



DNA Methylation Signatures as Biomarkers of Prior Environmental Exposures

Christine Ladd-Acosta^{1,2} · M. Daniele Fallin^{2,3}

Published online: 1 February 2019
© Springer Nature Switzerland AG 2019

Abstract

Purpose of Review This review demonstrates the growing body of evidence connecting DNA methylation to prior exposure. It highlights the potential to use DNA methylation patterns as a feasible, stable, and accurate biomarker of past exposure, opening new opportunities for environmental and gene-environment interaction studies among existing banked samples.

Recent Findings We present the evidence for association between past exposure, including prenatal exposures, and DNA methylation measured at a later time in the life course. We demonstrate the potential utility of DNA methylation-based biomarkers of past exposure using results from multiple studies of smoking as an example. Multiple studies show the ability to accurately predict prenatal smoking exposure based on DNA methylation measured at birth, in childhood, and even adulthood. Separate sets of DNA methylation loci have been used to predict past personal smoking exposure (postnatal) as well. Further, it appears that these two types of exposures, prenatal and previous personal exposure, can be isolated from each other. There is also a suggestion that quantitative methylation scores may be useful for estimating dose. We highlight the remaining needs for rigor in methylation biomarker development including analytic challenges as well as the need for development across multiple developmental windows, multiple tissue types, and multiple ancestries.

Summary If fully developed, DNA methylation-based biomarkers can dramatically shift our ability to carry out environmental and genetic-environmental epidemiology using existing biobanks, opening up unprecedented opportunities for environmental health.

Keywords DNA methylation · Biomarker · Past exposure · Environmental exposure · Prenatal smoking · EWAS · Epigenomic

Introduction

Heritability analyses for most complex disorders show that at least some portion of disease liability is due to environmental

factors [1], often a large component of risk. The specific health consequences of environmental exposures have been well established for many toxicants and outcomes [2, 3]. Yet, many environmental risk factors have not yet been discovered, despite evidence that they play a role in disease. Environmental epidemiology's goal of identification and characterization of non-heritable risk factors is critical, as these factors provide actionable insights about modifiable causes of disease that can lead to better prediction, prevention, treatment, and policy.

A major limitation to further discovery in environmental epidemiology has been the need for timing-specific exposure information and prospective outcome data. This is a great challenge, particularly for exposures influencing risk on outcome years to decades later, and for exposures that are difficult to measure or occur prior to feasible study enrollment, such as prenatal or preconception exposures. Some prospective cohort studies do begin prior to pregnancy, or early in pregnancy, and follow new babies through life (e.g. [4–9]). However, these study designs take years to accumulate outcomes, often with attrition or low enrollment numbers given the timing of

This article is part of the Topical Collection on *Genetic Epidemiology*

✉ Christine Ladd-Acosta
claddac1@jhu.edu

✉ M. Daniele Fallin
dfallin@jhu.edu

¹ Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe St, W6509, Baltimore, MD 21205, USA

² Wendy Klag Center for Autism and Developmental Disabilities, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

³ Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, 624 North Broadway, HH850, Baltimore, MD 21205, USA

enrollment and the length of commitment. Retrospective measurement of exposure is notoriously difficult, given the potential for recall bias in self-report, the lack of information in administrative data such as electronic health records, particularly for toxicants, and the short half-lives of many toxicants—such that biomarker measurement weeks or years later is irrelevant to amounts of exposure at the time of vulnerability.

Thus, there is a critical need in environmental epidemiology for measurement tools that can accurately capture past exposure, particularly prenatal and early life exposures. One emerging area of promise is the ability to measure toxicant content of shed baby teeth, available at middle childhood, but able to inform exposures that occurred in utero [10, 11]. While this is a promising avenue, it does require availability of baby teeth and is to date, relatively expensive with few labs able to perform detailed measurement. Among the other emerging options is the potential for blood, or other readily available tissue samples, to provide past exposure proxy information. This could be transformational for environmental epidemiology and genetic epidemiology. If one can use biosamples already in biobanks, such as UK biobank [12] or the vast genetic consortia banks (e.g. [13].) to estimate prior exposure with accuracy, there would be ample power to ask environmental exposure questions not previously possible and to truly integrate genetic and environmental information in these large sample sets.

One promising possibility for a blood (or convenience tissue)-based biomarker of past exposure that could enable environmental and gene-environmental work in existing biosamples is the potential for DNA methylation patterns to mark prior exposure. As we show in this review, there is now a substantial body of evidence that DNA methylation measured in blood, and other tissues, is associated with prior exposure, and that this association may be strong enough to enable an accurate predictor of exposure that is timing and toxicant specific. More work must be done to establish such biomarkers for specific exposure, but here, we show evidence from discovery epigenome-wide association studies (EWAS) for several exposures and timing, paving the way for such biomarker development. Such discoveries must be further evaluated in prediction models to establish their biomarker utility. As an example, we elaborate on the work done with the association between prenatal smoking exposure and DNA methylation patterns, which has moved from EWAS discovery to biomarker development. The results show promising accuracy, reproducibility, specificity to exposure, and persistence over many years. We also discuss DNA methylation patterns as a cumulative exposure biomarker, or biomarker of aging, through what has been termed “DNA methylation clocks”. Through this review, we hope to present these findings as examples of the opportunities that exist for environmental and genetic-environmental epidemiology through DNA methylation-

based biomarkers and call for more work to be done in the field to realize this potential.

Suitability of DNA Methylation as a Biomarker of Past Exposure

DNA methylation is a type of epigenetic mark with several inherent properties that make it well suited for exposure biomarker purposes. DNA methylation involves the covalent addition of a methyl or hydroxymethyl group to cytosine nucleotides in human DNA, and thus, it is relatively stable and not easily degraded with long-term storage. It also does not require any burdensome up front sample collection or processing methods. These properties are particularly important when considering new methods to extract past exposure information from existing biobanks and repositories. While chemically stable, DNA methylation is a dynamic process that can be modified by environmental context and over time, a critical feature of any exposure biomarker. It provides a mechanism for cells and organisms to respond to their environment without changing the DNA sequence. Finally, because DNA methylation is quantitative in nature, it may capture “biological dose” and/or effects of exposure mixtures.

There are several advantages to using DNA methylation as a biomarker of exposure relative to prospectively or retrospectively collected exposure data, metabolites, gene expression, or objective wearable devices. More traditional exposure ascertainment methods can pose several problems. Prospective collection of exposure data is ideal but is costly and can be inefficient for diseases with lower prevalence rates or those with long lag times between exposure and development of disease. Retrospective collection of exposure data is subject to recall bias or misclassification and is impossible to collect for certain exposures (e.g., metal toxicants). The emergence of objective wearable devices can overcome many of these issues but have only recently come online, and thus, do not enable utilization of existing large-scale biobanks. Use of molecular biomarkers of exposure has been mainstream for decades. For some exposures, metabolites have been the gold standard measurement tool to collect accurate highly reliable information about exposure. For example, cotinine, a major metabolite of nicotine, is widely recognized as the optimal collection metric to obtain smoking status [14, 15]. Untargeted metabolomic assays also have the potential to capture exposure mixtures and quantities. However, one of the major limitations to using metabolites as biomarkers of past exposure is their short half-life. The half-life of most metabolites, including cotinine, is on the order of hours to days [16–18]. Metabolites collected from untargeted assays can also be sensitive to dietary intake differences and sample collection protocols that may vary within and across large biobanks. Laboratory and analytic methods to best address

these issues are still under development. Exposure-related transcriptome changes have also been observed. Isolating high-quality RNA suitable for gene expression profiling can be challenging in an epidemiologic and biobank resource setting because it is less stable than DNA and more subject to degradation with longer-term storage or suboptimal collection protocols. New molecular biomarkers that are long-lived, specific, stable, and that can be reliably measured in existing banked samples are needed; as evidenced in detail below, DNA methylation meets these criteria.

DNA Methylation Is Associated with Past Exposure, Across Multiple Domains

With the emergence of affordable genome-scale epigenetic technologies, it is now feasible to measure DNA methylation in a large number of samples and perform epigenome-wide association studies (EWAS) to discover methylation differences, at specific CpG sites in the genome, associated with particular exposures or outcomes [19]. This technological advance, coupled with a strong interest in identifying molecular changes related to environmental exposures has led to a rapid increase in environmental epigenomic studies. A wide range of exposures have now been linked to epigenetic changes in studies where both types of data were measured at the same time; these have been extensively reviewed elsewhere [20–22]. In this review, we focus on EWAS showing DNA methylation patterns, measured across the lifespan, reflect past exposures. As summarized in Table 1, methylation changes have been linked to past exposure, across a wide range of environmental domains.

Prenatal Exposure to Smoking and Alcohol Several EWAS have identified site-specific changes in DNA methylation levels at birth related to prenatal exposure to maternal smoking [24•, 25, 26•, 32•, 33•] and alcohol use [37] (Table 1). Several genomic regions have shown suggestive differences in cord blood DNA methylation levels related to maternal drinking habits during early pregnancy [37]. However, studies of prenatal alcohol exposure and DNA methylation are limited by sample size and window of pregnancy timing. Additional genome-wide significant findings may emerge with increased sample sizes and/or more resolved alcohol exposure metrics in the future. For prenatal smoking exposure, site-specific changes in DNA methylation have been detected in peripheral blood obtained from infants [27], older children [26•, 30•, 31••, 32•, 33•], and adolescents [32•]. Associations between later life blood DNA methylation and prenatal smoking exposure persist even after adjusting for postnatal and personal smoking exposures [32•, 33•]. Smoking and drinking are thought to have similar social determinants and correlated patterns of use; however, the

associated DNA methylation findings published to date have not been consistent across these exposures, indicating that DNA methylation signatures may be exposure-specific and not merely capturing a social determinant construct [26•, 31••, 37].

Nutrition and Supplementation As shown in Table 1, a number of studies have observed DNA methylation changes in samples collected—from birth through adulthood—related to differences in peri- and prenatal exposure to nutrient intake and nutritional supplements [39–46]. Differences in maternal nutrient intake during peri-conception and pregnancy through diet and food availability have been linked to DNA methylation changes, at specific genes, in blood and buccal samples obtained from their offspring at birth, infancy, and childhood [41–45]. A number of studies have leveraged data from cohorts dating back to the 1960s when the first randomized control trials were carried out to assess the impact of folic acid and/or docosahexaenoic acid (DHA) supplementation on birth and child outcomes. Saliva DNA methylation profiles collected in 47-year-old adult offspring of the Aberdeen Folic Acid Supplementation Trial (AFAST) participants showed differences related to whether their mothers received folic acid supplementation during pregnancy or were in the placebo group [39]. A randomized controlled trial for docosahexaenoic acid (DHA), an omega-3 fatty acid, observed differentially methylated genomic regions among infants whose mothers received DHA relative to those that did not receive the supplement. Furthermore, the methylation differences were also shown to be present in peripheral blood samples collected at 5 years of age [46].

Prenatal Toxicant Exposures In the past year, DNA methylation changes have been linked to air pollutant exposure in the prenatal time period (Table 1). More specifically, a multi-study EWAS meta-analysis identified CpG loci showing significant methylation changes in cord blood, at birth, related to prenatal nitrogen dioxide (NO₂) exposure levels. Interestingly, prenatal NO₂-associated methylation changes were also observed in peripheral blood obtained from older children. The NO₂ exposure levels at the time of blood sample collection in the older children were substantially lower than those the children experienced during pregnancy, arguing that their presence in childhood samples was not likely due to continued postnatal exposure or current NO₂ exposure status [52]. More evidence in this area is likely to transpire as additional studies with unified prenatal air pollutant and DNA methylation data emerge. In addition to site-specific changes in DNA methylation, a significant global decrease in the total genomic amount of 5-hydroxymethyl, a specific type of DNA methylation, was observed in birth and early childhood blood samples among children with elevated prenatal exposure to mercury [54].

Table 1 Summary of discovery EWAS and look-up replication studies showing exposure-related DNA methylation patterns are present and can be detected long after an exposure occurred

Exposure domain	Timing of:		Tissue	Scale and purpose	Reference	Notes
	Exposure	DNAm measure				
Behavior/lifestyle <i>Smoking</i>	Prenatal	Birth	Placenta Cord blood	Genome-scale discovery EWAS discovery EWAS discovery Meta-EWAS discovery	[23] [24] [25] [26•]	
		Infancy	Blood spots: Guthrie cards	Lookup of birth EWAS meta-results	[27]	
	Birth and 18 months	Childhood	Cord and peripheral blood	Candidate gene, replication and DNAm persistence	[28]	Showed longitudinal persistence of prenatal smoking DNAm marks
			Buccal epithelium	Targeted, gene discovery	[29]	
	Prenatal	Birth, childhood, adolescence	Peripheral blood	EWAS discovery	[30•]	
			Peripheral blood	Lookup of birth EWAS in older children	[26•, 31••]	
			Cord blood, peripheral blood	EWAS discovery in cord blood, lookup results in blood collected later in life	[32•]	Showed longitudinal persistence of DNAm marks at age 7 and 17, even after adjusting for postnatal and personal exposure
			Peripheral blood	Replication of adolescent EWAS results in an independent sample across 3 timepoints	[33•]	Adjusted for postnatal personal and parental smoking
	Adult	Later adulthood	Peripheral blood	EWAS discovery in adolescent blood	[33•] [34]	Adjusted for postnatal personal and parental smoking
			Buccal epithelium	EWAS discovery	[34]	
Peripheral blood			EWAS discovery Meta-EWAS discovery	[35]	Show methylation trajectories in former smokers do not return to non-smoker levels even after 30yrs	
Nutrition and supplementation <i>Folic acid</i>	Prenatal	Peripheral blood	EWAS discovery and replication in different Race	[36]	Exposure from randomized folic acid supplementation trial - 1960's	
		Cord blood	Meta-EWAS discovery	[37]	Seasonally driven nutrition differences - The Gambia	
Maternal diet <i>Maternal diet</i>	Prenatal	Peripheral blood	Meta-EWAS discovery current use prediction	[38]	Seasonally driven nutrition differences – The Gambia	
		Saliva	EWAS discovery	[39]		
Maternal diet <i>Maternal diet</i>	Prenatal	Peripheral blood	Candidate gene discovery	[40, 41]		
		Buccal	Candidate gene discovery	[42, 43] [44]		

Table 1 (continued)

Exposure domain	Timing of:		Tissue	Scale and purpose	Reference	Notes
	Exposure	DNAm measure				
<i>Methyl-group donor intake</i>		Birth	Cord blood	Candidate gene discovery	[45]	
<i>Docosahexaenoic acid (DHA)</i>	Prenatal	Neonates	Guthrie blood cards	EWAS discovery	[46]	From DOMInO randomized controlled trial for DHA supplementation
<i>Short-term high-fat overfeeding diet</i>	Adulthood	Childhood Adults, 6–8 weeks later	Peripheral blood Skeletal muscle	Lookup at later time point for persistent changes Targeted gene discovery	[47]	Diet-related DNA methylation levels were still present months after returning to a normal diet
Adversity						
<i>Dietary (famine)</i>	Prenatal	Adults, at age 60 Adulthood	Peripheral blood Peripheral blood	Candidate gene discovery Candidate gene replication	[48, 49] [50]	Data from the Dutch Hunger Winter study Bangladesh sample
<i>Stress (maltreatment)</i>	Childhood	Later Childhood	Saliva	Candidate gene discovery	[51]	Showed 6 month longitudinal changes in DNA methylation among 3–5 year olds with severe maltreatment exposure
Toxicants						
<i>Air pollutants (nitrogen dioxide)</i>	Prenatal	Birth Childhood	Cord blood Peripheral blood	EWAS discovery Lookup of birth EWAS results in older children	[52] [52]	Associations with prenatal NO ₂ levels not explained by postnatal NO ₂ exposure near time of sample collection.
<i>Lead</i>	Cumulative Lifelong	Adulthood	Peripheral Blood	DNAm global burden (LINE-1)	[53]	
<i>Mercury</i>	Prenatal	Birth and childhood	Cord and peripheral blood	DNAm global burden (LC-MS/MS)	[54]	
Health conditions						
<i>Maternal early pregnancy BMI</i>	Prenatal	Birth	Cord blood	Meta-EWAS discovery	[55]	
<i>Maternal eating disorder</i>	Prenatal	Adolescence	Peripheral blood	Lookup of meta-EWAS results in adolescence	[55]	
	Prenatal	Birth	Cord blood	Global genome DNAm burden	[56]	This could reflect dietary deficiency exposure

BMI body mass index, DNAm DNA methylation, EWAS epigenome-wide association study

Prenatal Exposure to Adversity Several social adversity exposures have been associated with long-term changes in DNA methylation (Table 1), although they have mainly focused on candidate genes. For example, candidate gene-based work, from the historic Dutch Hunger Winter study, revealed that DNA methylation levels at the *IGF2* gene locus differ significantly between individuals with prenatal exposure to the 1944–1945 famine relative to their unexposed same-sex siblings [48]. These changes were detected in blood samples provided 60 years after their prenatal exposure to famine. Exposure to severe maltreatment during early childhood has also been linked to methylation changes in saliva. Significant decreases in DNA methylation at the *NR3C1* gene locus were observed among preschool age children exposed to stress/maltreatment in the 6 months prior to biospecimen collection compared to unexposed children with similar economic status [51].

Maternal Conditions in Pregnancy There is also evidence that exposure to adverse maternal health conditions during pregnancy is related to methylation changes at birth through adolescence (Table 1). A meta-analysis of 19 cohorts reported 86 site-specific changes in DNA methylation, in cord blood, related to maternal body mass index (BMI) at the start of pregnancy [55]. Of those, 72 sites showed a similar association, direction, and magnitude of effect in peripheral blood samples obtained in adolescence [55]. DNA methylation levels among infants born to women with an active eating disorder during pregnancy differed from those whose mothers had an active eating disorder (ED) prior to conception and non-ED controls [56].

Adult Exposures and Later Measurement Several studies have reported long-lasting DNA methylation patterns in later adulthood biospecimens related to past earlier adulthood exposures. Similar to prenatal exposures, most findings to date are for behavioral and lifestyle types of exposures including smoking and alcohol use (Table 1). This is likely due to lack of unified exposure and methylation data in the same samples for other, more difficult, to obtain exposures. In world-wide population samples, meta-EWAS have identified thousands of loci where peripheral blood methylation levels differ by current, former, and never smoker status [34–36]. Joehanes et al. found that methylation values among former smokers that quit smoking 30 years prior to collection of methylation measurements in blood samples still had not reached levels comparable to individuals that never smoked [35]; thus, DNA methylation changes associated with past exposures can be long-lived. Further, smoking-related methylation values appear to capture additional valuable information about past exposures: time since quitting and number of pack-years smoked [34–36]. This has important implications for the potential to use DNA methylation signatures to serve not only as a simple

dichotomous exposure biomarker but also as a biomarker that can be used to determine specific windows and doses of exposure. Similar differences in methylation related to smoking status, time since quitting, and pack-years have also been documented in buccal samples [34], another highly accessible and available tissue source. However, a comparison of DNA methylation patterns among hundreds of former drinkers compared to never drinkers, ~4 years after alcohol cessation, showed only marginal differences between the two exposure groups [38]. Epigenetic changes related to nutrition in adults have also been observed (Table 1). Males exposed to a short-term high-fat overfeeding diet showed epigenetic changes that persisted for 6–8 weeks after the men resumed their normal diets [47].

Longitudinal DNAm Data To date, three studies have reported repeated measures of DNA methylation and associations with exposure information; two were focused on DNA methylation signatures of prenatal smoking exposure and the third examined the effects of maltreatment. Longitudinal analysis of methylation profiles at prenatal smoking-associated CpG sites showed similar differences in DNAm related to prenatal smoking status at 18 months [28], 7, and 17 years of age [32•] even after accounting for any postnatal smoking exposures in the older children [32•]. However, in adolescence, there were three CpG sites that showed reversion back to methylation levels observed among adolescence with no prenatal exposed to maternal smoking [57•]. This suggests that signatures of prenatal exposure developed solely in cord blood samples may fail to account for important differences in methylation stability in the postnatal period. Thus, the development of a robust epigenetic biomarker of past exposure will need to take this into account and evaluate methylation patterns at multiple post-exposure time points. The third study examined baseline and longitudinal changes in saliva methylation levels over a period of 6 months, among preschool age children, to assess the effects of maltreatment (at baseline) on methylation at *NR3C1* [51]. Children with no history of maltreatment showed little variation in methylation across the two time points. However, children with a history of maltreatment had significantly higher levels of methylation at baseline and significantly decreased methylation 6 months later. This suggests looking for differences in methylation variation among exposed and unexposed individuals, as opposed to mean methylation shifts, may be a fruitful and important avenue for future studies.

Cumulative Exposures, Aging, and Epigenetic “Clocks”

In addition to serving as a biomarker for discrete intervals of exposure, DNA methylation signatures have also been reported

to capture continuous cumulative levels of exposures including toxicant and behavioral. For example, measures of global DNA methylation levels in LINE-1 elements were significantly decreased among men with increased cumulative exposure to lead, as assayed via patella bone K-Xray which is a well-established traditional biomarker of long-term lead exposure [58]. In addition, several studies of adult smokers have consistently demonstrated that DNA methylation patterns at specific sites accurately reflect the cumulative amount and duration of current and prior smoking [34–36]. A number of DNA methylation “clocks” have been developed to reflect gestational [59–61], pediatric [62], and adult [63–68] chronologic ages, a type of demographic exposure, that can also be thought of as a cumulative exposure. These methylation clocks have been widely used to predict a number of adverse health outcomes demonstrating the utility of DNA methylation exposure biomarkers in epidemiology studies, more broadly [69–72]. For example, the adult-derived epigenetic clock has been shown to better predict all-cause mortality than examination of traditional risk factors or chronological age [73].

Biomarkers Require Predictive Modeling Beyond EWAS Discovery Analyses

EWAS findings continue to emerge and provide valuable insights into the biologic targets of environmental exposures. However, the main output from EWAS is not directly informative or useful as a predictive biomarker. Results are typically per-CpG, rather than a collective “signature”. Further, discovery analyses typically rely on general associations between exposed versus unexposed samples. A predictive modeling approach is needed to develop a useful biomarker. Accuracy parameters such as sensitivity, specificity, and area under the ROC curve (AUC) are more relevant for biomarker development [74, 75]. Further, a collection of CpGs associated with the particular exposure will necessarily have better predictive properties than a single CpG. Selection of this collective list, modeling of the prediction algorithm, and evaluation of prediction performance are necessary. This approach has been taken in the development of epigenetic clocks described above. Choices for CpG selection include simply taking all CpGs meeting a particular statistical threshold in EWAS, or building machine-learning models using techniques such as support vector machines or elastic net [76••]. Prediction algorithms can then include all CpGs equally, or weighted by their association with the exposure, or other characteristics. The output may be a probabilistic exposure membership (dichotomous, with associated probability), or a methylation-based exposure “score” [57••, 77].

Prenatal Smoking as an Example

For the most well-studied and replicated exposure—prenatal smoking—work in this area has already begun and can be used as an exemplary model for the field to be extended to other types of exposures. The first site-specific differences in DNA methylation related to prenatal exposure to smoking were reported in 2012 by Joubert et al. [24], where EWAS revealed 26 CpG sites with exposure-associated DNA methylation differences achieving genome-wide significance. Not long after, studies emerged replicating the findings in additional birth samples and adding a hand full of new loci [25, 32•, 33•]. Many also showed similar DNA methylation patterns associated with prenatal smoking exposure, but when measured in blood samples from older children, ranging in age from 5 to 17 years [30•, 31••, 32•, 33•], even after accounting for parental and personal postnatal smoking exposures [32•, 33•].

Ladd-Acosta et al. [31••] were the first to use predictive modeling to evaluate how well DNA methylation levels, measured in blood samples from 5-year-old children, at the originally reported 26 CpG sites associated with prenatal smoking exposure, could predict prenatal exposure to smoking from childhood, rather than cord blood. Their support vector machine classifier, with 10-fold cross validation, predicted the children’s exposure to sustained active maternal smoking in pregnancy with 87% accuracy when compared to maternal report of smoking during pregnancy (Table 2). Receiver operating characteristic (ROC) curves also showed that the specificity of the model was high; prediction of prenatal smoking exposure using permuted random sets of 26 loci never achieved greater than 60% accuracy and the prenatal smoking classifier was not able to predict exposure to maternal alcohol or medication use with higher than 56% accuracy [31••]. The following year, Reese et al. [77] developed a single numeric methylation score, based on DNA methylation measured in blood, and showed good correspondence to prenatal cotinine levels consistent with sustained exposure to active maternal smoking. In an independent test set of cord blood samples, the methylation score was able to predict prenatal exposure to sustained smoking with 91% overall accuracy [77] (Table 2). A recent cord blood methylation meta-analysis, spanning 13 world-wide studies and 6685 samples, showed consistency with previous findings and expanded the set of loci significantly associated with prenatal smoking from dozens to 2965 CpG sites [26•]. Nominally significant differences in methylation were also observed in older children ($n = 3187$) for every CpG site identified at birth [26•]. More recently, Richmond et al. [57••] developed a methylation-based smoking score using meta-EWAS findings and evaluated its ability to predict prenatal smoking exposure in an independent set of blood samples collected 30 years after pregnancy (Fig. 1; Table 2). The first score they derived was based on 568 loci that reached

Table 2 DNA methylation-based biomarkers of exposure to smoking

Study	Timing of: Smoking exposure		Persistence	Tissue	Model	Predictive accuracy	Notes
	Smoking exposure	DNAm measure					
Reese et al. (2017) [77]	Prenatal	Birth	Days-months	Cord blood	Score: linear combination of 28 loci and LASSO regression coefficients	AUC ^a = 0.90	
Ladd-Acosta et al. (2016) [31••]		Early childhood (age 5)	5 years	Peripheral blood	SVM classifier: 26 loci	AUC ^a = 0.87	
Richmond et al. (2018) [57••]		Adulthood (age 30)	~ 30 years	Peripheral blood	Score: weighted ^e sum of 568 loci ^d	AUC ^a = 0.69	
		Adulthood (age 30)	~ 30 years	Peripheral blood	Score: weighted ^e sum of 19 ^c loci	AUC ^a = 0.72	
		Adulthood (age 30)		Peripheral blood	Score: weighted ^e sum of 2623 ^f loci	AUC ^a = 0.57	Shows current personal smoking score is not a good predictor of prenatal smoking exposure.
Shenker et al. (2013) [78•]	Adult	Later adulthood		Peripheral blood	GLM using 4 loci	AUC ^b = 0.83 AUC ^c = 0.97	DNAm biomarker better predictor than cotinine (AUC = 0.47) DNAm correlated with time since quitting and duration of smoking

DNAm DNA methylation, GLM generalized linear model, SVM support vector machine

^a Predicting prenatally exposed versus unexposed

^b Predicting former vs never smokers

^c Predicting current versus never smokers

^d 568 DNAm loci originally identified in cord blood

^e DNAm loci originally identified in blood during middle childhood

^f DNAm loci identified in adult blood as predictive of current smoking

^g Weights determined by per CpG effect size/average effect size for all measured CpG sites

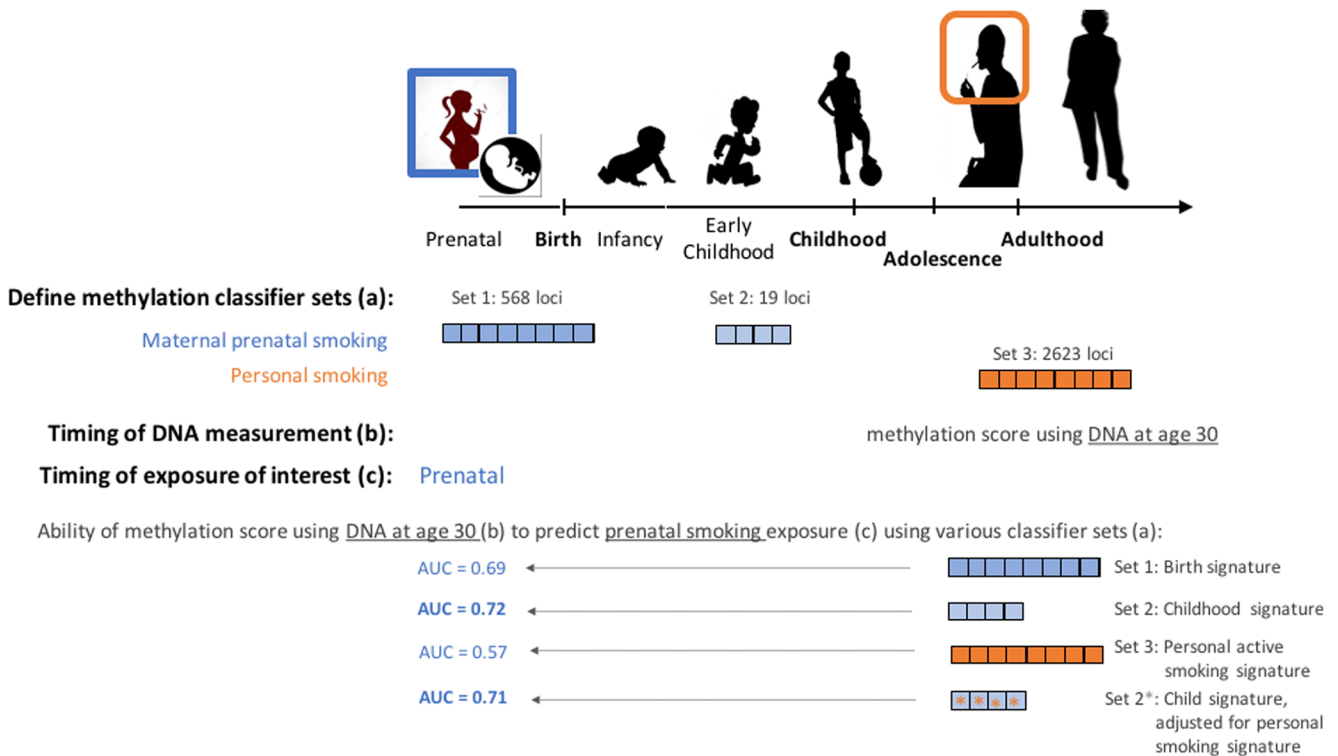


Fig. 1 DNA methylation biomarkers, regardless of timing of sample collection, can be used to predict prenatal smoking. As reported in Richmond et al. [57••], adult biosamples can accurately predict prenatal smoking, even after accounting for post-natal (own) smoking. Pre-

defined sets of CpG DNA methylation loci can be used for prediction. Derived reference sets from infant cord blood and from middle childhood blood are available (top). The CpG set derived from childhood samples achieves slightly better prediction parameters (bottom)

genome-wide significance in cord blood at birth (associated with prenatal smoking exposure) and a second score was based on 19 sites detected in blood from older children at genome-wide significance (associated with prenatal smoking) [26•]. Given the age of the participants at time of blood collection and methylation measurements, it is possible that the offspring themselves smoked; therefore, the authors also computed a methylation score for personal (postnatal) smoking exposure using 2623 sites identified as significantly associated with current smoking status in a large adult smoking meta-analysis [35]. As shown in Fig. 1 and Table 2, the classification accuracy of the prenatal exposure methylation score, based on 30-year-old adult blood specimens, was highest when using the 19 locus methylation score method that had been derived using middle childhood methylation data (AUC = 0.72). Somewhat unexpectedly, the cord blood-derived score had a lower overall prediction accuracy (AUC = 0.69). This highlights the importance of including childhood samples in discovery EWAS and for including loci identified in childhood samples in prenatal biomarker development, if later life biosamples are the intended use. Importantly, they also showed that current smoking exposure scores cannot predict prenatal smoking exposure with high accuracy (AUC = 0.57). Thus, these classifiers appear specific to prenatal exposure. This is consistent with previous observations that there is some, but not complete, overlap of loci

associated with prenatal smoking exposure and personal adolescent or adult smoking exposures [33•, 35].

Finally, separate DNA methylation patterns have been shown to predict prior adult personal smoking exposure. A 4-CpG model using predictive generalized linear models has been shown to predict prior personal smoking status among adults [78•]. The 4-locus model was highly accurate in an independent test sample with an AUC = 0.83 [78•] (Table 2). Furthermore, they showed DNA methylation is a better long-term biomarker of exposure than cotinine. The prediction model using cotinine levels was able to accurately predict former adulthood smoking in only 47% of the samples compared to 83% when DNAm was used as a biomarker of personal smoking history [78•] (Table 2). While associations between DNAm levels and specific dose, duration, and time since quitting have been observed in adults [34–36], these more detailed exposure classes have not been pursued in published predictive analyses to date.

Need for Additional Evidence

The smoking exposure examples demonstrate the potential for DNA methylation-based biomarkers of prior exposure. Multiple studies show the ability to accurately predict prenatal exposure based on DNA methylation measured at birth, in

childhood, and even adulthood. Separate sets of DNA methylation loci can be used to accurately predict past personal adult exposure as well. Further, it appears that these two types of exposures, prenatal and previous personal exposure, can be isolated from each other. There is also a suggestion that quantitative methylation scores may be useful for estimating dose. If fully developed, such biomarkers, across multiple exposures and DNA measurement windows, can dramatically shift our ability to carry out environmental and genetic-environmental epidemiology using existing biobanks. However, much more work must be done. First, studies must move from site-by-site discovery EWAS approaches to classification approaches. The field must establish best practices for selecting CpGs that create accurate and generalizable classifiers. Multiple feature selection algorithms are available, and multiple metrics of predictive accuracy exist. The influence of QC pipelines on accuracy must also be considered, as has been done in other omics classifier work [79]. Perhaps most importantly, the accuracy and utility of DNA methylation biomarkers of exposure must be explored across ancestries and tissue matrices. Because DNA methylation at many CpG sites is, in part, genetically controlled [80, 81], it is likely that DNA methylation signatures of exposure may vary by ancestry. Additionally, the effects of environmental exposures on the epigenome can be influenced by underlying genotypes [82–86]. Genetic heterogeneity is likely to be particularly important among genes that establish, maintain, and regulate DNA methylation as well as for genes involved in exposure metabolism and detoxification. Thus, studies that assess potential genetic modification of epigenetic signatures of exposure are also needed. Tissue type will also play a critical role. While it is not necessary that a biomarker be on the causal path of an exposure to the ultimate health outcome of interest, it may still be true that different DNA methylation sites show predictive accuracy in different cell types. This is because the base level and variability of DNA methylation varies by cell type, and thus, the opportunity for additional variation that captures exposure is likely to be heterogeneous across tissue types. This has already been established for epigenetic clocks, where patterns from single tissue types do not fully overlap in their age prediction accuracy [65]. These caveats to not diminish enthusiasm for this potentially influential area for epidemiology, but do call attention to the rigorous work ahead.

Conclusions The ability to obtain measures of environmental exposures in existing samples and biobanks will enable new large-scale analyses to investigate modifiable environmental risk factors for disease as well as their interaction with genes. Both inherent properties and empiric evidence support the potential for DNA methylation to serve as a stable, long-term biomarker of past exposures across a range of environmental domains. Predictive models and methylation-based exposure scores are emerging and have shown high accuracy in

their ability to predicting former prenatal and adulthood personal smoking exposures. To fully realize the potential of DNA methylation as exposure biomarkers, continued large-scale EWAS and development of predictive models, across time points, tissue types, and ancestry are needed.

Compliance with Ethical Standards

Conflict of Interest M. Daniele Fallin and Christine Ladd-Acosta each declare no potential conflicts of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
 - Of major importance
1. Wang K, Gaitsch H, Poon H, Cox NJ, Rzhetsky A. Classification of common human diseases derived from shared genetic and environmental determinants. *Nat Genet.* 2017;49(9):1319–25.
 2. Samet JM. Tobacco smoking: the leading cause of preventable disease worldwide. *Thorac Surg Clin.* 2013;23(2):103–12.
 3. NTP monograph on health effects of low-level lead. *NTP Monogr.* 2012(1):xiii, xv–148.
 4. Wang G, Divall S, Radovick S, Paige D, Ning Y, Chen Z, et al. Preterm birth and random plasma insulin levels at birth and in early childhood. *JAMA.* 2014;311(6):587–96.
 5. Newschaffer CJ, Croen LA, Fallin MD, Hertz-Picciotto I, Nguyen DV, Lee NL, et al. Infant siblings and the investigation of autism risk factors. *J Neurodev Disord.* 2012;4(1):7.
 6. Oken E, Baccarelli AA, Gold DR, Kleinman KP, Litonjua AA, de Meo D, et al. Cohort profile: project viva. *Int J Epidemiol.* 2015;44(1):37–48.
 7. Jaddoe VW, et al. The generation R study: design and cohort profile. *Eur J Epidemiol.* 2006;21(6):475–84.
 8. Magnus P, Birke C, Vejrup K, Haugan A, Alsaker E, Daltveit AK, et al. Cohort profile update: the Norwegian mother and child cohort study (MoBa). *Int J Epidemiol.* 2016;45(2):382–8.
 9. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort profile: the 'children of the 90s'—the index offspring of the Avon longitudinal study of parents and children. *Int J Epidemiol.* 2013;42(1):111–27.
 10. Claus Henn B, Austin C, Coull BA, Schnaas L, Gennings C, Horton MK, et al. Uncovering neurodevelopmental windows of susceptibility to manganese exposure using dentine microspatial analyses. *Environ Res.* 2018;161:588–98.
 11. Andra SS, Austin C, Arora M. Tooth matrix analysis for biomonitoring of organic chemical exposure: current status, challenges, and opportunities. *Environ Res.* 2015;142:387–406.

12. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779.
13. Cross-Disorder Group of the Psychiatric Genomics, C, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013;45(9):984–94.
14. Benowitz NL. Biomarkers of environmental tobacco smoke exposure. *Environ Health Perspect*. 1999;107(Suppl 2):349–55.
15. Benowitz NL. Cotinine as a biomarker of environmental tobacco smoke exposure. *Epidemiol Rev*. 1996;18(2):188–204.
16. Lee DH, Jacobs DR Jr. Methodological issues in human studies of endocrine disrupting chemicals. *Rev Endocr Metab Disord*. 2015;16(4):289–97.
17. Johns LE, Cooper GS, Galizia A, Meeker JD. Exposure assessment issues in epidemiology studies of phthalates. *Environ Int*. 2015;85:27–39.
18. Christensen JM. Human exposure to toxic metals: factors influencing interpretation of biomonitoring results. *Sci Total Environ*. 1995;166:89–135.
19. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet*. 2011;12(8):529–41.
20. Bakulski KM, Fallin MD. Epigenetic epidemiology: promises for public health research. *Environ Mol Mutagen*. 2014;55(3):171–83.
21. Burris HH, Baccarelli AA. Environmental epigenetics: from novelty to scientific discipline. *J Appl Toxicol*. 2014;34(2):113–6.
22. Cortessis VK, Thomas DC, Levine AJ, Breton CV, Mack TM, Siegmund KD, et al. Environmental epigenetics: prospects for studying epigenetic mediation of exposure-response relationships. *Hum Genet*. 2012;131(10):1565–89.
23. Suter M, Ma J, Harris AS, Patterson L, Brown KA, Shope C, et al. Maternal tobacco use modestly alters correlated epigenome-wide placental DNA methylation and gene expression. *Epigenetics*. 2011;6(11):1284–94.
24. Joubert BR, Håberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect*. 2012;120(10):1425–31.
25. Markunas CA, Xu Z, Harlid S, Wade PA, Lie RT, Taylor JA, et al. Identification of DNA methylation changes in newborns related to maternal smoking during pregnancy. *Environ Health Perspect*. 2014;122(10):1147–53.
26. Joubert BR, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am J Hum Genet*. 2016;98(4):680–96 **Largest epigenome-wide association study for prenatal smoking exposure to date, consisting of 6,685 samples from 13 studies. Identified thousands of loci showing DNA methylation changes in cord blood, at birth, related to *in utero* exposure to smoking.**
27. Hannon E, et al. Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Med*. 2018;10(1):19.
28. Novakovic B, Ryan J, Pereira N, Boughton B, Craig JM, Saffery R. Postnatal stability, tissue, and time specific effects of AHRR methylation change in response to maternal smoking in pregnancy. *Epigenetics*. 2014;9(3):377–86.
29. Breton CV, Byun HM, Wenten M, Pan F, Yang A, Gilliland FD. Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. *Am J Respir Crit Care Med*. 2009;180(5):462–7.
30. Breton CV, et al. Prenatal tobacco smoke exposure is associated with childhood DNA CpG methylation. *PLoS One*. 2014;9(6):e99716 **Provides evidence that DNA methylation patterns present in child DNA reflect prenatal exposure to smoking.**
31. Ladd-Acosta C, et al. Presence of an epigenetic signature of prenatal cigarette smoke exposure in childhood. *Environ Res*. 2016;144(Pt A):139–48 **Reported prenatal smoking associated methylation patterns, originally detected in an independent birth sample, are also present in childhood. They were the first to report prenatal exposure to smoking can be accurately (AUC=0.87) predicted using DNA methylation patterns in the blood of 5 year old children.**
32. Richmond RC, et al. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Hum Mol Genet*. 2015;24(8):2201–17 **Repeated biosampling in children, from birth to age 17, enabled examination of long-term persistence of prenatal smoking associated methylation changes in the same individuals over time. Significant differences in methylation were observed at multiple loci even after adjusting for postnatal household and personal exposures.**
33. Lee KW, et al. Prenatal exposure to maternal cigarette smoking and DNA methylation: epigenome-wide association in a discovery sample of adolescents and replication in an independent cohort at birth through 17 years of age. *Environ Health Perspect*. 2015;123(2):193–9 **Shows DNA methylation changes related to prenatal exposure can be detected in adolescence. Also replicated findings in an independent sample at birth and ages 7 and 17.**
34. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, et al. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum Mol Genet*. 2012;21(13):3073–82.
35. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet*. 2016;9(5):436–47.
36. Lee MK, Hong Y, Kim SY, London SJ, Kim WJ. DNA methylation and smoking in Korean adults: epigenome-wide association study. *Clin Epigenetics*. 2016;8:103.
37. Sharp GC, Arathimos R, Reese SE, Page CM, Felix J, Küpers LK, et al. Maternal alcohol consumption and offspring DNA methylation: findings from six general population-based birth cohorts. *Epigenomics*. 2018;10(1):27–42.
38. Liu C, Marioni RE, Hedman ÅK, Pfeiffer L, Tsai PC, Reynolds LM, et al. A DNA methylation biomarker of alcohol consumption. *Mol Psychiatry*. 2018;23(2):422–33.
39. Richmond RC, Sharp GC, Herbert G, Atkinson C, Taylor C, Bhattacharya S, et al. The long-term impact of folic acid in pregnancy on offspring DNA methylation: follow-up of the Aberdeen folic acid supplementation trial (AFASST). *Int J Epidemiol*. 2018;47:928–37.
40. Steegers-Theunissen RP, Obermann-Borst SA, Kremer D, Lindemans J, Siebel C, Steegers EA, et al. Periconceptional maternal folic acid use of 400 microg per day is related to increased methylation of the IGF2 gene in the very young child. *PLoS One*. 2009;4(11):e7845.
41. Dominguez-Salas P, Moore SE, Baker MS, Bergen AW, Cox SE, Dyer RA, et al. Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nat Commun*. 2014;5:3746.
42. Silver MJ, Kessler NJ, Hennig BJ, Dominguez-Salas P, Laritsky E, Baker MS, et al. Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. *Genome Biol*. 2015;16:118.
43. Waterland RA, Kellermayer R, Laritsky E, Rayco-Solon P, Harris RA, Travisano M, et al. Season of conception in rural Gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet*. 2010;6(12):e1001252.
44. Pauwels S, Ghosh M, Duca RC, Bekaert B, Freson K, Huybrechts I, et al. Maternal intake of methyl-group donors affects DNA methylation of metabolic genes in infants. *Clin Epigenetics*. 2017;9:16.

45. Pauwels S, Ghosh M, Duca RC, Bekaert B, Freson K, Huybrechts I, et al. Dietary and supplemental maternal methyl-group donor intake and cord blood DNA methylation. *Epigenetics*. 2017;12(1):1–10.
46. van Dijk SJ, Zhou J, Peters TJ, Buckley M, Sutcliffe B, Oytam Y, et al. Effect of prenatal DHA supplementation on the infant epigenome: results from a randomized controlled trial. *Clin Epigenetics*. 2016;8:114.
47. Jacobsen SC, Brøns C, Bork-Jensen J, Ribel-Madsen R, Yang B, Lara E, et al. Effects of short-term high-fat overfeeding on genome-wide DNA methylation in the skeletal muscle of healthy young men. *Diabetologia*. 2012;55(12):3341–9.
48. Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, et al. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet*. 2009;18(21):4046–53.
49. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A*. 2008;105(44):17046–9.
50. Finer S, Iqbal MS, Lowe R, Ogunkolade BW, Pervin S, Mathews C, et al. Is famine exposure during developmental life in rural Bangladesh associated with a metabolic and epigenetic signature in young adulthood? A historical cohort study. *BMJ Open*. 2016;6(11):e011768.
51. Parent J, Parade SH, Laumann LE, Ridout KK, Yang BZ, Marsit CJ, et al. Dynamic stress-related epigenetic regulation of the glucocorticoid receptor gene promoter during early development: the role of child maltreatment. *Dev Psychopathol*. 2017;29(5):1635–48.
52. Gruzieva O, Xu CJ, Breton CV, Annesi-Maesano I, Antó JM, Auffray C, et al. Epigenome-wide meta-analysis of methylation in children related to prenatal NO₂ air pollution exposure. *Environ Health Perspect*. 2017;125(1):104–10.
53. Wright RO, Schwartz J, Wright RJ, Bollati V, Tarantini L, Park SK, et al. Biomarkers of lead exposure and DNA methylation within retrotransposons. *Environ Health Perspect*. 2010;118(6):790–5.
54. Cardenas A, Rifas-Shiman SL, Godderis L, Duca RC, Navas-Acien A, Litonjua AA, et al. Prenatal exposure to mercury: associations with global DNA methylation and Hydroxymethylation in cord blood and in childhood. *Environ Health Perspect*. 2017;125(8):087022.
55. Sharp GC, Salas LA, Monnereau C, Allard C, Yousefi P, Everson TM, et al. Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: findings from the pregnancy and childhood epigenetics (PACE) consortium. *Hum Mol Genet*. 2017;26(20):4067–85.
56. Kazmi N, Gaunt TR, Relton C, Micali N. Maternal eating disorders affect offspring cord blood DNA methylation: a prospective study. *Clin Epigenetics*. 2017;9:120.
- 57.●● Richmond RC, et al. DNA methylation as a marker for prenatal smoke exposure in adults. *Int J Epidemiol*. 2018; **Showed methylation scores obtained from DNA collected at age 30, can predict prenatal exposure to smoking with 72% accuracy. Also showed loci associated with postnatal personal smoking are not good predictors of prenatal smoking exposure (AUC=0.57), suggesting methylation patterns differ by exposure window.**
58. Leggett RW. An age-specific kinetic model of lead metabolism in humans. *Environ Health Perspect*. 1993;101(7):598–616.
59. Simpkin AJ, Suderman M, Howe LD. Epigenetic clocks for gestational age: statistical and study design considerations. *Clin Epigenetics*. 2017;9:100.
60. Knight AK, Craig JM, Theda C, Bækvad-Hansen M, Bybjerg-Grauholm J, Hansen CS, et al. An epigenetic clock for gestational age at birth based on blood methylation data. *Genome Biol*. 2016;17(1):206.
61. Bohlin J, Håberg SE, Magnus P, Reese SE, Gjessing HK, Magnus MC, et al. Prediction of gestational age based on genome-wide differentially methylated regions. *Genome Biol*. 2016;17(1):207.
62. Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, et al. Age-associated DNA methylation in pediatric populations. *Genome Res*. 2012;22(4):623–32.
63. Lin Q, Weidner CI, Costa IG, Marioni RE, Ferreira MRP, Deary IJ, et al. DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy. *Aging (Albany NY)*. 2016;8(2):394–401.
64. Weidner CI, et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol*. 2014;15(2):R24.
65. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115.
66. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda SV, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013;49(2):359–67.
67. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, et al. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*. 2012;11(6):1132–4.
68. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, et al. Epigenetic predictor of age. *PLoS One*. 2011;6(6):e14821.
69. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol*. 2015;16:25.
70. Knight AK, Conneely KN, Smith AK. Gestational age predicted by DNA methylation: potential clinical and research utility. *Epigenomics*. 2017.
71. Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H, et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol*. 2016;17(1):171.
72. Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schonfels W, Ahrens M, et al. Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci U S A*. 2014;111(43):15538–43.
73. Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai PC, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)*. 2016;8(9):1844–65.
74. Bossuyt PM, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.
75. Phan JH, Kothari S, Wang MD. omniClassifier: a desktop grid computing system for big data prediction modeling. *ACM BCB*. 2014;2014:514–23.
- 76.●● Zhuang J, Widschwendter M, Teschendorff AE. A comparison of feature selection and classification methods in DNA methylation studies using the Illumina Infinium platform. *BMC Bioinformatics*. 2012;13:59 **Developed a methylation score in cord blood at birth that reflects prenatal exposure to sustained smoking. The score was able to predict exposure status in an independent birth sample with 90% accuracy.**
77. Reese SE, Zhao S, Wu MC, Joubert BR, Parr CL, Håberg SE, et al. DNA methylation score as a biomarker in newborns for sustained maternal smoking during pregnancy. *Environ Health Perspect*. 2017;125(4):760–6.
- 78.●● Shenker NS, et al. *DNA methylation as a long-term biomarker of exposure to tobacco smoke*. *Epidemiology*. 2013;24(5):712–6 **First to show generalized linear model, using methylation levels at 4 CpG sites, can accurately predict previous personal smoking history among adults.**
79. Xu J, Thakkar S, Gong B, Tong W. The FDA's experience with emerging genomics technologies-past, present, and future. *AAPS J*. 2016;18(4):814–8.
80. Andrews SV, Ellis SE, Bakulski KM, Sheppard B, Croen LA, Hertz-Picciotto I, et al. Cross-tissue integration of genetic and

- epigenetic data offers insight into autism spectrum disorder. *Nat Commun.* 2017;8(1):1011.
81. Smith AK, Kilaru V, Kocak M, Almlı LM, Mercer KB, Ressler KJ, et al. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics.* 2014;15:145.
 82. Ren JC, Wu YX, Wu Z, Zhang GH, Wang H, Liu H, et al. MTHFR gene polymorphism is associated with DNA Hypomethylation and genetic damage among benzene-exposed Workers in Southeast China. *J Occup Environ Med.* 2018;60(4):e188–92.
 83. Zhang GH, Lu Y, Ji BQ, Ren JC, Sun P, Ding S, et al. Do mutations in DNMT3A/3B affect global DNA hypomethylation among benzene-exposed workers in Southeast China?: effects of mutations in DNMT3A/3B on global DNA hypomethylation. *Environ Mol Mutagen.* 2017;58(9):678–87.
 84. Declerck K, Remy S, Wohlfahrt-Veje C, Main KM, van Camp G, Schoeters G, et al. Interaction between prenatal pesticide exposure and a common polymorphism in the PON1 gene on DNA methylation in genes associated with cardio-metabolic disease risk-an exploratory study. *Clin Epigenetics.* 2017;9:35.
 85. Aarabi M, San Gabriel MC, Chan D, Behan NA, Caron M, Pastinen T, et al. High-dose folic acid supplementation alters the human sperm methylome and is influenced by the MTHFR C677T polymorphism. *Hum Mol Genet.* 2015;24(22):6301–13.
 86. Sipahi L, Wildman DE, Aiello AE, Koenen KC, Galea S, Abbas A, et al. Longitudinal epigenetic variation of DNA methyltransferase genes is associated with vulnerability to post-traumatic stress disorder. *Psychol Med.* 2014;44(15):3165–79.