely KN. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. BMC Genome. 2014;15:145.

33. Teh AL, Pan H, Chen L, Ong ML, Dogra S, Wong J, MacIsaac JL, Mah SM, McEwen LM, Saw SM, et al. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. Genome Res. 2014;24(7):1064–74.

34. Zhang X, Moen EL, Liu C, Mu W, Gamazon ER, Delaney SM, Wing C, Godley LA, Dolan ME, Zhang W Linking the genetic architecture of cytosine modifications with human complex traits. Hum Mol Genet. 2014.

35. Sun YV, Turner ST, Smith JA, Hammond PI, Lazarus A, Van De Rostyne JL, Cunningham JM, Kardia SL. Comparison of the DNA methylation profiles of human peripheral blood cells and transformed B-lymphocytes. Hum Genet. 2010;127(6):651–8.

36. Allis CD. Epigenetics. New York: CSHL Press; 2007.

37. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):R115.

38. Gamazon ER, Badner JA, Cheng L, Zhang C, Zhang D, Cox NJ, Gershon ES, Kelsoe JR, Greenwood TA, Nievergelt CM, et al. Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. Mol Psychiatry. 2013;18(3):340–6.

39. 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(1):161–76.

40. Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. Lancet. 1986;1(8479):507–8.

41. Gray R, Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. Bone Marrow Transplant. 1991;7(Suppl 3):9–12.

42. 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(44):17046–9.

43. Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, Kobor MS. Factors underlying variable DNA methylation in a human community cohort. Proc Natl Acad Sci U S A. 2012; 109(Suppl 2):17253–60.

44. Szyf M. The early life environment and the epigenome. Biochim Biophys Acta. 2009;1790(9):878–85.

45. Labonte B, Suderman M, Maussion G, Navaro L, Yerko V, Mahar I, Bureau A, Mechawar N, Szyf M, Meaney MJ, et al. Genome-wide epigenetic regulation by early-life trauma. Arch Gen Psychiatry. 2012;69(7):722–31.

46. Pogribny IP, Rusyn I. Environmental toxicants, epigenetics, and cancer. Adv Exp Med Biol. 2013;754:215–32.

47. Waterland RA, Jirtle RL. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol Cell Biol. 2003;23(15):5293–300.

48. • Ronn T, Volkov P, Davegardh C, Dayeh T, Hall E, Olsson AH, Nilsson E, Tornberg A, Dekker Nitert M, Eriksson KF et al. A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. PLoS Genet. 2013; 9(6):e1003572. *This study was the fist MWAS to identify DNA methyaltion in adipose tissue associated with exercise.*

49. Barres R, Yan J, Egan B, Treebak JT, Rasmussen M, Fritz T, Caidahl K, Krook A, O'Gorman DJ, Zierath JR. Acute exercise remodels promoter methylation in human skeletal muscle. Cell Metab. 2012;15(3):405–11.

50. •• Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H Tobacco-smoking-related differential DNA methylation: 27 K discovery and replication. Am J Hum Genet. 2011; 88(4):450–457. *This study was the first MWAS to identify and replicate DNA methylation in F2RL3 associated with tobacco smoking.*

51. 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;21:3073–82.

52. Monick MM, Beach SR, Plume J, Sears R, Gerrard M, Brody GH, Philibert RA. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. Am J Med Genet B Neuropsychiatr Genet. 2012;159B(2):141–51.

53. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, Belvisi MG, Brown R, Vineis P, Flanagan JM. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet. 2013;22(5):843–51.

54. • Sun YV, Smith AK, Conneely KN, Chang Q, Li W, Lazarus A, Smith JA, Almli LM, Binder EB, Klengel T et al. Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans. Hum Genet. 2013; 132(9):1027–1037. *This study was the first to demonstrate consistent smoking-related DNA methylation across racial groups. It was the largest MWAS in samples with non-European ancestry up to date.*

55. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS ONE. 2013;8(5):e63812.

56. Dogan MV, Shields B, Cutrona C, Gao L, Gibbons FX, Simons R, Monick M, Brody GH, Tan K, Beach SR, et al. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. BMC Genome. 2014;15:151.

57. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, Davey Smith G, Hughes AD, Chaturvedi N, Relton CL. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. Clin Epigenetics. 2014;6(1):4.

58. Harlid S, Xu Z, Panduri V, Sandler DP, Taylor JA. CpG sites associated with cigarette smoking: analysis of epigenome-wide data from the sister study. Environ Health Perspect. 2014; 122(7):673–8.

59. • Besingi W, Johansson A Smoke-related DNA methylation changes in the etiology of human disease. Hum Mol Genet. 2014; 23(9):2290–2297. *This study reported different epigenetic association patterns between tobacco smoking versus smoking-less tobacco use.*

60. • Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, Huang Z, Hoyo C, Midttun O, Cupul-Uicab LA et al. 450 K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during

pregnancy. Environ Health Perspect. 2012; 120(10):1425–1431. *This MWAS was among the first to identify DNA methylation of newborns associated with maternal smoking behavior.*

61. Markunas CA, Xu Z, Harlid S, Wade PA, Lie RT, Taylor JA, Wilcox AJ Identification of DNA Methylation Changes in Newborns Related to Maternal Smoking during Pregnancy. Environ Health Perspect. 2014.

62. Zhang Y, Yang R, Burwinkel B, Breitling LP, Holleczek B, Schottker B, Brenner H. F2RL3 methylation in blood DNA is a strong predictor of mortality. Int J Epidemiol. 2014;43:1215–25.

63. Kawajiri K, Fujii-Kuriyama Y. Cytochrome P450 gene regulation and physiological functions mediated by the aryl hydrocarbon receptor. Arch Biochem Biophys. 2007;464(2):207–12.

64. Swedenborg E, Pongratz I. AhR and ARNT modulate ER signaling. Toxicology. 2010;268(3):132–8.

65. 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(2):131–7.

66. Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, Flanagan JM, Vineis P. DNA methylation as a long-term biomarker of exposure to tobacco smoke. Epidemiology. 2013;24(5):712–6.

67. Breitling LP, Salzmann K, Rothenbacher D, Burwinkel B, Brenner H. Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease. Eur Heart J. 2012;33:2841–8.

68. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci USA. 2005;102(30):10604–9.

69. Kendler KS, Neale MC, Sullivan P, Corey LA, Gardner CO, Prescott CA. A population-based twin study in women of smoking initiation and nicotine dependence. Psychol Med. 1999;29(2):299–308.

70. MacGregor AJ, Antoniades L, Matson M, Andrew T, Spector TD. The genetic contribution to radiographic hip osteoarthritis in women: results of a classic twin study. Arthritis Rheum. 2000;43(11):2410–6.

71. Silman AJ, Newman J, MacGregor AJ. Cigarette smoking increases the risk of rheumatoid arthritis. Results from a nationwide study of disease-discordant twins. Arthritis Rheum. 1996;39(5):732–5.

72. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature. 2010;465(7299):721–7.

73. Wong WS, Roubal I, Jackson DB, Paik WN, Wong VK. Outsourced teleradiology imaging services: an analysis of discordant interpretation in 124,870 cases. J Am Coll Radiol. 2005;2(6):478–84.

74. Bell JT, Spector TD. A twin approach to unraveling epigenetics. Trends Genet. 2011;27(3):116–25.

75. Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S, Aden D, Daunay A, Busato F, Mein CA, Manfras B, et al. Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. PLoS Genet. 2011;7(9):e1002300.

76. Kuratomi G, Iwamoto K, Bundo M, Kusumi I, Kato N, Iwata N, Ozaki N, Kato T. Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins. Mol Psychiatry. 2008;13(4):429–41.

77. Petronis A, Gottesman II, Kan P, Kennedy JL, Basile VS, Paterson AD, Popendikyte V. Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? Schizophr Bull. 2003;29(1):169–78.

78. Dempster EL, Pidsley R, Schalkwyk LC, Owens S, Georgiades A, Kane F, Kalidindi S, Picchioni M, Kravariti E, Toulopoulou T, et al. Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. Hum Mol Genet. 2011;20(24):4786–96.

79. Wong CC, Meaburn EL, Ronald A, Price TS, Jeffries AR, Schalkwyk LC, Plomin R, Mill J Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. Mol Psychiatry. 2013.

80. Heyn H, Carmona FJ, Gomez A, Ferreira HJ, Bell JT, Sayols S, Ward K, Stefansson OA, Sandoval J et al DNA methylation profiling in breast cancer discordant identical twins identifies DOK7 as novel epigenetic biomarker. Carcinogenesis. 2012.

81. Baranzini SE, Mudge J, van Velkinburgh JC, Khankhanian P, Khrebtukova I, Miller NA, Zhang L, Farmer AD, Bell CJ, Kim RW, et al. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. Nature. 2010;464(7293):1351–6.

82. Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, Berdasco M, Fraga MF, O'Hanlon TP, Rider LG, et al. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. Genome Res. 2010;20(2):170–9.

83. Gervin K, Vigeland MD, Mattingsdal M, Hammero M, Nygard H, Olsen AO, Brandt I, Harris JR, Undlien DE, Lyle R. DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. PLoS Genet. 2012;8(1):e1002454.

84. Carlin JB, Gurrin LC, Sterne JA, Morley R, Dwyer T. Regression models for twin studies: a critical review. Int J Epidemiol. 2005;34(5):1089–99.

85. • Sun YV, Davis RL Rapid collection of biospecimens by automated identification of patients eligible for pharmacoepigenetic studies. J Personal Med. 2013; 3(4):263–274. *This study implemented a novel longitudinal design for recruitment and blood collection in the pharmacoepigenetic study.*

86. Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A. The blood exposome and its role in discovering causes of disease. Environ Health Perspect. 2014;122:769–74.

87. Miller GW, Jones DP. The nature of nurture: refining the definition of the exposome. Toxicol Sci. 2014;137(1):1–2.

88. Petersen AK, Zeilinger S, Kastenmuller G, Romisch-Margl W, Brugger M, Peters A, Meisinger C, Strauch K, Hengstenberg C, Pagel P, et al. Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. Hum Mol Genet. 2014;23(2):534–45.

89. Jin SG, Kadam S, Pfeifer GP. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res. 2010;38(11):e125.

90. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010;466(7310):1129–33.