ARTICLE

Coffee consumption is associated with DNA methylation levels of human blood

Yu-Hsuan Chuang¹, Austin Quach², Devin Absher³, Themistocles Assimes⁴, Steve Horvath^{*,2,5,6} and Beate Ritz^{*,1,6}

Beneficial health effects have been attributed to coffee consumption, but it is not yet known whether epigenetics may have a role in this process. Here we associate epigenome-wide DNA methylation levels to habitual coffee consumption from two studies with blood (2100 and 215 participants), and one with saliva samples (256 participants). Adjusting for age, gender, and blood cell composition, one CpG (cg21566642 near *ALPPL2*) surpassed genome-wide significance ($P = 3.7 \times 10^{-10}$) and from among 10 additional CpGs significant at $P \leq 5.0 \times 10^{-6}$, six were located within 1500 bps of a transcriptional start site. Results for these 11 top-ranked CpGs remained significant after further adjusting for smoking. Also, methylation levels of another 135 CpGs were influenced by both coffee drinking and smoking ($P \leq 1.0 \times 10^{-7}$). Functional enrichment analysis suggested that coffee-associated CpGs were located near transcription factor binding ($P = 1.2 \times 10^{-6}$) and protein kinase activity genes ($P = 2.9 \times 10^{-5}$). Interestingly, when we stratified by menopausal hormone therapy (MHT), methylation differences with coffee consumption were observed only in women who never used MHT. We did not replicate any of the associations found in blood in our saliva samples, suggesting that coffee may affect DNA methylation levels in immune cells of the blood but not in saliva. *European Journal of Human Genetics* (2017) 25, 608–616; doi:10.1038/ejhg.2016.175; published online 15 February 2017

INTRODUCTION

Coffee is one of the most widely consumed beverages in the world and is believed to have potential health risks and benefits.¹ Coffee consumption has been linked to a wide range of health outcomes including cardiovascular, metabolic, and neurocognitive function. Heavy coffee consumption induces cardiovascular responses and insomnia,¹ but coffee consumption has also been associated with lower risk of type 2 diabetes, endometrial cancer,² and neurodegenerative diseases such as Parkinson's disease (PD) and Alzheimer's disease (AD).³ Caffeine is thought to prevent cognitive decline by inhibiting formation of beta-amyloid and by acting as an anti-inflammatory agent in AD,^{4,5} whereas in PD, it is thought to reduce neuroinflammation and lipid-mediated oxidative stress.^{6,7} AD and PD are slowly progressive diseases with a long prodromal phase, making it difficult to rule out reverse causality such that at risk individuals may decrease coffee intake due to development of sleep problems or loss of smell.⁸

Genomic studies identified eight genetic loci that have an influence on habitual coffee consumption, including some near *CYP1A2* and *AHR*, encoding the caffeine metabolizing enzyme Cytochrome P450 1A2 and a *CYP1A2* regulator Aryl Hydrocarbon Receptor, respectively.^{9,10} DNA methylation (DNAm) might act as a potential epigenetic mediator for caffeine's influence on health.¹¹ Mechanistic epigenetic studies of caffeine have mainly focused on animal models.^{12–14} For example, maternal prenatal caffeine intake increased methylation of the steroidogenic factor 1 promoter in fetal adrenal tissue in mice,¹² whereas caffeine elicited effects similar to acute

exercise in rat skeletal muscle tissue and resulted in lower DNAm levels in promoter regions of energy metabolism genes.¹⁵ Little is known whether epigenetic changes can be found in human due to their coffee consumption habits. Exploring whether coffee consumption affects DNAm can help identify epigenetic signatures and provide mechanistic insights for results from past epidemiological studies and possibly new insights into health risks or benefits of coffee consumption.

Here, for the first time we identified DNAm sites from a genome-wide screen that relate to habitual coffee consumption in humans. We conducted a meta-analysis of DNAm levels in blood samples from two different data sets: PD-free control subjects enrolled in the Parkinson's Environment and Genes (PEG first round, 2001–2007)^{16,17} study consisting of 215 non-Hispanic Caucasians, and women from the WHI consisting of 995 Caucasians, 431 Hispanics, and 674 African Americans. We also related coffee consumption to DNAm levels in saliva samples from 127 PD patients and 129 PD-free controls (age-, gender-, and ethnicity-matched) enrolled in the second round of the PEG study (2009-ongoing).^{18,19} Detailed information for each data set can be found in Supplementary Table S1 and in Methods section.

METHODS

Description of PEG1 subjects

Study population. This data set consists of 215 Caucasian population controls with complete information on coffee consumption and blood samples for DNA. The PEG1 study is a population-based case-control study in central California (Fresno, Kern, or Tulare Counties) recruiting subjects from 2001 to 2007. To be eligible, participants had to be residents of one of three central

¹Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, CA, USA; ²Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; ³HudsonAlpha Institute of Biotechnology, Huntsville, AL, USA; ⁴School of Medicine (Cardiovascular Medicine), Stanford University, Palo Alto, CA, USA; ⁵Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, CA, USA

⁶Joint senior authors.

*Correspondence: Dr S Horvath, Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095-7088, USA.

Tel: +1 310 825 9299; Fax: +1 310 267 2113; E-mail: shorvath@mednet.ucla.edu

or Dr B Ritz, Department of Epidemiology, Fielding School of Public Health, University of California, 650 Charles E. Young Drive, BOX 951772, Los Angeles, CA 90095-1772, USA.

Tel: +1 310 825 8579; Fax: +1 310 206 6039; E-mail: britz@ucla.edu

Received 20 July 2016; revised 30 September 2016; accepted 15 November 2016; published online 15 February 2017

Table 1 The top-ranked CpG sites associated with coffee consumption in blood with/without smoking adjustment

CpG	Gene	Chr.	Position (bp) ^a	Relation to		UCSC CpG island	UCSC RefGene group ^b	meta.			PEG1 Caucasian			WHI Caucasian-only, adjusted for age, cell counts (subset2)			WHI Hispanic-only, adjusted for age, cell counts (subset3)			WHI African American-only, adjusted for age, cell counts (subset4)		
				UCSC CpG island	UCSC RefGene group ^b			meta.Z	P-value	biCor	P-value	biCor	P-value	biCor	P-value	biCor	P-value	biCor	P-value	biCor	P-value	
(a) Adjusting for chronological age, gender, and blood cell counts																						
1	cg21566642	Intergenic, near ALPPL2	2	233284661	Island	TSS1500	-6.26	3.73E-10	-0.29	3.60E-05	-0.14	1.34E-05	-0.07	1.28E-01	-0.10	1.33E-02						
2	cg20333292	GPR132	14	105532012	Island	TSS1500	5.05	4.33E-07	0.28	7.43E-05	0.05	8.67E-02	0.13	6.87E-03	0.07	9.10E-02						
3	cg21163128	BSCL2	11	62477362	Island	TSS1500	5.00	5.70E-07	0.28	6.77E-05	0.10	2.40E-03	0.09	6.23E-02	0.04	2.80E-01						
4	cg23303782	GRK5	10	120967744	S_Shore	Body	4.81	1.49E-06	0.29	3.55E-05	0.06	4.37E-02	0.09	6.39E-02	0.06	1.14E-01						
5	cg26105150	FSTL5	4	163085403		TSS1500	4.78	1.77E-06	0.27	1.35E-04	0.05	8.98E-02	0.16	6.78E-04	0.02	5.45E-01						
6	cg19723563	PTHLH	12	28123034	Island	TSS200	4.74	2.13E-06	0.17	2.14E-02	0.07	2.20E-02	0.13	5.22E-03	0.08	3.77E-02						
7	cg17928869	Intergenic, between EIF1 and KRT42P	17	39822542	S_Shore		4.68	2.86E-06	0.23	1.35E-03	0.08	1.17E-02	0.11	2.11E-02	0.05	1.93E-01						
8	cg12866551	MALRD1	10	20019641			4.66	3.12E-06	0.23	1.48E-03	0.07	2.04E-02	0.14	2.75E-03	0.03	4.21E-01						
9	cg15722372	PSMD8	19	38865020	N_Shore	TSS200	4.65	3.26E-06	0.20	4.55E-03	0.05	9.92E-02	0.10	4.67E-02	0.11	4.94E-03						
10	cg19974428	TMEM130	7	98468047	Island	TSS1500	4.65	3.34E-06	0.22	2.29E-03	0.07	3.67E-02	0.11	2.64E-02	0.07	5.55E-02						
11	cg15140902	FLJ22536	6	21667815	S_Shore	Body	4.60	4.14E-06	0.12	1.09E-01	0.10	1.52E-03	0.10	4.12E-02	0.09	1.72E-02						
(b) Adjusting for chronological age, gender, blood cell counts, and smoking																						
1	cg21163128	BSCL2	11	62477362	Island	TSS1500	4.95	7.36E-07	0.30	1.87E-05	0.09	7.19E-03	0.09	6.96E-02	0.04	2.82E-01						
2	cg08119527	Intergenic, near PODXL	7	131340667			4.84	1.28E-06	0.30	2.81E-05	0.09	4.68E-03	0.09	6.27E-02	0.03	4.44E-01						
3	cg26331135	CNTN4	3	2144068	S_Shelf	5'UTR	4.71	2.46E-06	0.31	1.31E-05	0.04	1.92E-01	0.13	7.96E-03	0.04	2.91E-01						
4	cg20333292	GPR132	14	105532012		TSS1500	4.62	3.85E-06	0.24	7.22E-04	0.06	8.21E-02	0.12	1.36E-02	0.06	1.04E-01						
5	cg08311403	ROBO3	11	124735215	Island	TSS200	4.57	4.98E-06	0.16	3.04E-02	0.07	3.16E-02	0.06	2.48E-01	0.14	2.67E-04						

Abbreviations: ALPPL2, alkaline phosphatase, placental-like 2; BSCL2, Berardinelli-Seip congenital lipodystrophy 2 (seipin); CNTN4, Contactin 4; EIF1, eukaryotic translation initiation factor 1; FLJ22536, miscRNA; FSTL5, follistatin-like 5; GPR132, G Protein-Coupled Receptor 132; GRK5, G protein-coupled receptor kinase 5; KRT42P, keratin 42 pseudogene; MALRD1, MAM and LDL receptor class A domain containing 1; PODXL, podocalyxin-like; PSMD8, proteasome (prosome, macropain) 26 S subunit, non-ATPase; 8, PTHLH, parathyroid hormone-like hormone; ROB03, roundabout guidance receptor 3; TMEI130, transmembrane protein 130; TSS, transcription start site; UTR, untranslated region.
List of CpGs associated with coffee consumption, (in/near) gene, chromosome, and CpG island location, gene region, (Stouffer's test) Z-value and P-value from meta-analysis, a robust correlation coefficient (known as biweight midcorrelation) and P-value for daily coffee consumption (in last 3 months in WHI or 12 months in PEG1) and DNA methylation levels within four population subsets.
^aLocation is based on NCBI genome build 37.
^bTSS500: within 1500 bps of a TSS, TSS200: within 200 bps of a TSS.

California counties, had to have lived in California for at least 5 years, and to be at least 35 years of age.¹⁶ Population controls were identified from Medicare lists and also using residential property tax assessor records. Potential controls were screened for eligibility by mail or telephone, and only one person per household was allowed to enroll.^{18,19} The study was approved by the UCLA Institutional Review Board, and informed consent was obtained from all subjects.

Exposure assessment. Standardized interviews were conducted to obtain information on demographics, lifetime caffeinated beverage consumption, smoking, and menopausal hormone therapy (MHT) histories. In the interview, information on the frequency and amount of caffeinated beverage consumption at different periods of lifetime were collected: young adult <25 years, adult 25–44 years, middle-aged 45–64 years, and senior ≥65 years. We used this information to calculate weighted average daily coffee consumption. Only caffeinated coffee consumed during the past 12 months prior to the date of blood draw contributed to our exposure measures.

Description of PEG2 subjects

Study population. This data set consists of data from 127 PD patients and 129 population controls with complete information on coffee consumption and saliva samples. We extracted DNA from recruited for the PEG2 study which started in 2009 (ongoing). PD patients were identified using the California PD Registry for the three target counties in central California. Those who lived in the study area were eligible and were mailed invitations, and those who agreed

were examined by a UCLA movement disorder specialist who applied UK Brain Bank and Gelb diagnostic criteria.^{20,21} Controls selection was based on the same criteria as in the PEG1 study, but only used tax assessor records to identify residents whom we recruited at the door step. Saliva samples selected from PEG2 participants were matched on age, gender, and race for cases and controls.

Exposure assessment. Methods for assessing exposure were identical to the methods employed in the PEG1 study. Phenotype data and DNAm data of the PEG studies are available at GEO accession database GSE72775 (blood) and GSE78874 (saliva).

WHI subjects description

Study population. This dataset consists of a subgroup of 2100 women (995 Caucasians, 431 Hispanics, and 674 African Americans) with complete information on coffee consumption as well as genome-wide DNAm data from blood drawn at baseline. The WHI is a multi-center study launched in 1993, which enrolled postmenopausal women aged 50–79 years into either one or more randomized clinical trials or an observational study.²² These women were originally selected from two WHI subcohorts for a nested genomic case-control study of coronary heart disease (CHD) with genome-wide genotype and cardiovascular disease-related biomarker data.²³ Thus, 50% ($n = 1053$) of these WHI women were eventually diagnosed with CHD; however, disease status has no effect on DNAm level measured at baseline. The two cohorts are: (1) the WHI SNP Health Association Resource (SHARe) cohort, which includes

Figure 1 Blood DNA methylation levels associated with coffee consumption adjusted for age, gender, and blood cell counts. **(a)** Manhattan plot of the meta-analysis methylation association P -values adjusted for chronological age, gender, and blood cell counts. The line indicates P -value threshold of 10^{-7} . One CpG on chromosome 2 passed this threshold. The y axis corresponds to negative $\log_{10}$ transformed meta- P -value. The x axis refers to chromosome number, and X and Y chromosomes. **(b)** Distributions of CpGs relative to CpG island and gene regions for all 450 k CpGs on the microarray and the 11 most significant coffee-associated CpGs listed in Table 1a. P -values were obtained by Fisher's test for comparing proportions. A full color version of this figure is available at the *European Journal of Human Genetics* journal online.

genotyping data from ~8500 African American and ~3500 Hispanic women through WHI core study M5-SHARE (www.whi.org/researchers/data/WHIStudies/StudySites/M5) as well as information on biomarker through WHI Core study W54-SHARE (.../data/WHIStudies/StudySites/W54); (2) the two European Americans Hormonal Therapy (EA HT) trials selected for GWAS and biomarkers in core studies W58 (.../data/WHIStudies/StudySites/W58) and W63 (.../data/WHIStudies/StudySites/W63).

Exposure assessment. Information on demographics, smoking history, and MHT was obtained using a structured questionnaire at baseline. Food frequency questionnaires were used to collect information on daily coffee or tea (all types) consumption in the past 3 months prior to baseline. Our exposure measures were directly taken from answers provided in response to the questionnaire.

DNA extraction and genome-wide DNA methylation analysis

DNAm data were obtained from the Infinium HumanMethylation450 Bead-Chip (Illumina, San Diego, CA, USA) using DNA samples extracted from peripheral blood cells and leukocytes in saliva. Methylation β values ranging from 0 (unmethylated) to 1 (fully methylated) were used for analysis.²⁴

Statistical analyses

The raw methylation data were preprocessed using the background normalization method from the Genome Studio software (Illumina, San Diego, CA, USA). To assess correlations between continuous coffee consumption (cup/day) and site-specific DNAm levels, biweight midcorrelation (bicor) was applied in a genome-wide screen. In the main correlation analysis using DNAm levels from blood, potential confounders such as age at blood draw, gender, and blood cell counts were adjusted for by regressing out the effects of these factors and retaining the residuals. Smoking status (ever vs never) was further adjusted for

in ancillary analyses. We used the Houseman algorithm in the *minfi* R package and epigenetic clock software for estimating blood cell counts.^{25–27} All blood analyses were stratified by ethnicity, thus four subsets were generated: PEG1 Caucasians PD-free controls, WHI Caucasians, WHI Hispanics, and WHI African Americans. In order to obtain an overall *P*-value across the four subsets, we conducted a meta-analysis using Stouffer's method for combining *Z*-values (meta-*Z*), that is, $\sum z_i/\sqrt{4}$. The corresponding two-sided *P*-values (meta-*P*-value) were calculated under the assumption of a normal distribution. These approaches were also applied to identify smoking-associated CpGs and CpGs influenced by both coffee and smoking. We then identified the top-ranked coffee-associated CpGs by meta-*P*-value, and applied functional enrichment analysis on 2124 genes identified from the top 3000 most significant coffee-associated CpGs (meta-*P*-value threshold $\sim 1.1 \times 10^{-3}$) using the online bioinformatics tool – the Database for Annotation, Visualization and Integrated Discovery (DAVID v.6.7, NIAID/NIH, Bethesda, MD, USA). We further conducted MHT-stratified meta-analysis for the top 11 coffee-associated CpGs using the WHI data in order to investigate the modifying effect MHT has on the coffee–DNAm association. In the analysis using DNAm levels from saliva in PEG2, potential confounders such as age at saliva collection, gender, and ethnicity were adjusted for as above. Analyses and scatter plots were created using the WGCNA package in R v.3.1.2 (R Development Core Team 2016, Vienna, Austria), whereas Manhattan plots of epigenome-wide association study (EWAS) *P*-values were generated with the *qqman* package. QQ-plots of EWAS *P*-values were also generated in R, and lambda, that is, median(X^2)/0.454, were calculated to identify potential inflation.

RESULTS AND DISCUSSION

Coffee consumption and DNA methylation levels in blood

In our EWAS study, we analyzed methylation levels of ~486 k CpGs on the Illumina 450 K array. Since many CpGs exhibit strong

Table 2 Smoking adjusted results and saliva results for the 11 top-ranked CpG sites in Table 1a

						Blood samples		Saliva samples						
						PEG1 and WHI adjusted for age, gender, cell counts, smoking (N = 2297)		PEG2 adjusted for age, gender, and race (N = 256)			PEG2 adjusted for age, gender, race, and PD status (N = 256)			
Relation to UCSC CpG						UCSC RefGene		meta.						
CpG	Gene	Chr.	Position (bp) ^a	island	group ^b	biCor	P-value	biCor	Z-value	P-value	biCor	Z-value	P-value	
1	cg21566642	Intergenic, near ALPPL2	2	233284661	Island		−3.46	5.37E ^{−04}	−0.10	−1.65	9.87E ^{−02}	−0.09	−1.37	1.71E ^{−01}
2	cg20333292	GPR132	14	105532012		TSS1500	4.62	3.85E ^{−06}	0.06	0.90	3.71E ^{−01}	0.03	0.46	6.46E ^{−01}
3	cg21163128	BSCL2	11	62477362	Island	TSS1500	4.95	7.36E ^{−07}	−0.01	−0.11	9.14E ^{−01}	−0.03	−0.47	6.42E ^{−01}
4	cg23303782	GRK5	10	120967744	S_Shore	Body	4.27	1.92E ^{−05}	0.01	0.14	8.89E ^{−01}	0.04	0.58	5.62E ^{−01}
5	cg26105150	FSTL5	4	163085403		TSS1500	4.34	1.42E ^{−05}	0.02	0.33	7.44E ^{−01}	−0.01	−0.21	8.32E ^{−01}
6	cg19723563	PTHLH	12	28123034	Island	TSS200	4.22	2.43E ^{−05}	0.03	0.42	6.76E ^{−01}	0.00	−0.02	9.80E ^{−01}
7	cg17928869	Intergenic, between EIF1 and KRT42P	17	39822542	S_Shore		4.06	4.94E ^{−05}	0.04	0.60	5.48E ^{−01}	0.02	0.29	7.73E ^{−01}
8	cg12866551	MALRD1	10	20019641			4.47	7.78E ^{−06}	−0.06	−0.94	3.48E ^{−01}	−0.06	−0.90	3.68E ^{−01}
9	cg15722372	PSMD8	19	38865020	N_Shore	TSS200	4.44	9.01E ^{−06}	−0.01	−0.10	9.17E ^{−01}	−0.02	−0.36	7.17E ^{−01}
10	cg19974428	TMEM130	7	98468047	Island	TSS1500	4.04	5.42E ^{−05}	−0.03	−0.49	6.23E ^{−01}	−0.05	−0.86	3.92E ^{−01}
11	cg15140902	FLJ22536	6	21667815	S_Shore	Body	4.04	5.39E ^{−05}	0.00	0.05	9.61E ^{−01}	−0.02	−0.25	8.03E ^{−01}

Abbreviations: ALPPL2, alkaline phosphatase, placental-like 2; BSCL2, Berardinelli-Seip congenital lipodystrophy 2 (seipin); EIF1, eukaryotic translation initiation factor 1; FLJ22536, miscRNA; FSTL5, follistatin-like 5; GPR132, G protein-coupled receptor 132; GRK5, G protein-coupled receptor kinase 5; KRT42P, keratin 42 pseudogene; MALRD1, MAM and LDL receptor class A domain containing 1; PD, Parkinson's disease; PSMD8, proteasome (prosome, macropain) 26 S Subunit, non-ATPase, 8; PTH1H, parathyroid hormone-like hormone; TMEM130, transmembrane protein 130; TSS, transcription start site.

List of CpGs associated with coffee consumption, (in/near) gene, chromosome, and CpG island location, gene region, (Stouffer's test) Z-value and P-value from meta-analysis for the PEG1 and WHI studies, a robust correlation coefficient (known as biweight midcorrelation) Z-value and P-value for daily coffee consumption (in last 3 months in WHI or 12 months in PEG) and DNA methylation levels.

^aLocation is based on NCBI genome build 37.

^bTSS500: within 1500 bps of a TSS, TSS200: within 200 bps of a TSS.

pairwise correlations, the Bonferroni-corrected significance threshold of $\alpha = 0.05/500\,000 = 1 \times 10^{-7}$ was considered overly conservative; we used a modified threshold of $P < 5 \times 10^{-6}$ to evaluate genome-wide significance in our study. In the PEG1 and WHI data sets, adjusting for chronological age, gender, and imputed blood cell counts, we identified one CpG with genome-wide Bonferroni-corrected significance: cg21566642 near the *ALPL2* gene (meta- $P = 3.7 \times 10^{-10}$). Ten additional CpGs surpassed the significance threshold of $P < 5.0 \times 10^{-6}$ (Table 1a and Figure 1a) and are located in/near the genes *GPR132*, *BCL2*, *MALRD1*, *GRK5*, *PSMD8*, *FSTL5*, *PTHLH*, and so on. (Table 1a).

The top-ranked CpGs appear to be linked to genes involved in lipid metabolism^{28–31} and immune response (RefSeq, July 2008). For instance, the protein encoded by *GPR132* is a receptor for oxidized free fatty acids and is a treatment target for diabetes because of its role in lipid metabolism and antioxidant activity.²⁸ In a mixed race study of atherosclerosis, *GPR132* was found to be hypomethylated among low socioeconomic status (SES) individuals with increased inflammatory activity compared with high SES individuals.²⁹ *BCL2* encodes the transmembrane protein 'seipin' residing in the endoplasmic reticulum. Variants in *BCL2* cause congenital generalized lipodystrophy, characterized by the loss of adipose tissue and severe insulin resistance.³⁰ *MALRD1* encodes yet another lipid-related gene that has been shown

to regulate bile acid and lipid levels in the enterohepatic system.³¹ Genes related to immune response include *GRK5* and *PSMD8*. The protein encoded by *GRK5* regulates polymorphonuclear leukocyte motility, whereas *PSMD8* encodes an immune-proteasome component related to major histocompatibility (MHC) class I antigen processing and presentation (provided by RefSeq, July 2008). These coffee-associated CpGs were mostly located within 200–1500 bps upstream of a transcription start site of a gene, that is, promoter region (Fisher's $P = 0.03$, Figure 1b). Additional stratified analyses for the PEG and WHI samples, as well as comparisons of effect sizes between genders or ethnicities are provided in Supplementary Tables S2, S3 and Table 1.

Many studies have focused on associations between DNAm and smoking, and it is well known that a subgroup of coffee consumers is more likely to smoke. The highest correlation we observed between coffee intake and smoking in any of our cohorts was $r = 0.31$ among PEG1 Caucasian controls ($P = 3.4 \times 10^{-6}$). Due to the common co-exposure to coffee and smoking, any adjustment for smoking is expected to affect associations between DNAm and coffee consumption. Indeed, smoking adjustment reduced the statistical significance of cg21566642 (meta- $P = 5.4 \times 10^{-4}$, Table 2 and Figure 2a) to less than the genome-wide threshold. However, associations between the 11

Figure 2 Blood DNA methylation levels associated with coffee consumption adjusted for age, gender, blood cell counts, and smoking. (a) Manhattan plot of the meta-analysis methylation association P -values adjusted for chronological age, gender, blood cell counts, and smoking. The line indicates P -value threshold of 10^{-7} —no CpG passed this threshold. Y axis corresponds to negative $\log_{10}$ transformed meta- P -value. X axis refers to chromosome number, X and Y chromosomes. (b) Distributions of CpGs relative to CpG island and gene regions for all 450 k CpGs on the microarray and the five most significant coffee-associated CpGs listed in Table 1b. P -values were obtained by Fisher's test for comparing proportions. A full color version of this figure is available at the *European Journal of Human Genetics* journal online.

top-ranked CpGs and coffee consumption were still preserved after smoking adjustment (meta. $P < 0.05/11 = 4.5 \times 10^{-3}$, Table 2). Further, we identified methylation differences for 135 CpGs associated with both coffee drinking and smoking (meta. $P \leq 1.0 \times 10^{-7}$, Supplementary Table S4). After smoking adjustment, the most significant differentially methylated genes were *BSCL2* and *GPR132*, along with *CNTN4* and *ROBO3* that appear to be involved in axonal navigation. (meta. $P \leq 5 \times 10^{-6}$, Table 1b).

It is worth noting that genetic variants previously linked to coffee consumption in GWAS did not reach the significance threshold of $P \leq 1.0 \times 10^{-3}$ in our study, specifically *AHR*, *CYP1A1*, *CYP1A2*, *NRCAM*, and *ADORA2A*.^{9,10,32–34} However, our study corroborated the importance of the *STK11* gene (cg24145685: meta. $Z = 4.54$, meta. $P = 5.7 \times 10^{-6}$; after smoking adjustment: meta. $Z = 3.99$ and meta. $P = 6.6 \times 10^{-5}$), which encodes a member of the serine/threonine kinase family and interacts with another gene (*CAB39L*) identified in a previous GWAS focused on coffee consumption.³²

As mentioned above, previous studies reported reduced risks of developing PD and AD with habitual coffee consumption.^{35,36} To the best of our knowledge, our blood tissue data did not include AD or PD patients. Surprisingly, we found some CpGs located near genes linked with familial forms of PD associated with coffee consumption: *GBA* (meta. $P = 7.9 \times 10^{-5}$), *PARK2/Parkin* (meta. $P = 7.3 \times 10^{-4}$), and *PINK1* (meta. $P = 8.9 \times 10^{-4}$). Similarly, some GWAS-identified loci for AD were also associated with coffee intake: *PICALM* (meta. $P = 1.3 \times 10^{-5}$), *CLU* (meta. $P = 6 \times 10^{-4}$), and *EDC3* (meta. $P = 1.1 \times 10^{-4}$).^{37,38} The *PARK2* gene encodes an ubiquitin protein ligase called Parkin that targets proteins for degradation in the proteasome. Pathways related to Parkin include oxidative stress, Class I MHC antigen processing and presentation, and alpha-synuclein signaling.³⁹ The *EDC3* gene is also of interest as it is located near *CYP1A1/CYP1A2* and the enzymes encoded by them may interact with coffee consumption in reducing AD risk.³⁸

Table 3 Functional enrichment analysis for top 3000 most significant coffee-associated CpG sites in 2124 genes (meta. P -value cutoff $\sim 1.1 \times 10^{-3}$)

Rank	Category	Term	P-value	Bonferroni	Benjamini	FDR	Overlap genes (n)	Fold enrichment
1	GOTERM_MF_FAT	GO:0008134 ~ transcription factor binding	1.24E-06	1.51E-03	1.51E-03	2.01E-03	92	1.65
2	GOTERM_MF_FAT	GO:0003713 ~ transcription coactivator activity	2.06E-05	2.48E-02	3.59E-03	3.34E-02	45	1.94
3	GOTERM_MF_FAT	GO:0003712 ~ transcription cofactor activity	3.06E-05	3.66E-02	4.13E-03	4.95E-02	66	1.67
4	GOTERM_MF_FAT	GO:0004672 ~ protein kinase activity	2.88E-05	3.45E-02	4.37E-03	4.66E-02	99	1.50
5	GOTERM_MF_FAT	GO:0004674 ~ protein serine/threonine kinase activity	6.33E-05	7.42E-02	6.99E-03	1.02E-01	74	1.58

List of functional categories that coffee methylation-related genes are enriched in P -values, Bonferroni-corrected or Benjamini-adjusted P -value, FDR, number of overlapping genes, and fold enrichment.

Table 4 Stratified analysis of the 11 coffee-associated CpGs found in blood (in Table 1a) by MHT using WHI data only, adjusted for chronological age and blood cell counts

						WHI all ethnicities, adjusted for age and cell counts (N = 2100)		WHI all ethnicities MHT user, adjusted for age and cell counts (N = 1031) ^c		WHI all ethnicities MHT non-user, adjusted for age and cell counts (N = 1012) ^c	
				Relation to UCSC CpG UCSC RefGene							
CpG	Gene	Chr.	Position (bp) ^a	island	group ^b	meta. Z	meta. P-value	meta. Z	meta. P-value	meta. Z	meta. P-value
1	cg21566642	Intergenic, near ALPL2	2	233284661	Island	-3.47	5.23E-04	-2.15	3.19E-02	-4.48	7.43E-06
2	cg20333292	GPR132	14	105532012		4.49	7.23E-06	2.69	7.24E-03	4.09	4.38E-05
3	cg21163128	BSCL2	11	62477362	Island	4.34	1.41E-05	2.97	3.02E-03	3.51	4.44E-04
4	cg23303782	GRK5	10	120967744	S_Shore	3.90	9.53E-05	2.22	2.65E-02	3.20	1.37E-03
5	cg26105150	FSTL5	4	163085403		4.43	9.57E-06	2.78	5.47E-03	3.24	1.18E-03
6	cg19723563	PTHLH	12	28123034	Island	4.03	5.48E-05	2.27	2.34E-02	3.35	8.22E-04
7	cg17928869	Intergenic, between EIF1 and KRT42P	17	39822542	S_Shore	3.82	1.33E-04	1.76	7.88E-02	3.81	1.37E-04
8	cg12866551	MALRD1	10	20019641		4.47	7.97E-06	2.84	4.47E-03	2.72	6.58E-03
9	cg15722372	PSMD8	19	38865020	N_Shore	2.72	6.47E-03	1.60	1.11E-01	2.58	9.87E-03
10	cg19974428	TMEM130	7	98468047	Island	3.55	3.80E-04	2.01	4.41E-02	3.06	2.18E-03
11	cg15140902	FLJ22536	6	21667815	S_Shore	4.39	1.15E-05	2.46	1.40E-02	4.61	3.99E-06

Abbreviations: ALPL2, alkaline phosphatase, placental-like 2; BSCL2, Berardinelli-Seip congenital lipodystrophy 2 (seipin); EIF1, eukaryotic translation initiation factor 1; FLJ22536, miscRNA; FSTL5, follistatin-like 5; GPR132, G protein-coupled receptor 132; GRK5, G protein-coupled receptor kinase 5; KRT42P, keratin 42 pseudogene; MALRD1, MAM and LDL receptor class A domain containing 1; MHT, menopausal hormone therapy; PSMD8, proteasome (prosome, macropain) 26 S subunit, non-ATPase; PTHLH, parathyroid hormone-like hormone; TMEM130, transmembrane protein 130; TSS, transcription start site.

List of the 11 coffee-associated CpGs from Table 1a, (in/near) gene, chromosome, and CpG island location, (Stouffer's test) Z-value and P-value from meta-analysis using WHI data, which consist of three subsets: Caucasians, Hispanics, and African Americans, stratified by MHT.

^aLocation is based on NCBI genome build 37.

^bTSS500: within 1500 bps of a TSS, TSS200: within 200 bps of a TSS.

^cMissing MHT information: WHI Caucasian (N=35), Hispanic (N=11), African American (N=11).

Using the 2124 genes linked with the 3000 most significant coffee-associated CpGs ($P \leq 1.1 \times 10^{-3}$ without smoking adjustment) in gene set enrichment analysis, we identified 2901 CpGs (in/near 2058 genes) that were hypermethylated in habitual coffee drinkers, whereas only 3% (99 CpGs in/near 66 genes) were hypomethylated. Results of the DAVID functional analysis showed that these coffee-associated genes are enriched in functional categories of transcription factor binding ($P = 1.2 \times 10^{-6}$, Table 3) and protein kinase activity ($P = 2.9 \times 10^{-5}$). The enriched biological terms remained statistically significant after correcting for multiple comparisons (Benjamini-adjusted $P < 0.05$).

It has previously been suggested that the potential protective action of coffee on PD in women may be abrogated by postmenopausal estrogen use.^{40,41} Interestingly, when we stratified the participants from the WHI by MHT, we observed significant associations between coffee consumption and the 11 top CpG sites from Table 1a only in women who never used MHT and not in MHT users (Table 4).

Coffee consumption and DNA methylation levels in saliva

We also evaluated associations between coffee and genome-wide DNAm levels in saliva provided by 256 participants with and without PD enrolled in the PEG2 study. After adjustment for chronologic age,

gender, and ethnicity, no CpGs achieved genome-wide significance ($P \leq 10^{-7}$, Figure 3a). When examining the 11 most significant coffee-associated CpGs we previously identified in blood, none of the significant associations were preserved in saliva (Table 2). Moreover, we did not observe positive correlations between meta Z-values for blood and Z-values for saliva (Figure 3b). Further adjustment for PD status did not change these results (Table 2), suggesting that PD status did not affect DNAm levels in saliva for the coffee-related CpGs identified in blood. After adjusting for smoking, there appeared to be a significant correlation for coffee-related DNAm in blood and in saliva tissues, but in the opposite direction of what we expected (Figure 4b and Supplementary Table S5). This can be explained by the 'regression to the mean' effect, suggesting DNAm data from saliva did not replicate associations we found in blood.

Potential limitations

Our study has some potential limitations: first, we have a small amount of uncertainty regarding the reported ethnicity since we do not have genome-wide SNP data to compute the genetic principal components. However, ethnicity information in the studies we included have been carefully verified using 37 Ancestry Informative

Figure 3 Saliva DNA methylation levels associated with coffee consumption adjusted for age, gender, and ethnicity. (a) Manhattan plot of the methylation association P -values adjusted for chronological age, gender, and ethnicity. The line indicates the P -value threshold of 10^{-7} ; no CpGs passed this threshold. The y axis corresponds to negative $\log_{10}$ transformed meta- P -value. The x axis refers to chromosome number, and X and Y chromosomes. (b) Correlation between Z-values from biweight midcorrelations between DNA methylation and coffee consumption in blood and saliva for 50 most hypermethylated CpGs and 50 most hypomethylated CpGs in blood. The x axis corresponds to the meta Z-values adjusted for age, gender, and blood cell counts from PEG1 and WHI. The y axis corresponds to the Z-values adjusted for age, gender, and ethnicity from PEG2. A full color version of this figure is available at the *European Journal of Human Genetics* journal online.

Figure 4 Saliva DNA methylation levels associated with coffee consumption adjusted for age, gender, ethnicity, and smoking. **(a)** Manhattan plot of the methylation association P -values adjusted for chronological age, gender, ethnicity, and smoking. The line indicates P -value threshold of 10^{-7} -no CpG passed this threshold. Y axis corresponds to negative $\log_{10}$ transformed meta- P -value. X axis refers to chromosome number, X and Y chromosomes. **(b)** Correlation between Z-values from bi-weighted midcorrelations between DNA methylation and coffee consumption in blood and saliva for 50 most hypermethylated CpGs and 50 most hypomethylated CpGs in blood. X axis corresponds to the meta Z-values adjusted for age, gender, blood cell counts, and smoking from PEG1 and WHI. Y axis corresponds to the Z-values adjusted for age, gender, ethnicity, and smoking from PEG2. A full color version of this figure is available at the *European Journal of Human Genetics* journal online.

Markers. Moreover, we stratified by ethnicity to address the possibility of ethnic confounding. Second, stratification by gender in the PEG1 study removed the significance of some of the associations between coffee consumption and methylation loci. However, this might be due to the decreased sample size imparted by stratification. Third, similar to other DNAm studies, lambda values in this study were inflated; therefore, we should interpret P -values with caution. However, our findings in Caucasians were replicated in other ethnic groups giving some validation to the results. In addition, it remains debatable whether to present QQ-plots in DNAm studies, as CpGs are highly correlated and the distributional assumptions made in GWAS may not be met in EWAS (Supplementary Figure S1).

CONCLUSIONS

In summary, in peripheral blood mononuclear cells we identified CpGs located near 11 genes that were associated with habitual coffee consumption based on the significance threshold (meta. $P \leq 5.0 \times 10^{-6}$) while adjusting for age, gender, and blood cell composition. Moreover, these correlations remained significant after further adjustment for smoking. Furthermore, many differentially

methyated CpGs are located in/near genes reported to be associated with coffee-related chronic diseases or the common neurodegenerative diseases PD and AD for which coffee consumption has been suggested to be protective. Our results point to possible mechanisms through which coffee consumption may have beneficial effects and possibly may confer risk reduction. The measures of habitual coffee consumption we used in this study were based on recall over a short period (last 3 months in WHI or 12 months in PEG1); however, in PEG1 reported lifetime coffee consumption was highly correlated with coffee consumption reported for the past 12 months ($\text{cor} = 0.87$), suggesting that coffee consumption is consistent across time. Moreover, this is a mixed race and mixed gender study, therefore the coffee associations with DNAm levels in blood appear to extend to both genders and different ethnic groups, even though in women, results also seemed to depend on MHT use. Finally, our study suggests that while coffee affects DNAm levels in blood this does not seem to extend to saliva-derived tissue.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

Y-HC was funded by the Burroughs Wellcome Fund Inter-school Training Program in Chronic Diseases. Dataset1 (the PEG study) was supported by NIEHS; ES98-05-030-03A and PHS-ES012078 (BR). Dataset2 (the WHI study) was supported by NIH/National Heart, Lung, and Blood Institute (NHLBI) 60442456 BAA23 (TA, DA, SH). The WHI program is funded by the NHLBI, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. We thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: <http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Long%20List.pdf>.

- 1 Butt MS, Sultan MT: Coffee and its consumption: benefits and risks. *Crit Rev Food Sci Nutr* 2011; **51**: 363–373.
- 2 Giri A, Sturgeon SR, Luisi N, Bertone-Johnson E, Balasubramanian R, Reeves KW: Caffeinated coffee, decaffeinated coffee and endometrial cancer risk: a prospective cohort study among US postmenopausal women. *Nutrients* 2011; **3**: 937–950.
- 3 Qi H, Li S: Dose-response meta-analysis on coffee, tea and caffeine consumption with risk of Parkinson's disease. *Geriatr Gerontol Int* 2014; **14**: 430–439.
- 4 Chen JF, Xu K, Petzer JP *et al*: Neuroprotection by caffeine and A(2A) adenosine receptor inactivation in a model of Parkinson's disease. *J Neurosci* 2001; **21**: RC143.
- 5 Arendash GW, Schleif W, Rezai-Zadeh K *et al*: Caffeine protects Alzheimer's mice against cognitive impairment and reduces brain beta-amyloid production. *Neuroscience* 2006; **142**: 941–952.
- 6 Ferrari CC, Tarelli R: Parkinson's disease and systemic inflammation. *Parkinsons Dis* 2011; **2011**: 436813.
- 7 Farooqui T, Farooqui AA: Lipid-mediated oxidative stress and inflammation in the pathogenesis of Parkinson's disease. *Parkinsons Dis* 2011; **2011**: 247467.
- 8 Wirdefeldt K, Adami HO, Cole P, Trichopoulos D, Mandel J: Epidemiology and etiology of Parkinson's disease: a review of the evidence. *Eur J Epidemiol* 2011; **26**(Suppl 1): S1–58.
- 9 Cornelis MC, Monda KL, Yu K *et al*: Genome-wide meta-analysis identifies regions on 7p21(AHR) and 15q24(CYP1A2) as determinants of habitual caffeine consumption. *PLoS Genet* 2011; **7**: e1002033.
- 10 Cornelis MC, Byrne EM, Esko T *et al*: Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption. *Mol Psychiatry* 2015; **20**: 647–656.
- 11 Petronis A: Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* 2010; **465**: 721–727.
- 12 Ping J, Wang JF, Liu L *et al*: Prenatal caffeine ingestion induces aberrant DNA methylation and histone acetylation of steroidogenic factor 1 and inhibits fetal adrenal steroidogenesis. *Toxicology* 2014; **321**: 53–61.
- 13 Buscariollo DL, Fang X, Greenwood V, Xue H, Rivkees SA, Wendler CC: Embryonic caffeine exposure acts via A1 adenosine receptors to alter adult cardiac function and DNA methylation in mice. *PLoS ONE* 2014; **9**: e87547.
- 14 Wu DM, He Z, Ma LP, Wang LL, Ping J, Wang H: Increased DNA methylation of scavenger receptor class B type I contributes to inhibitory effects of prenatal caffeine ingestion on cholesterol uptake and steroidogenesis in fetal adrenals. *Toxicol Appl Pharmacol* 2015; **285**: 89–97.
- 15 Barres R, Yan J, Egan B *et al*: Acute exercise remodels promoter methylation in human skeletal muscle. *Cell Metab* 2012; **15**: 405–411.
- 16 Kang GA, Bronstein JM, Masterman DL, Redelings M, Crum JA, Ritz B: Clinical characteristics in early Parkinson's disease in a central California population-based study. *Mov Disord* 2005; **20**: 1133–1142.
- 17 Narayan S, Liew Z, Paul K *et al*: Household organophosphorus pesticide use and Parkinson's disease. *Int J Epidemiol* 2013; **42**: 1476–1485.
- 18 Costello S, Cockburn M, Bronstein J, Zhang X, Ritz B: Parkinson's disease and residential exposure to maneb and paraquat from agricultural applications in the central valley of California. *Am J Epidemiol* 2009; **169**: 919–926.
- 19 Narayan S, Sinsheimer JS, Paul KC *et al*: Genetic variability in ABCB1, occupational pesticide exposure, and Parkinson's disease. *Environ Res* 2015; **143**: 98–106.
- 20 Hughes AJ, Daniel SE, Kilford L, Lees AJ: Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* 1992; **55**: 181–184.
- 21 Gelb DJ, Oliver E, Gilman S: Diagnostic criteria for Parkinson disease. *Arch Neurol* 1999; **56**: 33–39.
- 22 Anonymous A: Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials* 1998; **19**: 61–109.
- 23 Curb JD, McTiernan A, Heckbert SR *et al*: Outcomes ascertainment and adjudication methods in the Women's Health Initiative. *Ann Epidemiol* 2003; **13**: S122–S128.
- 24 Dunning M, Barbosa-Morais N, Lynch A, Tavare S, Ritchie M: Statistical issues in the analysis of Illumina data. *BMC Bioinformatics* 2008; **9**: 85.
- 25 Houseman EA, Accomando WP, Koestler DC *et al*: DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012; **13**: 86.
- 26 Jaffe AE, Irizarry RA: Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 2014; **15**: R31.
- 27 Horvath S: DNA methylation age of human tissues and cell types. *Genome Biol* 2013; **14**: R115.
- 28 Vangaveti V, Shashidhar V, Jarrod G, Baune BT, Kennedy RL: Free fatty acid receptors: emerging targets for treatment of diabetes and its complications. *Ther Adv Endocrinol Metab* 2010; **1**: 165–175.
- 29 Stringhini S, Polidoro S, Sacerdote C *et al*: Life-course socioeconomic status and DNA methylation of genes regulating inflammation. *Int J Epidemiol* 2015; **44**: 1320–1330.
- 30 Miranda DM, Wajchenberg BL, Calsolari MR *et al*: Novel mutations of the BSCL2 and AGPAT2 genes in 10 families with Berardinelli-Seip congenital generalized lipodystrophy syndrome. *Clin Endocrinol (Oxf)* 2009; **71**: 512–517.
- 31 Vergnes L, Lee JM, Chin RG, Auwerx J, Reue K: Diet1 functions in the FGF15/19 enterohepatic signaling axis to modulate bile acid and lipid levels. *Cell Metab* 2013; **17**: 916–928.
- 32 Amin N, Byrne E, Johnson J *et al*: Genome-wide association analysis of coffee drinking suggests association with CYP1A1/CYP1A2 and NRCAM. *Mol Psychiatry* 2012; **17**: 1116–1129.
- 33 Byrne EM, Johnson J, McRae AF *et al*: A genome-wide association study of caffeine-related sleep disturbance: confirmation of a role for a common variant in the adenosine receptor. *Sleep* 2012; **35**: 967–975.
- 34 Hamza TH, Chen H, Hill-Burns EM *et al*: Genome-wide gene-environment study identifies glutamate receptor gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee. *PLoS Genet* 2011; **7**: e1002237.
- 35 Hernan MA, Takakouche B, Caamano-Isorna F, Gestal-Otero JJ: A meta-analysis of coffee drinking, cigarette smoking, and the risk of Parkinson's disease. *Ann Neurol* 2002; **52**: 276–284.
- 36 Barranco Quintana JL, Allam MF, Serrano Del Castillo A, Fernandez-Crehuet Navajas R: Alzheimer's disease and coffee: a quantitative review. *Neurol Res* 2007; **29**: 91–95.
- 37 Harold D, Abraham R, Hollingworth P *et al*: Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 2009; **41**: 1088–1093.
- 38 Victor Junji Yamamoto VdJrdP, Orestes Vicente Forlenza, Bernardo dos Santos, Daniel Shikanai Kerr: Association study in Alzheimer's disease of single nucleotide polymorphisms implicated with coffee consumption. *Arch Clin Psychiatry* 2015; **42**: 69–73.
- 39 Andersen JK: Oxidative stress in neurodegeneration: cause or consequence? *Nat Med* 2004; **10**(Suppl): S18–S25.
- 40 Ascherio A, Weisskopf MG, O'Reilly EJ *et al*: Coffee consumption, gender, and Parkinson's disease mortality in the cancer prevention study II cohort: the modifying effects of estrogen. *Am J Epidemiol* 2004; **160**: 977–984.
- 41 Palacios N, Gao X, McCullough ML *et al*: Caffeine and risk of Parkinson's disease in a large cohort of men and women. *Mov Disord* 2012; **27**: 1276–1282.

Supplementary Information accompanies this paper on *European Journal of Human Genetics* website (<http://www.nature.com/ejhg>)