

Temporal Stability and Determinants of White Blood Cell DNA Methylation in the Breakthrough Generations Study

James M. Flanagan¹, Mark N. Brook², Nick Orr³, Katarzyna Tomczyk³, Penny Coulson², Olivia Fletcher³, Michael E. Jones², Minouk J. Schoemaker², Alan Ashworth³, Anthony Swerdlow^{2,4}, Robert Brown^{1,5}, and Montserrat Garcia-Closas^{2,3}

Abstract

Background: Epigenome-wide association studies (EWAS) using measurements of blood DNA methylation are performed to identify associations of methylation changes with environmental and lifestyle exposures and disease risk. However, little is known about the variation of methylation markers in the population and their stability over time, both important factors in the design and interpretation of EWAS. We aimed to identify stable variable methylated probes (VMP), i.e., markers that are variable in the population, yet stable over time.

Methods: We estimated the intraclass correlation coefficient (ICC) for each probe on the Illumina 450K methylation array in paired samples collected approximately 6 years apart from 92 participants in the Breakthrough Generations Study. We also evaluated relationships with age, reproductive and hormonal history, weight, alcohol intake, and smoking.

Results: Approximately 17% of probes had an ICC > 0.50 and were considered stable VMPs (stable-VMPs). Stable-VMPs

were enriched for probes located in "shores" bordering CpG islands, and at approximately 1.3 kb downstream from the transcription start site in the transition between the unmethylated promoter and methylated gene body. Both cross-sectional and longitudinal data analyses provided strong evidence for associations between changes in methylation levels and aging. Smoking-related probes at 2q37.1 and *AHRR* were stable-VMPs and related to time since quitting. We also observed associations between methylation and weight changes.

Conclusion: Our results provide support for the use of white blood cell DNA methylation as a biomarker of exposure in EWAS.

Impact: Larger studies, preferably with repeated measures over time, will be required to establish associations between specific probes and exposures. *Cancer Epidemiol Biomarkers Prev*; 24(1); 221–9. ©2014 AACR.

Introduction

Epigenetic profiles, including methylation of the 5-carbon of cytosines, may help to unravel the pathogenesis of numerous complex diseases, in particular for diseases that have an environmental component to their etiology. Epigenetic modifications, such as DNA methylation or histone modifications, are key determinants of chromatin structure, genomic stability, and gene expression. These modifications are maintained during cell division and when perturbed play a key role in cancer development (1). Recent genomic approaches have identified thousands of genes that are aberrantly methylated in cancer, and associated with risk of

recurrence and prognosis (2, 3). Epigenetic changes could also be a biologic indicator of lifetime accumulation of environmental exposures related to early-life exposures (4), aging (5), hormones (6), alcohol (7), and smoking (8, 9). Previously, it was thought that DNA methylation was a stable yet reversible modification (10). Supporting this, we and others have reported stable methylation changes associated with smoking in white blood cell (WBC) DNA of former smokers many years after smoking cessation (9, 11), and several loci have been reported to associate with developmental exposure to famine many years after the exposure (4, 12). However, there is also evidence in the literature that DNA methylation can change rapidly during development (13) and in response to many different exposures (14–17).

Several studies have reported epigenetic traits in WBC DNA as potential cancer risk markers for solid tumors, including total genomic 5-methylcytosine (18, 19) candidate genes (20–23) and epigenome-wide association studies (EWAS), to identify blood-based DNA methylation signatures of risk in ovarian cancer (24), bladder cancer (25), head and neck squamous cell carcinomas (26), and breast cancer (9, 27). Although EWAS report intriguing associations, replication in larger studies with prediagnostic samples is required.

Few studies have investigated temporal stability of WBC DNA methylation in serial samples from the same individuals (28–30), which is an important step before utilization of a biomarker in population-based studies. Given that only methylation traits that

¹Epigenetics Unit, Department of Surgery and Cancer, Imperial College London, London, United Kingdom. ²Division of Genetics and Epidemiology, The Institute of Cancer Research, London, United Kingdom. ³Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, United Kingdom. ⁴Division of Breast Cancer Research, The Institute of Cancer Research, London, United Kingdom. ⁵Division of Medicine, The Institute of Cancer Research, London, United Kingdom.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Author: James M. Flanagan, Imperial College London, 4th Floor IRDB, Hammersmith Campus, Du Cane Road, London W12 0NN, UK. Phone: 44-20-75942127; Fax: 44-20-75942129; E-mail: j.flanagan@imperial.ac.uk

doi: 10.1158/1055-9965.EPI-14-0767

©2014 American Association for Cancer Research.

are stable over time can be well characterized using a single blood sample, it is important to establish the temporal stability of methylation markers in WBC DNA. The aim of our study was to identify stable variable methylated probes (VMP), i.e., markers that are variable in the population, yet stable over time, and to investigate associations with age, reproductive and hormonal history, and lifestyle factors. Cross-sectional and longitudinal analyses were carried out based on 92 paired serial blood samples collected at two time points approximately six years apart, from women participating in the Breakthrough Generations Study (BGS).

Materials and Methods

Study subjects

The study subjects were a sample of 92 participants in the BGS, a large general population cohort consisting of approximately 110,000 women enrolled in the UK from 2003 to 2011 (31). Eligible women were those enrolled in 2004 and who met the following criteria: had a blood sample and questionnaire collected at baseline (T0) and again at follow-up six years later (T1), 35 to 84 years old at enrollment, be free of any cancer before second blood collection, have blood samples received at processing lab no more than one day after collection, being white, and not known to have a relative in the eligible population. This identified a total of 532 eligible women. This population was further restricted to those reporting no problems at collection or processing (e.g., lipemic, hemolyzed, clotted samples), and with time between the receipt of the two samples ranging from 5.5 to 6.5 years. This resulted in 418 women eligible to be included in the analyses. We obtained a random sample of these 418 women to fill in two 96-well plates with paired T0 and T1 DNA samples from each woman, and obtained methylation data from 92 pairs of samples. Paired samples were processed on the same chips to minimize intraindividual batch effects. Sample storage time in liquid nitrogen tanks at -80°C was 7.1 years for T0 (range, 6.8–7.7) and 1.2 years for T1 (range, 0.6–1.3). For quality control (QC), we included duplicate samples from four study subjects that showed strong correlation ($R^2 > 0.90$) across probes that had methylation levels between 30% and 70%. This correlation compared to $R^2 < 0.69$ in the unrelated pairs. We also included reference DNA samples 100% and 0% methylated (in duplicate), which were used to filter poor performing probes (see below).

Illumina 450K beadchip methylation analysis

DNA samples were extracted from buffy coats using DNA Blood Mini Kits (Qiagen). Bisulphite conversion of 600 ng of each DNA sample was performed using the EZ-96 DNA Methylation-Gold Kit according to the manufacturer's protocol (Zymo Research). Then, 4 μL of bisulphite-converted DNA was used for hybridization on the Infinium HumanMethylation 450 BeadChip, following the Illumina Infinium HD Methylation protocol (Illumina). Hybridization, scanning, and raw data processing were performed by UCL Genomics (www.genomics.ucl.ac.uk/). The intensities of the images were extracted using the GenomeStudio (v.2011.1) Methylation module (1.9.0) software (Illumina), which normalizes within-sample data using different internal controls that are present on the HumanMethylation 450 BeadChip and internal background

probes. The methylation score for each CpG was represented as a β value according to the fluorescent intensity ratio with any values between 0 (unmethylated) and 1 (completely methylated). Raw microarray data, processed data, and a table of all intraclass correlation coefficient (ICC) and interquartile range (IQR) for each probe are available from Gene Expression Omnibus (accession GSE61151).

Statistical analysis

Raw methylation data were exported from GenomeStudio as background subtracted β values with corresponding detection P values. Quality control exclusion criteria were as follows: probes not detected in the population (detection P value > 0.05), samples with $>1\%$ of probes not detected, and probes not detected in $>20\%$ of samples. A total of 484,954 probes and all samples passed these initial quality control measures. The resulting data sets included samples from 92 individuals at T0 and a repeated sample from the same 92 individuals at T1. Data were preprocessed using peak-based correction (32) and COMBAT to adjust for batch effects (33). Individual probes were further filtered based on the following criteria: (i) 100% methylated reference $\beta > 0.80$ and 0% unmethylated reference $\beta < 0.20$ (excluding 40,470 probes), and (ii) annotated SNPs located within probe sequence (excluding 89,474 probes), leaving a total of 355,633 probes for further analysis.

DNA methylation values were coded as M values ($\log_2(\beta/1 - \beta)$). We used the ICC to identify markers that are variable in the BGS population yet stable over time. The ICC is defined as the ratio of the between-subject variance to the total variance (between-subject + within-subject variance that reflects both temporal and assay variation). Stable-VMPs were defined as those with ICC > 0.50 , i.e., probes that showed a larger variation between individuals than within individuals. IQR of $<1\%$ was used to define probes that were considered to have low variability across individuals.

Multivariate linear regression models were used to evaluate associations between methylation M values and the following variables (in relation to time of blood collection, unless otherwise specified): age, age at menarche, parity, age at first birth, menopausal status, age at menopause, oral contraceptive (OC) use, hormone replacement therapy (HRT) use, smoking status, years since quitting smoking, weight, alcohol use, and sample storage time, evaluated at time of each blood collection. Each pair-wise combination of probe and exposure was evaluated in separate regression models adjusted for age and geographical location (defined by postcodes into regions of South, East Anglia, Midlands, North, Scotland, Wales). In addition, we performed surrogate variable analyses (SVA; refs. 34, 35) and analyses adjusted by cell type using the "methylSpectrum" method (36). Briefly, methylSpectrum calculates the WBC subtype proportions in a sample of whole blood by comparing methylation signatures in the dataset of interest with methylation signatures in a reference dataset where the WBC subtype proportion is known. Cross-sectional data analyses were performed separately for T0 and T1. Longitudinal data analysis was used to evaluate determinants of changes in methylation across both time points. Random effects linear regression models for longitudinal data were used to evaluate relationships between changes in age and weight with changes in methylation levels over time. Unlike cross-sectional analyses at T0 or T1, longitudinal data

analyses control for unmeasured confounders that vary between subjects but not within subjects over time.

Quantile–quantile (Q-Q) plots were used to compare the observed and expected distribution of P values under the null hypothesis of no association. Q-Q plots provide a quick visual guide to assess departures from the null hypothesis across all the statistical tests performed for each of the probes. This is evidenced by a departure from the diagonal line that was measured by the inflation factor λ . Volcano plots were used to examine the relationship between regression coefficients and the corresponding P values from linear regression analyses. Volcano plots are used to visually identify probes that are significant and have a large effect, i.e., those probes in the upper left or right quadrants. Enrichment of stable-VMPs in specific probe types was tested using a 1df χ^2 test dichotomizing the probes into above and below the threshold for stable-VMPs and into the different probe categories tested. Differences in methylation β values or ICC values between two groups of probes were tested using the Wilcoxon signed rank-sum test. For these exploratory enrichment analyses, we considered $P < 10^{-4}$ as statistically significant. For analyses to confirm previously reported associations between methylation levels and age or smoking, we used the conventional $P < 0.05$ as the level of significance.

Statistical analyses were conducted using Stata/MP 12.1 for Windows and the following R packages: "icc" command from the "irr" package for ICC calculations, "dmrFind" command from the "Charm" package (37) for linear regression analyses including

SVA, "plm" package for longitudinal data analyses (38), and qqplot function from the "stats" package for Q-Q plots (<http://cran.r-project.org/>).

Results

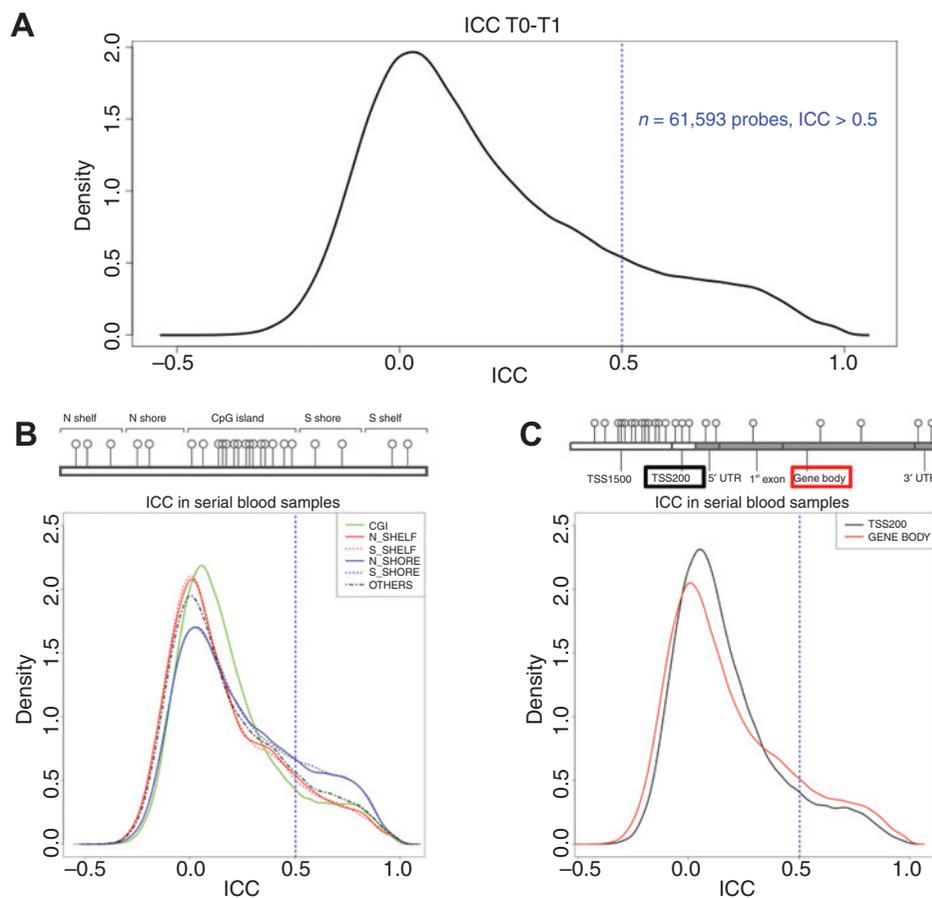
Identification of stable-VMPs in the Illumina 450K array

We defined stable-VMPs as those that had ICC > 0.50 because this implies larger variation between individuals than within individuals over time. Using this threshold, we identified 61,593 stable-VMPs of 355,633 probes that had passed QC and filtering (Fig. 1A). These constitute 13% of all probes or 17% of probes after excluding probes with SNPs and failed reference samples. These stable-VMPs were significantly ($P < 1 \times 10^{-4}$) enriched for CpG island "shores" (2 kb up- and downstream of CpG islands) compared with other regions (Fig. 1B), and for gene body compared with promoter probes (probes at 200 bp from transcription start sites, TSS200; $P < 1 \times 10^{-4}$; Fig. 1C).

A smoothed plot of the mean ICC and mean methylation level for each probe against the genomic location of the probe in relation to the TSS is shown in Fig. 2. This showed a peak of probes with higher ICC at approximately 1.3 kb downstream from the TSS (Wilcoxon signed rank-sum test, $P < 10^{-4}$ compared with TSS200 probes), corresponding to a mean methylation level of 50%, halfway between low promoter methylation and high intragenic methylation. We observed this for both plus strand and minus strand genes.

Figure 1.

Analysis of temporal variation of 450K probes in 92 serial WBC DNA samples in the BGS. A, distribution of ICC for 450K. Using a threshold of ICC > 0.50 , we identified 61,593 probes that were considered to be both variable and stable over time (stable-VMPs). B, these stable-VMP probes were enriched for probes in "shore" regions (blue) compared with "shelf" (red), CpG islands (green), and other regions (black). C, ICC distribution for probes in the promoter (TSS200) and gene body showed enrichment for gene body probes.



Determinants of WBC DNA methylation

Table 1 shows the distribution of subject characteristics at baseline (T0) and follow-up (T1), and whether there were changes in these variables over time. The median (range) age increased from 53 years (35–77) in T0 to 59 years (41–83) in T1, and the percentage of postmenopausal women increased from 45% to 66%. The median increase in weight from T0 to T1 was 1 kg.

Q-Q plots showed strong evidence for an excess of associations beyond that expected by chance for methylation levels and aging, both in cross-sectional analyses at T0, T1 and in longitudinal data analyses (T0: $\lambda = 1.19$, T1: $\lambda = 1.23$, longitudinal: $\lambda = 1.34$; Fig. 3A–C). SVA and cell-type adjustments did not explain the excess of associations.

Funnel plots suggested that highly significant probes tended to be associated with increases rather than decreases in methylation with aging (Fig. 3D and E). Regression coefficients for age in the cross-sectional and longitudinal data analyses were highly correlated ($\rho = 0.67$ for T0 and $\rho = 0.68$ for T1; Fig. 3D and E). Longitudinal data regression coefficients for 148 of 151 probes with data available that have recently been reported to be associated with age (30)

had $P < 0.05$ and showed coefficients highly correlated with coefficients in the original publication (Supplementary Fig. S1 and Supplementary Table S1). The regression coefficients from longitudinal analyses shown in tables are the average effect of age on methylation levels when age changes across time and between subjects by one year, i.e., they reflect both changes between subject and across time.

Q-Q plots for weight showed evidence for an association with methylation levels, particularly after adjustment for cell type (Supplementary Fig. S2). Regression coefficients for weight in the cross-sectional and longitudinal data analyses were highly correlated ($\rho = 0.68$ for T0 and $\rho = 0.68$ for T1), and volcano plots for weight suggested a tendency for probes with higher coefficients and lower P values to show associations with increased rather than decreased methylation with weight. The probes with the strongest evidence for associations with weight ($P < 10^{-5}$) are shown in Table 2.

Q-Q plots showed little evidence for an association between smoking status or time since quitting smoking and methylation levels (Supplementary Fig. S3). Of 29 probes that had been previously associated with smoking status at the time of analysis (9, 39, 40), three probes at 2q37.1 and one probe at

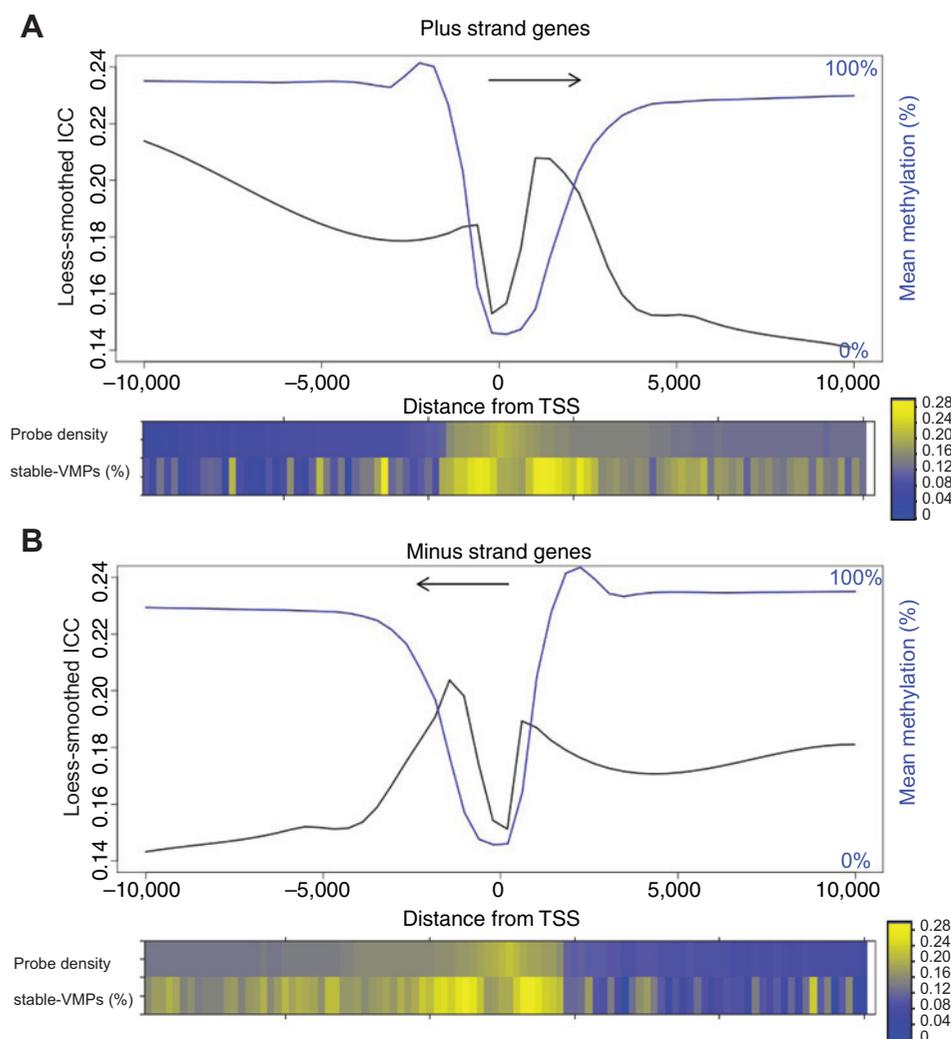


Figure 2.

Distribution of average ICC and mean methylation percentage according to genomic location. A, in black, the Loess-smoothed average of ICC across 95,277 probes associated with plus strand genes (genes oriented in the rightward direction in relation to the genome location) and within ± 10 kb of the TSS. In blue, the Loess-smoothed mean methylation of the same probes. Panel below shows a heatmap of scaled probe density and the proportion of stable-VMPs (ICC > 0.5) in each 200-bp window from -10 kb to $+10$ kb. B, in black, the Loess-smoothed average of ICC across 92,964 probes associated with minus strand genes (genes oriented in the leftward direction in relation to the genome location) and within ± 10 kb of the TSS. In blue, the Loess-smoothed mean methylation of the same probes. Probes in "shores" are not included. Panel below shows a heatmap of scaled probe density and the proportion of stable-VMPs (ICC > 0.5) in each 200-bp window from -10 kb to $+10$ kb.

Table 1. Characteristics of women participating in the study at two points in time, T0 and T1, approximately 6 years apart

	T0	T1
Individuals, <i>N</i>	92	92
Age at blood collection, years, median (range)	53 (35–77)	59 (41–83)
Age at menarche, years, median (range)	13 (10–17)	N/A
Menopausal status		
Premenopausal, <i>N</i> (%)	33 (36%)	13 (14%)
Postmenopausal (excluding current HRT users), <i>N</i> (%)	41 (45%)	61 (66%)
Undetermined/on HRT, <i>N</i> (%)	18 (20%)	18 (20%)
Age at menopause (postmenopausal only, excluding current HRT users), years, median (range)	50 (29–56)	52 (40–58)
Parity		
Nulliparous, <i>N</i> (%)	14 (15%)	No change
Parous, <i>N</i> (%)	78 (85%)	No change
Age at first birth (parous only), years, median (range)	25 (18–39)	No change
OC use		
Never, <i>N</i> (%)	21 (23%)	20 (22%)
Former, <i>N</i> (%)	70 (76%)	62 (67%)
Current, <i>N</i> (%)	1 (1%)	0 (0%)
Missing, <i>N</i> (%)	0 (0%)	10 (11%)
HRT use (postmenopausal only)		
Never, <i>N</i> (%)	25 (46%)	37 (51%)
Former, <i>N</i> (%)	16 (30%)	24 (33%)
Current, <i>N</i> (%)	13 (24%)	7 (10%)
Missing, <i>N</i> (%)	0 (0%)	4 (6%)
Weight, kg, median (range)	65 (49–127)	66 (49–121)
Weight change between T0 and T1, kg, median (range)		1 (–16–18)
Body mass index, median (range)	24 (20–51)	25 (19–49)
Regular alcohol use		
No, <i>N</i> (%)	18 (20%)	7 (8%)
Yes, <i>N</i> (%)	74 (80%)	74 (80%)
Missing, <i>N</i> (%)	0 (0%)	11 (12%)
Duration of alcohol use, years, median (range)	31 (2–44)	N/A
Smoking status		
Never, <i>N</i> (%)	61 (66%)	61 (66%)
Former, <i>N</i> (%)	28 (30%)	29 (32%)
Current, <i>N</i> (%)	3 (3%)	2 (2%)
Time since quitting, former smokers, years, median (range)	26 (4–47)	31.5 (10–53)
Age when started smoking (ever smokers), years, median (range)	17 (14–21)	N/A
WBC subtype, percentage, median (range)		
Lymphocytes		
T cells CD8 ⁺	9 (3–21)	10 (3–30)
T cells CD4 ⁺	15 (2–28)	16 (1–26)
Nature killer cells	4 (0–14)	4 (0–11)
B cells	8 (4–13)	8 (5–12)
Monocytes	10 (7–18)	10 (7–16)
Granulocytes	52 (38–65)	51 (38–66)

Abbreviation: N/A, not assessed.

AHRR showed significant associations with smoking status at $P < 0.05$ both at T0 and T1, in the same direction as previously reported (i.e., lower methylation levels for ever than never smokers; Table 3). Our data also suggested increased levels of methylation with longer time since quitting in former smokers, although not all of the probes were statistically significant (Table 3). Other probes also showed associations in the same direction as previously reported; however, they did not reach $P < 0.05$ in either T0 or T1 (Supplementary Table S2).

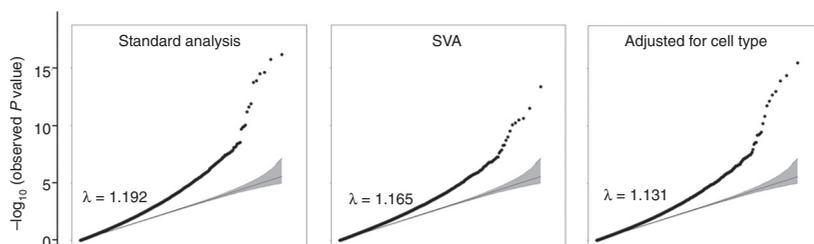
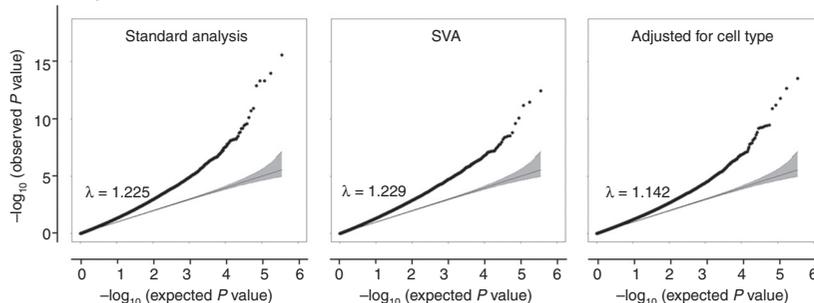
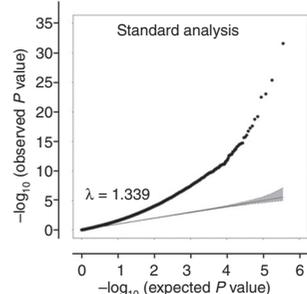
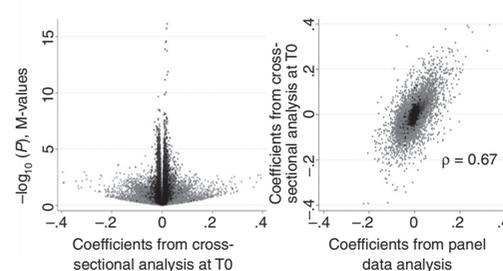
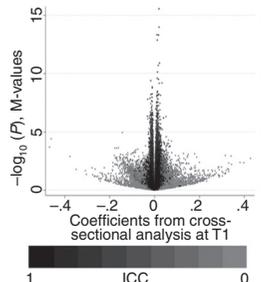
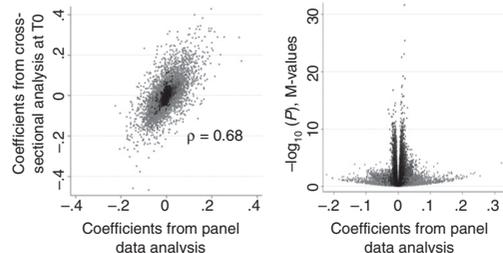
Examination of Q-Q plots for cross-sectional analyses of methylation levels and other subject characteristics at T0 and T1 showed little evidence for associations with methylation levels for most variables, with some indication of an excess signal at the tail for alcohol use (Supplementary Fig. S4).

Discussion

We have shown that approximately 17% ($N = 61,593$) of probes on the Illumina 450K methylation array can be considered stable-

VMPs, and thereby representing good candidates for EWAS using WBC DNA that incorporate a single time-point measure of DNA methylation levels. The remaining probes represent either markers with low variation across individuals or markers that vary over time within an individual. There was considerable enrichment for stable-VMPs, approximately 1.3 kb downstream of the transcription start sites of most genes. This is a genic category that is poorly covered by the Illumina 450K array for many genes and, therefore, would be better interrogated in future studies using unbiased genome-wide bisulphite sequencing. Nevertheless, the genes for which data are available for this category suggest that interindividual variation is highest at intermediate methylation levels (41). We have observed this for both plus strand and minus strand genes, supporting the hypothesis that the transition from promoter to gene body is the genomic region of most intraindividual variability (41). There could be several possible mechanisms for why there is increased variation in this genic region and how this may influence variability in gene expression that should be tested in future research. These include modified

Flanagan et al.

A Cross-sectional analyses at T0**B** Cross-sectional analyses at T1**C** Longitudinal analyses for changes between T0 and T1**D** Cross-sectional analyses at T0**E** Cross-sectional analyses at T1**F** Longitudinal analyses for changes between T0 and T1**Figure 3.**

Associations between DNA methylation probes and age at blood collection. Q-Q plot for the relationship between age and DNA methylation M values at T0 (A), T1 (B), and longitudinal data (C), using standard analyses, SVA, and cell-type adjustment. Volcano plots for regression coefficients for the relationship between age and DNA methylation M values and the corresponding $-\log_{10}(P \text{ value})$ at T0 (D), T1 (E), and longitudinal data (F). Grayscale (light-dark) shows ICC values (low-high). Scatter plots of regression coefficients from cross-sectional (D, E) and longitudinal analyses (F) of age and DNA methylation M values are shown.

transcription efficiency, different levels of sense and antisense mRNA, or altering (or being altered by) the variability in local histone marks that are focused on this promoter proximal intragenic region, such as H3K36me3 (42, 43).

There is considerable evidence from monozygotic twin studies (44, 45) and cross-sectional studies (5, 46, 47) that the epigenome changes with age. This phenomenon has been described as epigenetic drift (48–50) and suggests a general deregulation in maintaining the fidelity of tissue-specific epigenetic patterns over time. Q-Q plots showed a large excess of associations between methylation and age beyond that expected by chance. This could reflect a real biologic effect of age or a systematic error not accounted for in the analyses. The excess of associations was present in longitudinal data analyses that

account for unmeasured time-invariant confounders, as well as in cross-sectional analyses after adjustment for unmeasured confounding using SVA and after blood cell-type correction, suggesting these might be real biologic effects. Our data also suggested a tendency of associations between increased rather than decreased methylation levels and increasing age that needs confirmation. Although this report does not include replication of age associations in independent populations, our data did provide strong confirmation of age associations previously reported by Florath and colleagues (30), both in cross-sectional and longitudinal data analyses. A limitation of our analysis, however, is the short time interval between blood collections, with a change in age of only about 6 years. Additional work is needed to identify specific probes related to age and to gain

Table 2. ICC and association between weight and methylation levels for probes with $P < 10^{-5}$

Probe	Gene	Type of probe	Position (HG b37)	ICC	Cross-sectional, T0		Cross-sectional, T1		Longitudinal	
					β^a	P	β^a	P	β^a	P
cg08536358	PDHA2	II	96760014	0.01	-0.842	3.4×10^{-7}	-0.014	0.214	-0.400	7.1×10^{-7}
cg12641838	PSMB9	I	32822013	0.02	-0.164	3.1×10^{-6}	-0.022	0.252	-0.094	1.8×10^{-6}
cg14233050	PXN	II	120651871	0.00	-0.826	5.7×10^{-7}	0.007	0.256	-0.383	2.3×10^{-6}
cg12354986	TRIOBP	I	38142174	0.25	-0.047	4.2×10^{-4}	-0.055	0.001	-0.052	2.6×10^{-6}
cg05838155	FBXL20	I	37557276	0.09	-0.029	0.003	-0.023	0.005	-0.028	5.4×10^{-6}
cg17501210	RPS6KA2	II	166970252	0.74	-0.031	0.002	-0.035	0.001	-0.039	8.2×10^{-6}
cg10453777	13q21.2	II	62245229	0.14	-0.010	0.129	-0.027	2.3×10^{-5}	-0.020	9.5×10^{-6}

NOTE: Bolded results have consistent direction of association for cross-sectional analyses in T0 and T1, and longitudinal analyses for weight change between T0 and T1.

^a β coefficient per 5 kg change in weight (kg).

insight into the biologic mechanisms underlying this association between methylation and increasing age.

We and others have previously shown that methylation in blood DNA is influenced by environmental exposures such as smoking (9, 10, 39, 40). In the present study, we have confirmed associations between smoking habits and three probes at 2q37.1 and one probe at *AHRR* previously reported by Shenker and colleagues (9). Methylation levels in these probes were related to time since quitting smoking in former smokers and had ICC > 0.5, suggesting that methylation levels in stable-VMPs could be used as biomarkers of exposures occurring in the past. Previous studies have also shown associations between different measures of body size and methylation in blood DNA (29, 51–53). Our data provided support for associations between methylation levels and increasing weight, with a tendency for decreased rather than increased methylation and typically low ICC, suggesting these markers are unstable over time. Of the other exposures evaluated, alcohol use showed the strongest evidence for associations with methylation levels in WBC DNA. However, replication in future studies is required to establish associations with specific probes. In addition, although analyses were adjusted by age, larger studies will be required to better assess the potential confounding and effect modification of age given its strong influence on methylation levels. The biologic mechanisms underlying associations between past lifestyle exposures such as former smoking, weight changes, and alcohol use are unclear. In particular, a large proportion of WBCs are short lived [approximately 50% of WBCs in our study samples were granulocytes (Table 1) that are primarily formed by neutrophils with an average lifespan of 1 to 4 days, whereas 38% were lymphocytes that have an average lifespan of one week to a few months]. Therefore, a potential mechanism for long-lasting effects on methylation would be for exposures to induce methylation changes in the blood stem cell population that persist in each subsequent differentiated cell population. However, currently

there is no evidence yet to support this stem cell hypothesis. Furthermore, it should be noted that methylation variability in WBC DNA may not reflect variability in other tissues or cell types.

We used several approaches to account for measured as well as unmeasured confounding factors, including longitudinal data analyses, SVA, and cell-type corrections in cross-sectional analyses. Generally, cross-sectional and longitudinal analyses gave similar results, and SVA and blood cell-type correction had modest impact in reducing potential inflations of test statistics in Q-Q plots, on occasions even resulting in overcompensation. Thus, although taking into consideration that WBC contains a mixture of cell types is important, this and other reports (27, 54) have demonstrated that adjusting for blood cell subtype using current methods does not always have a substantial impact on results from EWAS. Using direct measures of blood cell composition or accounting for health conditions or medication use that could affect blood cell composition could be useful; however, these analyses were not possible in our sample. Therefore, although we used the best available methods to account for possible biases, it is possible that signals seen in Q-Q plots reflect biases unaccounted by analyses rather than real biologic effects. Future developments in statistical methods for the analysis of methylation data might be required to better account for inflation of test statistics observed in EWAS.

In conclusion, we have identified over 61,000 stable-VMPs in the Illumina 450K methylation array that represent good candidates for EWAS using a single blood sample per individual. Probes that are not considered stable-VMPs because of substantial temporal variation may still be informative if they vary across individuals, but may require serial DNA samples to capture the temporal variability when evaluating associations with other phenotypes. Probes that have low variability across individuals (IQR < 1%, $n = 74,666$ probes) may be considered uninformative for EWAS. Although this array is still widely used, the genomic coverage is limited (~2% of

Table 3. Association between smoking habits and methylation markers with $P < 0.05$ for associations with smoking status at T0 and T1, and corresponding ICC

Probe	Cytoband	Gene	Position (HG b37)	ICC	Association with smoking status (ever/never)				Association with years since stop smoking in former smokers			
					T0		T1		T0		T1	
					β	P	β	P	β per 5 years	P	β per 5 years	P
cg06644428 ^a	2q37.1		233284112	0.600	-0.205	0.043	-0.331	0.001	0.121	0.071	0.049	0.422
cg05951221 ^a	2q37.2		233284402	0.891	-0.209	0.001	-0.192	0.002	0.079	0.012	0.050	0.141
cg21566642 ^a	2q37.3		233284661	0.832	-0.167	0.003	-0.210	9×10^{-5}	0.070	0.037	0.031	0.238
cg05575921 ^{a,b,c}	5p15.33	<i>AHRR</i>	373378	0.723	-0.190	4×10^{-4}	-0.177	0.001	0.042	0.064	0.059	0.019

^aPreviously reported in ref. 9.

^bPreviously reported in ref. 40.

^cPreviously reported in ref. 39.

CpG sites) and only about 13% of the probes were identified as stable-VMPs. Thus, it is likely that many more regions across the genome will be identified as stable-variable regions using newer technologies such as whole-genome bisulphite sequencing. We also provided strong confirmation of associations between methylation levels in WBC DNA with age and past smoking. Further studies, preferably using serial blood collections and/or intervention studies, will be required to identify further methylation probes in blood that reflect changes in exposures.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: J.M. Flanagan, A. Swerdlow, R. Brown, M. Garcia-Closas

Development of methodology: J.M. Flanagan

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J.M. Flanagan, N. Orr, K. Tomczyk, M.E. Jones, M.J. Schoemaker, A. Ashworth, A. Swerdlow, M. Garcia-Closas

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J.M. Flanagan, M.N. Brook, M. Garcia-Closas

Writing, review, and/or revision of the manuscript: J.M. Flanagan, M.N. Brook, N. Orr, O. Fletcher, M.E. Jones, M.J. Schoemaker, A. Ashworth, A. Swerdlow, R. Brown, M. Garcia-Closas

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): P. Coulson, M.E. Jones, M.J. Schoemaker
Study supervision: J.M. Flanagan, R. Brown, M. Garcia-Closas

Acknowledgments

The authors thank the study participants, study staff, doctors, nurses, other health care staff, and data providers who have contributed to the study. They also thank UCL Genomics, in particular Kerra Pearce, who conducted the Illumina 450k array hybridizations. J.M. Flanagan holds an honorary contract with UCL Cancer Institute, Department of Cancer Biology.

Grant Support

J.M. Flanagan is funded by a Breast Cancer Campaign Fellowship and Cancer Research UK Programme A13086 (awarded to R. Brown and J.M. Flanagan), and M. Garcia-Closas is funded by Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research. This work was funded by the Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research (grant awarded to M. Garcia-Closas), and the BGS (awarded to A. Ashworth and A. Swerdlow). This was also funded by NHS to the NIHR Institute of Cancer Research/Royal Marsden Hospital Biomedical Research Centre.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 8, 2014; revised September 23, 2014; accepted October 14, 2014; published OnlineFirst November 4, 2014.

References

- Esteller M. Epigenetics in cancer. *N Engl J Med* 2008;358:1148–59.
- Flanagan JM, Cocciardi S, Waddell N, Johnstone CN, Marsh A, Henderson S, et al. DNA methylome of familial breast cancer identifies distinct profiles defined by mutation status. *Am J Hum Genet* 2010;86:420–33.
- Fackler MJ, Umbricht CB, Williams D, Argani P, Cruz LA, Merino VF, et al. Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. *Cancer Res* 2011;71:6195–207.
- 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;105:17046–9.
- Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 2010;20:434–9.
- Li S, Hursting SD, Davis BJ, McLachlan JA, Barrett JC. Environmental exposure, DNA methylation, and gene regulation: lessons from diethylstilbestrol-induced cancers. *Ann N Y Acad Sci* 2003;983:161–9.
- Christensen BC, Kelsey KT, Zheng S, Houseman EA, Marsit CJ, Wrensch MR, et al. Breast cancer DNA methylation profiles are associated with tumor size and alcohol and folate intake. *PLoS Genet* 2010;6:e1001043.
- Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet* 2011;88:450–7.
- Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum Mol Genet* 2013;22:843–51.
- Esteller M, Fraga MF, Guo M, Garcia-Foncillas J, Hedenfalk I, Godwin AK, et al. DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis. *Hum Mol Genet* 2001;10:3001–7.
- Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS ONE* 2013;8:e63812.
- 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;18:4046–53.
- Chiachiera F, Piunti A, Pasini D. Epigenetic methylations and their connections with metabolism. *Cell Mol Life Sci* 2013;70:1495–508.
- Bruniquel D, Schwartz RH. Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. *Nat Immunol* 2003;4:235–40.
- Kangaspeska S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, et al. Transient cyclical methylation of promoter DNA. *Nature* 2008;452:112–5.
- Metivier R, Gallas R, Tiffoche C, Le Peron C, Jurkowska RZ, Carmouche RP, et al. Cyclical DNA methylation of a transcriptionally active promoter. *Nature* 2008;452:45–50.
- Barres R, Yan J, Egan B, Treebak JT, Rasmussen M, Fritz T, et al. Acute exercise remodels promoter methylation in human skeletal muscle. *Cell Metab* 2012;15:405–11.
- Moore LE, Pfeiffer RM, Poscablo C, Real FX, Kogevinas M, Silverman D, et al. Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the Spanish Bladder Cancer Study: a case-control study. *Lancet Oncol* 2008;9:359–66.
- Brennan K, Flanagan JM. Is there a link between genome-wide hypomethylation in blood and cancer risk? *Cancer Prevention Res (Philadelphia)* 2012;5:1345–57.
- Flanagan JM, Munoz-Alegre M, Henderson S, Tang T, Sun P, Johnson N, et al. Gene-body hypermethylation of ATM in peripheral blood DNA of bilateral breast cancer patients. *Hum Mol Genet* 2009;18:1332–42.
- Brennan K, Garcia-Closas M, Orr N, Fletcher O, Jones M, Ashworth A, et al. Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. *Cancer Res* 2012;72:2304–13.
- Widschwendter M, Apostolidou S, Raum E, Rothenbacher D, Fiegl H, Menon U, et al. Epigenotyping in peripheral blood cell DNA and breast cancer risk: a proof of principle study. *PLoS ONE* 2008;3:e2656.
- Iwamoto T, Yamamoto N, Taguchi T, Tamaki Y, Noguchi S. BRCA1 promoter methylation in peripheral blood cells is associated with increased risk of breast cancer with BRCA1 promoter methylation. *Breast Cancer Res Treat* 2011;129:69–77.
- Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS ONE* 2009;4:e8274.
- Marsit CJ, Koestler DC, Christensen BC, Karagas MR, Houseman EA, Kelsey KT. DNA methylation array analysis identifies profiles of blood-derived

- DNA methylation associated with bladder cancer. *J Clin Oncol* 2011;29:1133–9.
26. Langevin SM, Koestler DC, Christensen BC, Butler RA, Wiencke JK, Nelson HH, et al. Peripheral blood DNA methylation profiles are indicative of head and neck squamous cell carcinoma: an epigenome-wide association study. *Epigenetics* 2012;7:291–9.
 27. Xu Z, Bolick SC, DeRoo LA, Weinberg CR, Sandler DP, Taylor JA. Epigenome-wide association study of breast cancer using prospectively collected sister study samples. *J Natl Cancer Inst* 2013;105:694–700.
 28. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, et al. Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 2008;299:2877–83.
 29. Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, Aspelund T, et al. Personalized epigenomic signatures that are stable over time and covary with body mass index. *Sci Translational Med* 2010;2:49ra67.
 30. Florath I, Butterbach K, Muller H, Bewerunge-Hudler M, Brenner H. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Hum Mol Genet* 2014;23:1186–201.
 31. Swerdlow AJ, Jones ME, Schoemaker MJ, Hemming J, Thomas D, Williamson J, et al. The breakthrough generations study: design of a long-term UK cohort study to investigate breast cancer aetiology. *Br J Cancer* 2011;105:911–7.
 32. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. *Epigenomics* 2011;3:771–84.
 33. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8:118–27.
 34. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genetics* 2007;3:1724–35.
 35. Leek JT, Storey JD. A general framework for multiple testing dependence. *Proc Natl Acad Sci U S A* 2008;105:18718–23.
 36. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform* 2012;13:86.
 37. Irizarry RA, Ladd-Acosta C, Carvalho B, Wu H, Brandenburg SA, Jeddeloh JA, et al. Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res* 2008;18:780–90.
 38. Croissant Y, Millo G. The plm Package. *J Stat Software* 2008;27.
 39. Joubert BR, Haberg 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;120:1425–31.
 40. Philibert RA, Beach SR, Lei MK, Brody GH. Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking. *Clin Epigenetics* 2013;5:19.
 41. Brennan K, Flanagan JM. Epigenetic epidemiology for cancer risk: harnessing germline epigenetic variation. *Methods Mol Biol* 2012;863:439–65.
 42. Hahn MA, Wu X, Li AX, Hahn T, Pfeifer GP. Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. *PLoS ONE* 2011;6:e18844.
 43. Shenker N, Flanagan JM. Intragenic DNA methylation: implications of this epigenetic mechanism for cancer research. *Br J Cancer* 2012;106:248–53.
 44. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* 2012;8:e1002629.
 45. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* 2005;102:10604–9.
 46. Johansson A, Enroth S, Gyllenstein U. Continuous aging of the human DNA methylome throughout the human lifespan. *PLoS ONE* 2013;8:e67378.
 47. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* 2010;20:440–6.
 48. Teschendorff AE, West J, Beck S. Age-associated epigenetic drift: implications, and a case of epigenetic thrift? *Hum Mol Genet* 2013;22:R7–R15.
 49. Wang SC, Oelze B, Schumacher A. Age-specific epigenetic drift in late-onset Alzheimer's disease. *PLoS ONE* 2008;3:e2698.
 50. West J, Widschwendter M, Teschendorff AE. Distinctive topology of age-associated epigenetic drift in the human interactome. *Proc Natl Acad Sci U S A* 2013;110:14138–43.
 51. Moleris A, Campion J, Milagro FI, Marcos A, Campoy C, Garagorri JM, et al. Differential DNA methylation patterns between high and low responders to a weight loss intervention in overweight or obese adolescents: the EVASYON study. *FASEB J* 2013;27:2504–12.
 52. Relton CL, Groom A, St Pourcain B, Sayers AE, Swan DC, Embleton ND, et al. DNA methylation patterns in cord blood DNA and body size in childhood. *PLoS ONE* 2012;7:e31821.
 53. Stepanow S, Reichwald K, Huse K, Gausmann U, Nebel A, Rosenstiel P, et al. Allele-specific, age-dependent and BMI-associated DNA methylation of human MCHR1. *PLoS ONE* 2011;6:e17711.
 54. Sun YV, Smith AK, Conneely KN, Chang Q, Li W, Lazarus A, et al. Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans. *Hum Genet* 2013;132:1027–37.

Cancer Epidemiology, Biomarkers & Prevention

Temporal Stability and Determinants of White Blood Cell DNA Methylation in the Breakthrough Generations Study

James M. Flanagan, Mark N. Brook, Nick Orr, et al.

Cancer Epidemiol Biomarkers Prev 2015;24:221-229. Published OnlineFirst November 4, 2014.

Updated version Access the most recent version of this article at:
doi:[10.1158/1055-9965.EPI-14-0767](https://doi.org/10.1158/1055-9965.EPI-14-0767)

Supplementary Material Access the most recent supplemental material at:
<http://cebp.aacrjournals.org/content/suppl/2014/11/05/1055-9965.EPI-14-0767.DC1>

Cited articles This article cites 53 articles, 10 of which you can access for free at:
<http://cebp.aacrjournals.org/content/24/1/221.full#ref-list-1>

Citing articles This article has been cited by 5 HighWire-hosted articles. Access the articles at:
<http://cebp.aacrjournals.org/content/24/1/221.full#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://cebp.aacrjournals.org/content/24/1/221>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.