

DNA Methylation in Whole Blood: Uses and Challenges

E. Andres Houseman¹ · Stephanie Kim² · Karl T. Kelsey³ · John K. Wiencke⁴

Published online: 17 April 2015
© Springer International Publishing AG 2015

Abstract Due to its convenience, the blood is commonly used in epigenomic studies, but its heterogeneous nature leads to interpretation difficulties, given the now widely recognized potential for confounding by cell composition effects. Many recent publications have reported significant associations between DNA methylation and a variety of health conditions or exposures. In this review, we summarize many of these recent publications, highlighting the findings in the context of potential cell composition effects, particularly findings that are indicative of immune response or inflammation. While there is substantial evidence for confounding by cell composition, there is nevertheless also evidence for differential DNA methylation suggestive of processes that are not cell mediated. We conclude that important biological insights

still may be gained from studying DNA methylation in whole blood, either by investigating the cell composition effects themselves or processes that demonstrate associations even after adjusting for cell composition effects.

Keywords Cell composition effect · Epigenetics · Epigenome-wide association · Epigenomics · Immune · Inflammation · Leukocyte

Introduction

DNA methylation, tightly associated with alterations in the nucleosome DNA scaffold and coordinated chromatin alterations, is partially responsible for coordination of gene expression in individual cells [1–3]. Consequently, in the last decade there has been an interest in studying associations between DNA methylation and a wide variety of phenotypes. Because access to blood specimens is typically much more convenient to obtain from human subjects, the bulk of published studies have used whole blood (sometimes referred to as peripheral blood). A wide variety of phenotypes and health conditions have been studied: aging [4–8], cancer [9–12], obesity [5, 13], cardiovascular disease [14], prenatal exposures/perinatal outcomes [15, 16], environmental exposures [17–24] (tobacco in particular [25, 26]), inflammatory diseases [27, 28], psychiatric conditions [21, 29–32], and fertility [33]. While many of these studies have used candidate gene approaches with bisulfite-pyrosequencing, an increasing number have conducted *epigenome-wide association studies* (EWAS) using commercially available microarrays such as the Infinium HumanMethylation450 BeadChip assay (“450K,” produced by Illumina, Inc.), its predecessor, the Infinium HumanMethylation27 BeadChip (“27K”), or an

This article is part of the Topical Collection on *Environmental Epigenetics*

✉ E. Andres Houseman
andres.houseman@oregonstate.edu

Stephanie Kim
stephanie_min@brown.edu

Karl T. Kelsey
karl_kelsey@brown.edu

John K. Wiencke
John.Wiencke@ucsf.edu

- ¹ School of Biological and Population Health Sciences, College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA
- ² Department of Epidemiology, Brown University School of Public Health, Providence, RI, USA
- ³ Department of Epidemiology, Brown University School of Public Health, Providence, RI, USA
- ⁴ Departments of Neurological Surgery, and Division of Neuroepidemiology, University of California San Francisco, San Francisco, CA, USA

older Illumina product based on the company's GoldenGate product.

Of recent concern has been the extent to which variation in DNA methylation is driven by *cell composition* effects rather than truly intranuclear processes. Normal tissue development, individual cellular differentiation, and cellular lineage determination are regulated by epigenetic mechanisms [2]. This necessarily means that DNA methylation shows substantial variation across tissue types [34, 35] as well as individual cell types, demonstrated particularly clearly among the distinct types of leukocytes [1, 36, 37, 38•, 39•, 40, 41]. Because variation in DNA methylation measured in the blood will necessarily reflect variation in constituent leukocytes, there is a concern that phenotype associations with cell composition will confound (or at least mediate) associations between DNA methylation and phenotype. An additional consideration is that endogenous and exogenous cellular stress can induce inflammatory signaling (arising, for example, in the endoplasmic reticulum of non-immune cells [42]); hence, changes in DNA methylation of stromal or non-immune specific cells affected by the malady of interest will almost certainly represent an important component of the immune response to the perturbed state of interest.

In this article, we review many of the recent studies that have examined associations with DNA methylation in the blood, highlighting cell composition effects through evidence of involvement in the immune or inflammatory responses and effects for which mediation by cell type can reasonably be ruled out. We then briefly discuss methods for mitigating the potential for confounding and potentially assessing mediation by cell composition.

Twin Studies

Several twin studies have examined DNA methylation in the blood. Kaminsky et al. (2009) applied a 12K CpG island microarray to assay whole blood DNA enriched for unmethylated cytosines using the methylation sensitive restriction enzyme HpaII [43]. The study compared 19 monozygotic (MZ) twin pairs and 20 dizygotic (DZ) twin pairs, matched for age, sex, and blood cell count (total, neutrophil fraction, and lymphocyte fraction) and found small but marginally significant differences in intraclass correlation coefficient (ICC) between the matched MZ and DZ twin pairs, suggesting that DNA methylation was slightly more concordant in MZ pairs than in DZ pairs, and thus the existence of very weak but possibly important genetic effects unexplained by proportions of major blood cell types. Boks et al. (2009) used the GoldenGate platform to analyze peripheral blood from 23 MZ twin pairs, 23 DZ twin pairs, and 96 controls matched for age and gender [6]. The study found age-related DNA methylation at loci

suggestive of potential cell composition effects (due to their involvement in immune or inflammatory processes): IL6, CARD15, PDGFRA, and NFKB1. However, the study also found other loci potentially independent of cell composition effects: ACVR1 and ELK. Li et al. (2013) used the 27K array to analyze peripheral blood in 22 MZ twin pairs [44], identifying 92 CpGs that significantly varied between twins within a pair and speculating that the differences were driven principally by immune function.

Associations With Genetic Variants and mRNA Expression

One detailed study used blood to investigate epigenetic associations with single nucleotide polymorphisms (SNPs) and with mRNA expression. Van Eijk et al. (2012) used the 27K platform to study whole blood from 72 male adults and 76 female adults, with a mean age of 52, focusing in particular on associations with SNPs and gene expression [45]. The authors used structural equation models to determine causal directionality, finding that the most common three-way association was the traditional model wherein genetic variants regulate DNA methylation, which in turn regulates expression. They also found that expression modules, defined as “clusters of expression probes,” differed substantially from DNA methylation modules [45]. The authors also used Gene Ontology (GO), an informative tool that describes gene products with associated biological functions and cellular processes, in order to analyze the multiple modules in various GO terms. Interestingly, expression modules were enriched for GO terms suggestive of immune processes but numerous others as well (e.g., those involving transcription and translation). In contrast, compared with expression modules, fewer methylation modules showed enrichment for GO terms, with 5 of 12 enriched GO terms suggestive of immune response.

Aging

Several studies have investigated epigenetic associations with aging. In a candidate gene study, Alexeeff et al. (2013) pyrosequenced selected genomic targets in whole blood from 789 elderly participants of the Normative Aging Study [46]. Significant associations were found for loci mapped to INFG, IL6, TLR2, and iNOS(NOS2), suggestive of immune/inflammatory processes. The authors also found associations with Alu and LINE1 repetitive elements. Similarly, Madrigano et al. (2012) pyrosequenced selected genomic targets in whole blood from 784 elderly participants of the Normative Aging Study [7], finding strongly significant age associations with loci mapped to ICAM, IL6, INFG, iNOS (NOS2), and TLR2, suggestive of immune/inflammatory

processes, but also found associations with loci mapped to genes that were not reflective of immune response: CROT, F3, GCR (NR3C1), and OGG. Note that F3 is involved in clotting and hemostasis and, thus, potentially reflective of signaling processes within the blood. Using the 450K array, Harris et al. (2012) examined mononuclear cells (MCs) from 55 children with Crohn's disease [27], finding a single differentially methylated locus mapped to TEPP. The authors report that TEPP (testis, prostate, and placenta-expressed protein) is poorly expressed in whole blood, though they interpret the small but significant differences as questionable and potentially explained by differences in immune subsets. Alisch et al. (2012) applied the 27K assay to peripheral blood from 398 boys aged 3–17 years, confirming results via 450K in a second pediatric population (78 participants aged 1–16 years) as well as set of 1158 adult subjects [4]. The study reported enrichment for GO terms reflective of both developmental processes and immune function. Finally, Almen et al. (2014) examined age and obesity interactions using the 27K array to assay the blood from 46 adults [5]. Because interaction drove the authors' filtering criteria, the loci demonstrating differential methylation were mapped principally to genes involved in metabolic function.

Cancer

Cancer biomarker studies have been well represented among investigations of DNA methylation in the blood. In a candidate gene study, Cassinotti et al. (2012) analyzed whole blood from 30 colorectal cancer (CRC) patients, 30 patients with adenomatous polyps, and 30 controls [9]. Samples were digested with the restriction enzyme *Hin6I*, and PCR was used to measure DNA methylation at loci mapped to 56 genes. Of six gene promoters selected as members of a biomarker panel for differentiating CRC patients from controls, one (PAX5) was reflective of immune response, while the others reflected cell cycle, tumor suppressor, or oncogene activity [CYCD2(CCND2), HIC1, RASSF1A, RB1, and SRBC(PRKCD8P)]. Of three genes selected for a biomarker panel to differentiate controls from patients with adenomatous polyps (HIC1, MDG1, RASSF1A), none were strongly suggestive of immune response. Flanagan et al. (2009) used a custom tiling array covering 17 candidate genes to analyze peripheral blood from 14 bilateral breast cancer patients and 14 matched controls, validating results via pyrosequencing in 190 cases and 190 controls [10]. The authors found that methylation differences were driven primarily by intragenic repetitive elements, one element associated strongly with lower ATM mRNA levels, thus reflecting a signal independent of cell composition effects. Using 27K, Teschendorff et al. (2009) analyzed peripheral blood from 148 healthy individuals and 113 age-matched pre-treatment ovarian cancer cases,

developing a biomarker to distinguish cases from controls [12]. Results were highly suggestive of immune response (which the authors link to aging processes), but also developmental pathways potentially independent of cell composition effects. Similarly, using 27K, Marsit et al. (2011) analyzed the blood from 112 bladder cancer patients and 118 controls, finding that differentially methylated loci were enriched for pathways suggestive of immune response or developmental pathways [11].

Fertility

The epigenetics of infertility has also been studied using blood. Friemel et al. (2014) applied the 450K array to peripheral blood from 30 infertile males aged 27–42 years (median age 35.5) and 10 fertile males aged 21–52 years old (median age 39.5) [33]. They found differential methylation for PIWI L1 and PIWIL2, both genes important in spermatogenesis and the former of which may regulate hematopoietic stem cells. Significant loci were enriched for the *MHC class II* GO term and HLA genes, reflective of immune activity.

Pregnancy and Birth

Many groups have used blood to study epigenetic processes involved in various outcomes related to pregnancy, birth, and early childhood. Martino et al. (2011) conducted a longitudinal study using MC samples from seven females, collecting cord blood at birth and following up with blood samples through 5 years [47]; all samples were arrayed via the 27K array. Loci showing significant longitudinal change were enriched for cell surface receptor and signal transduction terms reflective of changes in immune response. While the authors applied FACS to MCs to determine fractions of major cell types (CD4⁺ T cells, CD8⁺ T cells, B cells, and monocytes), analyzed the resultant cell fractions for longitudinal changes, and demonstrated change over time, the authors did not include the fractions as potential confounders in analysis of DNA methylation. Relton et al. (2012) applied the GoldenGate assay to blood from 178 birth cohort subjects to investigate associations with body composition measures at 9 years of age [48]. The most robust association was found for CpGs mapped to ALPL, which is principally involved with bone density/skeletal growth. Other genes to which significant CpGs were mapped include CASP10, CDKN1C, EPHA1, HLA-DOB, IRF5, MMP9, MPL, and NID1, two of which implicate immune function (HLA-DOB and IRF5) and one involved in hemostasis (MPL). Liu et al. (2014) studied 308 African-American mother-infant pairs assaying cord blood via 27K [49]. Loci showing significant associations with maternal pre-conception BMI were enriched for several

infection and inflammation pathways, but also tumorigenesis and apoptosis pathways. Morales et al. (2014) applied the GoldenGate assay to cord blood samples from 258 birth cohort subjects, studying associations with pre-pregnancy BMI and gestational weight gain [50]. Among the 44 loci most significantly associated with weight gain were several immune and inflammation related genes (IL16, IL1B, IL8, NFKB1). NFKB1 in particular was selected as one of four genes the authors report as functionally relevant (others being MMP7, KCNK4, and TRPM5). Non et al. (2014) assayed cord blood via 450K, comparing 13 mothers with non-medicated depression or anxiety, 22 mothers taking SSRIs, and 23 controls [30]. GO terms enriched for loci differing between controls and non-medicated mothers with depression or anxiety consisted mostly of terms related to transcription and translation of DNA, although methylation differences were small; the authors found no loci associated with SSRI use. White et al. (2013) assayed blood via 27K to compare 14 pregnant women having preeclampsia to 14 normotensive controls [15]. None of the 19 top hits were suggestive of immune or inflammatory processes, although none were significant after adjusting for multiple comparisons. Sanders et al. (2013) studied cadmium exposure among 17 mothers with blood assayed using the methylated CpG island recovery assay (MIRA) [22]. Enriched GO terms were not only reflective of cell cycle and cancer pathways but also OF immune response.

Environmental Exposures

In addition to the cadmium study mentioned above, numerous publications report epigenetic associations in the blood with environmental exposures. Bind et al. (2012) studied 704 elderly male subjects (mean age 73.2) by pyrosequencing candidate genes in the blood and examining associations with traffic-related pollutants [18]. They found significant interactions of DNA methylation and air pollution on C-reactive protein and fibrinogen for loci mapped to TLR2 (suggestive of immune effects) as well as for loci mapped to F3 (related to hemostasis) and for Alu and LINE-1 repeats. Madrigano et al. (2012) pyrosequenced 1377 blood samples for loci mapped to iNOS (NOS2) and GCR/NR3C1 [21]. Acknowledging a potential cell composition effect, the authors found associations with black carbon and PM_{2.5} for NOS2 (reflective of immune effects) but not for GCR. Kile et al. (2013) pyrosequenced NOS2 and repetitive elements in the blood from 38 welders, finding PM_{2.5} associations with NOS2 but not with repeats [19]. Alegria-Torres et al. (2013) pyrosequenced several genes in the blood from 39 male brick manufacturers, finding associations with polycyclic aromatic hydrocarbons (measured in urine) and DNA methylation of loci mapped to IL12, TNFA, p53, and Alu repeats [17]; note that IL12 and TNFA reflect immune response. Tarantini et al. (2013) pyrosequenced several

targets in the blood from 63 steel workers, studying association between DNA methylation and particulate matter assumed to be rich in metals [23]. They found associations not only between PM₁₀ and DNA methylation at loci mapped to NOS3 (reflective of immune response) but also between zinc exposure and methylation at loci mapped to EDN1 (not reflective of immune response).

Several studies have, in particular, investigated associations with exposures to tobacco smoke. Using the 27K array, Breitling et al. (2011) studied associations between DNA methylation and smoking in the blood from 177 subject [25]. Associations were found for loci mapped to F2RL3 (reflective of hemostasis) and replicated in 316 independent samples. Via 450K, Zellinger et al. (2013) studied smoking associations using the blood from 1793 subjects [26]. They found associations at loci mapped to AHRR (related to detoxification and not to immune response) and replicated the association in 479 independent samples. Somewhat related to smoking, Qiu et al. (2012) studied chronic obstructive pulmonary disease (COPD), using 27K to study the blood from two family-based cohorts ($n=1085$, $n=369$), investigating associations between DNA methylation and COPD [51]. They report associations for loci mapped to SERPINA1 (related to hemostasis) and FUT7. Neither gene is reflective of immune response or inflammation, although they report that 349 significant loci were enriched for GO terms reflective of these processes, as well as others (wound healing and coagulation cascades as well as response to stress and external stimuli).

Psychiatry

Finally, several studies have investigated associations between DNA methylation and psychiatric conditions. Using the 27K array, Nishioka et al. (2012) compared 18 schizophrenics with 15 controls [29], finding 603 differentially methylated CpG sites, many of these mapped to genes critical in neuronal differentiation and related to other psychiatric disorders, as well as genes functionally related to those previously found to be differentially methylated in schizophrenic patients. Enriched GO terms emphasized transcription factor binding and nucleotide binding, but neither immune response nor inflammation. In a candidate gene study employing pyrosequencing, Rusiecki et al. (2013) compared 75 post-traumatic stress disorder (PTSD) cases with 75 controls [31]; they found differences in loci mapped to IL18 (reflective of immune response) and to H19. In another candidate gene study pyrosequencing 82 candidate genes, Zhang et al. (2013) studied childhood adversity in alcohol-dependent patients, stratified by race (African-American vs. European American), for a total of 518 cases and 369 controls [32]. Significant loci were mapped to metabolic genes or those related to neurotransmission, but the custom panel was heavily biased towards such genes.

Table 1 Overview of various studies' findings related to DNA methylation associations that may or may not indicate for cell composition effects

Source	Topic of study	Population	Genes of interest (examples)
Studies that have found DNA methylation associations that are reflective of immune response/inflammation (indicative of potential confounding by cell composition)			
Boks et al. (2009) [6]	Twin studies	23 MZ twins, 23 DZ twins, 96 matched controls	IL6, CARD15, PDGFRA, NFKB1
Li et al. (2013) [44]	Twin studies	13 female MZ twins, 9 male MZ twins	EFNB1, G6PD
Van Eijk et al. (2012) [45]	Genetic variants/gene expression	148 healthy adults of Dutch ancestry	N/A ^a
Alexeeff et al. (2013) [46]	Aging	789 elderly participants of the Normative Aging Study	INFG, IL6, TLR2, iNOS (NOS2)
Madrigano et al. (2012) [7]	Aging	784 elderly participants of the Normative Aging Study	ICAM, IL6, INFG, iNOS (NOS2), TLR2
Harris et al. (2012) [27]	Aging	55 children with Crohn's disease	TEPP
Alisch et al. (2012) [4]	Aging	398 healthy males (3–17 years of age) from Simons Simplex Collection	N/A ^b
Teschendorff et al. (2009) [12]	Cancer	148 healthy post-menopausal women, 113 age-matched pre-treatment ovarian cancer cases	INS, IL-2, BRAF, MAPK1
Marsit et al. (2011) [11]	Cancer	112 bladder cancer cases, 118 controls	GATA1, IRF1, IRF7, NFAT
Friemel et al. (2014) [33]	Fertility	30 infertile men with normal CFTR and AZF tests, 10 fertile male controls	PIWIL1, PIWIL2, HLA-DRB1, HLA-DQB1
Martino et al. (2011) [47]	Pregnancy/birth	7 females (followed from birth to 5 years)	CALCA, DNNT, CHRNE
Relton et al. (2012) [48]	Pregnancy/birth	178 birth cohort subjects from Avon Longitudinal Study of Parents and Children	HLA-DOB, IRF5, MPL
Liu et al. (2014) [49]	Pregnancy/birth	308 African-American mother-infant pairs	WNT16, ERBB2, DOK2, PLAC1
Morales et al. (2014) [50]	Pregnancy/birth	258 birth cohort subjects from Avon Longitudinal Study of Parents and Children	IL16, IL1B, IL8, NFKB1
Sanders et al. (2013) [22]	Pregnancy/birth	17 mother-infant pairs (with maternal cadmium exposure level of 0.2 µg/L)	NCF4, MTF1, PRR13
Bind et al. (2012) [18]	Environmental exposures	704 elderly men from Veterans Administration Normative Aging Study	TLR2, ICAM-1, F3
Madrigano et al. (2012) [21]	Environmental exposures	1377 blood samples from 699 elderly male participants in the VA Normative Aging Study	iNOS (NOS2), GCR/NR3C1
Kile et al. (2013) [20]	Environmental exposures	38 boilermaker construction workers	iNOS (NOS2)
Alegria-Torres et al. (2013) [17]	Environmental exposures	39 male brick manufacturers	IL12, TNFA
Tarantini et al. (2013) [23]	Environmental exposures	63 steel workers	NOS3
Rusiecki et al. (2013) [31]	Psychiatry	75 post-traumatic stress disorder (PTSD) cases, 75 controls	IL18, H19
Studies that have found no DNA methylation associations with immune response/inflammation (indicative of independence of cell composition effects or potential for other biological bias)			
Madrigano et al. (2012) [7]	Aging	784 elderly participants of the Normative Aging Study	CROT, F3, GCR (NR3C1), and OGG
Almen et al. (2014) [5]	Aging	24 obese and 22 lean female adults from Latvian Genome Data Base	RNH1, ADCY1, BRUNOL6
Cassinotti et al. (2012) [9]	Cancer	30 colorectal cancer (CRC) patients, 30 patients with adenomatous polyps, 30 controls	CYCD2(CCND2), HIC1, RASS F1A, MDG1
Flanagan et al. (2009) [10]	Cancer	14 bilateral breast cancer patients, 14 matched controls	ATM, PGR, CDH1
White et al. (2013) [15]	Pregnancy/birth	14 pregnant women with preeclampsia, 14 matched normotensive controls	RIN2b, GABRA1, PCDHB7, and BEX1
Tarantini et al. (2013) [23]	Environmental exposures	63 steel workers	EDN1
Zellinger et al. (2013) [26]	Environmental exposures	1793 subjects from KORA F4 cohort	AHRR
Qiu et al. (2012) [51]	Environmental exposures	1085 subjects from the International COPD Genetics Network (ICGN), 369 subjects	SERPINA1, FUT7

Table 1 (continued)

Source	Topic of study	Population	Genes of interest (examples)
Nishioka et al. (2012) [29]	Psychiatry	from the Boston Early-Onset COPD Study (EOCOPD) 18 cases with first-episode schizophrenia, 15 controls	HTR1E, COMTD1
Zhang et al. (2013) [32]	Psychiatry	518 cases of alcohol dependence and 369 controls	ALDH1A1, CART, CHRNA5, HTR1B

^a This study has determined five methylation modules associated with immune-related Gene Ontologies (GO)

^b This study has determined ten modules associated with immune-related Gene Ontologies (GO)

Overview of Studies That Assay Blood Without Adjusting for Cell Composition

Most of the studies reviewed above found associations near genes that were reflective of immune response or inflammation (Table 1). In addition, many studies found associations near genes that were involved in hemostasis/coagulation, which are coordinated and closely regulated by immunologically active cells and related cytokines. Most of the studies that report no associations with immune or inflammation processes used panels that were heavily biased towards other biological processes (e.g., the GoldenGate Cancer Panel measures methylation across CpG sites in promoter regions of genes with known implications in cancer, and thus, it may not provide an unbiased assessment of the epigenome). The evident associations near genes reflective of immune response or inflammation suggest the potential for phenotype associations with leukocyte cell composition. Consequently, it stands to reason that the associations reported in many of the studies reviewed above may be confounded by cell composition effects as well as, potentially, by localized expanded numbers of cells with activated immune signaling pathways (Fig. 1). This point has been made recently by several authors [52••, 53]; in particular, Jaffe and Irizarry (2013) suggest that many of the observed associations between DNA methylation and age may be substantially confounded by cell composition effects [52••] (Fig. 1a). On the other hand, a number of the studies found associations near genes that were not obviously related to immune function or inflammation, so there remains a great potential for epigenetic signals that arise independently of cell composition (Fig. 1c).

Strategies for Avoiding Confounding by Cell Composition

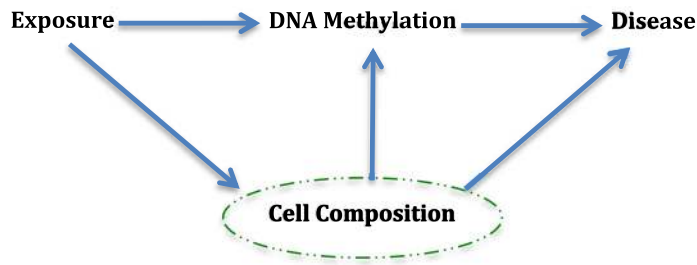
Multiple strategies have emerged for avoiding confounding by cell composition. The most direct method is to fractionate leukocytes and either to study a single cell type or, alternatively, to statistically adjust for directly measured cell counts or proportions. For example, Lam et al. (2013) argue that it is

critical to account for granulocyte proportion in EWAS studies, either by removing them and arraying MCs, or minimally, adjusting for them statistically [54]. Nestor et al. (2014) compared eight seasonal allergic rhinitis (SAR) patients with eight controls, arraying sorted CD4⁺ T cells using the 450K platform. The results showed clear separation in methylation profiles, and along with gene expression profiles, emphasized interleukin genes related to lymphocyte activation. However, we note that Th1 and Th2 cells differ in production of INFG, IL2, IL4, IL5, IL6, IL10, and the two cell types are differentially methylated in the promoter of INFG [41]. In general, lineage-specific DNA methylation regulates differentiation of T cell subsets, with DMRs present in cell type-specific genes (FOXP3, IL2RA, CTL4A, CD40LG, INFG) [40]. Thus, even in isolated cell types, DNA methylation associations may be confounded by subtle cell composition effects, including those related to cellular memory and activation state.

In general, cell sorting is difficult and error-prone. Though conventional complete blood cell (CBC) count methods are routine and inexpensive, they can only differentiate major leukocyte types (lymphocytes, monocytes, and neutrophils) in freshly isolated blood (Fig. 2a). More subtle distinctions, such as characterization of differences among Tregs, NK cells, or dendritic cells, can be made using FACS analysis on whole blood or blood cell fractions (Fig. 2b). However, FACS analyses require extensive logistical support and fresh blood samples and are costly; hence, FACS is infrequently used in large clinical or epidemiological studies. At the same time, activated immune cells will exist in many disease states; for example, activated NK cells, dendritic cells, monocytes, macrophages, etc. are the hallmarks of inflammatory conditions. While their numbers are likely to be small in the periphery, they will have very different DNA methylation signatures at particular loci. Consequently, if these cells are numerous enough, these differences will be detected as very small differences in the beta coefficients (mean methylation values).

One increasingly popular method of addressing cell composition effects is to adjust for them statistically. Liu et al. (2013) demonstrated that DNA methylation associations with rheumatoid arthritis are explained principally by cell composition effects [28•]. In this study, they imputed the proportions

a. Confounding by cell composition



b. Mediation Effects



c. Independent of cell composition effects



d. Reverse causality

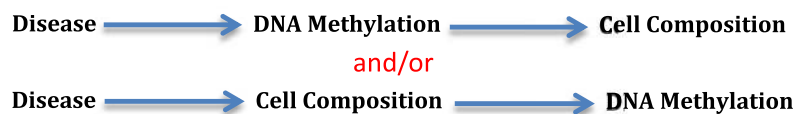


Fig. 1 Directed acyclic graphs (DAGs) of various scenarios in which cell composition can play a role in confounding or mediating the methylation associations, not being involved in the pathway or being involved in reverse causality. **a** Cell composition acts as a confounder of the association between DNA methylation and outcome of interest (i.e., certain disease). Varying amounts of different cell types may influence the differential methylation patterns expressed and also have direct impact (i.e., immune activation or inflammatory responses) related to disease. In most of these studies with implications for immune-related associations, cell composition effects have not been properly adjusted for and thus may

bias the methylation-associated results. **Dotted green lines** indicate the need to adjust for cell composition effects. **b** Cell composition effects are involved as mediators in the pathway associated with disease. Cell composition can either influence or be influenced by DNA methylation in its association with the outcome. **c** The association of DNA methylation patterns with disease may not be influenced by cell composition effects, but rather by other biological mechanisms. **d** Diseased states may potentially influence DNA methylation patterns and/or cell compositions in the blood and thus indicates potential limitation of studies due to reverse causality

of major cell types using a deconvolution algorithm whose use is becoming increasingly popular [38••]; this algorithm deploys information gained from a reference panel of sorted cell types, which in principle could be expanded to include profiles for rare leukocyte types, although the algorithm may be limited by the sensitivity of the assay used to measure DNA methylation (Fig. 2c). Another alternative is a new reference-free deconvolution method [56]; although its use seems promising, as evidenced by a recent study that used it to find biomarkers for Wilms tumors [56], its robustness remains to be confirmed.

Conclusions

Due to its convenience, the blood is commonly used in epigenomic studies, but its heterogeneous nature leads to interpretation difficulties. Many publications now report significant associations between DNA methylation and a variety of health conditions or exposures; typically, studies that have used genome-wide platforms have found associations indicative of immune response or inflammation, which almost certainly represent effects that are mediated by cell composition. Nevertheless, since many of these

a Complete Blood Count (CBC)

Test	Normal Levels	Certain Causes for Abnormal Results
White Blood Cell (WBC) <u>Specific Evaluation:</u> - Neutrophils (Absolute Neutrophil Count) - Lymphocytes (Absolute Lymphocyte Count) - Monocytes (Absolute Monocyte Count) - Eosinophil (Absolute Eosinophil Count) - Basophil (Absolute Basophil Count)	$3-10 \times 10^9$ cells/L*	<ul style="list-style-type: none"> • Autoimmune disorders • Viral infection • Inflammation • Leukemia
Red Blood Cell (RBC) <u>Specific Evaluation:</u> - Hemoglobin (Hb) - Hemocrit (Hct)	$4-6 \times 10^{12}$ cells/L*	<ul style="list-style-type: none"> • Dehydration • Anemia • Kidney dysfunction • Bone marrow damage
Platelets (PLT)	$100-500 \times 10^9$ cells/L*	<ul style="list-style-type: none"> • Viral Infection • Cancer • Anemia • Autoimmune disorders

*These normal level values are estimates for adults [59].

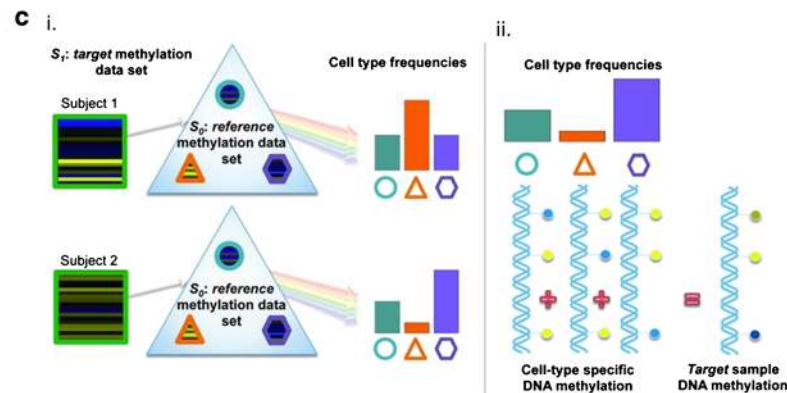
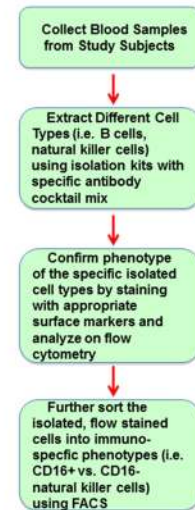
b Flow Cytometry/Fluorescence Activated Cell Sorter (FACS)

Fig. 2 Various strategies to control possible confounding by cell composition. **a** Complete blood count tests can be performed to determine the levels of white blood cells, red blood cells, and platelets. **b** Flow cytometry staining and fluorescence-activated cell sorter (FACS) may be performed to isolate specific immune cell types. Major limitations of these methods are that they are labor intensive and often costly. **c** Previously established statistical algorithms can be applied to control for cell composition effects (**c** adapted with permission from: Koestler et al.: Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a validation analysis. *Epigenetics* 8(8):816–26 (2013). doi: 10.4161/epi.25430) [55]. **i** By projecting the methylation values from an experimental data set to a reference library of DNA methylation signatures for major immune cell types (i.e., B cells, T

cells, granulocytes, monocytes, and NK cells), the estimates of specific cell(s) proportions in the blood can be determined. **ii** The methylation signatures for experimental samples are the weighted sum of the methylation signatures from distinct white blood cell types, where the weights are proportional to the specific cell-type frequencies in the blood. Illustration of the blood cell mixture deconvolution approach reproduced (and slightly altered) from [55]. The deconvolution approach involves (i) constrained projection of DNA methylation profiles from a target methylation data set (S1) onto a reference data set (S0), which is comprised of the DNA methylation signatures for isolated white blood cell types (*shapes* reflect different white blood cell types). The result is an estimate of the underlying distribution of cell proportions (circle, triangle, and hexagon) for each sample within S1

studies also report associations with other processes that are not easily explained as cell composition effects, the blood is still a valuable tissue to assay. Indeed, important biological insights may be gained from studying the cell composition effects themselves, not to mention their potential interactions with other processes.

Direct measurement of counts of various leukocyte cell types would be the ideal method for conducting such analysis; however, it is generally expensive and logistically difficult to measure these counts in a large study population beyond the standard CBC differential used clinically. Fortunately, there

are statistical methods available for generating approximate cell proportions imputed from reference data; while these are necessarily inferior to directly measured counts, they may often represent an acceptable tradeoff between bias and feasibility.

Clearly, DNA methylation studies of blood tissue and of all other tissues should include a detailed consideration of the epigenomic features that are intrinsic to the cells that make up the tissue. Our knowledge of the differentially methylated loci in leukocyte subtypes, although far more extensive than even a few years ago, is still incomplete. This fact means that

many DNA methylation associations attributed to intrinsic epigenetic processes are likely due to subtle effects on cell type composition involving cells that have yet to be characterized. Current human leukocyte libraries account for only approximately seven subtypes, but numerous specific subtypes may exist when one counts activation states of many common types [57, 58]. Memory versus naïve T or B cells, subsets of activated NK cells as well as different forms of dendritic and myeloid cells have yet to be studied. Current research is underway to characterize these types. As more data sets become available to characterize the epigenomic variation among different and less common functional subtypes of leukocytes, there will be an interest in applying these reference data in EWAS. However, success in such application will require the use of technologies that are more sensitive than the currently used microarray platforms, so we anticipate that in the future digital sequencing technologies will play a more prominent role in the conduct of such studies.

Compliance With Ethics Guidelines

Conflict of Interest E. Andres Houseman has a patent, US Application No. is 14/089,398 (filed 11/25/13).

Stephanie Kim declares that she has no conflict of interest.

Karl T. Kelsey has a patent, US Application No. is 14/089,398 (filed 11/25/13).

John K. Wiencke has a patent, US Application No. is 14/089,398 (filed 11/25/13).

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.

References

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

- Of importance
- Of major importance

1. Ji H, Ehrlich LI, Seita J, et al. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature*. 2010;467(7313):338–42. doi:10.1038/nature09367.
2. Khavari DA, Sen GL, Rinn JL. DNA methylation and epigenetic control of cellular differentiation. *Cell Cycle*. 2010;9(19):3880–3.
3. Natoli G. Maintaining cell identity through global control of genomic organization. *Immunity*. 2010;33(1):12–24. doi:10.1016/j.immuni.2010.07.006.
4. Alisch RS, Barwick BG, Chopra P, et al. Age-associated DNA methylation in pediatric populations. *Genome Res*. 2012;22(4):623–32.
5. Almén MS, Nilsson EK, Jacobsson JA, et al. Genome-wide analysis reveals DNA methylation markers that vary with both age and obesity. *Gene*. 2014;548(1):61–7.
6. Boks MP, Derks EM, Weisenberger DJ, et al. The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One*. 2009;4(8):e6767.
7. Madrigano J, Baccarelli A, Mittleman MA, et al. Aging and epigenetics. *Epigenetics*. 2012;7(1):63–70.
8. Sahin K, Yilmaz S, Gozukirmizi N. Changes in human sirtuin 6 gene promoter methylation during aging. *Biomed Rep*. 2014;2(4):574–8.
9. Cassinotti E, Melson J, Liggett T, et al. DNA methylation patterns in blood of patients with colorectal cancer and adenomatous colorectal polyps. *Int J Cancer*. 2012;131(5):1153–7.
10. Flanagan JM, Munoz-Alegre M, Henderson S, et al. Gene-body hypermethylation of ATM in peripheral blood DNA of bilateral breast cancer patients. *Hum Mol Genet*. 2009;18(7):1332–42.
11. Marsit CJ, Koestler DC, Christensen BC, et al. DNA methylation array analysis identifies profiles of blood-derived DNA methylation associated with bladder cancer. *J Clin Oncol*. 2011;29(9):1133–9. doi:10.1200/JCO.2010.31.3577.
12. Teschendorff AE, Menon U, Gentry-Maharaj A, et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One*. 2009;4(12):e8274. doi:10.1371/journal.pone.0008274.
13. Dick KJ, Nelson CP, Tsaprouni L, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet*. 2014;383(9933):1990–8. doi:10.1016/S0140-6736(13)62674-4.
14. Kim M, Long TI, Arakawa K, et al. DNA methylation as a biomarker for cardiovascular disease risk. *PLoS One*. 2010;5(3):e9692. doi:10.1371/journal.pone.0009692.
15. White WM, Brost B, Sun Z, et al. Genome-wide methylation profiling demonstrates hypermethylation in maternal leukocyte DNA in preeclamptic compared to normotensive pregnancies. *Hypertens Pregnancy*. 2013;32(3):257–69.
16. Joubert BR, Haberg SE, Nilsen RM, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect*. 2012;120(10):1425–31. doi:10.1289/ehp.1205412.
17. Alegria-Torres JA, Barretta F, Batres-Esquivel LE, et al. Epigenetic markers of exposure to polycyclic aromatic hydrocarbons in Mexican brickmakers: a pilot study. *Chemosphere*. 2013;91(4):475–80.
18. Bind M-A, Baccarelli A, Zanobetti A, et al. Air pollution and markers of coagulation, inflammation and endothelial function: associations and epigenome-environment interactions in an elderly cohort. *Epidemiology (Camb, Mass)*. 2012;23(2):332.
19. Kile ML, Fang S, Baccarelli AA, et al. A panel study of occupational exposure to fine particulate matter and changes in DNA methylation over a single workday and years worked in boilermaker welders. *Environ Heal*. 2013;12(1):47.
20. Kile ML, Houseman EA, Baccarelli AA, et al. Effect of prenatal arsenic exposure on DNA methylation and leukocyte subpopulations in cord blood. *Epigenetics*. 2014;9(5):774–82. doi:10.4161/epi.28153.
21. Madrigano J, Baccarelli A, Mittleman MA, et al. Air pollution and DNA methylation: interaction by psychological factors in the VA Normative Aging Study. *Am J Epidemiol*. 2012;176(3):224–32.
22. Sanders AP, Smeester L, Rojas D, et al. Cadmium exposure and the epigenome: exposure-associated patterns of DNA methylation in leukocytes from mother-baby pairs. *Epigenetics*. 2013;9(2):0–9.
23. Tarantini L, Bonzini M, Tripodi A, et al. Blood hypomethylation of inflammatory genes mediates the effects of metal-rich airborne pollutants on blood coagulation. *Occup Environ Med*. 2013. doi:10.1136/oemed-2012-101079.
24. Koestler DC, Avissar-Whiting M, Houseman EA, et al. Differential DNA methylation in umbilical cord blood of infants exposed to low levels of arsenic in utero. *Environ Health Perspect*. 2013;121(8):971–7. doi:10.1289/ehp.1205925.
25. Breitling LP, Yang R, Korn B, et al. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet*. 2011;88(4):450–7.

26. Zeilinger S, Kühnel B, Klopp N, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One*. 2013;8(5):e63812.
27. Harris RA, Nagy-Szakal D, Pedersen N, et al. Genome-wide peripheral blood leukocyte DNA methylation microarrays identified a single association with inflammatory bowel diseases. *Inflamm Bowel Dis*. 2012;18(12):2334–41.
28. • Liu Y, Aryee MJ, Padyukov L, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol*. 2013;31(2):142–7. doi:10.1038/nbt.2487. **This study provided the first example of EWAS analysis adjusting for cell proportions using inferred proportion estimates.**
29. Nishioka M, Bundo M, Koike S, et al. Comprehensive DNA methylation analysis of peripheral blood cells derived from patients with first-episode schizophrenia. *J Hum Genet*. 2012;58(2):91–7.
30. Non AL, Binder AM, Kubzansky LD, et al. Genome-wide DNA methylation in neonates exposed to maternal depression, anxiety, or SSRI medication during pregnancy. *Epigenetics*. 2014;9(7).
31. Rusiecki JA, Byrne C, Galdzicki Z, et al. PTSD and DNA methylation in select immune function gene promoter regions: a repeated measures case–control study of US military service members. *Front Psychiatry*. 2013;4.
32. Zhang H, Wang F, Kranzler HR, et al. Profiling of childhood adversity-associated DNA methylation changes in alcoholic patients and healthy controls. *PLoS One*. 2013;8(6):e65648.
33. Friemel C, Ammerpohl O, Gutwein J, et al. Array-based DNA methylation profiling in male infertility reveals allele-specific DNA methylation in PIWIL1 and PIWIL2 Fertil Steril. 2014;101(4):1097–1103.
34. Christensen BC, Houseman EA, Marsit CJ, et al. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet*. 2009;5(8):e1000602. doi:10.1371/journal.pgen.1000602.
35. Illingworth R, Kerr A, DeSousa D, et al. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol*. 2008;6(1):e22.
36. Accomando WP, Wiencke JK, Houseman EA, et al. Decreased NK cells in patients with head and neck cancer determined in archival DNA. *Clin Cancer Res*. 2012;18(22):6147–54. doi:10.1158/1078-0432.CCR-12-1008.
37. Accomando WP, Wiencke JK, Houseman EA, et al. Quantitative reconstruction of leukocyte subsets using DNA methylation. *Genome Biol*. 2014;15(3):R50. doi:10.1186/gb-2014-15-3-r50.
38. • Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinforma*. 2012;13:86. doi:10.1186/1471-2105-13-86. **This study proposes the widely-used reference-based algorithm for inferring cell proportion from DNA methylation.**
39. • Reinius LE, Acevedo N, Joerink M, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One*. 2012;7(7):e41361. **This study presents the most widely used reference data set for inferring cell proportions from DNA methylation.**
40. Schmidl C, Klug M, Boeld TJ, et al. Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res*. 2009;19(7):1165–74.
41. Yano S, Ghosh P, Kusaba H, et al. Effect of promoter methylation on the regulation of IFN- γ gene during in vitro differentiation of human peripheral blood T cells into a Th2 population. *J Immunol*. 2003;171(5):2510–6.
42. Hotamisligil GS. Endoplasmic reticulum stress and the inflammatory basis of metabolic disease. *Cell*. 2010;140(6):900–17.
43. Kaminsky ZA, Tang T, Wang SC, et al. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet*. 2009;41(2):240–5. doi:10.1038/ng.286.
44. Li C, Zhao S, Zhang N, et al. Differences of DNA methylation profiles between monozygotic twins' blood samples. *Mol Biol Rep*. 2013;40(9):5275–80.
45. van Eijk KR, de Jong S, Boks MP, et al. Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. *BMC Genomics*. 2012;13(1):636.
46. Alexeeff SE, Baccarelli AA, Halonen J, et al. Association between blood pressure and DNA methylation of retrotransposons and pro-inflammatory genes. *Int J Epidemiol*. 2013;42(1):270–80.
47. Martino DJ, Tulic MK, Gordon L, et al. Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans. *Epigenetics*. 2011;6(9):1085–94.
48. Relton CL, Groom A, Pourcain BS, et al. DNA methylation patterns in cord blood DNA and body size in childhood. *PLoS One*. 2012;7(3):e31821.
49. Liu X, Chen Q, Tsai HJ, et al. Maternal preconception body mass index and offspring cord blood DNA methylation: exploration of early life origins of disease. *Environ Mol Mutagen*. 2014;55(3):223–30.
50. Morales E, Groom A, Lawlor DA, et al. DNA methylation signatures in cord blood associated with maternal gestational weight gain: results from the ALSPAC cohort. *BMC Res Notes*. 2014;7(1):278.
51. Qiu W, Baccarelli A, Carey VJ, et al. Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Am J Respir Crit Care Med*. 2012;185(4):373–81.
52. •• Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15(2):R31. **This study demonstrates the potential for confounding by cell composition.**
53. Lam LL, Emberly E, Fraser HB, et al. Reply to Suderman et al.: importance of accounting for blood cell composition in epigenetic studies. *Proc Natl Acad Sci*. 2013;110(14):E1247.
54. Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics*. 2014. doi:10.1093/bioinformatics/btu029.
55. Koestler DC, Christensen B, Karagas MR, et al. Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a validation analysis. *Epigenetics*. 2013;8(8):816–26. doi:10.4161/epi.25430.
56. Charlton J, Williams RD, Weeks M, et al. Methylome analysis identifies a Wilms tumour epigenetic biomarker detectable in blood. *Genome Biol*. 2014;15(8):434.
57. Ornatsky O, Bandura D, Baranov V, Nitz M, et al. Highly multiparametric analysis by mass cytometry. *J Immunol Methods*. 2010;361(1–2):1–20. doi:10.1016/j.jim.2010.07.002.
58. Sekiguchi DR, Smith SB, Sutter JA, et al. Circulating lymphocyte subsets in normal adults are variable and can be clustered into subgroups. *Cytometry B Clin Cytom*. 2011;80(5):291–9. doi:10.1002/cyto.b.20594.