Fine-grained investigation of the relationship between human nutrition and global DNA methylation patterns

Fabrizia Noro[1] 0000-0002-4754-9303, Annalisa Marotta[1] 0000-0002-9217-1506, Marialaura Bonaccio[1] 0000-0002-8082-6552, Simona Costanzo[1] 0000-0003-4569-1186, Federica Santonastaso[1], Sabatino Orlandi[1] 0000-0002-9498-173X, Alfonsina Tirozzi[1] 0000-0002-6052-1807, Roberta Parisi[1] 0000-0003-3789-3490, Amalia De Curtis[1] 0000-0001-5104-0318, Mariarosaria Persichillo[1], Francesco Gianfagna[2,3] 0000-0003-4615-0816, Augusto Di Castelnuovo[2] 0000-0001-9767-7998, Maria Benedetta Donati[1] 0000-0003-1747-5443, Chiara Cerletti[1] 0000-0002-4759-3179, Giovanni de Gaetano[1] 0000-0002-7823-1402, Licia Iacoviello[1,3]* 0000-0003-0514-5885, Alessandro Gialluisi[1#] 0000-0002-7388-4463 and Benedetta Izzi[1#] 0000-0002-3649-7931 on behalf of the Moli-sani study Investigators.

[1]Department of Epidemiology and Prevention, IRCCS NEUROMED, Pozzilli, Italy.

[2]Mediterranea Cardiocentro, Napoli, Italy.

[3]EPIMED Research Center, Department of Medicine and Surgery, University of Insubria, Varese, Italy.

# These authors equally contributed to the manuscript

*__Corresponding author__:

Licia Iacoviello, MD, PhD

Department of Epidemiology and Prevention, IRCCS Neuromed, Via dell'Elettronica, 86077 - Pozzilli (IS), Italy

Phone: +39 0865915246

e-mail: licia.iacoviello@moli-sani.org

Abbreviations used: 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; CVD, cardiovascular disease; MD, Mediterranean diet; MDS, Mediterranean Diet Score; OD, Optical Density; SE, Standard Error; RF, Random Forest.

Acknowledgements

Moli-sani Study Investigators

The enrolment phase of the Moli-sani Study was conducted at the Research Laboratories of the Catholic University in Campobasso (Italy), the follow up of the Moli-sani cohort is being conducted at the Department of Epidemiology and Prevention of the IRCCS Neuromed, Pozzilli, Italy.

Steering Committee: Licia Iacoviello*°(Chairperson), Giovanni de Gaetano* and Maria Benedetta Donati*.

Scientific secretariat: Marialaura Bonaccio*, Americo Bonanni*, Chiara Cerletti*, Simona Costanzo*, Amalia De Curtis*, Augusto Di Castelnuovo[§], Francesco Gianfagna°[§], Mariarosaria Persichillo*, Teresa Di Prospero* (Secretary).

Safety and Ethical Committee: Jos Vermylen (Catholic Univesity, Leuven, Belgio) (Chairperson), Ignacio De Paula Carrasco (Accademia Pontificia Pro Vita, Roma, Italy), Antonio Spagnuolo (Catholic University, Roma, Italy).

External Event adjudicating Committee: Deodato Assanelli (Brescia, Italy), Vincenzo Centritto (Campobasso, Italy).

Baseline and Follow-up data management: Simona Costanzo* (Coordinator), Marco Olivieri (Associazione Cuore Sano, Campobasso, Italy), Teresa Panzera*.

Data Analysis: Augusto Di Castelnuovo[§] (Coordinator), Marialaura Bonaccio*, Simona Costanzo*, Simona Esposito*, Alessandro Gialluisi*, Francesco Gianfagna°[§], Emilia Ruggiero*.

Biobank and biochemical laboratory: Amalia De Curtis* (Coordinator), Sara Magnacca[§].

Genetic laboratory: Benedetta Izzi* (Coordinator), Annalisa Marotta*, Fabrizia Noro*, Roberta Parisi*, Alfonsina Tirozzi*.

Recruitment staff: Mariarosaria Persichillo* (Coordinator), Francesca Bracone*, Francesca De Lucia (Associazione Cuore Sano, Campobasso, Italy), Cristiana Mignogna*, Teresa Panzera*, Livia Rago*.

Communication and Press Office: Americo Bonanni*.

Regional Health Institutions: Direzione Generale per la Salute - Regione Molise; Azienda Sanitaria Regionale del Molise (ASReM, Italy); Molise Dati Spa (Campobasso, Italy); Offices of vital statistics of the Molise region.

Hospitals: Presidi Ospedalieri ASReM: Ospedale A. Cardarelli – Campobasso, Ospedale F. Veneziale – Isernia, Ospedale San Timoteo - Termoli (CB), Ospedale Ss. Rosario - Venafro (IS), Ospedale Vietri – Larino (CB), Ospedale San Francesco Caracciolo - Agnone (IS); Casa di Cura Villa Maria - Campobasso; Ospedale Gemelli Molise - Campobasso; IRCCS Neuromed - Pozzilli (IS).

*Department of Epidemiology and Prevention, IRCCS Neuromed, Pozzilli, Italy

°Department of Medicine and Surgery, University of Insubria, Varese, Italy

§Mediterranea Cardiocentro, Napoli, Italy

Baseline Recruitment staff is available at https://www.moli-sani.org/?page_id=173

ABSTRACT

Purpose: Nutrition is an important, modifiable, environmental factor affecting human health by modulating epigenetic processes, including DNA methylation (5mC). Numerous studies investigated the association of nutrition with global and gene-specific DNA methylation and evidences on animal models highlighted a role in DNA hydroxymethylation (5hmC) regulation. However, a more comprehensive analysis of different layers of nutrition in association with global levels of 5mC and 5hmC is lacking. We investigated the association between global levels of 5mC and 5hmC and human nutrition, through the stratification and analysis of dietary patterns into different nutritional layers: adherence to Mediterranean Diet (MD), main food groups, macronutrients and micronutrients intake.

Methods: ELISA technique was used to measure global 5mC and 5hmC levels in 1,080 subjects from the Moli-sani cohort**.** Food intake during the 12 months before enrolment was assessed by using the semi-quantitative EPIC food frequency questionnaire. Complementary approaches involving both classical statistics and supervised machine learning analyses were used to investigate the associations between global 5mC and 5hmC levels and adherence to Mediterranean diet, main food groups, macronutrients and micronutrients intake.

Results: We found that global DNA methylation, but not hydroxymethylation, was associated with daily intake of zinc and vitamin B3. Random Forests algorithms predicting 5mC and 5hmC through intakes of food groups, macronutrients and micronutrients revealed a significant contribution of zinc, while vitamin B3 was reported among the most influential features.

Conclusion: We found that nutrition may affect global DNA methylation, suggesting a contribution of micronutrients previously implicated as cofactors in methylation pathways.

Keywords: global DNA methylation, Mediterranean diet, micronutrients, food groups, zinc, vitamin B3.

INTRODUCTION

Since the characterization of DNA methylation abnormalities in several human diseases, including cancer [1] and cardiovascular disease (CVD) [2,3], identifying environmental factors which may epigenetically affect the genome has become of utterly importance. In this regard, nutrigenomics has helped to identify the role of nutrients in influencing gene regulation [4,5] through DNA methylation in several phases of life, [6] including childhood [7,8] and elderly [9].

Diet (high-fat, high-sugar) or food components (amino acids, bioactive compounds) can affect genome function and DNA methylation-dependent gene expression by influencing the folate-mediated one-carbon metabolism or the trans-methylation pathways [6]. Both polyphenols and vitamins (i.e. folate [10]), which are particularly present in healthy dietary patterns such as the Mediterranean Diet (MD) [11], are known to specifically act as epigenetic modulators by targeting the DNA methylation and DNA methyl-transferases pathway [12].

Thanks to the continuous development of different and more specific analytical technologies, both 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), resulting from 5mC oxidation via Ten-Eleven Translocation (TET) proteins

dependent demethylation [13], can be studied from a single locus scale to the genome-wide and global level [14], quantifying the average status of these modifications across the whole genome [15].

Global DNA methylation, an overall and accepted marker of environmental cues on the genome [16] has been evaluated by measuring 5mC and 5hmC via liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) or enzyme-linked immunosorbent assay (ELISA) [14], or, more frequently, via the characterization of the 5mC status at the Long Interspersed Element-1 (LINE-1) or the Short Interspersed Element (SINE)[15].

A number of population-based studies investigated the link between LINE-1 methylation and micro [17] and macronutrients intake, as well as food groups [18], up to specific dietary patterns [19]. Unlike 5mC, 5hmC has been less studied in the context of nutrition, although it is now accepted as the sixth DNA base in mammalian genomic DNA [20]. Indeed, it has been found widely distributed in many human tissues, especially in the brain [21]. Interestingly, a role for 5hmC in neuronal development has been recently demonstrated and genes that have acquired 5hmC during aging were associated with age-related neurodegenerative disorders [22]. A recent study showed that a high fat diet-induced metabolic disorder stimulates neural 5hmC remodelling in mice, with effects on mitochondrial dysfunction and neural impairment [23]. In the same line, Ciccarone et al. reported that the 5hmC levels are dynamically regulated in mice heart by a chronic high dietary fat intake, revealing a role of DNA hydroxymethylation in obesity-related heart pathophysiology [24].

Despite this experimental evidence, a more comprehensive population-based study analysing the relationship between the different layers of nutrition and global DNA methylation - considering both 5mC and 5hmC - is lacking. The assessment of global

5mC and 5hmC levels in nutritional studies could be important to identify potentially different global DNA methylation patterns in response to the intake of different nutrients. This could allow understanding the effect of individual food components or specific dietary patterns on human health and disease.

Here, we performed a fine-grained investigation of the relation between the global levels of 5mC and 5hmC and nutrition, through stratification and analysis of diet into three different nutritional layers: adherence to MD and intake of the main food groups, macronutrients and micronutrients, in a sub-cohort of the Italian Moli-sani study [25].

SUBJECTS AND METHODS

Study population

The study population was composed of subjects participating in the Moli-sani study (N=24,325; 49.20% men; ≥ 35 years) who were randomly recruited from the general population of Molise Region, between 2005 and 2010. The study design and procedures have been previously described [25,26]. For this study, we used data from a randomly selected sub-cohort of 1,160 subjects. Subjects with dietary questionnaires judged as unreliable by the interviewers or with missing values in the studied variables were excluded from the analysis.

The Moli-sani study complies with the Declaration of Helsinki and was approved by the Ethical Committee of the Catholic University in Rome, Italy. All participants provided written informed consent.

Dietary assessment

Food intake during the 12 months before enrolment was assessed by using the semi-quantitative EPIC food frequency questionnaire (FFQ) validated and adapted to the Italian population [27,28], for a total of 188-food items that were classified into 74 predefined food groups on the basis of similar nutrient characteristics or culinary usage. The EPIC questionnaire also allowed to compute the daily energy (Kcal/day) and alcohol intake (g/day) for the subjects assessed.

The Nutrition Analysis of FFQ (NAF) [29] was used to convert dietary data into frequencies of consumption and average daily quantities of food (g/day), macronutrients (g/day), micronutrients (mg/day or µg/day) and energy intake (kcal/day). NAF was linked to the Italian food composition tables (http://www.inran.it/646/tabelle_di_composizione_degli_alimenti.html).

Adherence to the traditional Mediterranean diet (MD) was determined through the Mediterranean Diet Score (MDS) developed by Trichopoulou et al. [30]. The MDS was obtained by assigning 1 point to healthy foods (fruits and nuts, vegetables, legumes, fish, cereals, monounsaturated (MUFA) to saturated fatty acid ratio (SFA)) whose consumption was above the sex-specific medians of intake of the Moli-sani study population, free from CVD, cancer and diabetes and then applied to the whole population; foods presumed to be detrimental (meat and dairy products) were scored positively if their consumption was below the median. All other intakes received 0 points. For ethanol, men who consumed 10–50 g/d and women who consumed 5–25 g/d received 1 point; otherwise, the score was 0. The MDS ranged from 0 to 9 (the latter reflecting maximal adherence).

DNA extraction and epigenetic measures

Buffy coat DNA was extracted through a silica matrix-based method, as described in [31]. Of the 1,160 DNA samples from the subjects selected from the Moli-sani cohort, 1,140 had good quality to perform the methylation analysis (see below).

Global levels of 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) were measured using the MethylFlash Global DNA Methylation (5mC) ELISA Easy Kit (colorimetric) and the MethylFlash Hydroxymethylated DNA 5-hmC Quantification Kit (colorimetric) (EpiGentek), according to the manufacturer's instructions. Quality control and statistical analyses of methylation measurements were carried out in R (The R Project, 2020; https://www.r-project.org/) [32].

Overall, 1,214 samples (including 1,140 original and 74 duplicate samples) were assessed for 5mC and 5hmC levels. Samples with absorbance Optical Density (OD) values below the mean of negative controls plus 2 Standard Deviations (SDs) for both 5mC and 5hmC were considered of bad quality and set to missing. Based on this criterion, we did not detect any bad quality sample for 5mC (OD > 0.089), while 7 samples were set to missing for 5hmC (OD > 0.099). After these filters, 1,140 and 1,135 unique samples were retained for 5mC and for 5hmC, respectively, which were standardized within plates. Additionally, outlier samples (i.e. with absolute values of standardized methylation levels above 3 Standard Deviations, 17 for 5mC and 2 for 5hmC) were removed from analyses, as well as 56 and 58 samples (respectively) corresponding to prevalent CVD cases in the extracted subcohort to avoid potential biases by reverse causality of CVD on methylation levels [33]. After QC, 1,067 samples with 5mC measures and 1,075 samples with 5hmC measures were left for the following analyses. Both measures showed distributions approaching normality (**Fig. S1a, b**).

Statistical Analyses

First, we analysed the association between adherence to Mediterranean diet [30] and standardized global methylation levels, adjusting for sex, age, energy intake (Kcal/day), educational level (none or primary/lower secondary/upper secondary/post-secondary school completed), white blood cell (granulocyte, monocyte and lymphocytes) fractions and for additional variables showing univariate trends of association with both exposure and outcome ($P<0.2$), which included smoking habits (subjects were assigned to three categories: smokers, ex-smokers, i.e. subjects who quitted at least one year before the interview, and non-smokers), leisure time physical activity (assessed through a structured questionnaire and expressed as daily energy expenditure in metabolic equivalent task-hours [MET-h/day] [34]), abdominal obesity based on waist-to-hip ratio, dyslipidaemia, cancer and, diabetes(waist circumference [cm] was measured in the middle between the 12^{th} rib and the iliac crest, while hip circumference [cm] was measured around the buttocks. Waist-to-hip ratio [WHR] was calculated, and the resulting measure of *abdominal obesity* was inferred as a dichotomous variable [Yes/No], defining as obese men with WHR ≥ 0.90 and women with WHR ≥ 0.85 [World Health Organization, 2011]. Prevalent diabetes, and dyslipidaemia were defined as dichotomous variables [Yes/No], based on the reported and verified use of specific drugs for their treatment, while prevalent CVD and cancer classification was based merely on self-report of medical history of the disease, possibly supported by medical documentation or by the use of specific drugs.

Then, we performed multivariable linear regressions to model 5mC and 5hmC as a function of daily intake of nutritional variables at three different layers. First, we tested association with the intake of eight food groups, namely vegetables, fruits,

cereals, fish, legumes, dairy products, meat (g/day), and the ratio between monounsaturated and saturated fats (MUFA-SFA ratio). Then we tested association with three main classes of macronutrients, including total proteins, lipids and available carbohydrates (g/day). Finally, we modelled the relation with the daily intake of seventeen different micronutrients, including Iron, Calcium, Sodium, Potassium, Phosphorus, Zinc, vitamin B1, B2, B3, B6, C and E (expressed in mg/day), as well as with the intake of vitamin B9, A1, D, Beta-carotene and Selenium (µg/day) (**Fig. 1**). All multivariable models were further adjusted for alcohol drinking habits (classifying subjects in current-/former-/occasional-/never drinkers- and treating missing values as an additional dummy class). These models were performed through lm() function in R, inputting all the nutritional variables of a given nutritional layer together. To avoid potential bias implied by multicollinearity, we carried out multivariable stepwise regressions through the stepAIC() function of the *MASS* package in R [35], with (default) "both" option. This kept within each model only those nutritional variables significantly contributing to an increase in the total variance explained by the model - in spite of the addition of a parameter to the regression – allowing to "clean" the models for potential collinearity bias introduced by the other nutritional variables.

To reduce the risk of detecting false positives – which is high in the presence of a large number of statistical tests [36] and to identify only robust associations between the multiple nutritional intakes tested and epigenetic modifications, we applied a correction for testing of multiple nutritional variables in the different layers, using a matrix spectral decomposition of their correlation matrices (Resumed in **Fig. S2**) in MatSpD (http://gump.qimr.edu.au/general/daleN/matSpD/) [37]. This did not detect any reduction in the number of latent variables to correct for at the food group level,

while two and seven main latent variables could be extracted from the macronutrients and micronutrients analysed, respectively. As for the methylation measures, we conservatively adjusted for two independent measures tested, in view of their moderate correlation (Pearson's r = 0.51). Therefore, a Bonferroni corrected statistical significance was set to $\alpha = 0.05/(8*2) = 3.1x10^{-3}$ for the analysis of food groups, $\alpha = 0.05/(2*2) = 0.012$ for macronutrients and $\alpha = 0.05/(7*2) = 3.6x10^{-3}$ for micronutrients. Since the three nutritional levels analysed are intertwined and the nutritional variables within each level cannot be considered fully independent, we did not correct significance thresholds for the number of nutritional levels or the total number of nutritional variables tested.

In linear models revealing significant associations (i.e. 5-mC vs micronutrients intake), we tested potential interactive effects of nutritional intakes, testing those micronutrients which showed the most significant and consistent associations both in classical statistical (linear regression) and in machine learning models (see below). Specifically, this hypothesis was tested for vitamin B3, first in a two-way interaction with zinc, and then in a three-way interaction with zinc and phosphorous.

Machine learning analyses

We aimed at identifying the most influential nutritional intakes in the prediction of 5mC and 5hmC within a non-linear setting, taking into account potential synergistic effects and more complex relationships. For this purpose, we built two random forest (RF) algorithms to predict the level of 5mC and 5hmC, respectively, based on food groups, macronutrient and micronutrient intakes tested above. RF algorithms represent supervised machine learning approaches based on the construction of

multiple decision trees to estimate a label as accurately as possible and are ideal in the presence of a high number of predictors (also known as features) [38].

Nutritional intakes underwent min-max normalization before analysis. The resulting dataset (N=1,067 and 1,075 for 5mC and 5hmC, respectively) was divided in a random training and a test set with a 70:30 ratio.Then we performed hyperparameter tuning through the train() function of the *caret* package (https://CRAN.R-project.org/package=caret), in a five-fold cross validation setting, to optimize the accuracy (R-squared) of the algorithm over two varying parameters: the number of variables randomly sampled as candidate predictors at each node split in the decision tree (mtry, varying between 1 and 15), and the number of trees to grow in the random forest (ntree alternative values: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000). Finally, we trained the optimized models within the training set (mtry=2, ntree=1000 for 5mC and mtry=7, ntree=300 for 5hmC), and built them through the randomForest() function of the homonymous package in R [39].

Then we used the optimized trained models to predict the labels (5mC and 5hmC) in the independent test sets, and performed a variable importance analysis within each model, through the importance() function. This reveals the importance of each intake variable i) based on permutation feature importance (PFI) analysis, shuffling measures of one nutrient intake at a time and then comparing the loss function (Mean Squared Error between actual and predicted label, or MSE) of the perturbed RF model with that of the full model (i.e. with no permuted feature). To make this analysis more inferential, we applied the PIMP() and PimpTest() functions of the *vita* package [40] to have a significance test for each feature importance. Only those nutritional intakes showing highest increase in MSE in permuted models and a significant importance P-value were considered as statistically influential on the

prediction of the methylation measures. For this analysis, the significance threshold was corrected for seventeen total latent intakes and two independent methylation measures tested, based on computations reported above ($\alpha = 0.05/((8+7+2)*2) = 1.5\times10^{-3}$).

RESULTS

The characteristics of the analysed sub-cohort (N=1,080 with at least an epigenetic measure available) are summarized in **Table 1**. Compared to the Moli-sani study, sex ratio was similar (48% men), but the analysed sub-cohort was slightly younger (mean (SD) age 54.9 (11.5) year vs 55.8 (12.0) years, $p < .0001$), due to the removal of prevalent CVD cases. Similarly, in the analyzed sub-cohort there was a lower prevalence of diabetes ($p = 0.02$) and hyperlipidaemia, as well as a higher calory intake and a slightly higher MDS ($p < 0.0001$). Overall, there was no systematic difference between the analysed sub-cohort and the whole Moli-sani population, except those due to removal of CVD cases. Raw univariate associations of 5mC and 5hmC with prevalent chronic health conditions are reported in Table S1.

We present below association p-values after Bonferroni correction, obtained by multiplying raw association p-values for the number of methylation measures (two) and of latent variables tested at each nutritional level (eight for food groups, two for macronutrients and seven for micronutrients), where applicable. Raw association p-values (before Bonferroni correction) are reported in the tables (see below). We observed an inverse although not significant association between the adherence to MD and global methylation levels (5mC) (standardized β (Standard Error) = -0.049 (0.028), $p = 0.16$). Multivariable association analyses modelling 5mC as a function of

the daily intake of eight different food groups did not reveal any statistically significant association surviving Bonferroni correction for multiple testing (**Table 2a**) as did the analysis of macronutrients (**Table 2b**). However, a negative association between global methylation and cereals intake approached statistical significance (β (SE) = -0.0011 (0.0004), p = 0.06). In the analysis of micronutrients we detected a positive, statistically significant association of methylation levels with daily intake of zinc (β (SE) = 0.072 (0.024), p = 0.04). Additional nominally significant associations were observed with the intake of vitamin B3 (-0.042 (0.016)) and phosphorus (-0.0005 (0.0002)), as well as vitamin D, sodium and vitamin B6 (see **Table 2c**). However, these did not survive correction for multiple testing ($\alpha = 3.6x10^{-3}$), which did not allow us to rule out a potential type I error (false positive) bias. Interaction analyses of the most associated micronutrient intakes revealed no significant associations of zinc*vitamin B3 (two-way) and of zinc*vitamin B3*phosphorus (three-way) interaction terms with 5mC (p = 0.49 and 0.77, respectively).

As for global hydroxymethylation (5hmC), no food group or macronutrient was retained in stepwise regression models. However, in the stepwise regression of micronutrients intake two variables were retained, namely sodium and iron, which, however, did not show any significant association with 5hmC levels (β (SE) = -9.0 (5.8)$\times 10^{-5}$, p = 0.84 and β (SE) = 0.033 (0.018), p = 0.49, respectively; see **Table 3**).

When we analysed non-linear relationships through independent RF algorithms for 5mC and 5hmC prediction, feature importance analysis revealed prominent intakes in the prediction of methylation measures. Vitamin B3 (niacin), phosphorus and vitamin B1 were the most important nutritional intakes in the prediction of 5mC, showing a >16% increase in the average loss function of the permuted algorithms compared to the original random forest (**Fig. 2a**). However, only vitamin B3 reached

376 statistical significance surviving correction for multiple testing ($p < 10^{-16}$). On the
377 other hand, total lipids intake was the most important variable in the prediction of
378 5hmC, but was associated only with a ~11% increase in the average MSE of
379 perturbed models (**Fig. 2b**), and did not reach statistical significance ($p > 0.05$). Still,
380 the original RF models deployed explained a relatively low fraction of variance both
381 for 5mC and 5hmC (R^2 of actual vs predicted measure in linear regression ≤ 1 %).

382

383 **DISCUSSION**

384 Our study shows that global DNA methylation but not hydroxymethylation, measured
385 in a general population sub-cohort of Italian adults, is associated with specific
386 micronutrient intakes, through complementary approaches involving both classical
387 statistics and supervised machine learning analyses.

388 Multivariable stepwise regressions revealed a significant positive association of 5mC
389 levels with the daily intake of zinc. Zinc is involved in a wide range of key biological
390 processes such as neurological function, reproduction, development [41,42], antiviral
391 [43] immunity and inflammation [44-46]. Moreover, it plays a role in inflammation-
392 related physiological processes - like aging [47] - and health conditions like
393 neurodegenerative disorders [48,49], diabetes [50], cardiovascular disease [51] and
394 cancer [52]. It has been suggested that the role of zinc in human pathology is
395 thought to be mainly dependent on its function as epigenetic regulator [53]. Indeed,
396 zinc has been identified as a regulatory component of the function of over 3,000
397 among transcription factors and enzymes [54,55] including the DNA
398 methyltransferases [56], responsible for the transfer of methyl groups to the DNA
399 strands [57]. A CXXC domain and a plant homeodomain region have been described

400 to be part of DNMT1 and DNMT3 protein structure, respectively [58,59]. These
401 regions depend upon zinc binding to make the DNMTs catalytically active [58,59].
402 Furthermore, zinc was found to be involved in methionine synthase and betaine
403 homocysteine methyltransferase [60,61], important in the regulation of DNA
404 methylation. Our data represent a step forward in the understanding of the effect of
405 zinc intake on DNA methylation and support the importance of nutritional
406 interventions as complementary disease treatment or as prevention strategy [62-65].
407 To better dissect the link between the global methylation and hydroxymethylation
408 levels and the intake of nutrients, we also analysed non-linear relationships using
409 supervised machine learning algorithms. Despite these models explained a small
410 fraction of variance in 5mC and 5hmC, these revealed an important contribution of
411 vitamin B3 intake to the prediction of 5mC levels. Vitamin B3 actually covers two
412 different compounds, namely nicotinic acid (pyridine-3-carboxylic acid) and
413 nicotinamide (nicotinic acid amide). Nicotinamide is biosynthetically converted to
414 nicotinamide adenine dinucleotide (NAD+), nicotinamide adenine dinucleotide
415 phosphate (NADP+) and their respective reduced forms (NAD(P)H). These cofactors
416 are central in cellular homeostasis and growth for their roles in many important
417 biological functions and redox reactions [66]. It is known that in humans the excess
418 of nicotinamide is degraded mainly through S-adenosylmethionine-dependent
419 methylation, catalysed by nicotinamide N-methyltransferase [67]. Therefore, the
420 excess of intake of nicotinamide may increase the consumption of methyl-group
421 resources and affect other S-adenosylmethionine-dependent methylation reactions
422 by competing for the limited methyl-group pool, possibly including DNA methylation
423 [68]. Accordingly, it has been observed that nicotinamide supplementation induces
424 epigenetic changes in developing rats [69] and its maternal supplementation causes

global DNA hypomethylation and gene expression changes in foetal rats [70]. Although this functional evidence provides further support to our observational finding, a more in-depth investigation on the role of niacin intake in regulating human DNA methylation and its possible effect on gene expression is needed, in light of its commonly used utilization to fortify foods like bread [71].

In spite of the partial discordance between the results of the multivariable regressions and of the random forest models, zinc was listed among the most predictive features also in machine learning analyses, while vitamin B3 was the second most associated micronutrient in linear 5mC prediction. Of note, this partial discrepancy may be well explained by the different settings and relationships modelled among the different nutritional intakes and 5mC, namely linear and analysing single nutritional levels in multivariable regressions *vs.* more complex and analysing all levels together in random forest approaches. Conversely, we observed no significant associations with global hydroxymethylation levels, neither in a linear nor in a non-linear setting. Since this represents the first attempt to test 5hmC for association with nutritional intakes in humans, we have no terms of comparison and further studies are needed to corroborate or confute this lack of evidence.

Our analysis revealed no significant association between the adherence to MD and global DNA methylation patterns, neither with 5mC nor with 5hmC. Adherence to healthy dietary patterns has been previously associated with LINE-1 methylation levels [18,19]. In particular, a cross-sectional study of 349 non-pregnant healthy women from Southern Italy, reported that the adherence to a dietary pattern characterized by a high intake of vegetables and fruits, was positively associated with LINE-1 methylation [19]. In the same study, the authors observed a significant positive correlation of LINE-1 methylation with “healthy” foods —such as wholemeal

bread, cereals, fish, fruit, raw and cooked vegetables, legumes, and soup— and a negative correlation with the intake of vegetable oil [19]. In another study analysing LINE-1 methylation measured in peripheral blood leukocytes from 161 healthy subjects [18], Zhang and colleagues showed that a “prudent” dietary pattern is associated with a lower prevalence of DNA hypomethylation. Furthermore, they observed that subjects with lower LINE-1 methylation consumed more saturated fats than those with higher levels [18]. Our data, in combination with some recent studies [72,73], might indicate that adherence to MD is rather affecting gene specific or repeated element DNA methylation than global DNA methylation as we measured. Although our study represents one of the largest and most comprehensive association analyses between human nutrition and global DNA methylation and hydroxymethylation patterns, it presents some limitations. First, the cross-sectional design does not allow inferring the causality links between nutritional intakes and methylation patterns. Second, we cannot exclude that by measuring global methylation via a different technique we would identify different relationships to the ones described in this study. Third, since we used only a global measure of DNA methylation/hydroxymethylation rather than focusing on specific genes, it is difficult to understand the functional meaning of these associations. However, this is to be intended only as a preliminary analysis of methylation patterns and their potential environmental influences in the Moli-sani cohort and longitudinal studies in larger sub-cohorts, focusing on specific genes, are underway. Still, the complementary approaches used here suggests that nutrition, in particular micronutrients intakes, may affect the global methylation status of DNA in humans. Functional studies are now warranted to better understand the role of both individual nutrients and of their

combination in specific dietary patterns, to better define their effect on DNA methylation and on related health conditions.

DECLARATIONS

Funding: the enrolment phase of the Moli-sani Study was supported by research grants from the Pfizer Foundation (Rome, Italy), the Italian Ministry of University and Research (MIUR, Rome, Italy)–Programma Triennale di Ricerca, Decreto no.1588 and Instrumentation Laboratory, Milan, Italy. FN, AG and BI were supported by Fondazione Umberto Veronesi.

Conflicts of interest

The authors declare that they have no conflict of interest.

Ethics approval and consent to participate

The Moli-sani study complies with the Declaration of Helsinki and was approved by the ethical committee of the Catholic University in Rome, Italy. All participants provided written informed consent.

Consent for publication

Not applicable.

Availability of data and material

The data underlying this article will be shared on reasonable request to the corresponding author. The data are stored in an institutional repository (https://repository.neuromed.it) and access is restricted by the ethical approvals and the legislation of the European Union.

Code availability

Not applicable.

Authors' contributions

BI, LI and AG designed the research; FN conducted the research; AM and FS contributed to the methylation experiments; MB, SC, AT, RP, ADeC and MP provided essential materials; AG, FS, FG and SO analysed data and performed statistical analysis; FN, BI and AG wrote the paper; BI, AG, and LI had primary responsibility for final content; CC, MBD, GdG, ADiC and LI conceived the Moli-sani study; All authors read and approved the final manuscript.

REFERENCES

1. Hoffmann MJ, Schulz WA (2005) Causes and consequences of DNA hypomethylation in human cancer. Biochem Cell Biol 83 (3):296-321. doi:10.1139/o05-036
2. Baccarelli A, Wright R, Bollati V, Litonjua A, Zanobetti A, Tarantini L, Sparrow D, Vokonas P, Schwartz J (2010) Ischemic heart disease and stroke in relation to blood DNA methylation. Epidemiology 21 (6):819-828. doi:10.1097/EDE.0b013e3181f20457
3. Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW (2010) DNA methylation as a biomarker for cardiovascular disease risk. PLoS One 5 (3):e9692. doi:10.1371/journal.pone.0009692
4. Caradonna F, Consiglio O, Luparello C, Gentile C (2020) Science and Healthy Meals in the World: Nutritional Epigenomics and Nutrigenetics of the Mediterranean Diet. Nutrients 12 (6). doi:10.3390/nu12061748
5. Friso S, Udali S, De Santis D, Choi SW (2017) One-carbon metabolism and epigenetics. Mol Aspects Med 54:28-36. doi:10.1016/j.mam.2016.11.007
6. Kadayifci FZ, Zheng S, Pan YX (2018) Molecular Mechanisms Underlying the Link between Diet and DNA Methylation. Int J Mol Sci 19 (12). doi:10.3390/ijms19124055
7. Pauwels S, Ghosh M, Duca RC, Bekaert B, Freson K, Huybrechts I, Langie SAS, Koppen G, Devlieger R, Godderis L (2017) Maternal intake of methyl-group donors affects DNA methylation of metabolic genes in infants. Clin Epigenetics 9:16. doi:10.1186/s13148-017-0321-y
8. Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, Slagboom PE, Heijmans BT (2009) DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Hum Mol Genet 18 (21):4046-4053. doi:10.1093/hmg/ddp353
9. Field AE, Robertson NA, Wang T, Havas A, Ideker T, Adams PD (2018) DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. Mol Cell 71 (6):882-895. doi:10.1016/j.molcel.2018.08.008
10. Crider KS, Yang TP, Berry RJ, Bailey LB (2012) Folate and DNA methylation: a review of molecular mechanisms and the evidence for folate's role. Adv Nutr 3 (1):21-38. doi:10.3945/an.111.000992
11. Bonaccio M, Pounis G, Cerletti C, Donati MB, Iacoviello L, de Gaetano G, Investigators M-SS (2017) Mediterranean diet, dietary polyphenols and low grade inflammation: results from the MOLI-SANI study. Br J Clin Pharmacol 83 (1):107-113. doi:10.1111/bcp.12924
12. Shankar S, Kumar D, Srivastava RK (2013) Epigenetic modifications by dietary phytochemicals: implications for personalized nutrition. Pharmacol Ther 138 (1):1-17. doi:10.1016/j.pharmthera.2012.11.002
13. Wu X, Zhang Y (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet 18 (9):517-534. doi:10.1038/nrg.2017.33
14. Li S, Tollefsbol TO (2020) DNA methylation methods: global DNA methylation and methylomic analyses. Methods. doi:10.1016/j.ymeth.2020.10.002
15. Vryer R, Saffery R (2017) What's in a name? Context-dependent significance of 'global' methylation measures in human health and disease. Clin Epigenetics 9:2. doi:10.1186/s13148-017-0311-0
16. Ferrari L, Carugno M, Bollati V (2019) Particulate matter exposure shapes DNA methylation through the lifespan. Clin Epigenetics 11 (1):129. doi:10.1186/s13148-019-0726-x
17. Perng W, Rozek LS, Mora-Plazas M, Duchin O, Marin C, Forero Y, Baylin A, Villamor E (2012) Micronutrient status and global DNA methylation in school-age children. Epigenetics 7 (10):1133-1141. doi:10.4161/epi.21915
18. Zhang FF, Morabia A, Carroll J, Gonzalez K, Fulda K, Kaur M, Vishwanatha JK, Santella RM, Cardarelli R (2011) Dietary patterns are associated with levels of global genomic DNA methylation in a cancer-free population. J Nutr 141 (6):1165-1171. doi:10.3945/jn.110.134536

19. Barchitta M, Maugeri A, Magnano San Lio R, Favara G, La Rosa MC, La Mastra C, Quattrocchi A, Agodi A (2019) Dietary Patterns are Associated with Leukocyte LINE-1 Methylation in Women: A Cross-Sectional Study in Southern Italy. Nutrients 11 (8). doi:10.3390/nu11081843
20. Shi DQ, Ali I, Tang J, Yang WC (2017) New Insights into 5hmC DNA Modification: Generation, Distribution and Function. Front Genet 8:100. doi:10.3389/fgene.2017.00100
21. Li W, Liu M (2011) Distribution of 5-hydroxymethylcytosine in different human tissues. J Nucleic Acids 2011:870726. doi:10.4061/2011/870726
22. Lopez V, Fernandez AF, Fraga MF (2017) The role of 5-hydroxymethylcytosine in development, aging and age-related diseases. Ageing Res Rev 37:28-38. doi:10.1016/j.arr.2017.05.002
23. Liu R, Chen L, Wang Z, Zheng X, Wang Y, Li H, Noda M, Liu J, Long J (2020) Downregulation of the DNA 5-hydroxymethylcytosine is involved in mitochondrial dysfunction and neuronal impairment in high fat diet-induced diabetic mice. Free Radic Biol Med 148:42-51. doi:10.1016/j.freeradbiomed.2019.12.042
24. Ciccarone F, Castelli S, Ioannilli L, Ciriolo MR (2019) High Dietary Fat Intake Affects DNA Methylation/Hydroxymethylation in Mouse Heart: Epigenetic Hints for Obesity-Related Cardiac Dysfunction. Mol Nutr Food Res 63 (4):e1800970. doi:10.1002/mnfr.201800970
25. Iacoviello L, Bonanni A, Costanzo S, De Curtis A, Di Castelnuovo A, Olivieri M (2007) The Moli-sani project, a randomized, prospective cohort study in the Molise region in Italy; design, rationale and objectives. Ital J Public Health 4:110-118. doi:10.2427/5886
26. Di Castelnuovo A, Costanzo S, Persichillo M, Olivieri M, de Curtis A, Zito F, Donati MB, de Gaetano G, Iacoviello L, Investigators M-SP (2012) Distribution of short and lifetime risks for cardiovascular disease in Italians. Eur J Prev Cardiol 19 (4):723-730. doi:10.1177/1741826711410820
27. Bonaccio M, Di Castelnuovo A, Costanzo S, Ruggiero E, De Curtis A, Persichillo M, Tabolacci C, Facchiano F, Cerletti C, Donati MB, de Gaetano G, Iacoviello L, Moli-sani Study I (2019) Chili Pepper Consumption and Mortality in Italian Adults. J Am Coll Cardiol 74 (25):3139-3149. doi:10.1016/j.jacc.2019.09.068
28. Pisani P, Faggiano F, Krogh V, Palli D, Vineis P, Berrino F (1997) Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres. Int J Epidemiol 26 Suppl 1:S152-160
29. Pala V, Sieri S, Palli D, Salvini S, Berrino F, Bellegotti M, Frasca G, Tumino R, Sacerdote C, Fiorini L, Celentano E, Galasso R, Krogh V (2003) Diet in the Italian EPIC cohorts: presentation of data and methodological issues. Tumori 89 (6):594-607
30. Trichopoulou A, Costacou T, Bamia C, Trichopoulos D (2003) Adherence to a Mediterranean diet and survival in a Greek population. N Engl J Med 348 (26):2599-2608. doi:10.1056/NEJMoa025039
31. Malferrari G, Monferini E, DeBlasio P, Diaferia G, Saltini G, Del Vecchio E, Rossi-Bernardi L, Biunno I (2002) High-quality genomic DNA from human whole blood and mononuclear cells. Biotechniques 33 (6):1228-1230. doi:10.2144/02336bm09
32. Team R (2020) A language and environment for statistical computing. Computing 1. doi:10.1890/0012-9658(2002)083[3097:CFHIWS]2.0.CO;2
33. Zhong J, Agha G, Baccarelli AA (2016) The Role of DNA Methylation in Cardiovascular Risk and Disease: Methodological Aspects, Study Design, and Data Analysis for Epidemiological Studies. Circ Res 118 (1):119-131. doi:10.1161/CIRCRESAHA.115.305206
34. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, O'Brien WL, Bassett DR, Jr., Schmitz KH, Emplaincourt PO, Jacobs DR, Jr., Leon AS (2000) Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc 32 (9 Suppl):S498-504. doi:10.1097/00005768-200009001-00009
35. Venables WN, Ripley BD (2002) Modern Applied Statistics with S. Fourth edn. Springer,
36. Noble WS (2009) How does multiple testing correction work? Nat Biotechnol 27 (12):1135-1137. doi:10.1038/nbt1209-1135

37. Li J, Ji L (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb) 95 (3):221-227. doi:10.1038/sj.hdy.6800717
38. Touw WG, Bayjanov JR, Overmars L, Backus L, Boekhorst J, Wels M, van Hijum SA (2013) Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? Brief Bioinform 14 (3):315-326. doi:10.1093/bib/bbs034
39. Liaw A, Wiener M (2002) Classification and Regression by randomForest. R News 2 (3):18-22
40. Janitza S, Celik E, Boulesteix A-L (2018) A computationally fast variable importance test for random forests for high-dimensional data. Advances in Data Analysis and Classification 12 (4):885-915. doi:10.1007/s11634-016-0276-4
41. Mattei D, Pietrobelli A (2019) Micronutrients and Brain Development. Curr Nutr Rep 8 (2):99-107. doi:10.1007/s13668-019-0268-z
42. Terrin G, Berni Canani R, Di Chiara M, Pietravalle A, Aleandri V, Conte F, De Curtis M (2015) Zinc in Early Life: A Key Element in the Fetus and Preterm Neonate. Nutrients 7 (12):10427-10446. doi:10.3390/nu7125542
43. Calder PC (2020) Nutrition, immunity and COVID-19. BMJ Nutr Prev Health 3 (1):74-92. doi:10.1136/bmjnph-2020-000085
44. Jarosz M, Olbert M, Wyszogrodzka G, Mlyniec K, Librowski T (2017) Antioxidant and anti-inflammatory effects of zinc. Zinc-dependent NF-kappaB signaling. Inflammopharmacology 25 (1):11-24. doi:10.1007/s10787-017-0309-4
45. Roohani N, Hurrell R, Kelishadi R, Schulin R (2013) Zinc and its importance for human health: An integrative review. J Res Med Sci 18 (2):144-157
46. Wessels I, Haase H, Engelhardt G, Rink L, Uciechowski P (2013) Zinc deficiency induces production of the proinflammatory cytokines IL-1beta and TNFalpha in promyeloid cells via epigenetic and redox-dependent mechanisms. J Nutr Biochem 24 (1):289-297. doi:10.1016/j.jnutbio.2012.06.007
47. Cabrera AJ (2015) Zinc, aging, and immunosenescence: an overview. Pathobiol Aging Age Relat Dis 5:25592. doi:10.3402/pba.v5.25592
48. Mezzaroba L, Alfieri DF, Colado Simao AN, Vissoci Reiche EM (2019) The role of zinc, copper, manganese and iron in neurodegenerative diseases. Neurotoxicology 74:230-241. doi:10.1016/j.neuro.2019.07.007
49. Narayanan SE, Rehuman NA, Harilal S, Vincent A, Rajamma RG, Behl T, Uddin MS, Ashraf GM, Mathew B (2020) Molecular mechanism of zinc neurotoxicity in Alzheimer's disease. Environ Sci Pollut Res Int 27 (35):43542-43552. doi:10.1007/s11356-020-10477-w
50. Dubey P, Thakur V, Chattopadhyay M (2020) Role of Minerals and Trace Elements in Diabetes and Insulin Resistance. Nutrients 12 (6). doi:10.3390/nu12061864
51. Gac P, Czerwinska K, Macek P, Jaremkow A, Mazur G, Pawlas K, Poreba R (2020) The Importance of Selenium and Zinc Deficiency in Cardiovascular Disorders. Environ Toxicol Pharmacol:103553. doi:10.1016/j.etap.2020.103553
52. Zhang Y, Tian Y, Zhang H, Xu B, Chen H (2020) Potential pathways of zinc deficiency-promoted tumorigenesis. Biomed Pharmacother 133:110983. doi:10.1016/j.biopha.2020.110983
53. Wong CP, Rinaldi NA, Ho E (2015) Zinc deficiency enhanced inflammatory response by increasing immune cell activation and inducing IL6 promoter demethylation. Mol Nutr Food Res 59 (5):991-999. doi:10.1002/mnfr.201400761
54. Andreini C, Banci L, Bertini I, Rosato A (2006) Counting the zinc-proteins encoded in the human genome. J Proteome Res 5 (1):196-201. doi:10.1021/pr050361j
55. Brito S, Lee MG, Bin BH, Lee JS (2020) Zinc and Its Transporters in Epigenetics. Mol Cells 43 (4):323-330. doi:10.14348/molcells.2020.0026
56. Lyko F (2018) The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. Nat Rev Genet 19 (2):81-92. doi:10.1038/nrg.2017.80
57. Ren W, Gao L, Song J (2018) Structural Basis of DNMT1 and DNMT3A-Mediated DNA Methylation. Genes (Basel) 9 (12). doi:10.3390/genes9120620

58. Pradhan M, Esteve PO, Chin HG, Samaranayke M, Kim GD, Pradhan S (2008) CXXC domain of human DNMT1 is essential for enzymatic activity. Biochemistry 47 (38):10000-10009. doi:10.1021/bi8011725
59. Suzuki M, Yamada T, Kihara-Negishi F, Sakurai T, Hara E, Tenen DG, Hozumi N, Oikawa T (2006) Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. Oncogene 25 (17):2477-2488. doi:10.1038/sj.onc.1209272
60. Castro C, Millian NS, Garrow TA (2008) Liver betaine-homocysteine S-methyltransferase activity undergoes a redox switch at the active site zinc. Arch Biochem Biophys 472 (1):26-33. doi:10.1016/j.abb.2008.01.017
61. Jing M, Rech L, Wu Y, Goltz D, Taylor CG, House JD (2015) Effects of zinc deficiency and zinc supplementation on homocysteine levels and related enzyme expression in rats. J Trace Elem Med Biol 30:77-82. doi:10.1016/j.jtemb.2014.10.013
62. Alexander J, Tinkov A, Strand TA, Alehagen U, Skalny A, Aaseth J (2020) Early Nutritional Interventions with Zinc, Selenium and Vitamin D for Raising Anti-Viral Resistance Against Progressive COVID-19. Nutrients 12 (8). doi:10.3390/nu12082358
63. Mossink JP (2020) Zinc as nutritional intervention and prevention measure for COVID-19 disease. BMJ Nutr Prev Health 3 (1):111-117. doi:10.1136/bmjnph-2020-000095
64. Mousavi SM, Hajishafiee M, Clark CCT, Borges do Nascimento IJ, Milajerdi A, Amini MR, Esmaillzadeh A (2020) Clinical effectiveness of zinc supplementation on the biomarkers of oxidative stress: A systematic review and meta-analysis of randomized controlled trials. Pharmacol Res 161:105166. doi:10.1016/j.phrs.2020.105166
65. Ramos-Lopez O, Milagro FI, Riezu-Boj JI, Martinez JA (2020) Epigenetic signatures underlying inflammation: an interplay of nutrition, physical activity, metabolic diseases, and environmental factors for personalized nutrition. Inflamm Res. doi:10.1007/s00011-020-01425-y
66. Makarov MV, Trammell SAJ, Migaud ME (2019) The chemistry of the vitamin B3 metabolome. Biochem Soc Trans 47 (1):131-147. doi:10.1042/BST20180420
67. Pissios P (2017) Nicotinamide N-Methyltransferase: More Than a Vitamin B3 Clearance Enzyme. Trends Endocrinol Metab 28 (5):340-353. doi:10.1016/j.tem.2017.02.004
68. Mrochek JE, Jolley RL, Young DS, Turner WJ (1976) Metabolic response of humans to ingestion of nicotinic acid and nicotinamide. Clin Chem 22 (11):1821-1827
69. Li D, Tian YJ, Guo J, Sun WP, Lun YZ, Guo M, Luo N, Cao Y, Cao JM, Gong XJ, Zhou SS (2013) Nicotinamide supplementation induces detrimental metabolic and epigenetic changes in developing rats. Br J Nutr 110 (12):2156-2164. doi:10.1017/S0007114513001815
70. Tian YJ, Luo N, Chen NN, Lun YZ, Gu XY, Li Z, Ma Q, Zhou SS (2014) Maternal nicotinamide supplementation causes global DNA hypomethylation, uracil hypo-incorporation and gene expression changes in fetal rats. Br J Nutr 111 (9):1594-1601. doi:10.1017/S0007114513004054
71. Titcomb TJ, Tanumihardjo SA (2019) Global concerns with B vitamin statuses: biofortification, fortification, hidden hunger, interactions, and toxicity. Comprehensive Reviews in Food Science and Food Safety 18 (6):1968– 1984
72. Arpon A, Riezu-Boj JI, Milagro FI, Marti A, Razquin C, Martinez-Gonzalez MA, Corella D, Estruch R, Casas R, Fito M, Ros E, Salas-Salvado J, Martinez JA (2016) Adherence to Mediterranean diet is associated with methylation changes in inflammation-related genes in peripheral blood cells. J Physiol Biochem 73 (3):445-455. doi:10.1007/s13105-017-0552-6
73. Ma J, Rebholz CM, Braun KVE, Reynolds LM, Aslibekyan S, Xia R, Biligowda NG, Huan T, Liu C, Mendelson MM, Joehanes R, Hu EA, Vitolins MZ, Wood AC, Lohman K, Ochoa-Rosales C, van Meurs J, Uitterlinden A, Liu Y, Elhadad MA, Heier M, Waldenberger M, Peters A, Colicino E, Whitsel EA, Baldassari A, Gharib SA, Sotoodehnia N, Brody JA, Sitlani CM, Tanaka T, Hill WD, Corley J, Deary IJ, Zhang Y, Schottker B, Brenner H, Walker ME, Ye S, Nguyen S, Pankow J, Demerath EW, Zheng Y, Hou L, Liang L, Lichtenstein AH, Hu FB, Fornage M, Voortman T, Levy D (2020) Whole Blood DNA Methylation Signatures of Diet Are Associated With Cardiovascular Disease Risk Factors and All-Cause Mortality. Circ Genom Precis Med 13 (4):e002766. doi:10.1161/CIRCGEN.119.002766

Table 1. Baseline characteristics. Characteristics of the subcohort sample with at least one methylation measure available (N=1,080) compared to the whole Moli-sani cohort (N=24,325)

	Subcohort			Whole Moli-sani cohort		
Variable	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
Age (years)	1,080	54.91	11.52	24,325	55.79	11.96
MDS	1,080	4.73	1.6	24,221	4.35	1.64
Physical activity (meth/d)	1,080	3.6	4.03	24,325	3.48	4.02
BMI (kg/m^2)	1,079	28.04	4.54	24,308	28.06	4.78
Energy intake (Kcal/d)	1,080	2210.19	682.57	24,225	2079.01	667.66
Abdominal Obesity (WHR)	1,079	0.92	0.07	24,297	0.92	0.08
Monocytes (%)	1,037	5.93	2.04	23,544	7.09	2.12
Granulocytes (%)	1,037	60.69	7.68	23,542	60.25	7.82
Lymphocytes (%)	1,037	33.33	7.39	23,545	32.63	7.34
Categorical variables	**N**	**n**	**%**	**N**	**n**	**%**
Males (n. %)	1,080	518	47.96	24,325	11.702	48.11
Education						
Primary	1,080	223	20.65	24,286	6.268	25.81
Lower secondary	1,080	285	26.39	24,286	6.742	27.76
Upper secondary	1,080	405	37.5	24,286	8.259	34.01
Post-secondary	1,080	167	15.46	24,286	3.017	12.42
Health conditions						
CVD	1,068	0	0	24,023	1.427	5.94
Cancer	1,076	35	3.25	24,198	788	3.26
Diabetes	1,065	38	3.57	24,017	1.214	5.05
Hyperlipidaemia	1,061	45	4.24	24,092	1.911	7.93
Drinking status (drinkers)						
Ever	1,080	151	13.98	24,325	6.156	25.31
Current	1,080	774	71.67	24,325	14.650	60.23
Former	1,080	96	8.89	24,325	1.032	4.24
Occasional	1,080	57	5.28	24,325	1.515	6.23
Missing	1,080	2	0.19	24,325	972	4
Smoker status (smokers)						
Ever	1,078	527	48.89	24,296	12.050	49.6
Current	1,078	263	24.4	24,296	5.582	22.97
Former	1,078	288	26.72	24,296	6.664	27.43

Abbreviations: MDS: Mediterranean Diet Score

Table 2. Results of the stepwise multivariable association models of global 5mC vs daily intake of **a)** eight food groups, **b)** three macronutrients and **c)** seventeen micronutrients.

a)

Food Group	Unit	Beta	SE	T-stat	Raw P-value	Bonferroni P-value
Cereals	g/day	-0.001	0.0004	-2.91	3.7×10^{-3}	0.06
Dairy products	g/day	-0.0004	0.0003	-1.53	0.13	1
MUFA-SFA ratio	NA	-0.16	0.11	-1.52	0.13	1
Vegetables	g/day	-	-	-	-	-
Fruits and nuts	g/day	-	-	-	-	-
Fish	g/day	-	-	-	-	-
Legumes	g/day	-	-	-	-	-
Meat	g/day	-	-	-	-	-

b)

Macronutrient	Unit	Beta	SE	T-stat	Raw P-value	Bonferroni P-value
Total Lipids	g/day	0.004	0.002	1.80	0.07	0.28
Available Carbohydrates	g/day	-	-	-	-	-
Total Proteins	g/day	-	-	-	-	-

c)

Micronutrient	Unit	Beta	SE	T-stat	Raw P-value	Bonferroni P-value
Zinc	**mg/day**	**0.07**	**0.02**	**2.96**	**3.1×10^{-3}**	**0.04**
Vitamin B3	mg/day	-0.04	0.02	-2.55	0.01	0.14
Phosphorus	mg/day	-0.0005	0.0002	-2.40	0.02	0.28
Vitamin D	mg/day	0.08	0.04	2.09	0.04	0.56
Sodium	mg/day	-0.0001	0.00006	-2.00	0.05	0.70
Vitamin B6	mg/day	0.28	0.12	1.97	0.05	0.70
Iron	mg/day	-	-	-	-	-
Calcium	mg/day	-	-	-	-	-
Potassium	mg/day	-	-	-	-	-
Vitamin B1	mg/day	-	-	-	-	-
Vitamin B2	mg/day	-	-	-	-	-
Vitamin C	mg/day	-	-	-	-	-
Vitamin B9	µg/day	-	-	-	-	-

Vitamin A1	µg/day	-	-	-	-	-
Beta-carotene	µg/day	-	-	-	-	-
Vitamin E	mg/day	-	-	-	-	-
Selenium	µg/day	-	-	-	-	-

Beta coefficients and their SE are reported as increase of global methylation (standardized % of CpG sites) per unitary increase of daily intake of each nutritional variable. Beta coefficient and the corresponding T-statistics and P-value are reported only for the nutritional variables that were retained in the stepwise regression. Variables for which no statistics is reported are those automatically excluded from predictors in the model since they do not represent a gain in the trade-off between goodness of fit and parsimony of the model. In other words, these variables did not significantly contribute to an increase in the total variance of 5mC, and were therefore not retained and tested in the final regression model. Statistically significant associations for each nutritional layer (surviving Bonferroni correction, i.e. Bonferroni p-value < 0.05) are highlighted in bold. Abbreviations: MUFA-SFA ratio, monounsaturated to saturated fat ratio; SE, standard error; T-stat = T statistics.

Table 3. Results of the stepwise multivariable association models of global 5-hmC vs daily intake of seventeen micronutrients.

Micronutrient	Unit	Beta	SE	T-stat	Raw P-value	Bonferroni P-value
Zinc	mg/day	-	-	-	-	-
Vitamin B3	mg/day	-	-	-	-	-
Phosphorus	mg/day	-	-	-	-	-
Vitamin D	mg/day	-	-	-	-	-
Sodium	mg/day	-0.00009	0.00006	-1.55	0.12	1
Vitamin B6	mg/day	-	-	-	-	-
Iron	mg/day	-0.03	0.02	-1.81	0.07	0.98
Calcium	mg/day	-	-	-	-	-
Potassium	mg/day	-	-	-	-	-
Vitamin B1	mg/day	-	-	-	-	-
Vitamin B2	mg/day	-	-	-	-	-
Vitamin C	mg/day	-	-	-	-	-
Vitamin B9	µg/day	-	-	-	-	-
Vitamin A1	µg/day	-	-	-	-	-
Beta-carotene	µg/day	-	-	-	-	-
Vitamin E	mg/day	-	-	-	-	-
Selenium	µg/day	-	-	-	-	-

Beta coefficients and their SE are reported as increase of global methylation (standardized % of CpG sites) per unitary increase of daily intake of each nutritional variable. Beta coefficient and the corresponding T-statistics and P-value are reported only for the nutritional variables that were retained in the stepwise regression. Variables for which no statistics is reported are those automatically excluded from predictors in the model since they do not represent a gain in the trade-off between goodness of fit and parsimony of the model. In other words, these variables did not significantly contribute to an increase in the total variance of 5hmC, and were therefore not retained and tested in the final regression model. Statistically significant associations for each nutritional layer (surviving Bonferroni correction, i.e. Bonferroni

p-value < 0.05) are highlighted in bold. Abbreviations: SE, standard error; T-stat = T statistics.

Fig. 1 Different nutritional scores and intakes analysed for association with global DNA methylation and hydroxymethylation levels in the present study.

The different nutritional strata tested for association with 5mC and 5hmC in the present study are illustrated. From top to bottom: adherence to Mediterranean Diet (courtesy of Oldways, www.oldwayspt.org), intake of main food groups, macronutrients and micronutrients. Abbreviations: 5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine

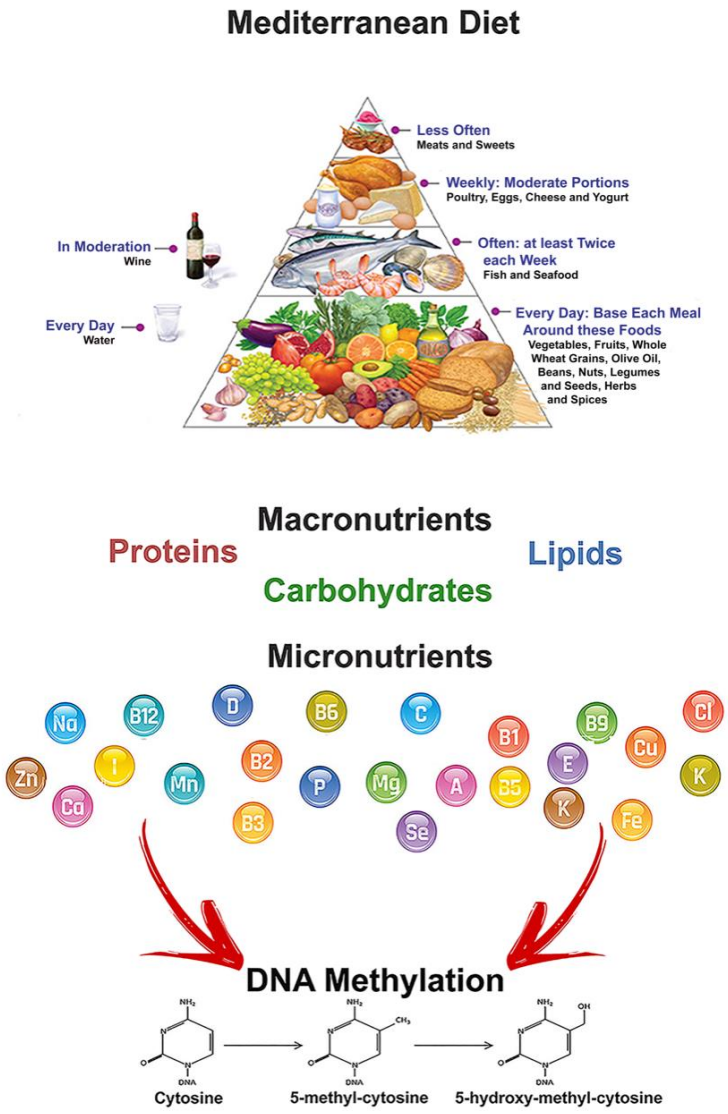

Fig. 2 Permutation Feature importance analysis of **a)** 5mC and **b)** 5hmC predictions through Random Forest algorithms. Loss drop after perturbations (defined as the average percentage increase in the Mean Squared Error of the permuted vs the baseline models) are reported for the ten most influential nutritional intakes within each analysis. Abbreviations: av. carbohydrates, available carbohydrates; MUFA-SFA ratio, monounsaturated to saturated fat ratio.