OPEN ACCESS Freely available online

PLOS ONE

Methylome Analysis and Epigenetic Changes Associated with Menarcheal Age

Christiana A. Demetriou[1,2]*, Jia Chen[3], Silvia Polidoro[4], Karin van Veldhoven[2], Cyrille Cuenin[5], Gianluca Campanella[2], Kevin Brennan[6], Françoise Clavel-Chapelon[7,8], Laure Dossus[7,8], Marina Kvaskoff[7,8], Dagmar Drogan[9], Heiner Boeing[9], Rudolf Kaaks[10], Angela Risch[11], Dimitrios Trichopoulos[12,13,14], Pagona Lagiou[12,13,15], Giovanna Masala[16], Sabina Sieri[17], Rosario Tumino[18], Salvatore Panico[19], J. Ramón Quirós[20], María-José Sánchez Perez[21,22], Pilar Amiano[22,23], José María Huerta Castaño[22,24], Eva Ardanaz[22,25], Charlotte Onland-Moret[26], Petra Peeters[26], Kay-Tee Khaw[27], Nick Wareham[27], Timothy J. Key[28], Ruth C. Travis[28], Isabelle Romieu[29], Valentina Gallo[2,30], Marc Gunter[2], Zdenko Herceg[4], Kyriacos Kyriacou[1], Elio Riboli[31], James M. Flanagan[6], Paolo Vineis[2]

1 Department of Electron Microscopy & Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus, **2** Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom, **3** Departments of Preventive Medicine, Paediatrics, and Oncological Sciences, Mount Sinai School of Medicine, New York, New York, United States of America, **4** Molecular and Genetic Epidemiology Unit, Human Genetics Foundation, Torino, Italy, **5** Epigenetics Group, International Agency for Research on Cancer, Lyon, France, **6** Epigenetics Unit, Department of Surgery and Cancer, Imperial College London, London, United Kingdom, **7** Institut National de la Santé et de la Recherche Médicale (INSERM), Centre for Research in Epidemiology and Population Health, Institut Gustave Roussy, Villejuif, France, **8** Nutrition, Hormones and Cancer Unit, Paris South University, Villejuif, France, **9** German Institute of Human Nutrition Potsdam-Rehbruecke, Department of Epidemiology, Nuthetal, Germany, **10** Department of Cancer Epidemiology, German Cancer Research Center [DKFZ], Heidelberg, Germany, **11** Department of Epigenomics and Cancer Risk Factors, German Cancer Research Center [DKFZ], Heidelberg, Germany, **12** Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America, **13** Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece, **14** Hellenic Health Foundation, Athens, Greece, **15** WHO Collaborating Center for Food and Nutrition Policies, Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, Athens, Greece, **16** Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute – ISPO, Florence, Italy, **17** Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy, **18** Cancer Registry and Histopathology Unit, "Civile - M.P. Arezzo" Hospital, ASP Ragusa, Italy, **19** Department of Clinical and Experimental Medicine, Federico II University, Naples, Italy, **20** Public Health and Planning Directorate, Asturias, Spain, **21** Andalusian School of Public Health, Granada, Spain, **22** CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain, **23** Public Health Division of Gipuzkoa, BioDonostia Research Institute, Health Department of Basque Region, San Sebastian, Spain, **24** Department of Epidemiology, Murcia Regional Health Council, Murcia, Spain, **25** Navarre Public Health Institute, Pamplona, Spain, **26** Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, the Netherlands, **27** MRC Epidemiology Unit, Cambridge Institute of Public Health, Cambridge, United Kingdom, **28** Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom, **29** Nutrition and Metabolism Section, International Agency for Research on Cancer, Lyon, France, **30** Centre for Primary Care and Public Health, Barts and the London School of Medicine, Queen Mary, University of London, London, United Kingdom, **31** School of Public Health, Imperial College London, London, United Kingdom

Abstract

Reproductive factors have been linked to both breast cancer and DNA methylation, suggesting methylation as an important mechanism by which reproductive factors impact on disease risk. However, few studies have investigated the link between reproductive factors and DNA methylation in humans. Genome-wide methylation in peripheral blood lymphocytes of 376 healthy women from the prospective EPIC study was investigated using LUminometric Methylation Assay (LUMA). Also, methylation of 458877 CpG sites was additionally investigated in an independent group of 332 participants of the EPIC-Italy sub-cohort, using the Infinium HumanMethylation 450 BeadChip. Multivariate logistic regression and linear models were used to investigate the association between reproductive risk factors and genome wide and CpG-specific DNA methylation, respectively. Menarcheal age was inversely associated with global DNA methylation as measured with LUMA. For each yearly increase in age at menarche, the risk of having genome wide methylation below median level was increased by 32% (OR:1.32, 95%CI:1.14–1.53). When age at menarche was treated as a categorical variable, there was an inverse dose-response relationship with LUMA methylation levels ($OR_{12–14vs.\leq 11\ yrs}$:1.78, 95%CI:1.01–3.17 and $OR_{\geq 15vs.\leq 11\ yrs}$:4.59, 95%CI:2.04–10.33; P for trend$<$0.0001). However, average levels of global methylation as measured by the Illumina technology were not significantly associated with menarcheal age. In locus by locus comparative analyses, only one CpG site had significantly different methylation depending on the menarcheal age category examined, but this finding was not replicated by pyrosequencing in an independent data set. This study suggests a link between age at menarche and genome wide DNA methylation, and the difference in results between the two arrays suggests that repetitive element methylation has a role in the association. Epigenetic changes may be modulated by menarcheal age, or the association may be a mirror of other important changes in early life that have a detectable effect on both methylation levels and menarcheal age.

Citation: Demetriou CA, Chen J, Polidoro S, van Veldhoven K, Cuenin C, et al. (2013) Methylome Analysis and Epigenetic Changes Associated with Menarcheal Age. PLoS ONE 8(11): e79391. doi:10.1371/journal.pone.0079391

Editor: Wei Yan, University of Nevada School of Medicine, United States of America

Received June 10, 2013; **Accepted** July 30, 2013; **Published** November 20, 2013

Funding: CAD received funding from EU-Europe aid grant CRIS 2009/223–507.The EPIC cohort is supported by the Europe Against Cancer Program of the European Commission (SANCO). The individual centres also received funding from: Denmark (Danish Cancer Society); France (Ligue centre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, and Institut National de la Santé et de la Recherche Médicale (INSERM)); Greece (Hellenic Ministry of Health, the Stavros Niarchos Foundation and the Hellenic Health Foundation); Germany (German Cancer Aid, German Cancer Research Center, and Federal Ministry of Education and Research (Grant 01-EA-9401)); Italy (Italian Association for Research on Cancer and the National Research Council); The Netherlands (Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, and Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF)); Spain (Health Research Fund (FIS) of the Spanish Ministry of Health (Exp 96/0032) and the participating regional governments and institutions); Sweden (Swedish Cancer Society, Swedish Scientific Council, and Regional Government of Skane); and the United Kingdom (Cancer Research UK and Medical Research Council UK and Breast Cancer Campaign). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: christiana.demetriou08@imperial.ac.uk

Introduction

In addition to genetic changes, epigenetic changes and particularly DNA methylation can play an important role in the aetiology of chronic diseases such as cancer [1–5]. Gene specific promoter methylation can silence genes involved in critical cellular processes such as cell cycle regulation, DNA repair or apoptosis. At the same time, genome wide hypomethylation and in particular reduced methylation in repetitive elements such as Long Interspersed Nuclear Element-1 (LINE-1) and Alu repeats has been associated with chromosomal instability and mutations leading to chronic disease [1,3,4,6]. Methylation changes are most evident in tissues such as tumour biopsies when compared to normal tissue. However, genome wide methylation changes in relation to disease have been observed in surrogate tissues such as Peripheral Blood Leukocyte (PBL) DNA. Aberrant methylation in PBLs has been previously associated with breast cancer [7,8], colorectal adenoma [9,10], gastric cancer [11], head and neck squamous cell carcinoma [12], and bladder cancer [13]. A recent meta-analysis of all relevant studies has shown that there is overall little evidence to support an association with cancer using surrogate assays [14]. The exception has been one study based on a large population-based case-control study, the Long Island Breast Cancer Study Project, LIBCSP, with over 2,100 peripheral blood samples, which revealed greater global and promoter specific methylation in PBLs of breast cancer cases using LUMA [15]. Although the potential influence of the disease onset on the methylome of blood DNA needs to be tested, these results suggest that methylation in PBLs DNA can serve as a biomarker for chronic diseases such as cancer; it also points to a role of aberrant methylation in carcinogenesis.

Environmental exposures influence epigenetic changes, including methylation levels, particularly *in utero* and in early life [16,17]. In fact, genomic methylation has been shown to differ with respect to several accepted disease risk factors. These include age, race, anthropometric measures, environmental exposures and dietary factors [18–22]. For example, a prudent dietary pattern characterized by high intake of vegetables and fruit was shown to be associated with a lower prevalence of genomic hypomethylation (17, 19). Also, alcohol drinking and low dietary folate were found to impact on genomic DNA methylation – genome wide and gene specific [8,19]. In addition, in a multiethnic birth cohort in New-York City [18], BMI was not found to be associated with DNA methylation, but elsewhere, in women of childbearing age, a higher BMI was associated with lower global methylation [23]. In line with the latter finding, Zhang et al. [22] showed that higher physical activity is associated with higher global methylation in a cancer free population.

Reproductive factors were also shown to impact on global DNA methylation. Terry et al. [18] showed that factors that impact on breast cancer risk, including a greater birth height, a later age at menarche, nulliparity, and a later age at first birth were associated with higher global DNA methylation levels, but these results were not replicated in other studies [8,24]. However, the studies that investigate reproductive factors and epigenetic alterations are few [25].

In the present study we aim to investigate the impact of a number of reproductive variables on DNA methylation in PBLs of healthy individuals. The relationship was first investigated with genomic DNA methylation measurements using LUminometric Methylation Assay (LUMA) in 376 women. LUMA is a cytosine extension assay where the ratio of DNA CpG site cleavage by methylation sensitive restriction endonucleases (*Hpa*II) to the cleavage from methylation insensitive endonucleases (*Msp*I) is used to determine % global methylation. *Hpa*II cleavage occurs most frequently in CpG island promoters and repetitive elements thus methylation at these sites heavily influences the LUMA methylation estimate. Subsequently, to replicate the findings observed with LUMA, whole genome methylation patterns were obtained using Illumina 450 K in an independent group of 332 women. The Illumina 450 K array covers 485,577 CpG sites, achieving a high coverage of the entire genome, excluding repetitive elements.

Results

LUMA Methylation is Associated with Menarcheal Age

Demographic, anthropometric, lifestyle, and reproductive characteristics for subjects included in Stage 1 are presented in Table 1. The median LUMA genome-wide methylation in these subjects was 71.7% and the standard deviation was 5.7%. Of all the anthropometric measures and lifestyle variables examined, only age at menarche was found to significantly differ across quartiles of percent genome wide methylation (Table 2). Higher genome wide methylation was associated with a younger age at menarche (Kruskal-Wallis P-value = 0.002), and this association was significant even after Bonferroni correction for multiple testing (Table 2). Age at blood collection, height, weight, BMI, physical activity, smoking status, daily alcohol, folate consumption, age at FFTP, menopausal status, parity, breastfeeding, oral contraceptive (OC) use, hormone replacement therapy (HRT) use and highest level of education achieved did not significantly differ between subjects in methylation quartiles (Table 2).

The association between age at menarche and methylation was further examined using logistic regression to adjust for potential confounders. When median methylation was used as a cut off, 194 subjects had methylation below median levels and 182 had methylation levels above median. Using these two classes, logistic regression showed that age at menarche was significantly associated with class occupancy. As shown in Table 3, for every yearly increase in age at menarche, the risk of having below median methylation was increased by 32% (OR: 1.32, 95%CI: 1.14–1.53). When age at menarche was treated as a categorical

Table 1. Subject demographic, anthropometric, lifestyle, and reproductive characteristics, by analysis stage.

Covariate	Metric	Stage 1	Stage 2	Stage 3
		(n = 376)*	(n = 332)*	(n = 195)
Age	Range	33.4–75.6	34–70	35–65
	Median	52.7	54	49
	Mean (SD$^{\pm}$)	52.9 (9.4)	52.5 (7.1)	49.4 (7.3)
Height	Range	136.8–185.0	139.5–177.5	137.5–176.0
	Median	160	159.3	159
	Mean (SD)	160.1 (6.7)	159.0 (6.4)	158.7 (6.7)
Weight	Range	39.6–110.2	42.8–106	44–103.5
	Median	64.5	63.5	62
	Mean (SD)	66.2 (11.2)	64.4 (11.2)	63.8 (9.8)
BMI$^{\pm}$				
$<$25 kg/m^2	n (%)	182 (48.4)	164 (49.4)	98 (50.3)
25–30 kg/m^2	n (%)	141 (37.5)	118 (35.5)	7035.9)
$\geq$30 kg/m^2	n (%)	53 (14.1)	50 (15.1)	25 (12.8)
Physical Activity	n (%)	**Inactive:** 34 (9.0)		
	n (%)	**Moderately Inactive:** 85 (22.6)		
	n (%)	**Moderately Active:** 210 (55.9)		
	n (%)	**Active:** 45 (12.0)		
	n (%)	**Missing:** 2 (0.5)		
	Range		1–5	0.5–30
	Median		3	8.5
	Mean (SD)		2.6 (0.8)	10.0 (6.9)
Smoking Status				
Current Smoker	n (%)	79 (21.1)	69 (20.9)	35 (17.9)
Former Smoker	n (%)	65 (17.4)	66 (20.0)	48 (24.6)
Never	n (%)	230 (61.5)	195 (59.1)	112 (57.5)
Daily alcohol consumption (g/day)	Range	0–51.2	0–88.7	0–62.6
	Median	3.5	1.9	3.3
	Mean (SD)	6.5 (8.4)	8.7 (13.1)	9.6 (13.0)
Daily folate consumption (μg/day)	Range	90.4–1113.0	45.3–586.2	52.6–644.8
	Median	268.5	236.1	264.1
	Mean (SD)	291.5 (107.9)	247.3 (82.0)	276.9 (95.4)
Age at Menarche				
$\leq$11 yrs	n (%)	72 (19.4)	62 (18.8)	47 (24.1)
12–14 yrs	n (%)	242 (65.2)	233 (70.9)	134 (68.7)
$\geq$15 yrs	n (%)	57 (15.4)	34 (10.3)	14 (7.2)
Age at FFTP$^{\pm}$				
$<$25 yrs	n (%)	147 (45.6)	118 (35.5)	65 (41.9)
25–30 yrs	n (%)	131 (40.7)	118 (35.5)	69 (44.5)
$>$30 yrs	n (%)	44 (13.7)	96 (29.0)	21 (13.6)
Parous				
No	n (%)	45 (12.0)	29 (8.8)	40 (20.5)
Yes	n (%)	330 (88.0)	301 (91.2)	155 (79.5)
Breastfeeding				
No	n (%)	109 (30.0)	108 (32.7)	75 (48.4)
Yes	n (%)	255 (70.0)	222 (67.3)	80 (51.6)
Menopausal Status				
Premenopausal	n (%)	175 (46.5)	155 (46.7)	90 (46.2)
Postmenopausal	n (%)	201 (53.5)	177 (53.3)	105 (53.8)

Table 1. Cont.

Covariate	Metric	Stage 1	Stage 2	Stage 3
		(n = 376)*	(n = 332)*	(n = 195)
HRT$^{\pm}$ use				
Ever	n (%)	29 (7.8)	13 (3.9)	34 (17.4)
Never	n (%)	343 (92.2)	317 (96.1)	158 (81.0)
OC$^{\pm}$ use				
Ever	n (%)	176 (46.9)	131 (39.7)	94 (48.2)
Never	n (%)	199 (53.1)	199 (60.3)	101 (51.8)

*Failure of category counts to add up to this value denotes missing values.
$^{\pm}$SD: Standard Deviation, BMI: Body Mass Index, FFTP: First Full Term Pregnancy, HRT: Hormone Replacement Therapy, OC: Oral Contraceptive.
doi:10.1371/journal.pone.0079391.t001

variable, there was an inverse dose-response relationship with methylation levels: for the age category 12–14 compared to ≤11 years the OR was 1.78 (95% CI: 1.01–3.17), and for the age category ≥15 compared to ≤11 years the OR was 4.59 (95% CI: 2.04–10.33) (P for trend<0.0001). These significant associations persisted even after adjustment for relevant confounders: centre, plate number, age at blood collection, height, weight, total physical activity, smoking status, daily alcohol consumption, and daily folate consumption.

Illumina 450 k Methylome Analysis Identifies an Epi-allele Associated with Menarcheal Age

In the second population group, 329 subjects (out of 332) had available information on age at menarche (Table 1). In contrast to LUMA global methylation, the median genome-wide methylation level using the 450 k ILLUMINA assay did not significantly differ between menarcheal age groups. Similarly, CpG island methylation and promoter methylation were not significantly different between subjects in different menarcheal age categories. However, there was a trend towards decreasing methylation with increasing age at menarche, consistent with the LUMA results (Figure 1).

When adjusting for case-control status, age, and position on the chip in a linear regression model with methylation M-values as a continuous outcome, and age at menarche as a categorical variable (>11 yrs vs. ≤11 yrs), age at menarche was significantly associated with methylation in a single CpG site (cg01339004), located on the body of the SMAD6 gene ($p<1.00\times10^{-7}$, genome-wide level significance) (Table 4, Figure 2). When only those subjects that remained healthy for at least 5 years following recruitment and blood collection were analysed, the same CpG site was found to be significantly associated with age at menarche ($p = 6.71\times10^{-8}$).

However, using bisulphite Pyrosequencing for the SMAD6 cg01339004 locus, we were unable to replicate this finding in an independent sample set using a generalized linear model while adjusting for the same confounders (n = 185, p = 0.07). Wilcoxon rank sum non-parametric test also did not reveal significantly differential methylation between the two age at menarche categories (p = 0.082) measured using bisuphite pyrosequencing (Figure 2).

Discussion

In this study, age at menarche was negatively associated with LUMA genome wide methylation in a statistically significant manner. The association of genome wide methylation with menarcheal age was the only strong and consistent association we found and remained unaltered after adjustment for relevant confounders. Previous study results on age at menarche and methylation were conflicting. Terry et al. [18] found that a later age at menarche was associated with higher genomic global methylation later on in adulthood, but DNA methylation was only assessed in 92 individuals and the authors used a different technique for measuring global methylation ([^{3}H]-methyl acceptance assay). On the other hand, Choi et al. [8] did not demonstrate a statistically significant association between menarcheal age and global DNA methylation using LINE1 methylation as a surrogate for global methylation.

The negative association between LUMA methylation and later age at menarche is counter-intuitive because (a) a later age at menarche is known to protect from breast cancer, and (b) lower global methylation is expected to increase genome instability and thus increase cancer risk [8,26]. However, our observation is consistent with the findings in the LIBCSP study, where breast cancer was associated with increased genome wide methylation as measured with LUMA [15]. This apparent paradox could be explained by LUMA's characteristics, i.e. broad coverage in CpG dense regions, such as promoters, and decreased coverage in the remaining genome [27]. Another potential explanation is that aberrant methylation associated with age at menarche is unrelated to the methylation changes relevant to breast cancer, or that the association with age at menarche is in fact confounded by other determinants of methylation levels.

Given the conflicting reports in the literature [8,15,18], we aimed to replicate, in a dataset with whole genome methylation data, the association between age at menarche and DNA methylation that we observed with the LUMA technology. This was done by using the robust Illumina technology. This approach also enabled the identification of specific genes that might be involved in the mechanistic pathways linking menarcheal age with disease. In contrast to the findings of LUMA, genome wide methylation in this second dataset did not significantly differ between subjects in different menarcheal age groups. However, there was a non-significant trend towards decreasing methylation with increasing age at menarche, consistent with the LUMA findings (Figure 1). The lack of association in this dataset could be caused by differences in coverage between the two assays used. LUMA assesses methylation of a specific restriction enzyme site (HpaII, CCGG), which occurs most frequently in CpG island promoters – also covered by the 450 K array – but also in repetitive elements. However, the Infinium HumanMethylation 450 BeadChip, due to its probe design, does not interrogate

Table 2. Anthropometric and lifestyle variables in healthy controls with respect to LUMA genome wide methylation quartiles (Stage 1).

Variable		Methylation Quartile: Cut-offs				
		1: 23.0–68.4%	2: 68.5–71.7%	3: 71.8–74.0%	4: 74.1–80.0%	p-value[a]
	Units	(n = 103)	(n = 91)	(n = 93)	(n = 89)	
Age at blood collection	Mean ± SD	52.9±8.5	52.2±8.5	53.1±10.0	53.6±10.7	0.863
Height	Mean ± SD	159.0±6.1	160.7±6.4	160.8±6.7	160.0±7.5	0.424
Weight	Mean ± SD	66.0±12.3	66.5±10.5	64.8±11.3	67.7±10.4	0.316
BMI[±]	Mean ± SD	26.2±5.1	25.8±4.1	25.1±4.4	26.5±4.4	0.112
Physical Activity						
Inactive	N, (%)	4 (3.9)	9 (9.9)	11 (11.8)	10 (11.2)	
Moderately Inactive	N, (%)	29 (28.2)	22 (24.2)	16 (17.2)	18 (20.3)	
Moderately Active	N, (%)	60 (58.3)	45 (49.4)	54 (58.1)	51 (57.3)	
Active	N, (%)	9 (8.7)	15 (16.5)	11 (11.8)	10 (11.2)	
Missing	N, (%)	1 (0.9)	0 (0.0)	1 (1.1)	0 (0.0)	0.404
Smoking Status						
Current Smoker	N, (%)	58 (56.3)	59 (64.8)	49 (52.7)	47 (52.9)	
Former Smoker	N, (%)	23 (22.3)	12 (13.2)	20 (21.5)	22 (24.7)	
Never	N, (%)	22 (21.4)	20 (22.0)	23 (24.7)	19 (21.3)	
Unknown	N, (%)	0 (0.0)	0 (0.0)	1 (1.1)	1 (1.1)	0.596
Alcohol consumption –lifetime average (g/day)	Mean ± SD	7.0±8.6	6.9±9.7	5.4±6.4	6.7±8.6	0.822
Dietary folate intake (g/day)	Mean ± SD	294.6±121.9	276.9±96.7	293.1±104.2	301.0±105.3	0.386
Age at menarche	Mean ± SD	13.2±1.7	13.3±1.7	12.5±1.4	12.8±1.7	**0.002***
Age at FFTP[±]	Mean ± SD	25.3±3.5	26.2±3.9	24.8±4.0	25.0±3.7	0.132
Menopausal Status						
Pre	N, (%)	42 (40.8)	47 (51.6)	46 (49.5)	40 (44.9)	
Post	N, (%)	60 (58.3)	43 (47.3)	45 (48.4)	49 (55.1)	
Surgical Post	N, (%)	1 (0.9)	1 (1.1)	2 (2.1)	0 (0.0)	0.545
Parous						
No	N, (%)	12 (11.7)	7 (7.7)	13 (14.0)	19 (21.3)	
Yes	N, (%)	90 (87.4)	84 (92.3)	79 (84.9)	70 (78.7)	
Unknown	N, (%)	1 (0.9)	0 (0.0)	1 (1.1)	0 (0.0)	0.075
Breastfeeding						
No	N, (%)	30 (29.1)	19 (20.9)	26 (28.0)	34 (38.2)	
Yes	N, (%)	70 (68.0)	70 (76.9)	62 (66.7)	53 (59.6)	
Unknown	N, (%)	3 (2.9)	2 (2.2)	5 (5.3)	2 (2.2)	0.086
OC[±] **use**						
No	N, (%)	50 (48.5)	53 (58.2)	47 (50.5)	49 (55.1)	
Yes	N, (%)	53 (51.5)	38 (41.8)	46 (49.5)	39 (43.8)	
Unknown	N, (%)	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.1)	0.512
HRT[±] **use**						
No	N, (%)	93 (90.3)	84 (92.3)	81 (87.1)	85 (95.5)	
Yes	N, (%)	3 (2.9)	6 (6.6)	10 (10.7)	4 (4.5)	
Unknown	N, (%)	7 (6.8)	1 (1.1)	2 (2.2)	0 (0.0)	0.399
Highest Education						
None	N, (%)	10 (9.6)	6 (6.6)	7 (7.4)	6 (6.8)	
Primary	N, (%)	45 (43.7)	41 (45.1)	36 (38.7)	35 (39.3)	
Technical/Professional	N, (%)	15 (14.6)	9 (9.9)	18 (19.4)	20 (22.5)	
Secondary	N, (%)	12 (11.7)	21 (23.1)	17 (18.3)	17 (19.1)	

Table 2. Cont.

Variable		Methylation Quartile: Cut-offs				
		1: 23.0–68.4%	2: 68.5–71.7%	3: 71.8–74.0%	4: 74.1–80.0%	p-value[a]
	Units	(n = 103)	(n = 91)	(n = 93)	(n = 89)	
University	N, (%)	12 (11.7)	13 (14.3)	14 (15.1)	10 (11.2)	
Unspecified	N, (%)	9 (8.7)	1 (1.0)	1 (1.1)	1 (1.1)	0.264

[a]For continuous variables, P-value was derived from Kruskal-Wallis test. For categorical variables, P-value was derived from a chi square test, with the exclusion of ''Unknown'' categories due to their small cell counts. Both reflect the association between quartiles of methylation and the investigated variables.
*Significant at the Bonferroni-corrected significance cut off (P = 0.003) for multiple comparisons.
±BMI: Body Mass Index, FFTP: First Full Term Pregnancy, HRT: Hormone Replacement Therapy, OC: Oral Contraceptive.
doi:10.1371/journal.pone.0079391.t002

repetitive elements which are found to be differentially methylated in many cases of neoplasia [28,29]. For example, satellite and SINE repeats were found to be enriched with hypomethylated Differentially Methylated Regions (DMRs) whereas LINE was enriched with hypermethylated DMRs in malignant peripheral nerve sheath tumours compared to normal Schwann cells [30]. If age at menarche is related to methylation patterns in these repetitive elements, the hypomethylation would not have been evident in the 450 K chip but it would have been detected in the LUMA assay. This suggests that the LUMA based association is being driven largely by methylation differences in repetitive elements, where age at menarche could have a greater effect.

In the locus by locus analysis, methylation of a single CpG site was shown to be associated with age at menarche. However, in an independent sample set, this finding was not replicated. Given the multiple comparisons in the locus by locus analyses in the Illumina dataset, one cannot rule out the possibility that this finding is the result of chance, and given that the independent sample set did not replicate this finding using an alternative method, we conclude that it is likely to be a false positive association. However, further validation in further independent data sets, with a greater sample size may increase the power sufficiently to detect possible associations between methylation of individual loci and age at menarche.

The mechanistic link which could explain the association between menarcheal age and genome-wide DNA methylation, but not in individual CpG loci is yet to be determined. However, endogenous oestrogen exposure is a strong candidate for epigenetic changes since an earlier age at menarche exposes a woman to a greater cumulative amount of endogenous oestrogens

Table 3. Logistic Regression for percent genome wide methylation (LUMA levels below vs. above median) by age at menarche as a categorical variable and other relevant confounders.

Variable	Methylation Median ± SD	Adjusted OR[a]	95% Confidence Interval	P-value
Center	NA	0.99	0.97–1.02	0.518
Plate number	NA	0.94	0.80–1.10	0.400
Age at blood collection (continuous)	NA	0.98	0.96–1.00	0.254
Height (continuous in cm)	NA	0.98	0.95–1.01	0.354
Weight (continuous in kg)	NA	1.01	0.98–1.03	0.453
Total physical activity index – sex specific (continuous activity categories)	NA	1.03	0.78–1.37	0.813
Smoking status				
Never	71.62±6.25	1.00		
Past smoker	72.18±4.51	0.85	0.46–1.55	0.592
Current smoker	70.73±5.17	0.98	0.56–1.72	0.954
Daily alcohol intake (continuous in g/day)	NA	1.01	0.99–1.04	0.354
Daily folate intake (continuous in μg/day)	NA	1.00	0.99–1.00	0.863
Age at menarche (continuous in years)	NA	**1.32**	1.14–1.53	**<0.0001***
Age at menarche (categorical)				
≤11 years old	72.59±4.49	**1.00**		
12–14 years old	71.62±5.93	**1.78**	1.01–3.17	**0.048***
≥15 years old	70.12±6.33	**4.59**	2.04–10.33	**<0.0001***
P for Trend			**<0.0001***	

[a]Each OR is adjusted for all other variables in the table.
*Significant at the 0.05 level.
doi:10.1371/journal.pone.0079391.t003

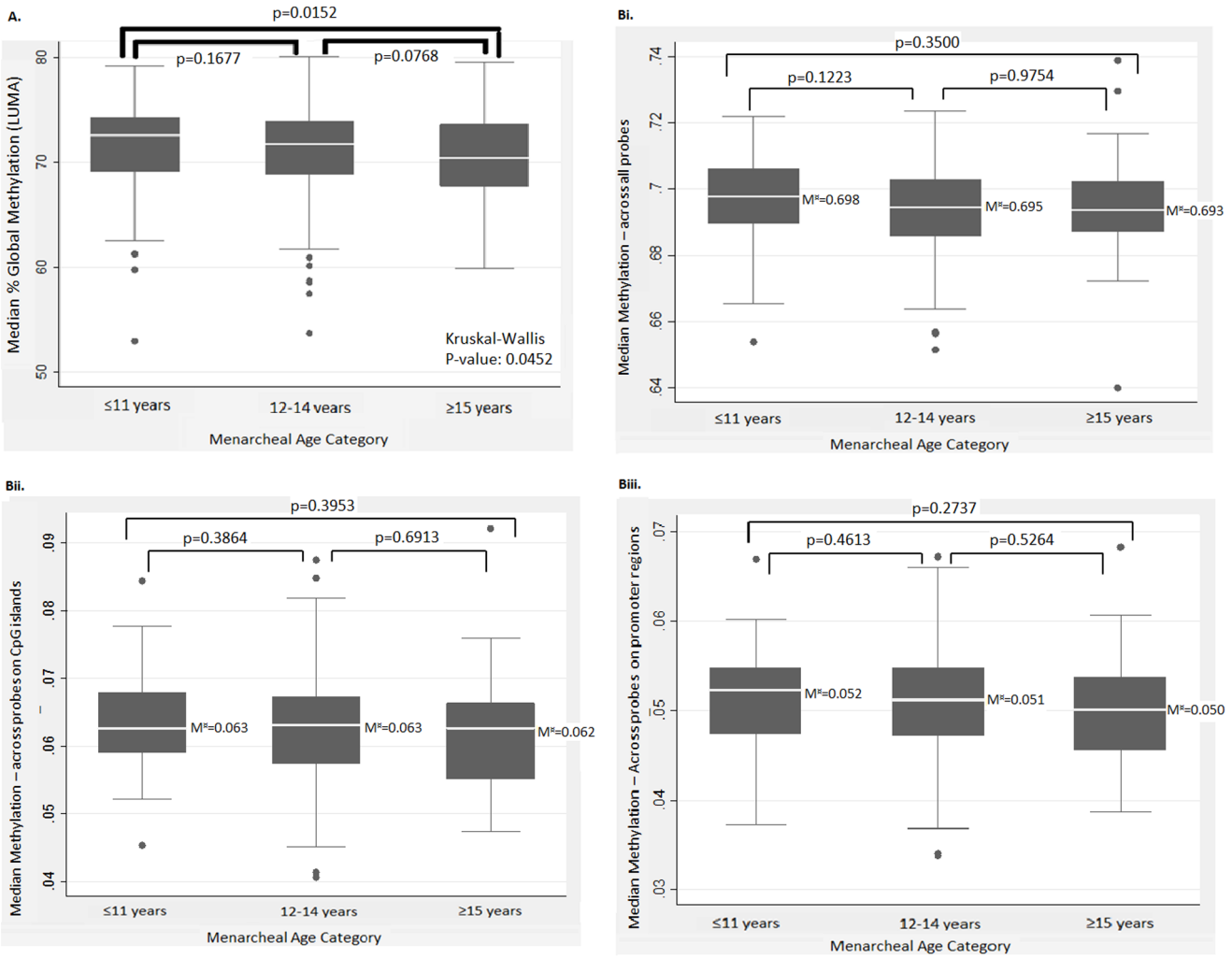

Figure 1. Boxplots of median genome-wide methylation between the three menarcheal age categories. A: Median % global methylation as measured with LUMA in Stage 1. Bi. Genome-wide methylation across all probes (averaged per individual). Bii. Genome-wide methylation across probes on CpG islands (averaged per individual). Biii. Genome-wide methylation across probes on promoter regions (averaged per individual). M^{α} = Median methylation value. p = p value from Wilcoxon rank-sum test comparisons.
doi:10.1371/journal.pone.0079391.g001

over her lifetime, and there have been reports which show that oestrogen impacts on DNA methylation. More specifically it was shown that oestrogen receptor (ER) positive breast tumour tissues have differential methylation at several CpG loci compared to ER negative tumours [24,31] and oestrogen induced breast tumours have differential DNA methylation patterns in ACI rat mammary gland tissue [32]. Further investigation into the role of oestrogen on repetitive element methylation is, therefore, warranted.

It is also possible that age at menarche is an indirect indicator of other macroscopic changes that may impact on DNA methylation.

Table 4. Significant CpG sites in a linear regression model.

TargetID	P-value$^{\propto}$	Q-value$^{\not\propto}$	Regression Coefficient$^{\approx}$	Chromosome number	Gene	CpG Position Relative to Gene	CpG Island's Name
cg01339004	8.83E-08	0.0392	−0.2765	15	SMAD6	Body	NA

Methylation treated as a continuous outcome (M-values: PBC and COMBAT on chip) and menarcheal age category (>11 vs. ≤11 years) treated as a categorical exposure. Adjusting for age at blood collection, case-control status, and position on the chip.
Analysis of all subjects or of only the 240 subjects that remained healthy for at least 5 years following recruitment yielded the same results.
$^{\propto}$P value from a liner regression model where methylation is treated as a continuous outcome (M-values: PBC and COMBAT on chip) and the effect of age at menarche as a categorical variable (>11 vs. ≤11 years), adjusted for age, case-control status, and chip position.
$^{\not\propto}$Q value: False Discovery Rate (FDR) corrected P-value.
$^{\approx}$The regression coefficient for each probe; change in methylation for having an age at menarche >11 years vs. ≤11 years.
doi:10.1371/journal.pone.0079391.t004

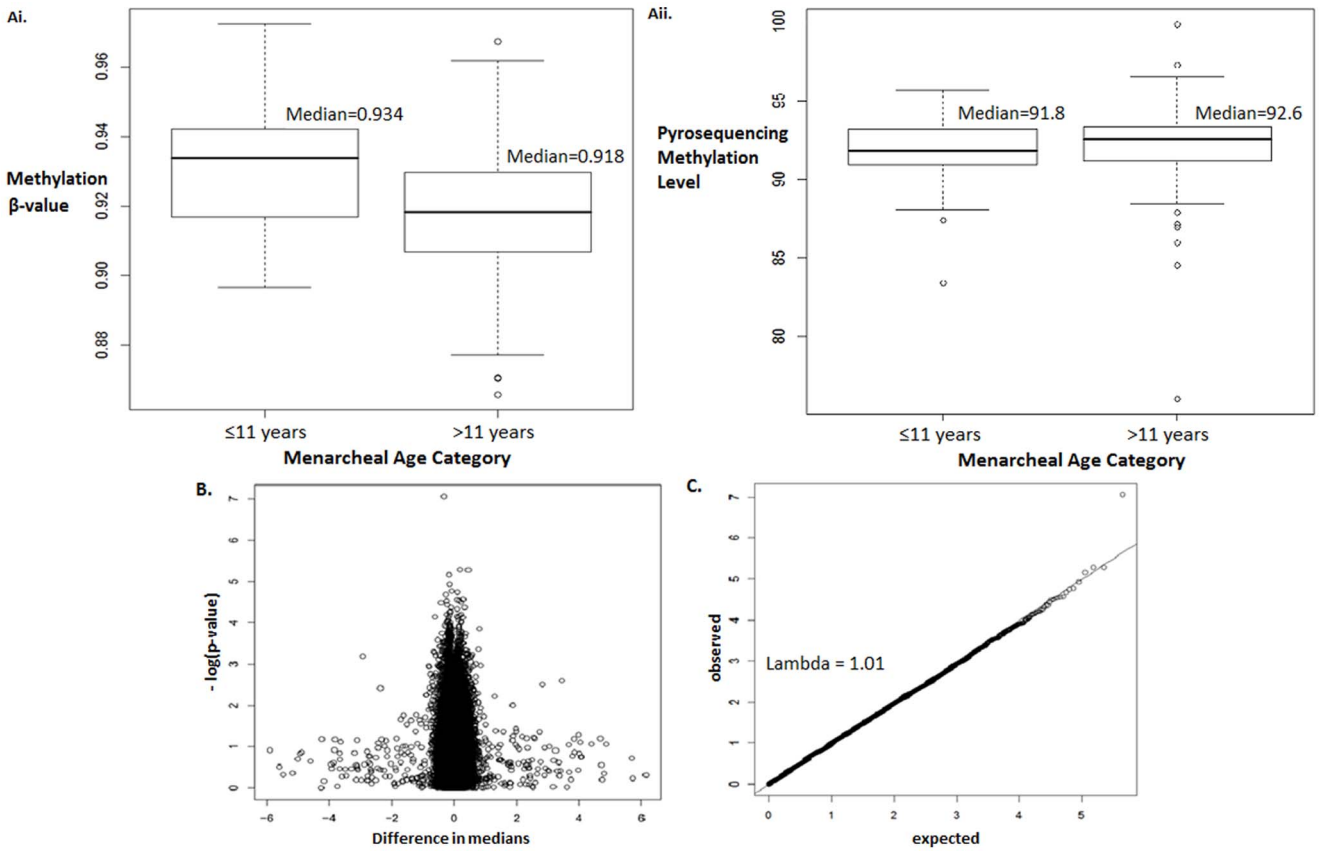

Figure 2. Analysis of SMAD6 cg01339004 probe methylation. Ai: Boxplot of β-value methylation of cg01339004 probe as measured with Illumina 450 k beadchip in Stage 2. Aii: Boxplot of methylation level of cg01339004 probe as measured with bisulphite pyrosequencing in Stage 3. B: Volcano plot: Difference in median methylation between the two menarcheal age groups (>11 (n = 268) vs. ≤11 years, (n = 62), against the –log(P-Value) of a linear regression analysis with methylation as a continuous outcome (M-values) and age at menarche (>11 vs. ≤11 years) as a categorical exposure, adjusting for age, case-control status, and chip position. C. Q-Q plot on P-values from a linear regression analysis with methylation as a continuous outcome (M-values) and age at menarche (>11 vs. ≤11 years) as a categorical exposure, adjusting for age, case-control status, and chip position.
doi:10.1371/journal.pone.0079391.g002

It is widely accepted that development is plastic and the sensitivity of the epigenetic system to environmental factors is heightened during periods of developmental plasticity such as childhood, adolescence and puberty [17]. Epigenetic modifications in response to environmental exposures at these critical periods are often subtle initially and even though they do not lead to phenotypic changes at the time of exposure, they may lead to increased risk of dysfunction and disease later on in life [17]. Trends in the past decades show a rapid shift towards an earlier age at menarche and this is more pronounced in developed countries [33]. This trend is too steep to be attributed to genetic changes. Instead, environmental exposures at the periods of developmental plasticity are likely to be the cause of the dramatic decrease in age at menarche. For example, childhood obesity disrupts the hormonal milieu leading to an increase in adipocyte secreted leptin, or in adrenal secreted androgens, all of which impact on menarcheal onset [34]. Pre or neo-natal nutrition as well as early life exposure to endocrine disrupting chemicals (EDCs) can also lead to hormonal imbalances impacting on age at menarche [17]. Given that these exposures occur at the periods when the epigenetic signature is more plastic, they might also lead to aberrant DNA methylation changes which will be inherited during cell divisions and be detectable years later. Therefore, the aberrant DNA methylation pattern observed in adulthood might not be related to menarcheal age per se but to an early life environmental exposure, like diet, that impacts both on age at menarche and on DNA methylation.

This study suggests an association between age at menarche and DNA methylation. All samples in this study were collected prior to the onset of disease, and the changes observed were present in the blood of individuals when they were still healthy, at least five years prior to their diagnosis, limiting the potential influence of the presence of cancer (reverse causality) on the methylome of blood DNA. In addition, the sample sizes examined –376 subjects for LUMA and 332 subjects for Illumina 450 K Methylation are fairly large datasets, allowing for sufficient power to detect significant methylation changes if present. However, one important limitation of our study was the lack of information on other early life exposures, therefore it was impossible to investigate whether such exposures confound the observed association between age at menarche and DNA methylation. Thus, this hypothesis needs to be further investigated in birth cohorts.

Overall, our results suggest that DNA methylation changes, particularly in repetitive elements, may be associated with menarcheal age. However, it is also possible that some important changes taking place in early life and which are associated with age at menarche – in particular nutrition – have a detectable effect on methylation levels.

Materials and Methods

Stage 1: Genome Wide Methylation with LUminometric Methylation Assay (LUMA)

Study participants. All participants signed an informed consent and the study protocol was approved by the Ethics Committee of the International Agency for Research on Cancer.

Epidemiologic data and blood samples collected from the European Prospective Investigation into Cancer and Nutrition (EPIC) were used. EPIC is an ongoing study designed to investigate diet, nutrition, lifestyle and environmental factors with respect to cancer incidence. The cohort consists of 519,978 participants from 23 centres in 10 European countries - Denmark, France, Germany, Greece, Italy, the Netherlands, Norway, Spain, Sweden and the United Kingdom. Information on lifestyle, diet, anthropometric measures and environmental exposures were collected using questionnaires at recruitment and were standardized across the different participating centres. Blood was also collected from the majority of subjects at recruitment [35]. For the LUMA investigation, 600 individuals – half breast cancer cases and half controls – from the EPIC cohort were chosen. Of these, 77 subjects were initially excluded: 1 subject was a duplicate, there was not enough DNA for 24 subjects, and 52 samples produced no or a weak signal. Of the remaining 523 subjects with reliable measurements, we investigated 376 women in this specific study, who remained free of cancer for at least 5 years following blood collection.

Genome wide DNA methylation. LUMA was used to quantify genome wide methylation levels [27,36] in PBLs in the blood of subjects, collected at recruitment. Genomic DNA was extracted using standard protocols. LUMA gives a measure of % global methylation using the ratio of DNA cleavage by methylation sensitive (HpaII) and methylation insensitive (MspI) restriction enzymes. In LUMA, polymerase extension assay by Pyrosequencing is employed to determine cleavage. The LUMA method was validated using DNA controls of known DNA methylation status [37]. In the assay, 5-Aza-dC treated and CpG methylated Jurkat genomic DNA (New England Biolabs, Ipswich, MA) were used as methylated and unmethylated control samples. Genome wide methylation is expressed as a percentage obtained from the equation [37]:

$$Genomewide Methylation(\%) = [1-((HpaII \sum G/\sum T) \div (MspI \sum G/\sum T))] \times 100.$$

Statistical analyses. We first compared the distribution of a number of anthropometric measures, reproductive factors, and lifestyle characteristics such as age at blood collection, height, weight, parity, age at first full term pregnancy, breastfeeding and hormone use across quartiles of percent global methylation. The quartile cut offs were the 25^{th}, 50^{th}, and 75^{th} percentile methylation values in controls. For continuous variables, the non-parametric equivalent of one-way analysis of variance (ANOVA), the Kruskal-Wallis test, was used as genome wide methylation was not normally distributed. For categorical variables, chi-square test was used.

The reproductive variables that were statistically differentially distributed between methylation quartiles were investigated further. The resulting profile of genome wide methylation distribution was skewed and several transformations failed to normalize it. In addition, various GLM models investigated failed to adequately describe the outcome distribution. As a result, the methylation outcome was dichotomized – above and below median methylation – and unconditional logistic regression was used to evaluate the association between exposure variables and DNA methylation, the latter being the dependent variable. All significant variables were included both as continuous and categorical when relevant (e.g. age at menarche $\leq$11 y, 12–14 y, $\geq$15 y). Based on the available literature, the logistic regression model was fully adjusted for centre, plate number, age at blood collection, height, weight, total physical activity, smoking status, daily alcohol consumption, and daily folate consumption. All confounders were entered into the model as continuous variables with the exception of smoking status which was treated as categorical – past, never, present.

All analyses were performed using STATA (Release 11; College Station, TX: StataCorp LP).

Stage 2: Locus-by-locus Analysis with Illumina 450 K to Replicate the Findings of LUMA Genome Wide Methylation Analysis

Study participants. All participants signed an informed consent and the study protocol was approved by the ethics committee of the Human Genetics Foundation (HuGeF).

The EPIC Italy sub-cohort consists of 32,578 female subjects recruited from 5 different centers – Varese, Turin, Florence, Naples, and Ragusa. From this subcohort, 166 breast cancer cases and 166 controls, matched on date of birth ($\pm$5 years), seasonality of blood draw, and date of recruitment were selected. However, since for this investigation case/control status is not the outcome and since at the time of blood collection all individuals were healthy, all 332 blood samples were treated as healthy blood. Nevertheless, given the long latency period of neoplasia, analyses were carried out on all subjects with age at menarche information (n = 329) as well as on only the subjects that remained healthy at least 5 years following recruitment and blood collection (n = 240).

Illumina 450 K methylation. DNA was extracted from buffy coats or blood cell fractions using the QIAsymphony DNA Midi Kit (Qiagen, Crawley, UK). 500 ng of DNA was bisulphite-converted with the EZ-96 DNA Methylation-Gold^{TM} Kit, used according to the manufacturer's protocol (Zymo Research, Orange, CA, USA). Next, the 450 K DNA methylation array by Illumina (Infinium HumanMethylation 450 BeadChip) was performed on 4 µl of bisulphite-converted DNA, following the Illumina Infinium HD Methylation protocol. This array includes 485,577 cytosine positions of the human genome (482,421 CpG sites (99.4%), 3091 non-CpG sites and 65 random SNPs; hereafter the term CpG will be used to refer to all of these, unless otherwise specified). Briefly, a whole genome amplification step was followed by enzymatic end-point fragmentation and hybridization to HumanMethylation 450 BeadChips at 48°C for 17 h, followed by single nucleotide extension. The incorporated nucleotides were labelled with biotin (ddCTP and ddGTP) and 2,4-dinitrophenol (DNP) (ddATP and ddTTP). After the extension step and staining, the BeadChip was washed and scanned using the Illumina HiScan SQ scanner. The intensities of the images were extracted using the GenomeStudio (v.2011.1) Methylation module (1.9.0) software, which normalizes within-sample data using different internal controls that are present on the HumanMethylation 450 BeadChip and internal background probes. The Infinium HumanMethylation 450 BeadChip data, for subjects with age at menarche information, were made available on the data repository Gene Expression Omnibus (GEO), with accession number GSE51057.

Statistical analyses. Methylated and unmethylated intensities for each probe were provided by GenomeStudio software. In

addition to the methylation value for each probe, corresponding detection p-values were also provided by GenomeStudio software (Illumina). The detection values indicate the confidence that can be placed on a β-value reading. As a first step, all readings with a p-value>0.05 were considered as non-detected so as to not influence downstream pre-processing and analyses.

Background noise correction was then performed as background fluorescence can contribute an additive error to each signal intensity leading to a reduced dynamic range for the methylation reading. Given that signal intensities can be red or green, and given the technical variation in fluorescent signal depending on the intensity colour, dye bias also had to be taken into account using the method described by Triche et al. [38]. The analysis of other classical quality control measures (such as staining, extension, hybridization, or bisulphite conversion) provided by GenomeStudio did not reveal any major quality issues.

The methylated and unmethylated intensities provided by GenomeStudio were used to calculate methylation β-values based on the equation:

$$Beta(\beta) = \frac{\max(M,0)}{\max(M,0) + \max(U,0) + 100}$$

where M is the intensity of the methylated signal and U the intensity of the unmethylated signal at each probe.

Beta values were later peak based corrected (PBC) as suggested by Dedeurwaerder et al. [39] in order to correct for the bias arising from the two different probe designs on the array. In order to correct for batch effects, COMBAT was then used [40,41]. Lastly, missing data were imputed using KNN (k-nearest neighbours) method once, implemented in knn.impute function from R-CRAN.

In order to replicate the results observed with LUMA, methylation beta values across all probes were averaged per individual to derive a measure of genome-wide methylation per subject. Similarly, probes in CpG islands and promoter regions were averaged per subject. Wilcoxon-rank sum tests were performed to examine whether genome-wide, CpG island and promoter methylation was significantly different between subjects in the three menarcheal age groups examined with LUMA (≤11 y, 12–14 y, ≥15 y).

In addition, locus by locus analysis of methylation was performed using a linear regression model. Quantile normalization was performed prior to regression using the R package "preprocessCore" from Bioconductor. In order to satisfy the normality assumptions of a linear regression model, beta methylation values were converted to M-values as described in [42] and entered into the model as a continuous outcome. M-values were also peak based corrected and COMBAT adjusted for chip number to correct for batch effects. Only age at menarche, the only significant reproductive variable in LUMA analysis, was investigated here. Age at menarche was treated as a categorical outcome (≤11 yrs vs. >11 yrs). This categorization was chosen since in the LUMA data, both menarcheal age categories above 11 yrs old were significantly associated with methylation when compared to a menarcheal age of ≤11 years.

Q-values, measuring the maximum False Discovery Rate (FDR) from the Benjamini and Hochberg method were derived for each probe analysis, and overall Type I error was controlled for by conservatively applying Bonferroni multiple testing correction. The per-test significance cut-off value was set to 1.00×10^{-7}. The analysis was performed first on all 329 subjects with age at menarche information and then repeated only on the 240 subjects that remained healthy for at least five years following recruitment. The linear model was adjusted for age, case-control status and chip position. All confounders were entered into the model as continuous variables with the exception of chip position.

Stage 3: Single Locus (SMAD6, cg01339004) Pyrosequencing Analysis to Replicate the Findings of Locus by locus Illumina 450 K Analysis

Study participants. The participants used in this stage were also subjects of the EPIC Italy sub-cohort. One hundred ninety-five women, with available information on age at menarche, were selected for bisulphite pyrosequencing based on sample availability from other studies. Eight of these subjects overlapped with the subjects used in Stage 2.

Bisulphite Pyrosequencing was used to quantify CpG specific methylation at the SMAD6 locus in these individuals using standard protocols [43]. The primers used for cg01339004 were Forward ([BIOTIN]–TGGTATAGTAGTGGTTTGGTATAAGAT), Reverse (TACCACCCACCCATTCACTCTATAA) and Sequencing Primer (TCTATAAATAAACAAACTAAAACC).

Statistical analyses. Out of the 195 samples analysed, for 10 samples the pyrosequencing results did not pass quality check. Wilcoxon rank sum non-parametric test was performed to examine whether SMAD6 cg01339004 methylation was different between age at menarche categories (≤11 vs. >11 years). In addition, a generalized linear regression model with SMAD6 cg01339004 methylation as a continuous outcome and with age at menarche as a categorical variable (≤11 vs. >11 years) was run to correct for confounding variables – age at blood collection and case-control status – as in the case of the Illumina 450 K analysis.

Acknowledgments

The authors would like to express their appreciation to Dr. Fulvio Ricceri at the Human Genetics Foundation in Torino for his invaluable help in data provision and analyses.

Author Contributions

Conceived and designed the experiments: CAD JC CC S. Polidoro KvV GC KB JMF KK ZH PV. Performed the experiments: FCC LD MK DD HB RK AR DT PL GM SS RT S. Panico JRQ MJS PA JMHC EA COM PP KTK NW TJK RCT IR VG MG ER. Analyzed the data: CAD JC CC S. Polidoro KvV GC KB JMF KK ZH PV. Wrote the paper: CAD JC CC S. Polidoro KvV GC KB JMF KK ZH PV.

References

1. Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell 128: 683–692. doi:10.1016/j.cell.2007.01.029.
2. Jovanovic J, Rønneberg JA, Tost J, Kristensen V (2010) The epigenetics of breast cancer. Mol Oncol 4: 242–254. doi:10.1016/j.molonc.2010.04.002.
3. Esteller M, Corn PG, Baylin SB, Herman JG (2001) A gene hypermethylation profile of human cancer. Cancer Res 61: 3225–3229.
4. Esteller M, Herman JG (2002) Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. J Pathol 196: 1–7. doi:10.1002/path.1024.
5. Portela A, Esteller M (2010) Epigenetic modifications and human disease. Nat Biotech 28: 1057–1068. doi:10.1038/nbt.1685.
6. Veeck J, Esteller M (2010) Breast cancer epigenetics: from DNA methylation to microRNAs. J Mammary Gland Biol Neoplasia 15: 5–17. doi:10.1007/s10911-010-9165-1.

7. Cho YH, Yazici H, Wu HC, Terry MB, Gonzalez K, et al. (2010) Aberrant promoter hypermethylation and genomic hypomethylation in tumor, adjacent normal tissues and blood from breast cancer patients. Anticancer Res 30: 2489–2496.
8. Choi JY, James SR, Link PA, McCann SE, Hong CC, et al. (2009) Association between global DNA hypomethylation in leukocytes and risk of breast cancer. Carcinogenesis 30: 1889–1897. doi:10.1093/carcin/bgp143.
9. Lim U, Flood A, Choi S, Albanes D, Cross AJ, et al. (2008) Genomic Methylation of Leukocyte DNA in Relation to Colorectal Adenoma Among Asymptomatic Women. Gastroenterology 134: 47–55. doi:10.1053/j.gastro.2007.10.013.
10. Pufulete M, Al-Ghnaniem R, Leather AJM, Appleby P, Gout S, et al. (2003) Folate status, genomic DNA hypomethylation, and risk of colorectal adenoma and cancer: a case control study. Gastroenterology 124: 1240–1248.
11. Hou L (2010) Blood leukocyte DNA hypomethylation and gastric cancer risk in a high-risk Polish population. International Journal of Cancer Journal International Du Cancer 127: 1866–1874. doi:10.1002/ijc.25190.
12. Hsiung DT, Marsit CJ, Houseman EA, Eddy K, Furniss CS, et al. (2007) Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. Cancer Epidemiol Biomarkers Prev 16: 108–114. doi:10.1158/1055-9965.EPI-06-0636.
13. Moore LE, Pfeiffer RM, Poscablo C, Real FX, Kogevinas M, et al. (2008) Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the Spanish Bladder Cancer Study: a case-control study. Lancet Oncol 9: 359–366. doi:10.1016/S1470-2045(08)70038-X.
14. Brennan K, Flanagan JM (2012) Is there a link between genome-wide hypomethylation in blood and cancer risk? Cancer Prev Res (Phila) 5: 1345–1357. doi:10.1158/1940-6207.CAPR-12-0316.
15. Xu X, Gammon MD, Hernandez-Vargas H, Herceg Z, Wetmur JG, et al. (2012) DNA methylation in peripheral blood measured by LUMA is associated with breast cancer in a population-based study. FASEB J 26: 2657–2666. doi:10.1096/fj.11-197251.
16. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16: 6–21. doi:10.1101/gad.947102.
17. Barouki R, Gluckman PD, Grandjean P, Hanson M, Heindel JJ (2012) Developmental origins of non-communicable disease: implications for research and public health. Environ Health 11: 42. doi:10.1186/1476-069X-11-42.
18. Terry MB, Ferris JS, Pilsner R, Flom JD, Tehranifar P, et al. (2008) Genomic DNA Methylation among Women in a Multiethnic New York City Birth Cohort. Cancer Epidemiology Biomarkers & Prevention 17: 2306–2310. doi:10.1158/1055-9965.EPI-08-0312.
19. Terry MB, Delgado-Cruzata L, Vin-Raviv N, Wu HC, Santella RM (2011) DNA methylation in white blood cells: association with risk factors in epidemiologic studies. Epigenetics 6: 828–837.
20. Scoccianti C, Ricceri F, Ferrari P, Cuenin C, Sacerdote C, et al. (2011) Methylation patterns in sentinel genes in peripheral blood cells of heavy smokers: Influence of cruciferous vegetables in an intervention study. epigenetics 6: 1114–1119. doi:10.4161/epi.6.9.16515.
21. Herceg Z (2007) Epigenetics and cancer: towards an evaluation of the impact of environmental and dietary factors. Mutagenesis 22: 91–103. doi:10.1093/mutage/gel068.
22. Zhang FF, Morabia A, Carroll J, Gonzalez K, Fulda K, et al. (2011) Dietary patterns are associated with levels of global genomic DNA methylation in a cancer-free population. J Nutr 141: 1165–1171. doi:10.3945/jn.110.134536.
23. Piyathilake C, Badiga S, Johanning G, Alvarez R, Partridge E (2011) Predictors and Health Consequences of Epigenetic Changes Associated with Excess Body Weight in Women of Child-bearing Age. Cancer Epidemiol Biomarkers Prev 20: 719–719. doi:10.1158/1055-9965.EPI-11-0094.
24. Christensen BC, Kelsey KT, Zheng S, Houseman EA, Marsit CJ, et al. (2010) Breast Cancer DNA Methylation Profiles Are Associated with Tumor Size and Alcohol and Folate Intake. PLoS Genet 6: e1001043. doi:10.1371/journal.pgen.1001043.
25. Teegarden D, Romieu I, Lelièvre SA (2012) Redefining the impact of nutrition on breast cancer incidence: is epigenetics involved? Nutr Res Rev 25: 68–95. doi:10.1017/S0954422411000199.
26. Widschwendter M, Jones PA (2002) DNA methylation and breast carcinogenesis. Oncogene 21: 5462–5482. doi:10.1038/sj.onc.1205606.
27. Karimi M, Johansson S, Stach D, Corcoran M, Grandér D, et al. (2006) LUMA (LUminometric Methylation Assay)–a high throughput method to the analysis of genomic DNA methylation. Exp Cell Res 312: 1989–1995. doi:10.1016/j.yexcr.2006.03.006.
28. Ehrlich M (2002) DNA methylation in cancer: too much, but also too little. Oncogene 21: 5400–5413. doi:10.1038/sj.onc.1205651.
29. Wilson AS, Power BE, Molloy PL (2007) DNA hypomethylation and human diseases. Biochim Biophys Acta 1775: 138–162. doi:10.1016/j.bbcan.2006.08.007.
30. Feber A, Wilson GA, Zhang L, Presneau N, Idowu B, et al. (2011) Comparative methylome analysis of benign and malignant peripheral nerve sheath tumors. Genome Res 21: 515–524. doi:10.1101/gr.109678.110.
31. Flanagan JM, Cocciardi S, Waddell N, Johnstone CN, Marsh A, et al. (2010) DNA methylome of familial breast cancer identifies distinct profiles defined by mutation status. Am J Hum Genet 86: 420–433. doi:10.1016/j.ajhg.2010.02.008.
32. Starlard-Davenport A, Tryndyak VP, James SR, Karpf AR, Latendresse JR, et al. (2010) Mechanisms of epigenetic silencing of the Rassf1a gene during estrogen-induced breast carcinogenesis in ACI rats. Carcinogenesis 31: 376–381. doi:10.1093/carcin/bgp304.
33. Tanner JM (1973) Trend towards Earlier Menarche in London, Oslo, Copenhagen, the Netherlands and Hungary., Published online: 11 May 1973; | doi:101038/243095a0 243: 95–96. doi:10.1038/243095a0.
34. Ahmed ML, Ong KK, Dunger DB (2009) Childhood obesity and the timing of puberty. Trends Endocrinol Metab 20: 237–242. doi:10.1016/j.tem.2009.02.004.
35. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, et al. (2002) European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr 5: 1113–1124. doi:10.1079/PHN2002394.
36. Karimi M, Johansson S, Ekström TJ (2006) Using LUMA: a Luminometric-based assay for global DNA-methylation. Epigenetics 1: 45–48.
37. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, et al. (2008) Intra-individual change over time in DNA methylation with familial clustering. JAMA 299: 2877–2883. doi:10.1001/jama.299.24.2877.
38. Triche TJ, Weisenberger DJ, Berg DVD, Laird PW, Siegmund KD (2013) Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucl Acids Res. Available: http://nar.oxfordjournals.org/content/early/2013/03/09/nar.gkt090. Accessed 2013 May 11.
39. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, et al. (2011) Evaluation of the Infinium Methylation 450 K technology. Epigenomics 3: 771–784. doi:10.2217/epi.11.105.
40. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, et al. (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 11: 733–739. doi:10.1038/nrg2825.
41. Bock C (2012) Analysing and interpreting DNA methylation data. Nature Reviews Genetics 13: 705–719. doi:10.1038/nrg3273.
42. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, et al. (2010) Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 11: 587. doi:10.1186/1471-2105-11-587.
43. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, et al. (2012) 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. doi:10.1093/hmg/dds488.