ARTICLE

Epidemiology and population health

Epigenome-wide association study of adiposity and future risk of obesity-related diseases

Gianluca Campanella[1] · Marc J. Gunter[2] · Silvia Polidoro[3] · Vittorio Krogh[4] · Domenico Palli [5] · Salvatore Panico[6] ·
Carlotta Sacerdote[3,7] · Rosario Tumino[8] · Giovanni Fiorito[3,9] · Simonetta Guarrera[3,9] · Licia Iacoviello[10] ·
Ingvar A. Bergdahl[11] · Beatrice Melin[12] · Per Lenner[12] · Theo M. C. M. de Kok[13] · Panagiotis Georgiadis[14] ·
Jos C. S. Kleinjans[13] · Soterios A. Kyrtopoulos [14] · H. Bas Bueno-de-Mesquita[1,15,16,17] · Karen A. Lillycrop[18] ·
Anne M. May[19] · N. Charlotte Onland-Moret[19] · Robert Murray[18] · Elio Riboli[1,20] · Monique Verschuren[21] ·
Eiliv Lund[22] · Nicolle Mode[22] · Torkjel M. Sandanger[22] · Valentina Fiano[23] · Morena Trevisan[23] · Giuseppe Matullo[3,9] ·
Philippe Froguel [24,25] · Paul Elliott [1,20] · Paolo Vineis[1,3,20] · Marc Chadeau-Hyam [1,20]

Abstract

Background Obesity is an established risk factor for several common chronic diseases such as breast and colorectal cancer, metabolic and cardiovascular diseases; however, the biological basis for these relationships is not fully understood. To explore the association of obesity with these conditions, we investigated peripheral blood leucocyte (PBL) DNA methylation markers for adiposity and their contribution to risk of incident breast and colorectal cancer and myocardial infarction.

Methods DNA methylation profiles (Illumina Infinium® HumanMethylation450 BeadChip) from 1941 individuals from four population-based European cohorts were analysed in relation to body mass index, waist circumference, waist-hip and waist-height ratio within a meta-analytical framework. In a subset of these individuals, data on genome-wide gene expression level, biomarkers of glucose and lipid metabolism were also available. Validation of methylation markers associated with all adiposity measures was performed in 358 individuals. Finally, we investigated the association of obesity-related methylation marks with breast, colorectal cancer and myocardial infarction within relevant subsets of the discovery population.

Results We identified 40 CpG loci with methylation levels associated with at least one adiposity measure. Of these, one CpG locus (cg06500161) in ABCG1 was associated with all four adiposity measures ($P = 9.07 \times 10^{-8}$ to 3.27×10^{-18}) and lower transcriptional activity of the full-length isoform of ABCG1 ($P = 6.00 \times 10^{-7}$), higher triglyceride levels ($P = 5.37 \times 10^{-9}$) and higher triglycerides-to-HDL cholesterol ratio ($P = 1.03 \times 10^{-10}$). Of the 40 informative and obesity-related CpG loci, two (in IL2RB and FGF18) were significantly associated with colorectal cancer (inversely, $P < 1.6 \times 10^{-3}$) and one intergenic locus on chromosome 1 was inversely associated with myocardial infarction ($P < 1.25 \times 10^{-3}$), independently of obesity and established risk factors.

Conclusion Our results suggest that epigenetic changes, in particular altered DNA methylation patterns, may be an intermediate biomarker at the intersection of obesity and obesity-related diseases, and could offer clues as to underlying biological mechanisms.

Joint first authors: Gianluca Campanella, Marc J. Gunter, Silvia Polidoro.

Joint last authors: Paul Elliott, Paolo Vineis, Marc Chadeau-Hyam.

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41366-018-0064-7) contains supplementary material, which is available to authorized users.

✉ Marc Chadeau-Hyam
m.chadeau@imperial.ac.uk

Extended author information available on the last page of the article

Introduction

Obesity is a growing global public health concern with an estimated 1.9 billion adults classified as overweight or obese worldwide [1]. Obesity is a risk factor for a variety of chronic diseases including type 2 diabetes, cardiovascular disease (CVD) and a number of malignancies such as postmenopausal breast, colorectal, liver, kidney and endometrial cancer [2, 3]. Obesity-associated metabolic abnormalities such as hyperinsulinemia, hyperglycaemia
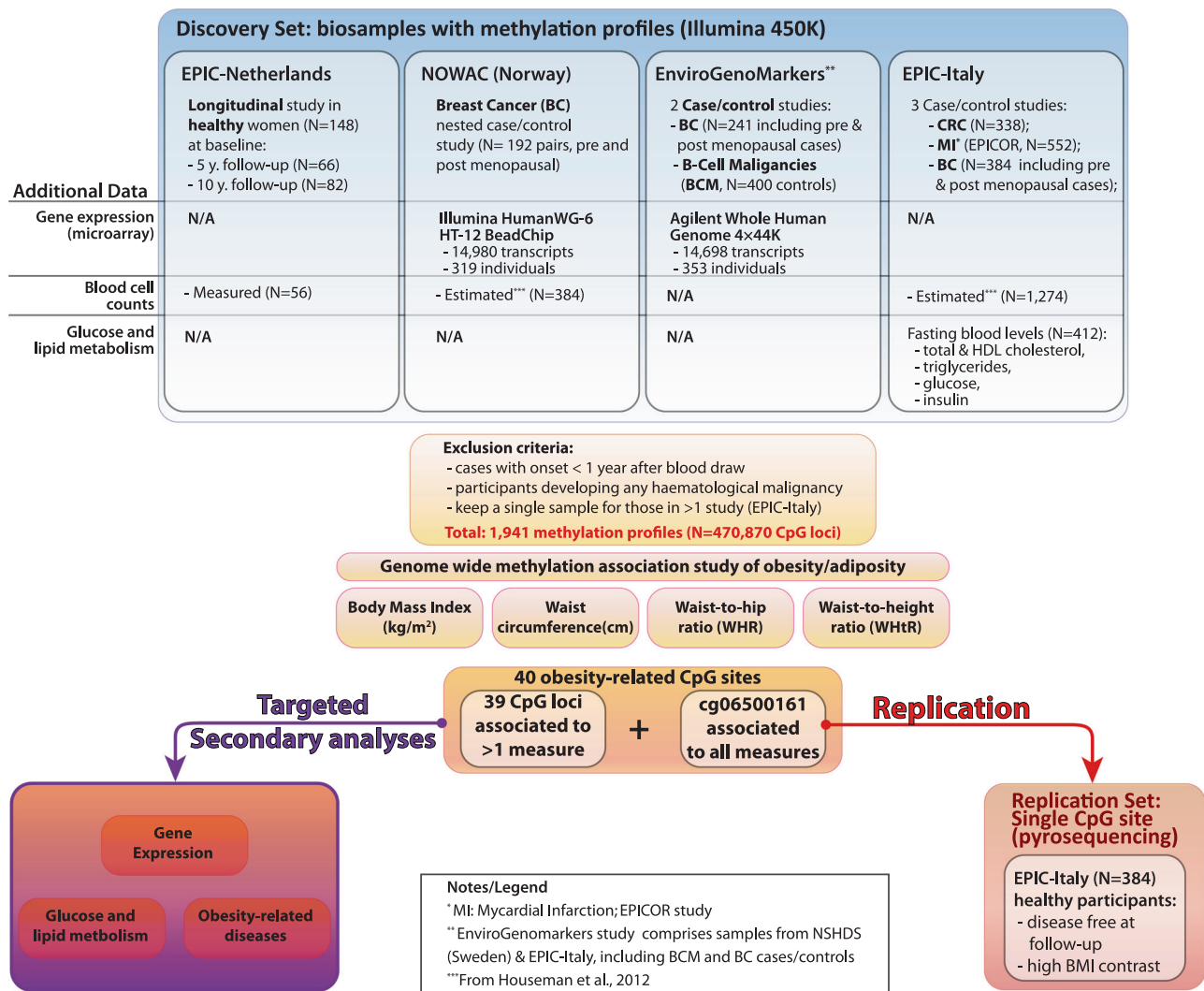
Discovery Set: biosamples with methylation profiles (Illumina 450K)

	EPIC-Netherlands	NOWAC (Norway)	EnviroGenoMarkers[**]	EPIC-Italy
	Longitudinal study in **healthy** women (N=148) at baseline: - 5 y. follow-up (N=66) - 10 y. follow-up (N=82)	**Breast Cancer (BC)** nested case/control study (N= 192 pairs, pre and post menopausal)	2 **Case/control** studies: - **BC** (N=241 including pre & post menopausal cases) - **B-Cell Malignacies** (**BCM**, N=400 controls)	3 Case/control studies: - **CRC** (N=338); - **MI**[*] (EPICOR, N=552); - **BC** (N=384 including pre & post menopausal cases);
Additional Data Gene expression (microarray)	N/A	**Illumina HumanWG-6 HT-12 BeadChip** - 14,980 transcripts - 319 individuals	**Agilent Whole Human Genome 4×44K** - 14,698 transcripts - 353 individuals	N/A
Blood cell counts	- Measured (N=56)	- Estimated[***] (N=384)	N/A	- Estimated[***] (N=1,274)
Glucose and lipid metabolism	N/A	N/A	N/A	Fasting blood levels (N=412): - total & HDL cholesterol, - triglycerides, - glucose, - insulin

Exclusion criteria:
- cases with onset < 1 year after blood draw
- participants developing any haematological malignancy
- keep a single sample for those in >1 study (EPIC-Italy)

Total: 1,941 methylation profiles (N=470,870 CpG loci)

Genome wide methylation association study of obesity/adiposity

Body Mass Index (kg/m²)	Waist circumference(cm)	Waist-to-hip ratio (WHR)	Waist-to-height ratio (WHtR)

40 obesity-related CpG sites

39 CpG loci associated to >1 measure	+	**cg06500161 associated to all measures**

Targeted Secondary analyses

Replication

Gene Expression

Glucose and lipid metbolism | **Obesity-related diseases**

Replication Set: Single CpG site (pyrosequencing)

EPIC-Italy (N=384) healthy participants:
- disease free at follow-up
- high BMI contrast

Notes/Legend
[*] MI: Mycardial Infarction; EPICOR study
[**] EnviroGenomarkers study comprises samples from NSHDS (Sweden) & EPIC-Italy, including BCM and BC cases/controls
[***]From Houseman et al., 2012

Fig. 1 Schematic representation of the study population and statistical analyses performed

and dyslipidaemia have also been associated with CVD and a number of malignancies [4]. However, the contribution of obesity and its physiological consequences to the development of these diseases remains poorly understood.

The aetiology of obesity is multifactorial, with both genetic and environmental determinants. Numerous genetic variants linked to obesity have been discovered by large meta-analyses of genome-wide association studies [5–7]; however, taken together, these variants explain only a modest fraction of the phenotypic variation, and most of them do not appear to directly modulate phenotypes potentially driving individual risk profiles for obesity-related diseases [8]. It has been hypothesised that adiposity influences DNA methylation and the identification of genetic loci that are epigenetically modified in obesity could help further understanding on the development of obesity and its pathophysiologic sequelae. Recently, several

epigenome-wide association studies have been published that have identified several hundred CpG sites that are statistically significantly associated with body mass index (BMI) [9] and with metabolic disturbances and subsequent risk of type 2 diabetes [10, 11]. To our knowledge, no studies have investigated the association of adiposity-related CpG loci and subsequent risk of other obesity-related diseases such as myocardial infarction, breast and colorectal cancer.

In this analysis, we obtained genome-wide DNA methylation levels in peripheral blood leucocytes (PBLs) from over 1900 individuals from four prospective cohorts in Italy, the Netherlands, Norway and Sweden. Using a meta-analytical framework, we tested the association of DNA methylation levels at each CpG locus with four measures of adiposity and investigated their association with transcriptomic profiles and subsequent development of myocardial infarction, breast and colorectal cancer.

Methods

Features of the study population and the overall analytical strategy are summarised in Fig. 1.

Discovery set

The study included participants from four large population-based cohorts: the Italian and Dutch components of the European Prospective Investigation into Cancer and Nutrition (EPIC-Italy, $N = 47{,}749$ [12] and EPIC-Netherlands, $N = 33{,}066$ [13]), the Norwegian Women and Cancer Study (NOWAC, $N = 53{,}363$) [14], and the Västerbotten Intervention Program (VIP) part of the Northern Sweden Health and Disease Study (NSHDS, $N = 101{,}790$) [15].

A total of 2560 genome-wide DNA methylation profiles from PBLs were available as part of:

(i) Three separate prospective case-control studies nested within EPIC-Italy on breast cancer ($N = 332$), colorectal cancer ($N = 338$), and myocardial infarction ($N = 552$);

(ii) Two separate prospective nested case-control from EPIC-Italy and NSHDS on breast cancer and B-cell malignancies (EPIC-Italy and NSHDS components of the EnviroGenoMarkers project, $N = 241$ and 400, respectively).

(iii) A longitudinal study in healthy women nested within EPIC-Netherlands ($N = 148$ pairs of baseline and follow-up samples, of which $N = 66$ at 5 years and $N = 82$ at 10 years from baseline).

(iv) A prospective case-control study nested within NOWAC on breast cancer ($N = 384$).

NOWAC study was approved by the Regional Committee for Medical and Health Research Ethics and the Norwegian Data Inspectorate. The EnviroGenomarkers project and its associated studies (EPIC-Italy and NHSDS) and experimental protocols were approved by the Regional Ethical Review Board of the Umeå Division of Medical Research, for NHSDS, and the Florence Health Unit Local Ethical Committee, for EPIC Italy, and by local review board for EPIC-NL. Other studies involving EPIC-Italy samples were approved by the HuGeF Ethics, Committee. All participants gave written informed consent.

For the EPIC participants who contributed to more than one case-control study, multiple DNA methylation profiles were available. We only retained the single profile that exhibited the best technical quality (based on control probes). Participants from prospective nested case-control studies who developed the relevant condition less than 1 year after blood draw (to guard against reverse causation), or who developed any kind of haematological malignancy at any time after enrolment, were excluded, as were those

with no anthropometric data. Characteristics of the resulting 1941 participants (588 men and 1353 women) are reported in Supplementary Table 1. Cancer cases were ascertained by linkage to national cancer registries, and myocardial infarction cases were confirmed by a cardiologist based on clinical records. Across all studies, completeness of follow-up was close to 100%, and cancer cases were histologically confirmed. Healthy controls were matched to cases by gender, time since blood sampling, and year of birth in order to control for effects of blood storage time and ageing.

Replication set

In order to validate the adiposity−methylation associations identified in the discovery set, we defined a replication sample ($N = 384$) from EPIC-Italy (independent of those individuals tested in the discovery sample). Samples exclusion criterion included: (i) prevalent dyslipidaemia or diabetes (self-reported); (ii) diagnosis of any non-haematological malignancy less than 5 years after blood draw; (iii) diagnosis of any haematological malignancy at any time after enrolment. Adiposity was assessed using a combination of BMI and waist circumference. To maximise contrast in obesity metrics and explore possible departure from a linear dose−response curve linking obesity and methylation levels, at high levels of obesity, we oversampled obese participants and randomly selected from the obese group 288 (75%) participants. The remaining 96 (25%) participants were sampled from the reference, normal-weight group. Samples that did not pass quality control procedures, as well as participants with incomplete anthropometric information, were excluded from further analyses, leaving a total of 358 participants (127 men and 231 women) whose characteristics are reported in Supplementary Table 2.

Adiposity parameters

Four adiposity measures were considered: BMI (kg/m^2), waist circumference (cm), waist to hip ratio (WHR), and waist to height ratio (WHtR). Anthropometric measurements concurrent with blood sampling were obtained by a trained nurse at enrolment for all EPIC and NSHDS participants and were self-reported for NOWAC. Waist circumference and related measures were only available for EPIC participants.

Biochemical and biological parameters

Fasting blood levels of circulating total and HDL cholesterol, triglycerides, glucose, and insulin were available in a subset of non-diabetic EPIC-Italy participants (independent of the samples used in the replication set) who declared not

to be taking lipid-lowering medications ($N = 412$), and for whom DNA methylation profiles were also available. Differential blood counts were available for a subset of EPIC-Netherlands participants at baseline ($N = 56$).

Laboratory analyses and data pre-processing

DNA methylation analyses

Genome-wide DNA methylation profiles were obtained using the Illumina Infinium® HumanMethylation450 (HM450) BeadChip assaying 470,870 autosomal CpG sites across the genome. Genomic DNA extracted from PBLs was bisulphite-converted using the Zymo Research EZ-96 DNA Methylation-Gold™ Kit, and hybridised to HM450 Bead-Chips according to the manufacturer's protocol. Microarrays were subsequently scanned using the Illumina HiScanSQ system, and raw intensity data were exported from Illumina GenomeStudio (version 2011.1). Control probes included in the microarray were used to assess bisulphite conversion efficiency and to exclude lower-quality samples from further analyses (probes with detection P-values greater than 0.05 were excluded). Data pre-processing was performed using in-house software written for the R statistical computing environment. In particular, for each sample and each probe, measurements were set to missing if obtained by averaging intensities over less than three beads, or if averaged intensities were below detection thresholds estimated from negative control probes. Background subtraction and dye bias correction (for probes using the Infinium II design) were also performed. DNA methylation levels at each locus were assumed to be proportional to the ratio of intensities arising from methylated cytosine residues over total intensities. Details of the DNA-methylation pre-processing procedures are given in the Appendix. DNA methylation levels in the replication sample were assessed using bisulphite pyrosequencing. Primers were designed using QIAGEN Pyro-Mark Assay Design (version 2.0). Preliminary PCR reactions were performed using the following cycling protocol: 95 °C for 10 min, followed by 45 cycles of denaturation at 95 °C for 30 s, annealing at 58 °C for 1 min, and extension at 72 °C for 1 min; a final extension was performed at 72 °C for 10 min. All samples were analysed two or three times (following bisulphite conversion) to assess and account for technical variability.

Gene expression analyses

Blood samples from the EnviroGenoMarkers and NOWAC studies additionally underwent genome-wide gene expression profiling using the Agilent Whole Human Genome 4×44 K Microarray [16], and the Illumina HumanWG-6 and HumanHT-12 BeadChip (NOWAC), respectively, as previously reported [17, 18]. Samples that failed quality control procedures were excluded from further analyses, leaving a total of 672 genome-wide gene expression profiles ($N = 353$ in EnviroGenoMarkers, and $N = 319$ in NOWAC) with associated DNA methylation data. For both studies, pre-processing steps included within- and between-array normalisation, followed by removal of un-annotated and non-specific probes, leaving a total of 14,698 transcripts for EnviroGenoMarkers, and 14,980 transcripts for NOWAC.

Statistical analyses

In the discovery set, separate analyses were conducted for each study and each adiposity measure. DNA methylation levels at each assayed CpG locus were modelled as the dependent variable in a generalised linear model accommodating beta-distributed responses [19], and for paired samples from EPIC-Netherlands, in a linear mixed model. All models were adjusted for microarray and position on the microarray, sex (if applicable), age at blood draw, and case-control status (if applicable). For each adiposity measure, estimated effect sizes (on the *logit* scale) and associated standard errors were obtained. Results across studies were combined in a meta-analytic framework using fixed effects models with inverse-variance weighting. For each CpG locus the model estimated, effect sizes, standard errors and corresponding P values as well as the I^2 heterogeneity statistic [20]. Associations were declared statistically significant based on their meta-analytic P value with a 5% Bonferroni-corrected significance threshold $\alpha = 0.05 / 470,870 \approx 1.06 \times 10^{-7}$. Sensitivity analyses included stratification by sex, adopting a 5% significance threshold after Bonferroni correction for the number of informative CpG loci identified in the meta-analysis for each adiposity measure. DNA methylation measurements in the replication sample consisted of $2-3$ technical replicates per participant and all were analysed together. In order to capture the possible technical variation within each participant, we used a linear mixed model with a participant random effect and adjusted for the same technical covariates as described above. CpG loci identified by the meta-analysis were only retained if they were not reported in the literature as cross-hybridising probes [21].

Informative CpG loci from the meta-analysis were further characterised by relating DNA methylation levels to transcriptional activity of the closest gene in EnviroGenoMarkers and NOWAC, using a $\log_2$-linear model adjusted for technical confounders, cohort and sex (EnviroGenoMarkers data only), age, case-control status, and BMI. As previously reported, linear models and linear mixed models (including a random intercept dependent on the main technical confounders) were used for the NOWAC and EnviroGenoMarkers data, respectively [16, 22].

Table 1 Associations between DNA methylation levels and BMI (A), waist circumference (B), WHR (C) and WHtR (D) identified in the discovery set.

	Chr	Location	Gene	N	β	SE(β)	P value	I^2
(A) BMI (kg/m^2)								
cg23172671	1	203,482,523		2072	9.46×10^{-3}	1.76×10^{-3}	7.80×10^{-8}	19.67%
cg16258657	3	29,338,542	*RBMS3*	2078	-8.31×10^{-3}	1.54×10^{-3}	6.59×10^{-8}	0.00%
cg15059608	3	62,305,003	*C3orf14*	2080	8.67×10^{-3}	1.58×10^{-3}	3.94×10^{-8}	30.98%
cg14492061	3	63,637,129	*SNTN*	2071	-9.42×10^{-3}	1.66×10^{-3}	1.36×10^{-8}	57.11%
cg18098839	3	167,742,700	*GOLIM4*	2085	-6.53×10^{-3}	1.16×10^{-3}	1.72×10^{-8}	0.00%
cg09831562	3	181,327,125	*SOX2OT*	2053	1.33×10^{-2}	2.13×10^{-3}	3.94×10^{-10}	0.00%
cg06164260	3	187,454,439	*BCL6*	2085	-4.03×10^{-3}	7.09×10^{-4}	1.27×10^{-8}	14.59%
cg11661512	5	170,864,692	FGF18	2083	-4.75×10^{-3}	8.12×10^{-4}	5.03×10^{-9}	60.48%
cg01538605	6	169,285,957		2085	-5.85×10^{-3}	9.84×10^{-4}	2.79×10^{-9}	0.00%
cg06087826	6	170,582,184		2083	-5.78×10^{-3}	1.05×10^{-3}	3.78×10^{-8}	11.59%
cg13286614	7	96,655,501	*DLX5*	2084	7.28×10^{-3}	1.28×10^{-3}	1.20×10^{-8}	54.63%
cg03619256	7	105,515,219	*ATXN7L1*	2085	-5.49×10^{-3}	8.24×10^{-4}	2.79×10^{-11}	0.00%
cg16797901	7	111,513,558	*DOCK4*	2061	-9.32×10^{-3}	1.73×10^{-3}	6.90×10^{-8}	29.37%
cg19249811	10	29,747,315	*SVIL*	1982	-3.23×10^{-2}	3.77×10^{-3}	1.00×10^{-17}	94.82%
cg17061862	11	9,590,431		2084	-7.23×10^{-3}	1.18×10^{-3}	9.16×10^{-10}	0.00%
cg00574958	11	68,607,622	*CPT1A*	2083	-7.62×10^{-3}	1.41×10^{-3}	6.15×10^{-8}	0.00%
cg27117792	12	102,330,180		2083	-9.90×10^{-3}	1.81×10^{-3}	4.29×10^{-8}	53.99%
cg19750657	13	38,935,967	*UFM1*	2083	8.25×10^{-3}	1.23×10^{-3}	2.15×10^{-11}	0.00%
cg02119938	15	78,505,051	*ACSBG1*	2071	-1.33×10^{-2}	2.48×10^{-3}	9.07×10^{-8}	54.49%
cg01115923	16	57,793,728	*KIFC3*	2085	-6.07×10^{-3}	1.00×10^{-3}	1.42×10^{-9}	20.33%
cg11024682	17	17,730,094	*SREBF1*	2085	5.36×10^{-3}	7.64×10^{-4}	2.29×10^{-12}	21.98%
cg24174557	17	57,903,544	*TMEM49*	2077	-8.36×10^{-3}	1.49×10^{-3}	2.14×10^{-8}	0.00%
cg04927537	17	76,976,091	*LGALS3BP*	2085	6.71×10^{-3}	1.26×10^{-3}	9.89×10^{-8}	54.38%
cg26950531	19	38,704,515	*DPF1*	2071	-1.44×10^{-2}	2.06×10^{-3}	3.40×10^{-12}	0.00%
cg27243685	21	43,642,366	*ABCG1*	2085	6.88×10^{-3}	1.18×10^{-3}	5.04×10^{-9}	48.25%
cg06500161	21	43,656,587	*ABCG1*	2085	6.88×10^{-3}	7.89×10^{-4}	2.76×10^{-18}	53.62%
(B) Waist circumference (cm)								
cg09831562	3	181,327,125	*SOX2OT*	1501	6.19×10^{-3}	1.02×10^{-3}	1.13×10^{-9}	0.00%
cg01538605	6	169,285,957		1531	-2.47×10^{-3}	4.63×10^{-4}	1.04×10^{-7}	0.00%
cg03619256	7	105,515,219	*ATXN7L1*	1531	-2.24×10^{-3}	3.91×10^{-4}	9.54×10^{-9}	0.00%
cg11376147	11	57,261,198	*SLC43A1*	1530	-2.20×10^{-3}	3.90×10^{-4}	1.53×10^{-8}	0.00%
cg19750657	13	38,935,967	*UFM1*	1530	3.38×10^{-3}	5.92×10^{-4}	1.16×10^{-8}	59.96%
cg07037944	15	64,290,807	*DAPK2*	1524	-2.73×10^{-3}	4.63×10^{-4}	3.65×10^{-9}	10.86%
cg04583842	16	88,103,117	*BANP*	1515	3.16×10^{-3}	5.84×10^{-4}	5.90×10^{-8}	0.00%
cg11024682	17	17,730,094	*SREBF1*	1531	1.99×10^{-3}	3.59×10^{-4}	3.01×10^{-8}	0.00%
cg24174557	17	57,903,544	*TMEM49*	1523	-3.94×10^{-3}	7.13×10^{-4}	3.26×10^{-8}	0.00%
cg26950531	19	38,704,515	*DPF1*	1520	-5.60×10^{-3}	1.00×10^{-3}	2.44×10^{-8}	50.19%
cg06500161	21	43,656,587	*ABCG1*	1531	3.17×10^{-3}	3.68×10^{-4}	8.36×10^{-18}	49.58%
cg21307484	22	37,546,220	*IL2RB*	1531	2.35×10^{-3}	4.19×10^{-4}	2.07×10^{-8}	0.00%
(C) WHR								
cg17901584	1	55,353,706	*DHCR24*	1531	-5.79×10^{-1}	1.07×10^{-1}	5.53×10^{-8}	0.00%
cg16246545	1	120,255,941	*PHGDH*	1531	-4.39×10^{-1}	7.49×10^{-2}	4.44×10^{-9}	0.00%
cg14476101	1	120,255,992	*PHGDH*	1527	-6.83×10^{-1}	1.02×10^{-1}	2.38×10^{-11}	0.00%
cg19472611	3	48,885,488	*PRKAR2A*	1513	-6.84×10^{-1}	1.27×10^{-1}	6.70×10^{-8}	12.56%

Table 1 (continued)

	Chr	Location	Gene	N	β	SE(β)	P value	I^2
cg25799109	3	57,102,900	*ARHGEF3 SPATA12*	1519	-6.08×10^{-1}	1.11×10^{-1}	4.14×10^{-8}	79.04%
cg20898587	7	51,969,244		1530	-3.08×10^{-1}	5.64×10^{-2}	4.60×10^{-8}	63.03%
cg06730756	8	142,198,993	*DENND3*	1478	-1.41	2.62×10^{-1}	8.25×10^{-8}	84.90%
cg11376147	11	57,261,198	*SLC43A1*	1530	-3.32×10^{-1}	5.91×10^{-2}	1.87×10^{-8}	41.25%
cg06500161	21	43,656,587	*ABCG1*	1531	3.65×10^{-1}	5.49×10^{-2}	3.07×10^{-11}	62.98%
(D) WHtR								
cg12593793	1	156,074,135		1530	-4.37×10^{-1}	8.21×10^{-2}	1.01×10^{-7}	0.00%
cg09831562	3	181,327,125	*SOX2OT*	1501	9.34×10^{-1}	1.61×10^{-1}	6.69×10^{-9}	0.00%
cg06164260	3	187,454,439	*BCL6*	1531	-2.75×10^{-1}	5.03×10^{-2}	4.59×10^{-8}	0.00%
cg03619256	7	105,515,219	*ATXN7L1*	1531	-3.79×10^{-1}	6.16×10^{-2}	7.55×10^{-10}	28.48%
cg11376147	11	57,261,198	*SLC43A1*	1530	-3.53×10^{-1}	6.16×10^{-2}	9.43×10^{-9}	0.00%
cg07037944	15	64,290,807	*DAPK2*	1524	-4.09×10^{-1}	7.35×10^{-2}	2.58×10^{-8}	4.90%
cg11024682	17	17,730,094	*SREBF1*	1531	3.10×10^{-1}	5.68×10^{-2}	4.77×10^{-8}	11.53%
cg24174557	17	57,903,544	*TMEM49*	1523	-6.29×10^{-1}	1.12×10^{-1}	2.29×10^{-8}	37.14%
cg05091997	17	60,897,721		1513	-9.63×10^{-1}	1.75×10^{-1}	3.97×10^{-8}	0.00%
cg26950531	19	38,704,515	*DPF1*	1520	-1.04	1.59×10^{-1}	7.51×10^{-11}	17.36%
cg26470501	19	45,252,955	*BCL3*	1531	-3.21×10^{-1}	5.49×10^{-2}	5.10×10^{-9}	0.00%
cg06500161	21	43,656,587	*ABCG1*	1531	5.07×10^{-1}	5.83×10^{-2}	3.32×10^{-18}	19.08%

The table reports the total number of samples N included in the meta-analysis, the beta regression model coefficient β measuring the percentage change (on the logit scale) in DNA methylation per unit increase in each obesity measure, and the associated standard error (SE), the P value for H_0: $\beta = 0$, and the I^2 heterogeneity statistic

Associations were declared statistically significant at a nominal 5% significance threshold.

Fasting levels of circulating total and HDL cholesterol, triglycerides, glucose, and insulin, as well as two further indicators of cardiometabolic risk (total-to-HDL cholesterol ratio and triglycerides-to-HDL cholesterol ratio) [23], and two indicators of β-cell function and insulin resistance (HOMA-%B and HOMA-IR [24]), were measured in the EPICOR study ($N = 412$, see Fig. 1). These were analysed in relation to DNA methylation levels at informative CpG loci using log-linear models adjusted for sex, age, case-control status (separately for cancers and myocardial infarction), BMI and WHR.

We also investigated the potential for bias induced by differential blood cell counts adjusting for measured counts that were available in 56 EPIC-Netherlands study participants. These analyses were performed for our informative CpG loci using linear models adjusted for age, and were corrected for multiple comparisons using a 5% Bonferroni corrected significance level based on the total number of informative CpG loci identified in the meta-analysis. In addition, we used the 'double bootstrap' approach by Houseman et al. [25] and DNA methylation profiles in purified leukocytes made available by Reinus et al. [26] to infer associations between adiposity measures and blood differentials estimated from DNA methylation data. Adiposity-related CpG loci were then analysed with respect to incident post-menopausal breast cancer (from EPIC-Italy, EnviroGenoMarkers, and NOWAC studies $N = 283$ cases/ 282 controls, see Fig. 1), colorectal cancer (from EPIC-Italy CRC study $N = 171$ cases/132 controls, see Fig. 1), and myocardial infarction (from EPICOR study $N = 131$ cases/ 195 controls, see Fig. 1) in a logistic regression model for case-control status adjusted for age, sex (except post-menopausal breast cancer), BMI and WHR, as well as disease-specific confounders (Supplementary Table 3).

Data and code availability

Most of the microarray data are already publicly available (accession numbers available upon request to the corresponding author). The remaining data could be accessed upon request to the originating cohorts. Access will be conditional to adherence to local ethical and security policy. R-codes used for the analyses presented in the paper are available upon request.

Results

Epigenome-wide association analysis of adiposity

The meta-analysis of studies in the discovery set identified a list of 40 unique informative CpG loci associated with one

Table 2 Associations of DNA methylation levels at informative CpG loci with *cis* transcriptional activity in: (A) EnviroGenoMarkers data ($N = 353$); (B) NOWAC data ($N = 319$).

mRNA	Gene	CpG locus	Association with DNA methylation			Association with BMI		
			β	SE(β)	P value	β	SE(β)	P value
(A)								
NM_001031847	*CPT1A*	cg00574958	−3.72	1.46	1.15×10^{-2}	2.75×10^{-2}	7.53×10^{-3}	3.00×10^{-4}
NM_016617	*UFM1*	cg19750657	2.68	1.13	1.89×10^{-2}	8.70×10^{-3}	1.16×10^{-2}	4.52×10^{-1}
NM_001005291	*SREBF1*	cg11024682	−1.90	0.82	2.10×10^{-2}	-9.25×10^{-3}	7.45×10^{-3}	2.15×10^{-1}
NM_004915	*ABCG1*	cg27243685	−3.55	1.12	1.73×10^{-3}	-2.02×10^{-2}	7.96×10^{-3}	1.18×10^{-2}
		cg06500161	−2.81	0.94	2.93×10^{-3}			
NM_003627	*SLC43A1*	cg11376147	−3.61	1.65	2.89×10^{-2}	7.04×10^{-3}	9.53×10^{-3}	4.61×10^{-1}
NM_006623	*PHGDH*	cg16246545	−2.60	0.77	9.92×10^{-4}	-3.03×10^{-3}	8.76×10^{-3}	7.29×10^{-1}
		cg14476101	−2.03	0.60	8.49×10^{-4}			
NM_001128615	*ARHGEF3*	cg25799109	1.05	0.51	4.22×10^{-2}	2.48×10^{-3}	8.47×10^{-3}	7.70×10^{-1}
(B)								
NM_020685	*C3orf14*	cg15059608	2.29	0.53	1.74×10^{-5}	-2.31×10^{-3}	5.23×10^{-3}	6.59×10^{-1}
NM_001876	*CPT1A*	cg00574958	3.39	1.27	7.90×10^{-3}	3.12×10^{-3}	7.70×10^{-3}	6.85×10^{-1}
NM_001031847	*CPT1A*	cg00574958	2.37	0.93	1.15×10^{-2}	2.26×10^{-4}	5.65×10^{-3}	9.68×10^{-1}
NM_001005291	*SREBF1*	cg11024682	−2.99	1.41	3.47×10^{-2}	-3.63×10^{-2}	1.10×10^{-2}	1.08×10^{-3}
NM_004915	*ABCG1*	cg27243685	−6.21	1.54	6.79×10^{-5}	-5.25×10^{-2}	1.04×10^{-2}	7.38×10^{-7}
		cg06500161	−5.37	1.27	2.89×10^{-5}			
NM_014326	*DAPK2*	cg07037944	−3.06	1.14	7.68×10^{-3}	-3.27×10^{-2}	9.03×10^{-3}	3.40×10^{-4}
NM_006623	*PHGDH*	cg16246545	−3.74	0.84	1.06×10^{-5}	-1.59×10^{-2}	9.21×10^{-3}	8.45×10^{-2}
		cg14476101	−2.80	0.61	6.18×10^{-6}			

Associations between transcription levels of the identified transcripts and BMI are also presented. The tables report the log$_2$-linear regression model coefficient β measuring the gene expression change associated with a one percentage point increase in DNA methylation, the associated standard error (SE), and the P value for H_0: $\beta = 0$

or more adiposity measure ($P < 1.06 \times 10^{-7}$, Table 1). In particular, after correction for multiple testing, 26 CpG loci were statistically significantly associated with BMI (Table 1A, P values ranging between 9.89×10^{-8} and 2.76×10^{-18}), of which 8 are directly associated to BMI and 18 were characterised by low-to-moderate heterogeneity across studies ($I^2 < 50\%$). Analyses for waist circumference and derived measures identified fewer associations: 12 for waist circumference (including 6 direct associations, Table 1B), 9 for WHR (including only one direct association, Table 1C), and 12 for WHtR (including three direct associations, Table 1D). All 12 CpG loci associated with WHtR exhibited low-to-moderate heterogeneity across studies ($I^2 < 50\%$). Methylation levels at a single CpG locus (cg06500161, *ABCG1*) were significantly (and positively) associated with all four adiposity measures (P values ranging between 3.07×10^{-11} and 2.76×10^{-18}). Associations with DNA methylation levels at cg06500161 (*ABCG1*) were confirmed in the replication experiment (P values 5.61×10^{-10}, 8.17×10^{-11}, 7.68×10^{-9}, and 4.71×10^{-11} for BMI, waist circumference, WHR and WHtR, respectively).

DNA methylation measurements obtained using bisulphite pyrosequencing were additionally used to examine the shape of the dose−response curve of DNA methylation levels with each of the four adiposity measures; this showed linear direct relationships (Supplementary Figure 1). Additional stratification by sex did not yield any substantial differences in the results (Supplementary Table 4). However, associations appeared stronger in women than in men. Among men, 15 of the CpG loci associated with BMI did not reach statistical significance; however, seven (nominally statistically significant) showed effect size estimates similar to those obtained in the main analysis. This might at least partially be attributed to the smaller number of men included in our study ($N = 588$ men, vs. $N = 1353$ women).

Using measured blood differentials available in 56 study participants from EPIC-Netherlands, sensitivity analyses showed evidence of potential confounding induced by leucocyte high turnover rates and inter-individual tissue heterogeneity [25, 27] for only one of our 40 informative CpG loci, cg27117792 (intergenic). We additionally

Table 3 Associations of DNA methylation levels at ($N = 40$) informative CpG loci with fasting blood levels of total and HDL cholesterol, triglycerides, glucose (all in mmol/L), and insulin (µIU/mL) after adjustment for sex, age, case-control status (separately for cancers and myocardial infarction).

CpG locus	Gene	Biomarker	Without adjustment for BMI−WHR			With adjustment for BMI−WHR		
			β	SE(β)	P value	β	SE(β)	P value
cg00574958	*CPT1A*	Triglycerides	−0.11	0.03	8.18×10^{-5}	−0.11	0.03	3.59×10^{-4}
		Triglycerides/HDL cholesterol	−0.09	0.02	$7.65E\times10^{-6}$	−0.09	0.02	3.92×10^{-5}
cg27243685	*ABCG1*	Triglycerides	0.09	0.02	4.52×10^{-5}	0.07	0.02	1.09×10^{-3}
		HDL cholesterol	−0.14	0.04	1.05×10^{-3}	−0.11	0.04	1.08×10^{-2}
		Triglycerides/HDL cholesterol	0.07	0.02	1.21×10^{-5}	0.06	0.02	4.40×10^{-4}
		Glucose	0.20	0.06	1.09×10^{-3}	0.16	0.06	8.31×10^{-3}
cg06500161	*ABCG1*	HDL cholesterol	−0.15	0.03	6.00×10^{-8}	-0.12	0.03	8.87×10^{-6}
		Triglycerides	0.08	0.01	1.93×10^{-9}	0.07	0.01	3.61×10^{-6}
		Total/HDL cholesterol	0.12	0.03	4.37×10^{-6}	0.09	0.03	1.45×10^{-3}
		Triglycerides/HDL cholesterol	0.07	0.01	1.59×10^{-11}	0.06	0.01	8.45×10^{-8}
cg17901584	*DHCR24*	Triglycerides	−0.11	0.03	8.18×10^{-5}	−0.11	0.03	3.59×10^{-4}
		Triglycerides/HDL cholesterol	−0.11	0.02	1.09×10^{-6}	−0.11	0.02	5.21×10^{-6}
cg11024682	*SREBF1*	Triglycerides	0.05	0.01	2.74×10^{-4}	0.04	0.01	8.22×10^{-3}
		Triglycerides/HDL cholesterol	0.03	0.01	8.51×10^{-4}	0.03	0.01	2.11×10^{-2}
cg09831562	*SOX2OT*	HDL cholesterol	−0.35	0.08	1.41×10^{-5}	−0.32	0.08	1.01×10^{-4}

The table reports the beta regression model coefficient β measuring the percentage change in DNA methylation (on the logit scale) per one unit increase in blood concentrations of each biomarker, the associated standard error (SE), and the P value for H_0: $\beta = 0$, with and without adjustment for BMI and WHR

estimated blood cell composition using the 'double bootstrap' approach by Houseman et al. [25] in EPIC-Italy and NOWAC. We did not identify any associations between adiposity measures and blood differentials estimated from DNA methylation data (all P values > 0.05, see Supplementary Table 5). Furthermore, using results from the genome-wide meta-analyses from the GIANT consortium [5–7] we did not identify any SNP, with minor allele frequencies above 1% in the European population [28], and located within 500 kb windows centred at each informative CpG locus, associated with BMI or WHR (lowest reported P values 6.36×10^{-5} and 5.42×10^{-6} for BMI and WHR, respectively). We believe these data argue against potential genetic confounding. In addition, using genotype data available for 552 EPIC-Italy participants from the EPICOR study, we investigated associations between DNA methylation levels and 38 assayed SNPs that were also reported in the most recent GIANT meta-analysis, separately and combined in a (partial) genetic risk score. As illustrated in Supplementary Figure 2, we did not identify any significant associations, which also supports the lack of (non-local) genetic confounding.

Associations with transcriptional activity

Among the 40 informative CpG loci, six exhibited DNA methylation levels that were consistently associated with transcriptional activity in PBLs for the nearest gene in both EnviroGenoMarkers and NOWAC data (Table 2). With the exception of DNA methylation levels at cg00574958, which were inversely associated with transcription levels of *CPT1A* in EnviroGenoMarkers, but positively in NOWAC, all such associations showed consistent directions in both studies. In particular, DNA methylation levels at cg11024682 were inversely associated with transcriptional activity of *SREBF1*, and DNA methylation levels at both pairs of informative CpG loci found in close proximity (cg27243685 and cg06500161 on *ABCG1*, and cg16246545 and cg14476101 on *PHGDH*) were consistently inversely associated with transcriptional activity of the corresponding genes. In addition, transcription levels of *ABCG1* were independently inversely associated with BMI ($P = 7.38\times10^{-7}$) in the NOWAC data; an association in the same direction was also found in the EnviroGenoMarkers data ($P = 1.18\times10^{-2}$).

Table 4 Associations between informative CpG sites and post-menopausal breast cancer (A), colorectal cancer (B) and myocardial infarction (C), for probes reaching nominal 0.05 significance level with adjustment for age, established risk factors for each disease, and BMI and WHR (Model 1).

(A)

Breast cancer	Model 1[a]				Model 2[b]		
	N	β	SE(β)	P value	β	SE(β)	P value
cg01115923	182	−0.05	0.02	8.03×10^{-3}	−0.05	0.02	5.96×10^{-3}
cg03619256	182	−0.04	0.02	1.86×10^{-2}	−0.05	0.02	1.12×10^{-2}
cg26470501	182	−0.04	0.02	2.13×10^{-2}	−0.04	0.02	1.15×10^{-2}
cg13286614	182	−0.07	0.03	2.17×10^{-2}	−0.07	0.03	2.12×10^{-2}
cg25799109	182	−0.08	0.04	2.99×10^{-2}	−0.08	0.04	2.79×10^{-2}
cg21307484	182	−0.04	0.02	3.45×10^{-2}	−	−	−

(B)

Colorectal cancer	Model 1[c]				Model 2[d]		
	N	β	SE(β)	P value	β	SE(β)	P value
cg21307484	276	−0.08	0.02	4.76×10^{-4}	−0.07	0.02	1.29×10^{-3}
cg11661512	275	−0.07	0.02	1.61×10^{-3}	−0.07	0.02	2.14×10^{-3}
cg26470501	276	−0.05	0.02	9.14×10^{-3}	−0.05	0.02	5.83×10^{-3}
cg26950531	276	−0.13	0.05	9.22×10^{-3}	−0.15	0.05	4.24×10^{-3}
cg14476101	275	−0.10	0.04	1.88×10^{-2}	−0.11	0.04	1.21×10^{-2}
cg16258657	276	−0.08	0.04	2.50×10^{-2}	−0.09	0.04	1.61×10^{-2}
cg06164260	276	−0.04	0.02	2.63×10^{-2}	−0.04	0.02	1.64×10^{-2}
cg19249811	273	0.20	0.09	3.56×10^{-2}	−	−	−
cg17901584	276	−0.08	0.04	3.63×10^{-2}	−0.09	0.04	1.85×10^{-2}

(C)

Myocardial infarction	Model 1[e]				Model 2[f]		
	N	β	SE(β)	P value	β	SE(β)	P value
cg12593793	401	−0.07	0.02	1.25×10^{-3}	−0.08	0.02	4.63×10^{-4}
cg13286614	402	−0.08	0.03	4.50×10^{-3}	−0.08	0.03	8.95×10^{-3}
cg17901584	402	−0.09	0.04	1.61×10^{-2}	−0.09	0.04	1.37×10^{-2}
cg26470501	402	−0.04	0.02	2.24×10^{-2}	−0.04	0.02	1.16×10^{-2}
cg27117792	402	−0.09	0.04	3.45×10^{-2}	−0.10	0.04	1.92×10^{-2}
cg06164260	402	−0.04	0.02	4.63×10^{-2}	−0.04	0.02	2.98×10^{-2}

The table reports the beta regression model coefficient β measuring the DNA methylation difference (on the logit scale) between cases and controls, the associated standard error (SE), and the P value for H_0: $\beta = 0$. Associations found significant at a Bonferroni-corrected significance level assuming 30 (the number of principal components needed to explain more than 95% of the variance) independent tests across the 40 actual tests are bolded in the table. Results for the model not adjusted for BMI and WHR (Model 2) are also reported

[a] Analyses are adjusted for age, BMI, WHR, alcohol, contraceptive use, HRT, smoking status, physical activity

[b] Analyses are adjusted for age, alcohol, contraceptive use, HRT, smoking status, physical activity

[c] Analyses are adjusted for sex, age, BMI, WHR, alcohol, processed meat, fibre, smoking status, physical activity

[d] Analyses are adjusted for sex, age, alcohol, processed meat, fibre, smoking status, physical activity

[e] Analyses are adjusted for sex, age, BMI, WHR, smoking, physical activity, TC/HDL ratio, Tg/HDL ratio (by fasting status)

[f] Analyses are adjusted for sex, age, smoking, physical activity, TC/HDL ratio, Tg/HDL ratio (by fasting status)

Associations with biomarkers of lipid and glucose metabolism

Statistically significant associations between methylation levels at the 40 informative CpG loci and biomarkers of lipid and glucose metabolism are reported in Table 3.

Associations were investigated without and with adjustment for BMI and WHR. While effect size estimates remained stable after this additional adjustment, some associations lost statistical significance. Adjusting for the effect of BMI and WHR, fasting blood levels of HDL cholesterol were associated with DNA methylation levels at cg0650016

(*ABCG1*) (the other CpG locus on ABCG1, cg27243685 was found associated in the model not adjusted for BMI and WHR), and cg09831562 (*SOX2OT*), though we did not identify any association with total cholesterol levels. Triglyceride levels were associated with DNA methylation levels at five informative CpG loci, cg00574958 (*CPT1A*), both CpG loci in *ABCG1* (cg06500161, cg27243685), cg17901584 (*DHCR24*), and cg11024682 (*SREBF1*). A significant association was found between total-to-HDL cholesterol ratio and DNA methylation levels at cg06500161 (*ABCG1*), and between triglycerides-to-HDL cholesterol ratio and five informative CpG loci: cg00574958 (*CPT1A*), cg27243685 and cg06500161 (both on *ABCG1*), cg17901584 (*DHCR24*), and cg11024682 (*SREBF1*, only for the model not adjusted for BMI and WHR). Finally, DNA methylation levels at cg06500161 (*ABCG1*) were associated with fasting blood glucose levels; no significant associations were found with insulin or HOMA index.

Associations of adiposity-related CpG loci with breast cancer, colorectal cancer and myocardial infarction

Among the 40 informative CpG loci, we identified one that was also associated with colorectal cancer (cg21307484 in *IL2RB*, Table 4B, Model 2). To investigate the potential for an independent effect of these markers, we ran similar models adjusted for BMI and WHR (Model 1). While the strength of the association with cg21307484 was slightly weakened, one additional association with colorectal cancer emerged involving methylation levels at cg11661512 in *FGF18* ($P = 1.6 \times 10^{-3}$). We also identified one CpG locus associated with myocardial infarction (cg12593793, Table 4C, Model 2). This locus was found to be inversely associated with smoking ($P = 2.93 \times 10^{-6}$ for current-to-never smoker comparison), even after adjustment for BMI and WHR (Model 1).

Discussion

We conducted a series of genome-wide association studies of DNA methylation levels for four adiposity measures in over 1900 individuals from large European population-based cohorts. Forty CpG loci were associated with at least one adiposity measure using a fixed effects meta-analysis. This model assumes that the obesity−methylation associations are similar across studies, which was supported by the modest heterogeneity estimates we obtained. We related DNA methylation levels to transcriptional activity in PBLs of the nearest gene and identified and replicated associations between higher DNA methylation levels at five CpG loci

and downregulation of *ABCG1*, *PHGDH*, and *SREBF1* gene expression. In addition, we found transcription levels of *ABCG1* to be inversely associated with BMI. Further, of the 40 adiposity-associated CpG loci, two were statistically significantly associated with subsequent colorectal cancer risk while one was related to myocardial infarction.

Of the 40 obesity-related CpG sites we report, 21 were also found to be associated with BMI at genome-wide significance level in a recent meta-analysis including more than 10,000 individuals [10]. Using a Bonferroni correction for 40 tests, we found that 26 of our obesity-related CpG sites, of which 18 were related to BMI in our data, replicated in the BMI meta-analysis. The 26 replicated CpG sites included both cg21307484, and cg12593793, which we found to be associated with colorectal cancer and myocardial infarction, respectively. Reciprocally, of the 187 BMI-related CpG that meta-analysis reported, 41 replicated in our data, and 2 were found to be associated with breast cancer, 2 with colorectal cancer, and 5 (including cg12593793) with myocardial infarction.

Of the 40 obesity-related CpG loci, only one (cg06500161, *ABCG1*) showed associations with all four adiposity measures [29]. DNA methylation levels at this locus were also associated with fasting triglyceride levels, HDL cholesterol, total-to-HDL cholesterol ratio, and triglycerides-to-HDL cholesterol ratio. These results are consistent with recent findings on the role of ABCG1 in cellular cholesterol efflux and reverse cholesterol transport [30] through removal of excess cholesterol and promotion of the maturation of nascent HDL particles to larger, cholesterol-rich lipoproteins [31]. *ABCG1* is highly expressed in cholesterol-loaded macrophages, where it guards against lipid overloading [32], and targeted disruption of *Abcg1* in mice results in substantial lipid accumulation in macrophages and multiple tissues [33]. *ABCG1* also plays a role in lipoprotein lipase regulation as demonstrated in knockdown experiments from cultured human macrophages [34], and may thus result in less efficient uptake of circulating triglycerides. Our results are in agreement with the 'conjoint trait' hypothesis of a combined low HDL cholesterol/high triglycerides phenotype that is regulated pleiotropically [35]. We also found an association between DNA methylation levels at one CpG locus on *CPT1A* and fasting triglyceride levels, consistent with recent reports in the literature [36–38] and with previous results in animal models [39]. DNA methylation levels at one CpG locus on *DHCR24* were associated with HDL cholesterol and triglycerides-to-HDL cholesterol ratio. Expression of this gene is known to be under epigenetic control [40], and its products are thought to mediate the anti-inflammatory effect of HDL cholesterol in endothelial cells [41, 42]; they also appear to regulate cellular response to oncogenic and oxidative stress [43].

Among the 40 informative CpG loci, two loci in *IL2RB* and *FGF18* and one CpG locus in an intergenic region of chromosome 1 were associated with colorectal cancer and myocardial infarction development, respectively, after controlling for anthropometric parameters and established risk factors. None of the informative CpG loci were associated with post-menopausal breast cancer following Bonferroni correction, though we note that the number of post-menopausal breast cancer cases included in the study was relatively small.

Differential DNA methylation levels in the *IL2RB* and *FGF18* genes were associated with risk of colorectal cancer, which suggests that changes in the activity of these genes may be contributing to the development of this malignancy. The *IL2RB* gene encodes the interleukin(IL)-2 receptor-β which binds IL-2, a cytokine which plays a major role in T-cell differentiation and proliferation. Overexpression of lymphocyte *IL2RB* has been linked to risk of colorectal cancer in small clinical case-control studies and was shown to predict risk when combined with six additional genes [44, 45]. Given that obesity is accompanied by alterations in T-cell populations, inflammation and immune dysregulation [46], our results provide a potential link between obesity, T-cell proliferative pathways and colorectal cancer development that warrants further study. The *FGF18* gene, which encodes the fibroblast growth factor 18, has also been previously linked to colorectal tumorigenesis [47]. FGF is a downstream target of the Wnt-signalling and β-catenin pathways which are central to cell growth and proliferation, and frequently mutated in colorectal cancer [48]. Alterations in the Wnt/β-catenin pathway have been associated with obesity-related disorders such as insulin resistance and type 2 diabetes [49, 50] and FGF18 interacts with PI3K/mTOR and growth factor signalling pathways. Our findings may provide new evidence that obesity is associated with specific changes in the Wnt/β-catenin signalling pathways, that could, in-turn, drive colorectal cancer development. Methylation levels at cg12593793 were found to be inversely associated with risk of myocardial infarction, independent of obesity and established risk factors, including smoking status. We may speculate that both smoking and adiposity influence myocardial infarction risk through a mechanism that alters methylation status at this locus.

In support of the biological basis of these relationships, some of the obesity-related CpG loci we report have been identified in previous studies, not only of adiposity measures, but also of lipid and glycaemic traits. In particular, cg06500161 (*ABCG1*) was found to be linked to incident type 2 diabetes in a recent study by Chambers et al. [11]; CpG loci on *CPT1A* and *PHGDH* both with adiposity measures [11] and the metabolic syndrome [51]; and *HIF3A*, *CPT1A* and *ABCG1* with BMI and waist circumference [52]. Nevertheless, we note that the recently

identified associations between BMI and DNA methylation at three CpG loci on *HIF3A* [9] did not reach genome-wide statistical significance in our meta-analysis. Inconsistencies between the prior data and the current analysis may partially reflect different modelling strategies: we opted for a more conventional parameterisation using DNA methylation levels as dependent variable, with adjustment for technical (microarray and position on the microarray) and other confounders.

We identified several CpG loci that are associated with adiposity, metabolic traits, and which appear to be associated with gene expression. Among these, two CpG loci in *IL2RB* and *FGF18*, and one intergenic CpG locus on chromosome 1 were associated with subsequent risk of colorectal cancer and myocardial infarction, respectively. These associations survived adjustment for adiposity measures and known risk factors, suggesting a potential role for DNA methylation as intermediate biomarker in the complex interplay between adiposity, metabolic health and risk of obesity-related diseases. However, because of the overlap between the methylation-obesity discovery and methylation-disease analyses, we cannot fully discard potential residual confounding in our BMI-adjusted analyses. In addition, as previously reported in a larger family-based study [9], and because our data do not arise from a structured population, our sample size renders any attempts to assess causality, for example using instrumental variable approaches, statistically underpowered and our analyses of the genetic data available in a fraction of our study population preclude the use of established BMI-related (combinations of) genetic variants as potential instruments in an attempt to infer causality.

Despite these limitations, we believe that our observational study data provide evidence for potential mechanisms involving DNA methylation alterations that may partly explain the association of obesity with obesity-related chronic diseases.

In conclusion, we have identified a number of CpG sites that are associated with adiposity and metabolic traits and which appear to have functional effects on gene expression. Among these, one was consistently associated to all four adiposity measures: cg06500161, *ABCG1*. In addition, two CpG sites in *IL2RB* and *FGF18* and one intergenic CpG locus on chromosome 1 were associated, independently of obesity and known risk factors, with subsequent risk of CRC and MI, respectively. This suggests a potential pleiotropic role for these epigenetic markers, and substantiates the involvement of DNA methylation in the complex interplay between adiposity, metabolic health, and risk of obesity-related diseases.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Stevens GA, Singh GM, Lu Y, Danaei G, Lin JK, Finucane MM et al. National, regional, and global trends in adult overweight and obesity prevalences. Popul Health Metr. 2012;10:16. https://doi.org/10.1186/1478-7954-10-22

2. Arnold M, Pandeya N, Byrnes G, Renehan AG, Stevens GA, Ezzati M, et al. Global burden of cancer attributable to high body-mass index in 2012: a population-based study. Lancet Oncol. 2015;16:36–46.

3. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. Lancet. 2014;384:755–65.

4. Samuel VT, Shulman GI. Mechanisms for insulin resistance: common threads and missing links. Cell. 2012;148:852–71.

5. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010;42:937–48.

6. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet. 2010;42:949–60.

7. Yang J, Loos RJF, Powell JE, Medland SE, Speliotes EK, Chasman DI, et al. FTO genotype is associated with phenotypic variability of body mass index. Nature. 2012;490:267–72.

8. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan Ja, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet. 2012;44:991–1005.

9. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S, et al. DNA methylation and body-mass index: a genome-wide analysis. Lancet. 2014;383:1990–8.

10. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2017;541:81–86.

11. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. Lancet Diabetes Endocrinol. 2015;3:526–34.

12. Palli D, Berrino F, Vineis P, Tumino R, Panico S, Masala G, et al. A molecular epidemiology project on diet and cancer: the EPIC-Italy Prospective Study. Design and baseline characteristics of participants. Tumori. 2003;89:586–93.

13. Beulens JWJ, Monninkhof EM, Verschuren WMM, van der Schouw YT, Smit J, Ocke MC, et al. Cohort profile: the EPIC-NL study. Int J Epidemiol. 2010;39:1170–8.

14. Lund E, Kumle M, Braaten T, Hjartaker A, Bakken K, Eggen E, et al. External validity in a population-based national prospective study—the Norwegian Women and Cancer Study (NOWAC). Cancer Causes Control. 2003;14:1001–8.

15. Hallmans G, Agren A, Johansson G, Johansson A, Stegmayr B, Jansson J-H, et al. Cardiovascular disease and diabetes in the Northern Sweden Health and Disease Study Cohort—evaluation of risk factors and their interactions. Scand J Public Health. 2003;61:18–24.

16. Chadeau-Hyam M, Vermeulen RCH, DGAJ Hebels, Castagne R, Campanella G, Portengen L, et al. Prediagnostic transcriptomic markers of chronic lymphocytic leukemia reveal perturbations 10 years before diagnosis. Ann Oncol. 2014;25:1065–72.

17. Dumeaux V, Borresen-Dale A-L, Frantzen J-O, Kumle M, Kristensen VN, Lund E. Gene expression analyses in breast cancer epidemiology: the Norwegian Women and Cancer postgenome cohort study. Breast Cancer Res. 2008;10:R13 https://doi.org/10.1186/bcr1859

18. DGAJ Hebels, Georgiadis P, Keun HC, Athersuch TJ, Vineis P, Vermeulen R, et al. Performance in omics analyses of blood samples in long-term storage: opportunities for the exploitation of existing biobanks in environmental health research. Environ Health Perspect. 2013;121:480–7.

19. Ferrari SLP, Cribari-Neto F. Beta regression for modelling rates and proportions. J Appl Stat. 2004;31:799–815.

20. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21:1539–58.

21. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8:203–9.

22. Dumeaux V, Olsen KS, Nuel G, Paulssen RH, Borresen-Dale AL, Lund E Deciphering normal blood gene expression variation—the NOWAC Postgenome Study. PLoS Genet. 2010;6:e1000873

23. Kannel WB, Vasan RS, Keyes MJ, Sullivan LM, Robins SJ. Usefulness of the triglyceride-high-density lipoprotein versus the cholesterol-high-density lipoprotein ratio for predicting insulin resistance and cardiometabolic risk (from the Framingham offspring cohort). Am J Cardiol. 2008;101:497–501.

24. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28:412–9.

25. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinform. 2012;13:86.

26. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén S-E, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS ONE. 2012;7:e41361.

27. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Greally JM, Gut I, et al. Recommendations for the design and analysis of epigenome-wide association studies. Nat Methods. 2013;10:949–55.

28. Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65.

29. Pfeifferm L, Wahl S, Pilling LC, Reischl E, Sandling JK, Kunze S et al. DNA methylation of lipid-related genes affects blood lipid levels. Circ Cardiovasc Genet. 2015;8:334–342

30. Li G, Gu H-M, Zhang D-W. ATP-binding cassette transporters and cholesterol translocation. IUBMB Life. 2013;65:505–12.

31. Vaughan AM, Oram JF. ABCA1 and ABCG1 or ABCG4 act sequentially to remove cellular cholesterol and generate cholesterol-rich HDL. J Lipid Res. 2006;47:2433–43.

32. Klucken J, Buchler C, Orso E, Kaminski WE, Porsch-Ozcurumez M, Liebisch C, et al. ABCG1 (ABC8), the human homolog of the Drosophila white gene, is a regulator of macrophage cholesterol and phospholipid transport. Proc Natl Acad Sci USA. 2000;97:817–22.

33. Kennedy MA, Barrera GC, Nakamura K, Baldan A, Tarr P, Fishbein MC, et al. ABCG1 has a critical role in mediating cholesterol efflux to HDL and preventing cellular lipid accumulation. Cell Metab. 2005;1:121–31.

34. Olivier M, Tanck MW, Out R, Villard EF, Lammers B, Bouchareychas L, et al. Human ATP-binding cassette G1 controls macrophage lipoprotein lipase bioavailability and promotes foam cell formation. Arterioscler Thromb Vasc Biol. 2012;32:2223–31.

35. Mahaney MC, Blangero J, Comuzzie AG, Vandeberg JL, Stern MP, Maccluer JW. Plasma HDL, cholesterol, triglycerides, and adiposity—a quantitative genetic test of the conjoint trait

36. Frazier-Wood AC, Aslibekyan S, Absher DM, Hopkins PN, Sha J, Tsai MY, et al. Methylation at CPT1A locus is associated with lipoprotein subfraction profiles. J Lipid Res. 2014;55:1324–30.

37. Gagnon F, Aissi D, Carrie A, Morange P-E, Tregouet D-A. Robust validation of methylation levels association at CPT1A locus with lipid plasma levels. J Lipid Res. 2014;55:1189–91.

38. Tobi EW, Goeman JJ, Monajemi R, Gu H, Putter H, Zhang Y, et al. DNA methylation signatures link prenatal famine exposure to growth and metabolism. Nat Commun. 2014;5:5592.

39. Stefanovic-Racic M, Perdomo G, Mantell BS, Sipula IJ, Brown NF, O'Doherty RM. A moderate increase in carnitine palmitoyltransferase 1a activity is sufficient to substantially reduce hepatic triglyceride levels. Am J Physiol—Endocrinol Metab. 2008;294:E969–77.

40. Drzewinska J, Walczak-Drzewiecka A, Ratajewski M. Identification and analysis of the promoter region of the human DHCR24 gene: involvement of DNA methylation and histone acetylation. Mol Biol Rep. 2011;38:1091–101.

41. McGrath KCY, Li XH, Puranik R, Liong EC, Tan JTM, Dy VM, et al. Role of 3β-hydroxysterol-Δ24 reductase in mediating antiinflammatory effects of high-density lipoproteins in endothelial cells. Arterioscler Thromb Vasc Biol. 2009;29:877–82.

42. Wu BJ, Chen K, Shrestha S, Ong KL, Barter PJ, Rye K-A. High-density lipoproteins inhibit vascular endothelial inflammation by increasing 3β-hydroxysterol-Δ24 reductase expression and inducing heme oxygenase-1. Circ Res. 2013;112:278–88.

43. Wu C, Miloslavskaya I, Demontis S, Maestro R, Galaktionov K. Regulation of cellular response to oncogenic and oxidative stress by seladin-1. Nature. 2004;432:640–5.

44. Chang YT, Huang CS, Yao CT, Su SL, Terng HJ, Chou HL, et al. Gene expression profile of peripheral blood in colorectal cancer. World J Gastroenterol. 2014;20:14463–71.

45. Marshall KW, Mohr S, Khettabi FE, Nossova N, Chao S, Bao W, et al. A blood-based biomarker panel for stratifying current risk for colorectal cancer. Int J Cancer. 2010;126:1177–86.

46. Nishimura S, Manabe I, Nagasaki M, Eto K, Yamashita H, Ohsugi M, et al. CD8+ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. Nat Med. 2009;15:914–20.

47. Shimokawa T, Furukawa Y, Sakai M, Li M, Miwa N, Lin Y-M, et al. Involvement of the FGF18 gene in colorectal carcinogenesis, as a novel downstream target of the β-catenin/T-cell factor complex. Cancer Res. 2003;63:6116–20.

48. Bienz M, Clevers H. Linking colorectal cancer to Wnt signaling. Cell. 2000;103:311–20.

49. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet. 2006;38:320–3.

50. Welters HJ, Kulkarni RN. Wnt signaling: relevance to beta-cell biology and diabetes. Trends Endocrinol Metab. 2008;19:349–55.

51. Das M, Sha J, Hidalgo B, Aslibekyan S, Do AN, Zhi D, et al. Association of DNA methylation at CPT1A locus with metabolic syndrome in the genetics of lipid lowering drugs and Diet Network (GOLDN) Study. PLoS ONE. 2016;11:e0145789.

52. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. Hum Mol Genet. 2015;24:4464–79.

Affiliations

Gianluca Campanella[1] · Marc J. Gunter[2] · Silvia Polidoro[3] · Vittorio Krogh[4] · Domenico Palli [5] · Salvatore Panico[6] · Carlotta Sacerdote[3,7] · Rosario Tumino[8] · Giovanni Fiorito[3,9] · Simonetta Guarrera[3,9] · Licia Iacoviello[10] · Ingvar A. Bergdahl[11] · Beatrice Melin[12] · Per Lenner[12] · Theo M. C. M. de Kok[13] · Panagiotis Georgiadis[14] · Jos C. S. Kleinjans[13] · Soterios A. Kyrtopoulos [14] · H. Bas Bueno-de-Mesquita[1,15,16,17] · Karen A. Lillycrop[18] · Anne M. May[19] · N. Charlotte Onland-Moret[19] · Robert Murray[18] · Elio Riboli[1,20] · Monique Verschuren[21] · Eiliv Lund[22] · Nicolle Mode[22] · Torkjel M. Sandanger[22] · Valentina Fiano[23] · Morena Trevisan[23] · Giuseppe Matullo[3,9] · Philippe Froguel [24,25] · Paul Elliott [1,20] · Paolo Vineis[1,3,20] · Marc Chadeau-Hyam [1,20]

1. Department of Epidemiology and Biostatistics, Imperial College London, London, UK

2. Section of Nutrition and Metabolism, International Agency for Research on Cancer (IARC), Lyon, France

3. Italian Institute for Genomic Medicine (IIGM), Turin, Italy

4. Fondazione IRCCS—Istituto Nazionale dei Tumori, Milan, Italy

5. Istituto per lo Studio e la Prevenzione Oncologica (ISPO Toscana), Florence, Italy

6. Department of Clinical Medicine and Surgery, University of Naples Federico II, Naples, Italy

7. Piedmont Reference Centre for Epidemiology and Cancer Prevention (CPO Piemonte), Turin, Italy

8. Cancer Registry and Histopathology Unit, Azienda Ospedaliera "Civile—M.P. Arezzo", Ragusa, Italy

9. Department of Medical Sciences, University of Turin, Turin, Italy

10. Department of Epidemiology and Prevention, IRCCS Istituto Neurologico Mediterraneo Neuromed, Pozzilli (IS), Italy

11. Department of Biobank Research, Umeå University, Umeå, Sweden

12. Department of Radiation Sciences, Oncology Umeå University, Umeå, Sweden

13. Department of Toxicogenomics, Maastricht University, Maastricht, The Netherlands

14. Institute of Biology, Medicinal Chemistry, and Biotechnology, National Hellenic Research Foundation, Athens, Greece

15. Department for Determinants of Chronic Diseases (DCD), National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

16. Department of Gastroenterology and Hepatology, University Medical Centre, Utrecht, The Netherlands

17. Department of Social and Preventive Medicine, University of Malaya, Kuala Lumpur, Malaysia

18. Centre for Biological Sciences, University of Southampton, Southampton, UK

19. Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, The Netherlands

20. MRC-PHE Centre for Environment and Health, Imperial College London, London, UK

21. Centre for Prevention and Health Services Research, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

22. Department of Community Medicine, University of Tromsø (UiT) —The Artic University of Norway, Tromsø, Norway

23. Department of Medical Sciences, Unit of Cancer Epidemiology–CERMS, University of Turin, Turin, Italy

24. CNRS UMR8199, Pasteur Institute of Lille, Lille University, Lille, France

25. Department of Genomics of Common Disease, Imperial College London, London, UK