







RESEARCH PAPER



## DNA methylation, colon cancer and Mediterranean diet: results from the EPIC-Italy cohort

Francesca Fasanelli <sup>a</sup>, Maria Teresa Giraudo<sup>b</sup>, Paolo Vineis<sup>c,d</sup>, Valentina Fiano<sup>a</sup>, Giovanni Fiorito<sup>d</sup>, Chiara Grasso <sup>a</sup>, Silvia Polidoro <sup>d</sup>, Morena Trevisan<sup>a</sup>, Sara Grioni <sup>e</sup>, Vittorio Krogh<sup>e</sup>, Amalia Mattiello<sup>f</sup>, Salvatore Panico<sup>f</sup>, Maria Concetta Giurdanella<sup>g</sup>, Rosario Tumino <sup>g</sup>, Laura De Marco<sup>a,h</sup>, Fulvio Ricceri <sup>i,j</sup>, and Carlotta Sacerdote<sup>\*h</sup>

<sup>a</sup>Unit of Cancer Epidemiology– CeRMS, Department of Medical Sciences, University of Turin, Turin, Italy; <sup>b</sup>Department of Mathematics, University of Turin, Turin, Italy; <sup>c</sup>Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Faculty of Medicine, Imperial College London, UK; <sup>d</sup>Molecular and Genetic Epidemiology Unit, Italian Institute for Genomic Medicine (IIGM), Turin, Italy; <sup>e</sup>Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; <sup>f</sup>Department of Clinical Medicine and Surgery, Federico II University, Naples, Italy; <sup>g</sup>Cancer Registry, Department of Prevention, ASP, Ragusa, Italy; <sup>h</sup>Unit of Cancer Epidemiology, Città della Salute e della Scienza University-Hospital and Center for Cancer Prevention (CPO), Turin, Italy; <sup>i</sup>Epidemiology Unit, Regional Health Services ASL TO3, Grugliasco, Italy; <sup>j</sup>Department of Clinical and Biological Sciences, University of Turin, Turin, Italy

### ABSTRACT

The biological mechanisms through which adherence to Mediterranean Diet (MD) protects against colon cancer (CC) are poorly understood. Evidence suggests that chronic inflammation may be implicated in the pathway. Both diet and CC are related to epigenetic regulation.

We performed a nested case-control study on 161 pairs from the Italian component of the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort, in which we looked for the methylation signals in DNA extracted from leucocytes associated with both CC and MD in 995 CpGs located in 48 inflammation genes. The DNA methylation signals detected in this analysis were validated in a subgroup of 47 case-control pairs and further replicated (where validated) in 95 new pairs by means of pyrosequencing.

Among the CpG sites selected *a-priori* in inflammation-related genes, seven CpG sites were found to be associated with CC status and with MD, in line with its protective effect. Only two CpG sites (cg17968347-*SERPINE1* and cg20674490-*RUNX3*) were validated using bisulphite pyrosequencing and, after replication, we found that DNA methylation of cg20674490-*RUNX3* may be a potential molecular mediator explaining the protective effect of MD on CC onset.

The use of a ‘meet-in-the-middle’ approach to identify the overlap between exposure and predictive markers of disease is innovative in studies on the relationship between diet and cancer, in which exposure assessment is difficult and the mechanisms through which the nutrients exert their protective effect is largely unknown.

### ARTICLE HISTORY

Received 5 December 2018  
Revised 24 May 2019  
Accepted 3 June 2019

### KEYWORDS

Colon cancer; Mediterranean diet; inflammation; DNA methylation; meet-in-the-middle

## Introduction


Epidemiological studies have unveiled a protective association between Mediterranean diet (MD) and incidence of cancer and of other major chronic diseases [1,2]. In the Italian branch of the European Prospective Investigation into Cancer and Nutrition (EPIC) study, increasing values of the Italian Mediterranean Index (IMI) were associated with a decreased risk of colon cancer (CC) as well as rectal cancer in both men and women [3]. A pooled analysis of three Italian case-control studies [4] and two

cohort studies [5,6] confirmed the favourable role of MD in reducing CC incidence. The protective mechanisms of MD have been attributed to a high content of antioxidants, olive oil, fibres, and a moderate alcohol intake [7]. The biological mechanisms through which MD protects against CC remain nonetheless poorly understood.

Adherence to MD has been previously hypothesized to protect against abdominal adiposity [8], and obesity is, in turn, a well-known risk factor for chronic diseases, specifically CC [9]. In a previous

**CONTACT** Carlotta Sacerdote  [carlotta.sacerdote@cpo.it](mailto:carlotta.sacerdote@cpo.it)  Unit of Cancer Epidemiology – Città della Salute e della Scienza University-Hospital, Center for Cancer Prevention (CPO), Via Santena 7, Turin 10126, Italy

\*Join last authors.

 Supplemental data for this article can be accessed [here](#).

© 2019 Informa UK Limited, trading as Taylor & Francis Group

study [10] we analysed adiposity as a potential mediator of the association between adherence to MD and CC, but we concluded that its reduction, due to MD, does not explain the protective effect.

An alternative hypothesis suggests that the causal pattern of the association between MD and CC could pass through chronic inflammation. In fact, randomized control trials [11] and observational studies [12] have shown that MD can attenuate the level of systemic inflammation while chronic inflammation has been shown to be itself a possible causative factor for a variety of cancer types, including CC. People with chronic inflammatory bowel diseases show indeed an increased risk of CC [13] and chronic aspirin use seems to reduce the risk of CC [14].

Both diet and CC are related to epigenetic regulation. Diet can perturb the way genes are controlled by epigenetic signals, particularly DNA methylation [15,16], which is, in turn, one of the regulatory mechanisms of systemic and local inflammation [17].

The aim of this investigation was to explore methylation changes of inflammation genes associated with both MD and CC status in peripheral blood cells of members of an Italian epidemiological cohort. Our purpose was to test the hypothesis that DNA methylation could be a mediator of the CC triggered by non-adherence to MD.

## Methods

### *Study sample and discovery set*

Data from the Italian component of the EPIC study (EPIC-Italy) [18], a total of 42,894 volunteers recruited from four centres in Italy (Varese, Turin, Naples, and Ragusa) in 1993–1998 were considered. Over a mean follow-up of 11.0 years (range 1.0–14.8), 313 cases of CC were diagnosed. Colon cancers were primary incident cases, identified as proximal (International Classification of Diseases for Oncology, third Edition [ICD-O-3] codes C18.0–C18.5; N = 121), distal (ICD-O-3 codes C18.6–C18.7; N = 154), and over-lapping or unspecified (ICD-O-3 codes C18.8–C18.9; N = 36) sites.

All participants in the study sample completed a validated semiquantitative food-frequency

questionnaire at enrolment. Blood samples were collected at recruitment and sent to local laboratories for processing and aliquot preparation.

A nested case-control study was conducted within EPIC-Italy cohort employing 169 incident CC cases diagnosed within follow-up and 169 matched controls selected at random from the participants at risk of CC at the time of the case diagnoses.

Controls were matched to cases by sex, date of birth (within 5 years), seasonality of blood sampling (autumn-winter/spring-summer) and study centre. Members of the cohort treated for diabetes or with a diagnosis of intestinal polyps at baseline were excluded from the case-control study, as well as those with a prevalent diagnosis of cancer (except non-melanoma skin cancer).

### *Epigenome-wide DNA methylation analysis*

Genomic DNA was extracted from buffy coats using the QIASymphony DNA Midi Kit (Qiagen, Hilden, Germany). Five hundred nanogram of DNA were bisulphite-converted using the EZ-96 DNA Methylation-Gold™ Kit (Zymo, California, USA) and hybridised to Infinium HumanMethylation450 BeadChips (Illumina, California, USA). Each chip was subsequently scanned using the Illumina HiScanSQ system, and sample quality was assessed using control probes on the microarrays. Raw intensity data were finally exported from Illumina GenomeStudio (version 2011.1). Each case-control pair was arranged randomly on the same chip.

Data pre-processing was carried out using an in-house software written for the R statistical computing environment (see [19] for a short description of the procedure). One hundred and sixty-one case-control pairs passed the pre-processing step for the successive analysis. DNA methylation was expressed as the ratio between the intensities of methylated cytosines and the total intensities ( $\beta$ -values).

### *Main exposure and other variables*

Italian Mediterranean Index (IMI), a summary measure of adherence to the MD, was used as a measure of the exposure [20]. Briefly, the

computation of this index is based on the intake of 11 food items: intakes of the Mediterranean foods pasta, Mediterranean vegetables (raw tomatoes, cooked leafy vegetables, raw leafy vegetables, onion or garlic, mixed salad, or mixed vegetables), fruits, legumes, olive oil, and fish of four 'non Mediterranean' foods (soft drinks, butter, red meat and potatoes); and of alcohol. If consumption of typical Mediterranean foods is in the third tertile of the distribution, the person receives 1 point; all other intakes receive 0 points. If consumption of non-Mediterranean foods is in the first tertile of the distribution the person receives 1 point. Alcohol receives 1 point for intake up to 12 g/day; abstainers and persons who consume >12 g/day receive 0 points. For the analysis, IMI was categorized in the following three classes: 0–2 (low adherence to MD), 3–4 (middle adherence to MD), and 5–11 (high adherence to MD).

We considered as further covariates: age, sex, centre, smoking status (never, former and current), total physical activity (inactive, moderately inactive, moderately active and active [21]), level of education (tertiles of the relative index of inequality RII [22]), body mass index (BMI) and cell types. The latter were estimated according to Houseman method [23].

### **Selection of CpG sites**

To study the methylation status of inflammation genes, data obtained from the epigenome-wide study were pruned selecting the CpGs located in a set of 48 inflammation genes (list in Supplementary Table 1). Such genes were selected through an extensive literature review to identify those related to inflammatory mechanisms involved in colon carcinogenesis.

Among the resulting 995 inflammation-related CpGs, only those with a mean difference (both positive and negative) in methylation percentage between cases and controls higher than 1% were considered in the association analysis (32 CpGs). The cutoff of 1% was chosen to increase the probability that the difference in methylation reflected an effective biological change.

A conditional logistic regression model with elastic net penalties [24] was employed to select the most important CpG sites among the 32 CpGs related to CC. DNA methylation levels of the

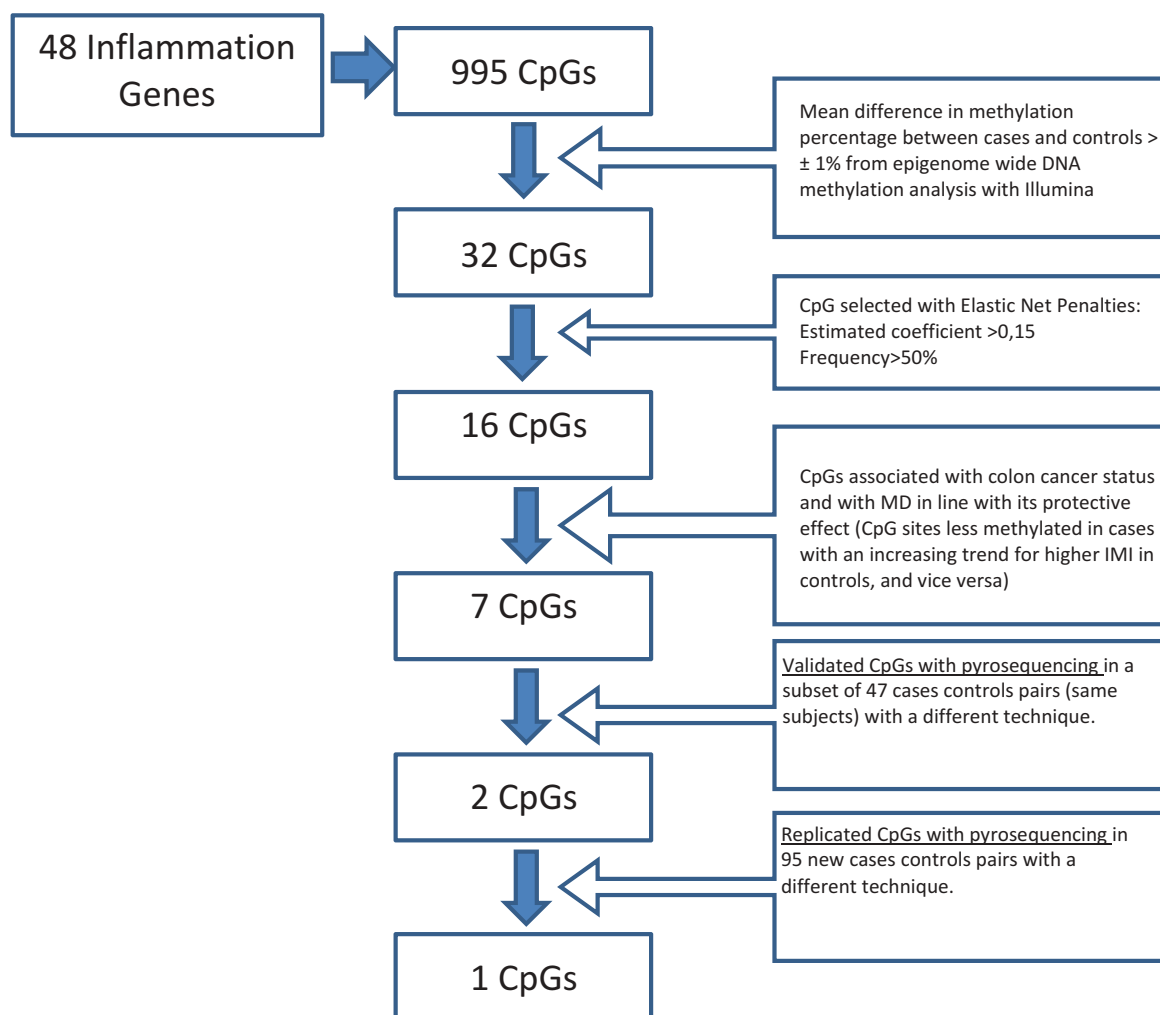
CpGs were standardized to 1 standard deviation computed on the control group. Elastic net (EN) [25] is a regularization and variable selection method, which retains the parsimony property of Lasso regression method [26], but encourages at the same time the grouping effect like Ridge regression [27]. We applied the cyclic coordinate descent algorithm [28], and we set the parameter controlling the trade-off between Lasso and Ridge penalties to 0.5. We used 10-fold cross-validation (CV) for the choice of the regularization parameter that characterizes the best model. To assess whether the associations found were stable in random subsets of the sample, 1000 EN models were fitted using each time 63.4% of the initial data. At the end, we obtained a ranked list of probes based on how many times they were included in models based on data subsets.

The CpG sites considered for further detailed analyses had to satisfy both the following criteria: i) CpGs selected by EN, with an estimated coefficient of the association with case-control status higher than 0.15 in absolute value (median of the distribution of the estimated coefficients) and ii) CpGs in the list of the most associated sites with a frequency higher than 50%. A diagram illustrating the selection of CpG for the analysis is shown in Figure 1.

### **Detailed analysis of the selected signals**

For each selected CpG, a conditional logistic regression model was fitted to estimate odds ratio of CC with DNA methylation levels included as independent variable (model A). A second model (model B) was fitted including additional adjustment for BMI, smoking status, total physical activity, level of education and IMI. The possible effect of cell composition on the results was assessed by further adding to model B the proportions of cell counts (model C). Sensitivity analyses were performed excluding cases where time elapsed between blood collection and diagnosis of CC was lower than 2 years or higher than 10 years.

The association between M-values of methylation (the logarithmic transform of  $\beta$ -values,  $M = \log_2(\beta/(1-\beta))$  [29]) and IMI was evaluated in the control group by fitting a linear mixed effect model with chip as random effect and IMI, sex, age, centre, BMI, smoking status, total physical



**Figure 1.** Diagram illustrating CpG selection.

Illumina = Infinium Human Methylation450 BeadChips IMI = Italian Mediterranean Index

activity, level of education and differential cell types as fixed effects. Linearity of trends across categories of IMI was tested by treating the categorical variable as continuous in the linear mixed effect model.

### **Validation, replication and meta-analysis**

A random selection of 47 case-control pairs from EPIC-Turin (validation set) was performed for validation with the pyrosequencing methodology. The CpGs for which adherence to MD conferred methylation levels that were protective on CC were selected for validation. This means that among the CpGs whose hypermethylation was protective on CC (odds ratio  $< 1$  for 1 standard deviation increase in methylation percentage) only those for which higher methylation levels corresponded to higher adherence

to MD were validated. Similarly, CpGs whose hypomethylation was associated with decreased CC risk and for which lower methylation levels were associated with higher adherence to MD were validated.

Another nested case-control study (replication set) was finally conducted employing 95 independent case-control pairs from EPIC-Italy. The validated signals were replicated in this sample using pyrosequencing.

Both in the validation and in the replication, the associations with DNA methylation levels obtained by pyrosequencing were analysed using models similar to those employed in the discovery phase (see details in the Tables). Additionally, a random effects meta-analysis of the discovery and replication studies was performed on the validated signals.

### Details on the pyrosequencing methodology

Pyrosequencing assay was performed on a PyroMark Q24 MDx system using PyroMark Gold Q24 Advanced reagents (Qiagen, Hilden Germany). Primers were designed according to PyroMark Assay Design software version 2.0 (Qiagen). PCR reaction was performed in a total volume of 35  $\mu$ l using the PyroMark PCR kit (Qiagen) containing 1X PCR Master Mix, 1X CoralLoad Concentrate, 0.2  $\mu$ M of each primer, and 1  $\mu$ l of bisulphite-converted DNA with the following cycling profile: 95°C for 10 min followed by 45 cycles of denaturation at 95°C for 30 sec, annealing at specific temperature for each gene (55°C for *RUNX3*; 50°C for *SERPINE1*) for 30 sec, extension at 72°C for 1 min. Extension at 72°C for 10 min was finally performed. The PCR product (15  $\mu$ l) was added to 19  $\mu$ l of distilled water and incubated under shaking with 40  $\mu$ l of binding buffer pH 7.6, containing 10mM Tris-HCl, 2 M NaCl, 1mM EDTA, and 1  $\mu$ l of sepharose beads covered by streptavidin. The PCR product was washed with ethanol 70%, denatured with NaOH 0.2 M and re-washed with Tris-Acetate 10 mM pH 7.6. Pyrosequencing reaction was performed in a total volume of 20  $\mu$ l, including 19.85  $\mu$ l of 20 mM Tris-Acetate, 5 mM MgAc<sub>2</sub> and 0.15  $\mu$ l of 50  $\mu$ M sequencing primer. Assays were created according to manufacturer's instruction. The nucleotide dispensation order was suggested by the software PyroMark Q24 Advanced version 3.0.0.

Methylation quantification was achieved using the provided software, and expressed for each DNA locus as percentage ratio of methylated cytosines of the sum of methylated and unmethylated cytosines. Positive controls for methylated [EpiTect Control DNA (human), methylated (Qiagen)] and unmethylated status [EpiTect Control DNA (human), unmethylated (Qiagen)] were included in each pyrosequencing run. Each sample was analysed twice in different runs and the average of the two results was computed. Adequacy of the results for each sample was achieved when difference in methylation percentage between runs was  $\leq 2\%$  and pyrograms resulted as 'passed'.

### Software for analysis

All statistical analyses were performed using R Statistical Software (The R Foundation for

Statistical Computing, Vienna, Austria) version 3.2.3 (2015–12-10) and Stata version 13 (StataCorp, College Station, TX, USA). A diagram illustrating the data sets used in the different stages of the analysis is shown in [Figure 2](#).

### Results

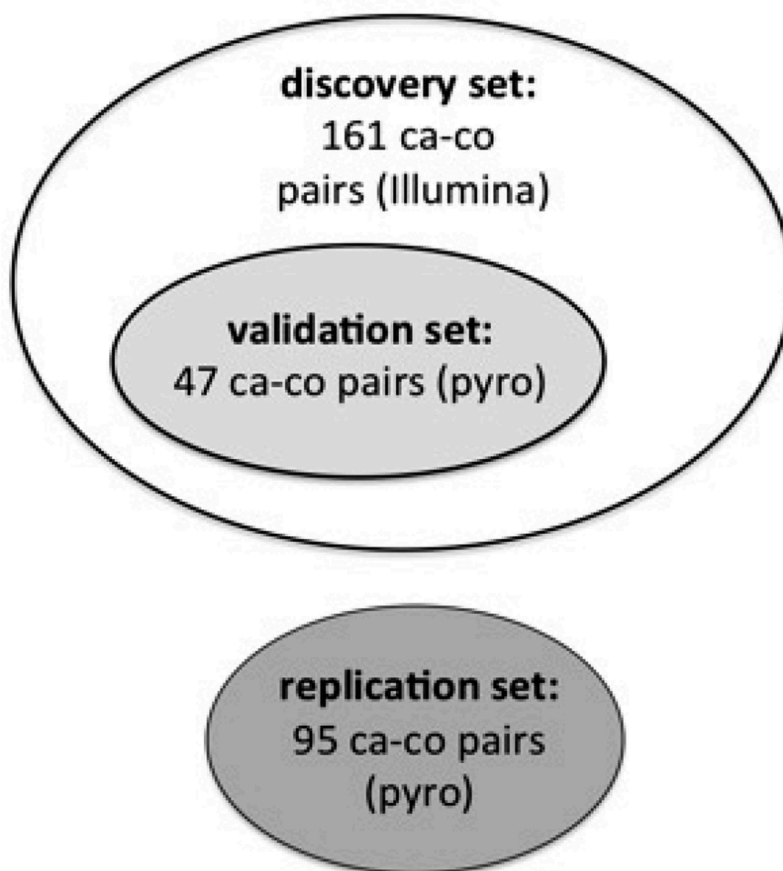
[Table 1](#) shows the baseline characteristics of the samples. Cases and controls showed differences only with respect to BMI. For completeness, the features of the validation and replication samples are reported in the Supplementary materials. Cases and controls showed differences in BMI and educational level in the validation set, while no differences were detected in the replication set.

The profile of parameter estimates plotted against the value of the regularization parameter and the CV curve of the EN conditional logistic model in the discovery analysis are reported in [Supplementary Figures 1 and 2](#). In particular, the CV error was minimized for a model with 26 predictors. [Table 2](#) shows the 16 predictors with an absolute value of the estimated coefficient higher than 0.15. The mean differences in methylation percentages were small for all these CpG sites ( $< 3\%$ ), except for cg12195446-*IRS2* (7.6%) and cg12252547-*MAL2* (5.1%). Ten CpGs were hypomethylated in cases. At least two probes were detected for each gene except for *RUNX1*, *STAT3*, *RUNX3* and *PTX3* for which only one probe was found.

A detailed analysis of the seven CpG sites selected for validation is reported in [Table 3](#) (results for the remaining CpG sites are reported in [Supplementary Table 4](#)). Evidence emerged of an association between all the CpGs and the case-control status also after adjustment for cell types and other confounding variables. In particular, for all the CpGs except cg08053846-*SERPINE1*, hypomethylation was associated with CC and methylation levels were increased for higher IMI categories.

Considering the validation dataset, only cg17968347-*SERPINE1* and cg20674490-*RUNX3* showed coherent associations with both CC and IMI ([Table 4](#) and [Figure 1](#)). Although these associations were at the limit of significance, their direction and magnitude were essentially the same. Of the two validated CpG sites, only for cg20674490-

## Datasets used in the analyses



**Figure 2.** Diagram illustrating the data sets used in the different stages of the analysis.

ca = casesco = controls Illumina = Infinium HumanMethylation450 BeadChips pyro = pyrosequencing.

*RUNX3* the association with CC showed coherent direction and similar magnitude in the replication sample (OR = 0.80, 95% CI 0.57, 1.13 in the replication set versus OR = 0.59, 95% CI 0.29, 1.19 in the validation set; Tables 4 and 5, model B). For this CpG the association with MD showed an increasing trend in the replication sample not found in the discovery and validation phases (Tables 3–5). The meta-analysis strengthened the association of CC with *RUNX3* higher methylation levels (OR<sub>META</sub> = 0.72, 95% CI 0.57, 0.91, *P*-value<sub>META</sub> = 0.006; see Figure 3).

### Discussion

The adherence to MD is known to have beneficial effects on human health, and in particular, a preventive effect on cancer. Several biological

mechanisms have been hypothesized to explain such protective effect, but the specific cellular pattern involved has not been experimentally found yet.

DNA methylation is likely to play a major role in carcinogenesis and it is probably influenced by diet [15,16]. Since it has been previously shown that MD may attenuate the level of chronic inflammation, which in turn has been related to an increased risk of CC, we hypothesized that the protective effect of adherence to the MD in CC is mediated by DNA methylation in genes that are involved in inflammation.

To test this hypothesis, we used the ‘meet-in-the-middle’ approach [30,31]. This approach is based on the assumption that identifying biomarkers associated with both a particular exposure and a certain disease strengthens the causal links between such

**Table 1.** Descriptive statistics of the case-control study nested in the EPIC-Italy cohort.

Variables	Controls	Cases	all	P-value*
N	161	161	322	
<b>Median of (IQR)</b>				
age, years	55 (9)	55 (9)	55 (9)	
BMI, kg/m <sup>2</sup>	25.01 (5)	26.42 (5)	25.85 (5)	0.0019
<b>Counts of</b>				
<i>Gender</i>				
Men	88 (55%)	88 (55%)	176 (55%)	
Women	73 (45%)	73 (45%)	146 (45%)	
<i>Centre</i>				
Varese	64 (40%)	64 (40%)	128 (40%)	
Ragusa	10 (6%)	10 (6%)	20 (6%)	
Turin	87 (54%)	87 (54%)	174 (54%)	
<i>Educational level</i>				0.078
1°tertile RII	51 (34%)	53 (36%)	104 (35%)	
2°tertile RII	54 (36%)	37 (25%)	91 (30%)	
3°tertile RII	45 (30%)	59 (39%)	104 (35%)	
<i>Total physical activity</i>				0.801
Inactive	34 (22%)	37 (24%)	71 (23%)	
moderately inactive	65 (42%)	69 (44%)	134 (43%)	
moderately active	36 (23%)	29 (19%)	75 (21%)	
Active	20 (13%)	20 (13%)	40 (13%)	
<i>Smoking status</i>				0.828
never smokers	71 (46%)	67 (43%)	138 (44%)	
former smokers	48 (31%)	53 (34%)	101 (33%)	
current smokers	36 (23%)	35 (23%)	71 (23%)	
<i>Fasting status</i>				0.360
Yes	119 (74%)	126 (78%)	245 (76%)	
No	42 (26%)	35 (22%)	77 (24%)	
<i>Cell Types</i>				
CD8T	8.1%	7.1%	7.5%	0.129
CD4T	13%	14%	14%	0.528
NK	7.4%	7.5%	7.5%	0.851
Bcell	5.2%	5.4%	5.3%	0.777
Mono	7.4%	7.1%	7.2%	0.543
Gran	63.0%	63.9%	63.2%	0.791

BMI = Body Mass Index; RII = Relative Index of Inequality; NK = Natural Killer; Bcell = B lymphocytes; Mono = Monocytes; Gran = Granulocytes.

\*differences between cases and controls are assessed using Mann Whitney Test for continuous variables and Chi-Square Test for categorical variables.

**Table 2.** CpG sites selected by EN with 10-fold CV with an estimated coefficient higher than 0.15 in absolute value in the case-control study nested in the EPIC-Italy cohort.

Probe Name	Estimated coefficient	Gene	Chromosome	Location	MAPINFO	Delta	Frequency
cg13104385	0.56	IL6	7	Body	22767384	0.020	62.4%
cg18773937	-0.37	IL1B	2	TSS1500	113594611	0.016	72.1%
cg12195446	0.37	IRS2	13	Body	110424497	0.076	64.6%
cg05265849	-0.34	IL6	7	Body	22767390	0.010	9.3%
cg02749784	-0.25	MAL2	8	TSS1500	120219927	0.018	72.4%
cg17968347	-0.23	SERPINE1	7	Body	100777740	0.012	55.7%
cg12252547	0.23	MAL2	8	TSS1500	120220032	0.051	57.3%
cg01265860	-0.23	RUNX1	21	Body	36256316	0.016	56.7%
cg24312520	-0.22	STAT3	17	Body	40489584	0.017	57.9%
cg16308790	-0.22	NFATC1	18	Body	77225973	0.011	51.2%
cg08053846	0.21	SERPINE1	7	TSS1500	100769605	0.018	64.4%
cg15363134	-0.20	NFATC1	18	Body	77161214	0.015	62.7%
cg20674490	-0.17	RUNX3	1	Body	25240932	0.019	50.9%
cg27026615	-0.17	PTX3	3	Body	157156326	0.021	43.6%
cg08510264	0.16	IRS2	13	First Exon	110438288	0.012	32.9%
cg06493806	0.15	NFATC1	18	Body	77278806	0.031	44.9%

Delta = mean absolute difference in methylation percentage between cases and controls.

Frequency = percentage of times the CpGs are selected when EN is applied on a random subset of the initial dataset.

The CpG sites with a frequency higher than 50% are indicated in red.

The CpG sites analysed in Table 3 (those with a frequency higher than 50% and an association with MD in the control group in line with its protective effect) are indicated in grey.

**Table 3.** Detailed analysis of the CpG sites Selected for Validation (those reported in Table 2 with a frequency higher than 50% and an association with MD in the control group in line with its protective effect).

<i>IL1B</i>	Association with colon cancer*						Association with MD (only control group)**			
	ca	co	OR	95% CI	p-value		IMI category	coef	95% CI	P-value
<i>cg18773937</i>										
model A	157	157	0.69	(0.53,0.89)	0.005	<b>linear model</b>	0–2 (20)	reference		
model B	134	134	0.61	(0.44,0.84)	0.003		3–4 (71)	0.25	(–0.13,0.63)	0.200
model C	134	134	0.59	(0.42,0.84)	0.003		5–11 (51)	0.46	(0.05,0.87)	0.029
<i>SERPINE1</i>									p-trend	0.025
<i>cg17968347</i>										
model A	161	161	0.78	(0.62,0.98)	0.030	<b>linear model</b>	0–2 (22)	reference		
model B	138	138	0.64	(0.47,0.86)	0.003		3–4 (71)	0.04	(–0.09,0.18)	0.552
model C	138	138	0.58	(0.42,0.81)	0.001		5–11 (52)	0.12	(–0.03,0.28)	0.122
<i>RUNX1</i>									p-trend	0.089
<i>cg01265860</i>										
model A	161	161	0.77	(0.61,0.98)	0.036	<b>linear model</b>	0–2 (22)	reference		
model B	138	138	0.72	(0.54,0.97)	0.030		3–4 (71)	0.11	(–0.06,0.27)	0.197
model C	138	138	0.57	(0.37,0.84)	0.006		5–11 (52)	0.13	(–0.05,0.31)	0.154
<i>STAT3</i>									p-trend	0.203
<i>cg24312520</i>										
model A	161	161	0.78	(0.62,0.98)	0.037	<b>linear model</b>	0–2 (22)	reference		
model B	138	138	0.81	(0.62,1.07)	0.142		3–4 (71)	–0.02	(–0.21,0.15)	0.762
model C	138	138	0.82	(0.61,1.12)	0.209		5–11 (52)	0.10	(–0.09,0.29)	0.323
<i>NFATC1</i>									p-trend	0.171
<i>cg15363134</i>										
model A	161	161	0.82	(0.67,0.99)	0.043	<b>linear model</b>	0–2 (22)	reference		
model B	138	138	0.76	(0.59,0.98)	0.032		3–4 (71)	0.22	(0.04,0.41)	0.018
model C	138	138	0.74	(0.56,0.98)	0.038		5–11 (52)	0.05	(–0.15,0.25)	0.617
<i>RUNX3</i>									p-trend	0.738
<i>cg20674490</i>										
model A	160	160	0.70	(0.55,0.91)	0.008	<b>linear model</b>	0–2 (22)	reference		
model B	137	137	0.66	(0.48,0.91)	0.012		3–4 (71)	0.15	(–0.10,0.20)	0.246
model C	137	137	0.64	(0.46,0.90)	0.009		5–11 (51)	0.09	(–0.19,0.38)	0.511
<i>SERPINE1</i>									p-trend	0.719
<i>cg08053846</i>										
model A	134	134	1.28	(1.02,1.60)	0.037	<b>linear model</b>	0–2 (22)	reference		
model B	114	114	1.27	(0.95,1.70)	0.103		3–4 (71)	–0.10	(–0.43,0.23)	0.548
model C	114	114	1.30	(0.96,1.74)	0.084		5–11 (51)	–0.20	(–0.56,0.15)	0.262
									p-trend	0.242

ca = cases; co = controls.

\*Associations between DNA methylation and colon cancer are assessed using conditional logistic regression models with DNA methylation levels included as an independent variable and standardized to 1 standard deviation.

model A = crude model (adjusted for study centre, sex, age and seasonality by design)

model B = model A adjusted also for BMI, smoking status, physical activity, level of education and IMI.

model C = model B adjusted also for differential cell types.

\*\* Associations between the M-values of DNA methylation sites and IMI are assessed using a multivariate linear-mixed effect model with chip as random effect and IMI, sex, age, centre, BMI, smoking status, total physical activity, level of education and the differential cell types as fixed effects.

exposure and the disease. The approach involves to combine, within a prospective study, a search for biomarkers modified in subjects who develop the disease and a search for links of such biomarkers to past environmental exposures. In our study, we considered MD as the exposure, CC as the disease and DNA methylation levels of inflammation genes as the candidate molecular biomarkers. Among the 995 inflammation CpGs studied, we identified probes in the *IL1B*, *SERPINE1*, *RUNX1*, *STAT3*, *NFATC1* and *RUNX3* genes as good candidate molecular biomarkers. We implemented the discovery phase with an

epigenome-wide assay that is a powerful tool to cover an increasing number of CpG sites but is inherently imprecise and noisy [32]. To overcome this problem we performed validation and replication using a locus-specific methylation technique. Among the seven probes selected for validation, only two were confirmed and only one of the two showed associations in line with our hypothesis in an independent sample (*RUNX3*).

The human runt-related transcription factor 3 (*RUNX3*) maps in chromosomal locus 1p36 in a region that is frequently deleted in many types of



**Table 4.** Validation analysis: associations between colon cancer/MD and DNA methylation levels at *SERPINE1* and *RUNX3* CpG sites obtained by pyrosequencing in the validation set.

<i>SERPINE1</i>	Association with colon cancer*						Association with MD (only control group)**			
	ca	co	OR	95% CI	P-value		IMI category	coef	95% CI	P-value
cg17968347										
model A	47	47	0.82	(0.59,1.15)	0.257	<b>linear model</b>	0–2 (6)	reference		
model B	42	42	0.69	(0.45,1.05)	0.080		3–4 (23)	0.30	(–0.08,0.69)	0.117
model C	42	42	0.80	(0.37–1.76)	0.588		5–11 (15)	0.39	(–0.02,0.80)	0.061
<i>RUNX3</i>									p-trend	0.091
cg20674490										
model A	47	47	0.74	(0.47,1.16)	0.196	<b>linear model</b>	0–2 (6)	reference		
model B	42	42	0.59	(0.29,1.19)	0.144		3–4 (23)	0.39	(–0.20,0.98)	0.188
model C	42	42	0.40	(0.14–1.08)	0.071		5–11 (15)	0.09	(–0.52,0.72)	0.756
									p-trend	0.627

ca = cases; co = controls.

\*Associations between DNA methylation and colon cancer are assessed using conditional logistic regression models with DNA methylation levels included as an independent variable and standardized to 1 standard deviation.

model A = crude model (adjusted for study centre, sex, age and seasonality by design).

model B = model A adjusted also for BMI, smoking status, physical activity, level of education and IMI.

model C = model B adjusted for differential cell types.

\*\* Associations between M-values of DNA methylation sites and IMI are assessed using linear models adjusting for age, sex, centre, BMI, smoking status, physical activity, level of education and differential cell types.

**Table 5.** Replication analysis: Associations between colon cancer/MD and DNA methylation levels at *SERPINE1* and *RUNX3* CpG sites obtained by pyrosequencing in the replication set.

<i>SERPINE1</i>	Association with colon cancer*						Association with MD (only control group)**			
	ca	Co	OR	95% CI	P-value		IMI category	coef	95% CI	P-value
cg17968347										
model A	93	93	1.08	(0.82,1.42)	0.573	<b>linear model</b>	0–2 (17)	reference		
model B	87	87	1.12	(0.79,1.56)	0.522		3–4 (44)	0.005	(–0.22,0.23)	0.961
							5–11 (27)	0.10	(–0.17,0.36)	0.431
<i>RUNX3</i>									p-trend	0.377
cg20674490										
model A	92	92	0.80	(0.60,1.07)	0.132	<b>linear model</b>	0–2 (17)	reference		
model B	86	86	0.80	(0.57,1.13)	0.210		3–4 (44)	0.29	(–0.12,0.70)	0.162
							5–11 (27)	0.38	(–0.09,0.85)	0.114
									p-trend	0.133

ca = cases; co = controls.

\*Associations between DNA methylation and colon cancer are assessed using conditional logistic regression models with DNA methylation levels included as an independent variable and standardized to 1 standard deviation.

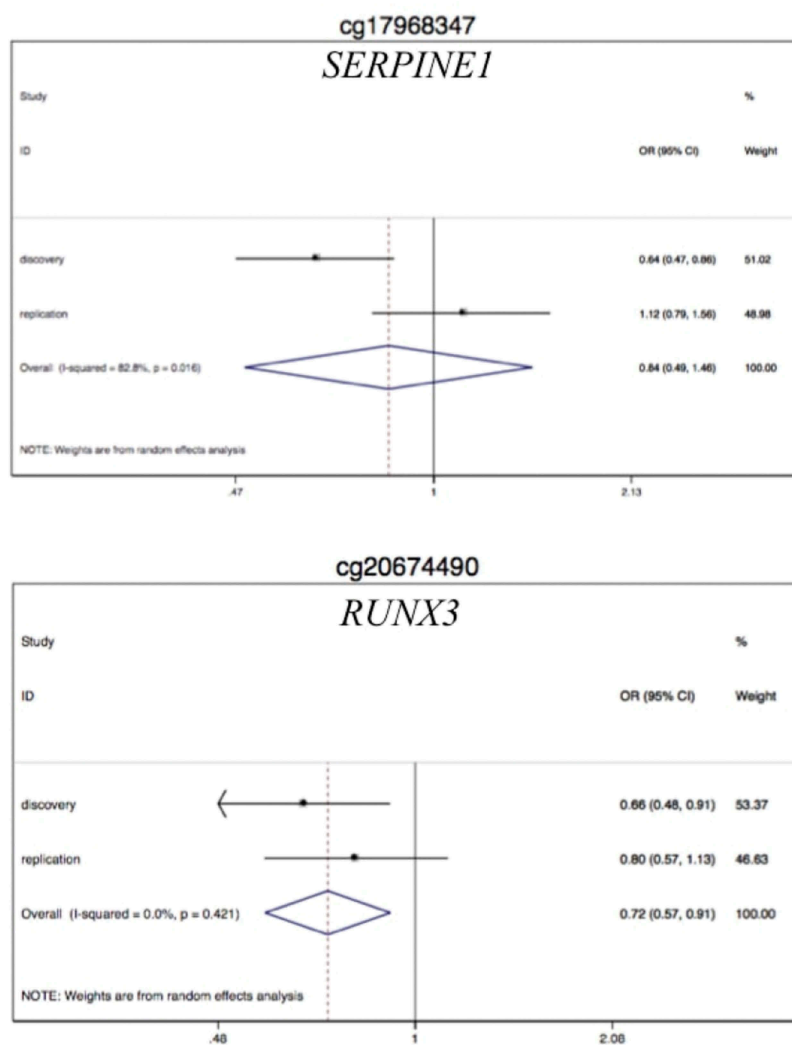
model A = crude model (adjusted for study centre, sex, age, year of recruitment and seasonality by design).

model B = model A also for BMI, smoking status, physical activity, level of education and IMI.

\*\* Associations between M-values of DNA methylation sites and IMI are assessed using linear models adjusting for age, sex, centre, BMI, smoking status, physical activity and level of education.

cancers. *RUNX3* has important functions in innate and adaptive immune cell types, in particular in inactivating *IL-23A* transcription, and has been associated with several immune-related diseases, in particular with an increased risk of ulcerative colitis [33] and Chron's disease [34]. Recent studies demonstrated that the control of immunity and inflammation exerted by *RUNX3* influences epithelial tumour development [35]. Furthermore, *RUNX3* is a tumour suppressor gene whose promoter hypermethylation was shown to be a key

mechanism of its inactivation [36]. In a study performed on 184 South Korean patients affected by gastric cancer, Zhang et al. found an association between the intake of different foods, in particular, fruit and egg, and *RUNX3* methylation [37]. At the opposite, no association among dietary items and *RUNX3* gene methylation has been found in a sample of 276 USA healthy women enrolled in a randomized controlled trial on breast cancer prevention [38]. For this reason, we think that *RUNX3* is a good candidate mediator between



**Figure 3.** Results of the mixed effects meta-analysis of the discovery and replication studies on the validated signals.

MD and CC, but further functional studies are needed to confirm this promising observation.

Few studies can be found in literature comparing different DNA methylation assays for biomarker development [39,40]. These studies show a good concordance between measurements of the two arrays, but they consider DNA methylation assessed in solid tissues or specific cell lines. To our knowledge, in only one study [41] based on DNA methylation of blood leukocyte, a validation of Illumina 450K array results with pyrosequencing technique is performed, but only five samples were analysed. In our study, we considered 94 samples for validation and 190 samples for replication. The fact that only one signal was confirmed emphasizes the importance of the validation and replication

phases employing an alternative methodology. These phases are essential in order to exclude technical errors and false-positive findings especially when the differences in methylation percentages are small most probably due to background noise.

Despite some limitations, this study has several strengths. First of all, we mention the use of the ‘meet-in-the-middle’ approach that we already applied in a similar study suggesting oxidative stress as the mediator in the association of air pollution and cardiovascular diseases [42]. In a study investigating the relation between diet and cancer, exposure assessment is always difficult and the specific mechanisms through which the nutrients exert their protective effect is largely unknown. By assessing the mediation role of

intermediate biomarkers we elucidated the biological mechanisms explaining this association.

Secondly, DNA methylation levels were measured in peripheral blood, providing a valuable source of information for low-grade inflammation. However, we were aware that heterogeneity in white blood cells could potentially confound DNA methylation measurements [43]. To address this problem we applied Houseman correction for cell composition in the association analyses [23].

Finally, we performed a case-control study nested in a prospective cohort, in which DNA methylation was assessed in peripheral blood collected at recruitment before the onset of cancer. The analysis of blood sample drawn years before the onset of disease prevents from reverse causality biases.

## Conclusions

In conclusion, our study is a first attempt to identify the biological mechanism behind the protection of MD on CC investigating the methylation levels of genes in circulating lymphocytes years before the onset of the disease. The results of the study indicate that DNA methylation of *RUNX3* gene may be a potential molecular mediator explaining the protective effect of MD on CC.

## Acknowledgments

The authors thank all participants in the Italian section of the EPIC study.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the Associazione Italiana per la Ricerca sul Cancro [IG2013-14410].

## ORCID

Francesca Fasanelli  <http://orcid.org/0000-0003-2340-2961>

Chiara Grasso  <http://orcid.org/0000-0002-2597-4990>

Silvia Polidoro  <http://orcid.org/0000-0003-2968-0575>

Sara Grioni  <http://orcid.org/0000-0002-5891-8426>

Rosario Tumino  <http://orcid.org/0000-0003-2666-414X>

Fulvio Ricceri  <http://orcid.org/0000-0001-8749-9737>

## References

- [1] Trichopoulou A, Costacou T, Bamia C, et al. Adherence to a Mediterranean diet and survival in a Greek population. *N Engl J Med.* 2003;348:2599–2608.
- [2] Sofi F, Cesari F, Abbate R, et al. Adherence to Mediterranean diet and health status: meta-analysis. *BMJ.* 2008;337:a1344.
- [3] Agnoli C, Grioni S, Sieri S, et al. Italian Mediterranean Index and risk of colorectal cancer in the Italian section of the EPIC cohort. *Int J Cancer.* 2013;132:1404–1411.
- [4] Rosato V, Guercio V, Bosetti C, et al. Mediterranean diet and colorectal cancer risk: a pooled analysis of three Italian case-control studies. *Br J Cancer.* 2016;115:862–865.
- [5] Torres Stone RA, Waring ME, Cutrona SL, et al. The association of dietary quality with colorectal cancer among normal weight, overweight and obese men and women: a prospective longitudinal study in the USA. *BMJ Open.* 2017;7:e015619.
- [6] Park SY, Boushey CJ, Wilkens LR, et al. High-quality diets associate with reduced risk of colorectal cancer: analyses of diet quality indexes in the multiethnic cohort. *Gastroenterology.* 2017;153:386–394.
- [7] Kontou N, Psaltopoulou T, Panagiotakos DB, et al. The Mediterranean diet in cancer prevention: a review. *J Med Food.* 2011;14:1065–1078.
- [8] Buckland G, Bach A, Serra-Majem L. Obesity and the Mediterranean diet: a systematic review of observational and intervention studies. *Obes Rev.* 2008;9(6):582–593.
- [9] Arnold M, Pandeya N, Byrnes G, et al. Global burden of cancer attributable to high body-mass index in 2012: a population-based study. *Lancet Oncol.* 2015;16:36–46.
- [10] Fasanelli F, Zugna D, Giraud MT, et al. Abdominal adiposity is not a mediator of the protective effect of Mediterranean diet on colorectal cancer. *Int J Cancer.* 2017;140:2265–2271.
- [11] Casas R, Sacanella E, Urpí-Sardà M, et al. Long-term immunomodulatory effects of a Mediterranean diet in adults at high risk of cardiovascular disease in the PREvención con DIeta MEDiterránea (PREDIMED) randomized controlled trial. *J Nutr.* 2016;146:1684–1689.
- [12] Koloverou E, Panagiotakos DB, Pitsavos C, et al.; ATTICA Study Group. Adherence to Mediterranean diet and 10-year incidence (2002–2012) of diabetes: correlations with inflammatory and oxidative stress biomarkers in the ATTICA cohort study. *Diabetes Metab Res Rev.* 2016;32:73–81.
- [13] Cannon J. Colorectal neoplasia and inflammatory bowel disease. *Surg Clin North Am.* 2015;95:1261–1269.
- [14] Santilli F, Boccatonda A, Davi G. Aspirin, platelets, and cancer: the point of view of the internist. *Eur J Intern Med.* 2016;34:11–20.

- [15] Scoccianti C, Ricceri F, Ferrari P, et al. Methylation patterns in sentinel genes in peripheral blood cells of heavy smokers: influence of cruciferous vegetables in an intervention study. *Epigenetics*. 2011;6:1114–1119.
- [16] Tammen SA, Friso S, Choi SW. Epigenetics: the link between nature and nurture. *Mol Aspects Med*. 2013;34:753–764.
- [17] Wang J, Hodes GE, Zhang H, et al. Epigenetic modulation of inflammation and synaptic plasticity promotes resilience against stress in mice. *Nat Commun*. 2018;9:477–491.
- [18] Palli D, Berrino F, Vineis P, et al. EPIC-Italy. A molecular epidemiology project on diet and cancer: the EPIC-Italy prospective study. Design and baseline characteristics of participants. *Tumori*. 2003;89:586–593.
- [19] Fasanelli F, Baglietto L, Ponzi E, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. *Nat Commun*. 2015 Dec;15(6):10192.
- [20] Agnoli C, Krogh V, Grioni S, et al. A priori-defined dietary patterns are associated with reduced risk of stroke in a large Italian cohort. *J Nutr*. 2011;141:1552–1558.
- [21] Wareham NJ, Jakes RW, Rennie KL, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaires used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutr*. 2003;6:407–413.
- [22] Ricceri F, Sacerdote C, Giraudo MT, et al. The association between educational level and cardiovascular and cerebrovascular diseases within the EPICOR study: new evidence for an old inequality problem. *PLoS One*. 2016;11:e0164130.
- [23] Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86–102.
- [24] Reid S, Tibshirani R. Regularization paths for conditional logistic regression: the clogitL1 package. *J Stat Softw*. 2014;58:1–23.
- [25] Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005;67:301–320.
- [26] Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol*. 1996;58:267–288.
- [27] Hoerl A, Kennard R. Ridge regression. *Encyclopedia of Statistical Sciences*. Vol. 8. New York: Wiley; 1998. p. 129–136.
- [28] Friedman HJ, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1–22.
- [29] Du P, Zhang X, Huang CC, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11:587–596.
- [30] Vineis P, Perera F. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiol Biomarkers Prev*. 2007;16:1954–1965.
- [31] Chadeau-Hyam M, Athersuch TJ, Keun HC, et al. Meeting-in-the-middle using metabolic profiling - a strategy for the identification of intermediate biomarkers in cohort studies. *Biomarkers*. 2011;16:83–88.
- [32] Michels KB, Binder AM, Dedeurwaerder S, et al. Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods*. 2013;10:949–955.
- [33] Guo C, Yao F, Wu K, et al. Chromatin immunoprecipitation and association study revealed a possible role of Runt-related transcription factor 3 in the ulcerative colitis of Chinese population. *Clin Immunol*. 2010;135:483–489.
- [34] Yamazaki K, Umeno J, Takahashi A, et al. A genome-wide association study identifies 2 susceptibility loci for Crohn's disease in a Japanese population. *Gastroenterology*. 2013;144:781–788.
- [35] Lotem J, Levanon D, Negreanu V, et al. Runx3 at the interface of immunity, inflammation and cancer. *Biochim Biophys Acta*. 2015;1855:131–143.
- [36] Goel A, Arnold CN, Tassone P, et al. Epigenetic inactivation of RUNX3 in microsatellite unstable sporadic colon cancers. *Int J Cancer*. 2004;112:754–759.
- [37] Zhang YW, Eom SY, Yim DH, et al. Evaluation of the relationship between dietary factors, CagA-positive *Helicobacter pylori* infection, and RUNX3 promoter hypermethylation in gastric cancer tissue. *World J Gastroenterol*. 2013;19:1778–1787.
- [38] Gillman AS, Gardiner CK, Koljack CE, et al. Body mass index, diet, and exercise: testing possible linkages to breast cancer risk via DNA methylation. *Breast Cancer Res Treat*. 2018;168:241–248.
- [39] Roessler J, Ammerpohl O, Gutwein J, et al. Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. *BMC Res Notes*. 2012;5:210–217.
- [40] BLUEPRINT consortium. Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. *Nat Biotechnol*. 2016;34:726–737.
- [41] Milenkovic D, Vanden Berghe W, Boby C, et al. Dietary flavanols modulate the transcription of genes associated with cardiovascular pathology without changes in their DNA methylation state. *PLoS One*. 2014;9:e95527.
- [42] Fiorito G, Vlaanderen J, Polidoro S, et al. EXPOsOMICS consortium. Oxidative stress and inflammation mediate the effect of air pollution on cardio- and cerebrovascular disease: A prospective study in nonsmokers. *Environ Mol Mutagen*. 2018;59:234–246.
- [43] Adalsteinsson BT, Gudnason H, Aspelund T, et al. Heterogeneity in white blood cells has potential to confound DNA methylation measurements. *PLoS One*. 2012;7:e46705.